Exploring emerging technological trajectories within the innovation ecosystem

By

Hassan Sherwani
Ph.D. Candidate in Management
Cycle XXXII

Supervisors:

Prof. Maria Isabella Leone
Dr. Yashar Ghiassi Farrokhfal

April 2021

Contents page

Thesis Overview

This thesis explores the notion of the ecosystem as related to innovation studies by relying upon literature review and empirical modes of inquiry. Prior studies have presented the concept of the ecosystem in a variety of approaches in management, information system, or governance domain, rendering the literature-rich but complex. Many scholars have explored the role of the ecosystem in managing the inter-dependency arising from the multiple overlapping and conflicting logics that come across the organization. Further studies have examined the ecosystem as an orchestration feature within the supply chain perspective, which can obfuscate firms' performance or reveal collaborative patterns of different organizations within that ecosystem. We argue that combining review methods and empirical approaches to explore innovations' role within ecosystems and complex network formation across other actors can reconcile the diverse findings while providing innovative avenues for future research.

The first chapter of this thesis aims to provide a critical review of prior studies on the ecosystem in innovation studies and systematize the diverse findings. Understanding the ecosystem context will provide helpful knowledge to organizations that are competing in such an interconnected system. In this paper, we attempt to provide this understanding by synthesizing scholars' work in the field of ecosystem innovation through a literature review. This study delves at how organizations handle interdependencies with other actors in their respective ecosystems, as well as how strategy formulation of such interrelationships affects innovation in general and value creation in specific.

In the second chapter, we explore emerging technological trends in the innovation ecosystem. We use the automotive ecosystem as a case study. While previous research scholarship primarily relies on firm-level analysis, we instead apply patent-analysis to study the technological development in the automotive ecosystem. This diverse ecosystem consists of a focal firm, suppliers, information and communication service providers, and key actors related to smart charging infrastructure. To empirically analyze this complex network, this study applies the topic modeling method using patents related to the automotive ecosystem registered at Worldwide Patent Statistical Database EPO's PATSTAT. The suggested approach's core is a generative model based on latent Dirichlet allocation (LDA), allowing identifying core topics related to the automotive ecosystem. We found that the technical topics and their trends generated by the suggested approach have significantly understood the technological landscape in a holistic ecosystem rather than a specific sector.

The third chapter explores the role of end customers in the ecosystem framework. The focus still lies on technology and its adoption; however, our focus of analysis is end customers for whom inter-dependent organizations tend to create value. Firms within the ecosystem may reduce this ambiguity by assessing the customer's sentiments when consumers ultimately decide to embrace or reject emerging technology. We identify the key factors that affect customers' opinions towards an emerging technology. Taking the ecosystem for Autonomous Vehicles as a case study, we attempt to recognize the risk and benefit perceptions that lead to the adoption of new technologies. We take a machine learning method for text classification and use extensive twitter data (455,727 tweets from June 2016 to January 2019) as our methodology. We find quantitatively and qualitatively that customers' risk and benefit perception is the key determinant. Exaggerated expectations of risk or benefit may lead to irrational behaviors. Organizations could determine customers' sentiments and implement their strategy for such emerging technologies to test waters.

Table 1: Brief description of all three papers

Title of papers	Type of study	Source of data	Level of analysis	Publication status
Innovation and value creation in business ecosystems	Review	Leading academic journals in innovation	Firm-level	Published [1]
Exploring emerging technological trends in Automotive ecosystem: a patent analysis	Empirical	Worldwide Patent Statistical Database EPO's PATSTAT www.epo.org	Firm-level	In process[2]
Analyzing the technological adoption through customers' perceived value: A case study for autonomous vehicle (AV)	Empirical	Twitter Search API www.developer.twitter.com	Customer& firm-level	In process[3]

[1] Sherwani, H. and Tee, R., 2018. Innovation and value creation in business ecosystems. In Learning and Innovation in Hybrid Organizations (pp. 13-32). Palgrave Macmillan, Cham.

[2] https://www.journals.elsevier.com/technological-forecasting-and-social-change/call-for-papers/leveraging-disruptive-technology-in-unlocking

[3] https://www.poms.org/2021/02/call_for_papers_special_issue_17.html

Chapter 1

Innovation and Value Creation in Business Ecosystems

Abstract

The notion of ecosystems in innovation literature identifies companies' networks that collectively deliver a holistic, interconnected technology framework that generates value for customers. Understanding the ecosystem context will provide helpful knowledge to organizations that are competing in such an interconnected system. In this paper, we attempt to provide this understanding by synthesizing scholars' work in the field of ecosystem innovation through a literature review. This paper focuses on how firms manage these interdependencies with other actors that emerge in business ecosystems and how such interdependencies' strategic management affects innovation and value creation in business ecosystems.

Keywords

Innovation ecosystems, network dynamics, value creation, innovation management.

Introduction

The ecosystem (i. e., ecological system) in the business context is a notion derived from the biological sciences. Biological ecosystem comprises a variety of interdependent species. Similarly, business ecosystems reflect interdependent networks of organizations. This approach expands the previous conception of value chains (Porter, 1980) to consider interdependencies and related organizational and institutional actors' coevolution. In such an ecosystem structure, each member contributes to the ecosystem's overall well-being and is dependent on other members for its survival. In turn, each member's survival and progress are influenced by the environment as a holistic entity in continuous evolution (Iansiti and Levien, 2004).

The field has grown from a network perspective to illustrate the combination of complex relationships and dimensions, such as technology and innovation interdependencies (Adner, 2012), knowledge generation among crucial players (Clarysse et al., 2014; Järvi et al., 2018) as well as platform ecosystems specific to the digital domain (Gawer and Cusumano, 2014). In such cases, ecosystem scholars have clarified various primary phenomena related to policy, technology, and innovation. Such processes include the management and orchestration of ecosystems (Ritala et al., 2013; Wareham et al., 2014; Williamson and De Meyer, 2012), the organization and relationships of ecosystems (Basole et al., 2015; Clarysse et al., 2014), as well as the advent of ecological innovation and technology (Ansari et al., 2016; Dattée et al., 2018). The spike in ecosystem study is recorded in several literature reviews, showing the broad range of theoretical approaches embraced by scholars (Aarikka-Stenroos and Ritala, 2017; de Vasconcelos Gomes et al., 2018; Tsujimoto et al., 2018). Jacobides et al. (2018) identified three streams of research. "A business ecosystem stream that focuses on a firm and its environment; an ecosystem innovation stream that focuses on a specific innovation or new value proposition and the constellation of stakeholders that endorse it; and an ecosystem stream platform that considers how actors organize around a platform."

From an ecosystem innovation viewpoint, recognizing ecosystem change's mannerism may provide helpful information about networks and organizations within these networks. There is an interdependence of companies with other entities in such dynamic network environments. The Innovation Ecosystem Research Stream focuses on focal innovation and components (upstream) and complements (downstream) that support it. It considers the ecosystem as "the collaborative arrangements through which firms combine their offerings into a coherent, customer-facing solution" (Adner, 2006). Focus is put on understanding how interdependent actors work to produce and commercialize inventions that favor the end-customer with the corollary that, if communication within the ecosystem is insufficient, innovations will fail (Kapoor & Lee, 2013). The main focus lies in the framework of innovations that allows consumers to utilize the end product with added value. Therefore, the essence of the ecosystem architecture is to create a connection between the core product, its components, and its complementary products/services ('complements'), which adds value to customers. Firms generating focal innovation may or may not be directly linked to complementary providers; the degree to which firms align across various agreements may influence their capacity to generate value for the end-customer (Adner, 2017). Here, the ecosystem is networking around an orchestrating network (Iyer et al., 2006), providing focal and complementary technologies. The research stream describes the relationship between the innovative focal organizations and their complementary technology (Kapoor & Lee, 2013). These partnerships allow both parties to discuss potential research and development and technology commercialization (Mäkinen & Dedehayir, 2012).

Another stream illustrates how information sharing influences inter-connected players' strength and affects the ecosystem (Alexy et al., 2013; Brusoni & Prencipe, 2013). Such inter-firm relationships provide health and sustainability to the ecosystem (West & Wood, 2013).

In this paper, we aim to illustrate the main characteristics and dynamics of innovation ecosystems. We conduct a literature review of innovation ecosystems, a study stream that is relatively new to the field of technology management and information systems. We refine scholars' work to date, drawing a conceptual intersection of business networks, knowledge integration, and innovation management. This study attempts to bring attention to the functions and roles of ecosystem innovation participants, factors that affect the evolution of ecosystems, the dynamics of ecosystem change, and the strategic considerations of ecosystem-based firms.

Theoretical Background

The term "ecosystem" has been of rising interest in strategy discussions (Moore, 1993; Iansiti and Levien, 2004; Clarysse et al., 2014; Adner, 2016). Businesses are moving towards a more networked approach that creates new opportunities and subsequent challenges (Adner, 2016). Hence, this idea of the ecosystem has sparked a rising interest among innovative organizations. It has raised awareness for how ecosystems affect their business models and value networks (Christensen and Rosenbloom, 1995), thereby improving overall value creation and value capture.

The ecosystem is a term that has been taken from biological sciences, which is a business context that refers to interdependent networks of organizations. As in natural ecosystems, each member would depend on other members to survive and learn to participate and contribute to the overall system. This scenario becomes relevant in the business context where each organization's survival, growth, and success is affected by the ecosystem holistically (Iansiti and Levien, 2004). In business, ecosystems outline the network of firms that collectively create an overall integrated technological system to create value for customers. Understanding ecosystems' mode in which their operations change may provide helpful information for firms that lie or associate themselves within these networked environments (Mäkinen & Dedehayir, 2012).

According to Moore (2005), markets, hierarchies, and ecosystems are the three main contemporary business components. These pillars, therefore, should provide the basis for companies to compete, regulate policy, and counter negative or antitrust actions. Most companies tend to adopt this approach; however, haphazard adoption may create issues. There is confusion related to boundary, overlap, redundancy, applicability, unit, and focus for analysis (Adner, 2016).

According to Moore (1993), "An economic community supported by a foundation of interacting organizations and individuals—the organisms of the business world. The economic community produces goods and services of value to customers, who are themselves members of the ecosystem. The member organisms also include suppliers, lead producers, competitors, and other stakeholders. Over time, they coevolve their capabilities and roles and align themselves with the directions set by one or more prominent companies. Those companies holding leadership roles may change over time. Still, the community values the function of ecosystem leader because it enables members to move toward shared visions to align their investments, and to find mutually supportive roles."

To elaborate this idea given by Moore (1993), business ecosystems are a network of interdependent organizations that co-ordinate to create success. Traditionally, companies are considered rivals where they fight each other for maximum market share. This concept of corporate rivalry has been contradicted as organizations in modern times work in different environments. Competition has different meanings. Firms tend to integrate competition and cooperation to produce more diverse value for the end customer. This practice is vital for competitors as they could rely on other firms to survive. It is worth noticing that business ecosystems consist of actors (Moore,1993) who participate within an organization. Besides, distribution channels and suppliers are considered part of the system as well. There are certain extended participants such as customers, standard bodies, and suppliers of complementary products. Some actors are thought to be an external influence on the system; however, they impact the main functions of the business. Examples are trade partners, unions, key investors, and regulatory bodies.

As mentioned earlier, an ecosystem describes an environment containing an organization. Though similar, the business environment and business ecosystems are not the same. It is also significantly different as it refers to the systematic nature of the total environment and key components making up that system. The ecosystem also addresses the internal evolutionary process through which an organization must go, its adoption capability in a transition period, and its coevolution and external environment. According to Hagel & Brown (2005), business ecosystems may explain a specific type of environment. In this case, clusters of companies focusing on a particular kind of business or technology might decide to locate their operations in close geographic proximity to each other.

To understand the role and dimension of business ecosystems, companies comprising these may be diverse and unique in their capabilities. However, they have come closer due to this business collaboration. An organization may go through intensive development of its infrastructure first to be competitive. This infrastructure includes various activities such as finance, accounting, legal issues, deployment of sales and marketing units, recruitment at executive and mid-level, and maintenance of relations with other partners. Through this consistent process of infrastructure development, there could be a chance of building up links within different units and departments. Now consider this scenario at a higher level where more companies and organizations act as actors at the individual level. The same phenomena of organizational structure will be followed but at a higher level. In this scenario, other institutions such as research centers, universities, governmental organizations, non-profit organizations may serve as an extra blend of network interaction (Hagel & Brown, 2005).

Business ecosystems' role is to surround, permeate, and reshape given markets and hierarchies (Moore, 2005). Most companies in a modern competitive environment emphasize efficiency and effectiveness as the basis of innovation. Profit margins are still necessary but, they are not the only criteria to compete. Companies have also realized that they cannot change the system by individual approach and innovation phenomena remain incomplete without collaboration with other companies. Firms embrace business ecosystems to coordinate innovation through the continuous evolution of multiple markets and hierarchies (Moore, 2005). It provides a win-win scenario for the customer as they could get maximum benefits through innovation. Furthermore, there are complementary innovations that need to be coevolved across company lines because there is no way that one firm could achieve all the required knowledge, technical resources, and managerial skills to fulfill the demand.

Organizations may decide to coordinate together but at the same time may wish to keep a certain level of autonomy. There is also a need for an agency whose agents are legally autonomous and not bound by employment relationships. (Gulati et al., 2012). Hence, These ecosystems have been described as a new type of organizational structure that combines open limits of membership with a highly stratified and more hierarchical decision-making process. . An agent could be an organization in itself. However, it can be taken as a unitary actor for analysis purposes, and this legally autonomous organization is called meta-organization. It may consist of networks of organizations or even individuals recognized by a system-level body. However, they are independent of authority-oriented employment relationships or contracts. These networks do not mean that organizations within this network would have unified goals.

Each organization may have its own goals; for instance, improving production quality could be important for one firm. Nevertheless, another firm might prioritize managing the levels of sales. Ecosystems allow each firm to fulfill its need under a unified network where it is unnecessary for constituent agents to share it. It is just like a traditional organization system in which individuals are free to have their priority. Hence, meta-organizations comprise firms where each agent has its motivation, goals, and incentive systems. These organizations are still different from traditional business settings. Meta-organizations are associated through authority-based contracts (Gulati et al., 2012). These contracts make all actors inside this infra-structure independent of each other at the firm level. They are connected through a network that allows them to stay connected and have complete autonomy.

To understand this concept, consider communities of economic agents where individual business activities could measure the overall community market value. For example, tech firms that make services for Apple iPod. They can be taken as the iPod business ecosystem. Another example is entertainment companies that choose to license music through iTunes or iPod-related music sites. In other words, a business ecosystem can also be conceived as a network of interdependent actors that collaborate and innovate (Moore, 2005).

Collaboration and knowledge integration

Ecosystems and the collaboration process seem a win-win for all. However, there could be some issues that could lead to undesired results. One of the main concerns in this network approach is knowledge complementarities. It can create interdependencies that need to be resolved (Thompson 1967). Knowledge adds a significant part in value creation; however, there could be severe barriers in transferring and replicating the knowledge. Hence, knowledge utilization matters. It is worth noticing at this point that it, in the broader sense, represents both 'explicit' knowledge and 'tacit' knowledge. Explicit knowledge refers to that which can be written down, whereas tacit knowledge cannot (Grant, 1996; Kikoski& Kikoski, 2004; Nonaka et al., 2000).

To further elaborate this concept, explicit knowledge is described as what can be encapsulated as a language or even a code. This coding style allows organizations to communicate, process, and store this set of knowledge conveniently. One example of explicit knowledge is patents or copyrights (Dalley and Hamilton, 2000). Through this process, explicit knowledge becomes a direct asset for the organization. On the other hand, tacit knowledge is personal and challenging to be codified and formalized, rooted in actions, procedures, commitment, value, and emotions (Kikoski& Kikoski, 2004). Tacit and explicit knowledge are both complementary, which means that both are equally required for knowledge creation. Knowledge is created through tacit and explicit interaction and not from either tacit or explicit knowledge alone. However, competitive advantage is achieved by the organization through tacit knowledge because explicit knowledge is known to every individual.

Polanyi (1969) described the significance of knowledge as the knowledge that is considered to be better explained than said. As far as an organization is concerned, whether a newly born startup or an established market player, every individual associated with the firm has a unique skill set. These skills are like an asset for the company, and every firm wishes to translate these skills into knowledge. One cannot codify these skills as they come along the hard way of individual focus, training, and experience. Others could learn them through the process of keen observation. In terms of an organization, whether it is a large company or a startup, each individual possesses skills that are unique and, once unlocked, can be a creative contribution to an organization (Kikoski& Kikoski, 2004).

It is crucial to notice that tacit knowledge indirectly plays its role in innovation, and hence, it is very significant at the organizational or even network level. This type of knowledge helps organizational activities and functions by creating new knowledge. This knowledge is called new as it has been extracted from individuals' skills and competence or group working within an organization. This knowledge plays its role in various applications such as new product development, novel business concepts, and procedures. These are the outcome of tacit knowledge and its adoption, and these results are the reason for innovation. Hence, tacit knowledge enables each skilled individual to contribute through novel ideas and concepts. Besides, it provides beneficial knowledge at the personal level available to others (Alwis & Hartmann, 2008; Kikoski& Kikoski, 2004). It is the same in a network scenario where new companies learn a lot from market-dominant players that would not have been possible without translating tacit knowledge in that respective ecosystem.

Knowledge management is directly related to the capability of any firm towards its information processing ability. Information processing setups are instrumental in knowledge utilization within firms (Tushman & Nadler, 1978). Information processing consists of information gathering, information interpretation, and information synthesizing. These all components act as a process of knowledge integration within an organization.

Knowledge conversion is an essential aspect of any enterprise (Nonaka et al., 2000). There are four modes of knowledge conversion;

i. Socialization: From tacit knowledge to tacit knowledge
ii. Externalization: From tacit knowledge to explicit knowledge
iii. Combination: From explicit knowledge to explicit knowledge
iv. Internalization: From explicit knowledge to tacit knowledge

Knowledge created by this spiral process can be valuable, as this created knowledge moves along the system (Nonaka et al., 2000). Tacit knowledge can be challenging as it comes through experience, and it can only evolve through more experience. It could be time-consuming and take place through trial and error procedures. Different companies use different methods towards socialization (e.g., the Kanban model in software development companies). The benefits of tacit knowledge are apparent, and hence, it should consist of high priority for organizations to motivate the creation of tacit knowledge (Alwis & Hartmann, 2008).

Knowledge sharing and transfer can be more involved in ecosystem networks as companies may not share the same motivation. They might contain some information within the walls of the respective organization. At an internal level of a firm, knowledge integration is significant as an individual does not possess enough cognitive power to contain it. Hence, it is not feasible for each individual and his/her ability to understand and learn the knowledge given by other specialists (Grant, 1996). That is why knowledge is shared in an organization.

Replication of knowledge integration across a meta-organization is somewhat tricky. In explicit knowledge, it is not easy to keep this knowledge safe enough through copyrights or patents. We observe many patents dispute in the regular business setting. It is hard for a startup to open up its explicit knowledge towards a network that openly. Most technological firms tend to be careful about sharing particular types of knowledge. That knowledge and tools are the only assets they possess, and hence they try to protect them. Joint ventures are one example where organizations share knowledge more freely, and as there is a win-win for both, they tend to cooperate more.

As far as tacit knowledge is concerned, it is even more challenging due to difficulty in knowledge transfer level. Companies within an ecosystem may have uneven market information and different levels of hierarchy. For understanding the integration process with tacit knowledge in consideration, there are two mechanisms. One is related to identifying the direction through which knowledge would be communicated between specialists and specialists in some other fields (Demsetz, 1991). For example, British Airways has global aircraft maintenance facilities. These main maintenance facilities include service and repair handled by a specialized host familiar with procedures and directives based on Federal Aviation Authority. Manufacturers give others services such as guidance and technical information. Hence, these rules, formulae, expert systems directives, policies, and procedures are tackled by the number of specialists and how they communicate to either non-specialists or those familiar with other aspects.

In other words, direction refers to codifying tacit knowledge into explicit rules and instructions useful for those who have partial or no knowledge. There is an issue; however, As Polanyi (1966) describes, there is a danger of losing some vital knowledge in this transition process. "We can know more than we can tell." Hence, converting tacit knowledge into explicit knowledge as a form of rules, directives, and policies could cause a certain degree of loss.

The second mechanism is organizational routines. It fills up the potential issues associated with the direction mechanism. According to March and Simon (1958), organizational routines touch on a mechanism for coordination independent of communicating the knowledge in the explicit form. It could depend on the number of activities by developing a fixed response to already defined responses or stimuli. In this process, the individual may create a particular pattern through which they may interact. This interaction allows the integration of their respective specialized knowledge without being converted to explicit knowledge. It certainly has an advantage over direction by having a great capacity to vary responses to a broader situation range. Besides, it could be more economical to apply in any organization or even in a network instead of documenting all the knowledge. This mechanism needs strong coordination, though.

As it is an informal procedure, it could not work in the absence of interaction among teams, commonly-developed roles, and training (Pentland and Rueter 1994). It might be a disadvantage for those companies who are not into an interactive environment and team-building process. Another problem is that there is a need for constant repetition and an approach that could be time-consuming in new collaborations. In the long term, this could help harness knowledge effectively, and integration could be more of a smooth procedure once firms could develop an interactive environment around (Grant, 1996).

According to Tushman and Nadler (1978), knowledge processing within a firm could be comprised of three components. They are knowledge gathering, knowledge interpretation, knowledge synthesis. Every component acts as a stage. At the first level, knowledge is collected through individuals and teams in an organization. This process of information gathering remains the same in the business networks context as well. The data could be collected from a given and desired organization where teams or individuals participate. This information is taken as an exploration. At this stage, given information does not have much meaning. At the second stage of knowledge interpretation, this knowledge starts getting used. At this final stage of knowledge synthesis, knowledge is exploited to benefit business ecosystems around knowledge hubs.

Knowledge interpretation is a process where information is processed, and it is valued. It is essential to know what information is valuable for the network and what information is not. Some of the data is discarded at this level. In the third stage of synthesis, knowledge is combined and integrated within networks to be used for future use. This information might be documented or stored in an information system.

As far as knowledge conceptualization is concerned, the knowledge transfer approach evolves in three ways. The first one is the traditional approach that assumes knowledge as distinct from practice. In this case, knowledge is taken as an object instead of a process. This object view states that knowledge is like a mental representation, which is then exhibited in terms of written words, representation, and routinized behavior (Nicolini, Gherardi, & Yanow, 2003). This view is interesting as it considers knowledge as a thing or object that acts like an asset with a significant company value. Like intellectual capital, every organization wants to enhance and grow it by focusing on creating, codifying, and capturing knowledge.

If knowledge transfer occurs at organizational levels, then the best approach is to create a network where participants could create a standard or feasible means of knowledge transfer. This approach is called the "syntactic" approach. It means it could be a common code, a language, a set of guidelines or procedures, or even a computer capability. For example, a standardized manual in a department could be viewed as the common mean of knowledge transfer and can ease the transfer of knowledge from an individual or firm. It is also useful in identifying issues that might slow down the process of information transfer. This approach is feasible if there is clarity about the problem statement and an agreement within the organization to deal with it (Weber & Khademian, 2008).

If, on the other hand, knowledge is not explicit and there is no actual identification system, then one could semantic view coined by Carlile (2002). This approach takes our problem statement from means for transferring information towards receipt of knowledge. It addresses the challenge that organizations might face in recognizing the role of interpretation while receiving and disseminating knowledge. The semantic approach acknowledges the differences within organizations at the individual or collective level. These differences could be at different levels, i.e., mid-level, managerial level, or executive levels. The nature of these differences may vary from experience, culture, language, and relationships among each other. It is essential to place these points of contrast and then organize a way out.

It is crucial to consider the relationship between knowledge and practice. It takes an approach called the "pragmatic view of knowledge" (Carlile, 2002; Weber & Khademian, 2008). According to this view, knowledge should be taken in the context of practice rather than a means of communication. It must be situated in a setting with geographic limits, a point in time, or a particular set of relationships. In simpler words, this knowledge comes into the evolving process by the experience of those who create or build this knowledge through practice.

Coming towards external and internal knowledge, there are three alternatives for knowledge transfer and integration as per Grant (1996). These are internalization within the firm, market contracts, and relational contracts. Market contracts are considered to be inefficient means for knowledge transfer having uncertainties overvaluation (Demsetz, 1991). These market contracts are helpful and are, in fact, efficient in the process of transferring knowledge when knowledge is layered within a product.

In the case of individual strategic alliances or broader networks, relational contracts are considered to be an immediate solution. If explicit knowledge could not be transferred efficiently through market contracts, then diffusion of its uncertainty over its applicability would not satisfy the internalization of this procedure within the firm. In such a scenario, networks (either individuals or firms) will be more suitable for transferring and integrating such knowledge (Grant, 1996).

While discussing the role of knowledge integration within firms, it is crucial to notice the speed with which such capabilities can be built and then extended. There is a danger that those relational contracts are not sufficiently efficient, and knowledge is not embodied within the product; those contracts might permit knowledge transfer relatively quickly. Hence, taking competitive advantage within the dynamic market setting, it is worth noticing that networks' critical merit would lie in giving the speed of access to new knowledge.

As it has been established that knowledge integration is one of the most essential and challenging aspects within a firm or a coordinating network. In recent times, most companies rely on technology as it could be useful to transfer knowledge among participating actors through information systems (Schau, Smith, & Schau, 2005). It is convenient to codify, communicate, assimilate, store and retrieve comparing the past scenarios. On the collaborative firm level, the main emphasis is on common interest, training, and background that participants of a network would use to facilitate the transfer and integration of knowledge (Weber & Khademian, 2008).

It is normal when actors within an ecosystem have a common focus, and then they would then share a common framework for understanding and utilizing the given information. However, Weber & Khademian (2008) argue that this is not the case in highly diverse settings. Knowledge integration is not that simple as information flowing through the network may have different uses, meanings, and even different values for groups or teams on the receiving end. That is why it is essential to distinguish between external and internal factors of a network. Certain aspects are overlapping, and they may lie in both internal and external states of an ecosystem.

Given the challenges we discussed regarding knowledge integration between firms, it plays a significant role in business ecosystems. Knowledge itself is considered an internal asset for any organization. All of its resources are connected and communicated through a well-organized knowledge management system. Then, there is external knowledge that comes from customers, suppliers, or even partners. It is of great value as this is the information that creates a high degree of innovation in the collaboration process. As different organizations have their own goals, they cannot share all the given knowledge. Therefore, firms need to agree on what type of knowledge can and should be transferred, processed, and utilized for knowledge integration.

Conceptual Framework

As we have discussed the significance of ecosystems and interaction of networks that may contain challenges and benefits of knowledge integration and resulting phenomena of innovation. We, therefore, propose a framework where ecosystems are composed of three components. They are business networks, knowledge integration, and innovation management. Value networks focus on the phenomena of creating value for customers. It revolves around the context of solving customer's problems (Christensen and Rosenbloom, 1995). A strategic management view suggests such networks as sources of competitive advantage for individual companies (Adner, 2012). Standard business practices indicate that companies are rivals to each other. They compete with each other for market share and try to gain a competitive edge over others. Simultaneously, it is essential to realize that competition can have different meanings in contemporary business settings. Firms tend to integrate competition and cooperation to produce more diverse value for the end customer. This practice is vital for competitors as they could rely on other firms to survive.

To understand the role and dimension of value networks in our framework, consider companies as diverse in their capabilities. However, they have come closer due to business collaboration. An organization may go through intensive development of its infrastructure first to be competitive. This infrastructure includes various activities such as finance, accounting, legal issues, deployment of sales and marketing units, recruitment at executive and mid-level, and maintenance of relations with other partners. Through this consistent process of infrastructure development, there could be a chance of building up links within different units and departments. Now consider this scenario at a higher level where more companies and each organization act like actors at individual levels. The same phenomena of organizational structure will be followed but at a higher level. In this scenario, other institutions such as research centers, universities, governmental organizations, and non-profit organizations may serve as an extra network interaction element. We have adopted Adner and Kapoor's (2010) graphical depiction of the corresponding business ecosystem in Figure 1.

--- Insert Figure 1 about here ---

Knowledge integration deals with creating new knowledge that could be shaped by signifying out the network nodes where this set of information is retained. The knowledge-based view focuses on generating new knowledge and technologies. One clear example of such knowledge-associated networks is open source communities. As far as the organization is concerned, irrespective of its size and market shares, every individual associated with the firm has a unique skill set. These skills are like an asset for the company, and every firm wishes to translate these skills into knowledge. These skills could be knowledge of information technology and specific tools, methods, or protocols for high-tech personnel. On the other side, personnel from human resource management might be experts in the team-building process. Some of these skills are not in written form, and a documented set cannot learn them. So, the problem is that one cannot codify these skills as they come along a hard way of individual focus, training, and experience. Hence, these skills remain like an open secret in an organization. Others could learn them through a process of keen observation. For companies working in the business ecosystem, each organization possesses a unique set of competence, and once unlocked, it can be a positive contribution to value networks.

Finally, the innovation management approach emphasizes knowledge integration (exploration) and fosters business ecosystems (exploitation) around knowledge hubs. According to (Tushman and Nadler, 1978), knowledge processing within a firm could be comprised of three components. They are knowledge gathering, knowledge interpretation, knowledge synthesis. Every component acts as a stage. At the first level, knowledge is collected through individuals and teams in an organization. This process of information gathering remains the same in the business networks context as well. The data could be collected from a given and desired organization where teams or individuals participate. This information is taken as an exploration. At this stage, given information does not have much meaning. At the second stage of knowledge interpretation, this knowledge starts getting used. At this final stage of knowledge synthesis, knowledge is exploited for the benefits of value networks. Silicon Valley could be an example of such networks. For such systems, financial networks that support main actors (companies, research centers, universities, tech developers) are considered key to success. The ecosystem view towards innovation management is described by Jacobides et al. (2018) in Figure 2.

--- Insert Figure 2 about here ---

It is interesting to view the role of actors involved with each of these components within the business ecosystem. The concept of the ecosystem could be view here as a whole. There is a different logic of action among all these categories. Each actor has a different interaction area between given component types and their relationships. Such actors are considered as dominant players (platform owners, cf. Gawer and Cusumano, 2002), and they play a crucial role in highlighting this interaction between ecosystems. These actors do overlap and interact between business, knowledge, and innovation elements, and hence, their activities are the essential part of value creation. Value creation is the outcome of these interactions among given networks in an ecosystem. Platforms, on the other hand, might be an interconnecting factor among these components as well. The platform can be an organization having complementary assets or technologies. Having these interconnecting actors and platforms, components within the ecosystem interact with each other and hence, evolve and emerge next to each other to provide value as a whole system.

The main benefit of the collaboration of these networks and interconnections is that it is established around value creation. Sometimes this value creation is linked to an immediate customer need. Consider, for example, the energy industry. There are smart grid devices through which even customers become part of value creation. They can give direct feedback to energy distributors, and hence, the whole ecosystem builds around value creation. There are some cases where this value creation does not come into use right at that time. Such value creation could be a form of innovation that may not directly be useful in one project but could be compatible with other future projects. For example, Apple's R&D teams were initially exploring tablets but ended up entering the phone market using the same multi-touch technology before launching tablet computing products.

Traditionally actors within a network operate in dimensions of the organization or the platform that is being used. However, suppose there is a shared platform (e.g., information system) in operation. In that case, such technological aspects and features will significantly impact the overall ecosystem and evolution of a given network (Thomas et al., 2014). Consider examples of the platform formed around dominant market players (Samsung, Apple, or even Nokia) though they are competitive.

For the knowledge integration process, the variety of complementary knowledge resources creates dependencies (Thompson 1967). This phenomenon could have both negative and positive aspects. As in the previous section, we discussed how knowledge integration could be challenging in a complex network of firms competing for more excellent customer value. There could be positive aspects for the organization as well. The firms within this network understand the significance of other firms and their capabilities. Sharing knowledge and collaborating through this phenomenon is the basis of the knowledge integration process. All these organizations have their motivation, yet they share a certain degree of knowledge to strengthen this bond.

The ecosystem consists of both providers and customers, and hence, benefits are for both parties from this collaboration. In innovation management, middle-level players or facilitators play the role of bridging actors of particular competence. It forms a platform within innovation ecosystems that is extremely helpful in interaction and building dependencies between organizations. A holistic view of the proposed conceptual framework is given n Figure 3.

--- Insert Figure 3 about here ---

Conclusions

To summarize, business ecosystems can be considered as global organizations that are not limited to specific geographic boundaries. This global dimension in ecosystems is interesting as it can improve the values of a product/ service for customers. On the other hand, there are certain challenges when it comes to knowledge management. Knowledge is no longer an organization's internal asset; it is an essential ingredient for networked organizational systems. Knowledge management is mostly viewed at the local level of the ecosystem, i.e., first, an organization develops knowledge integration within its units. Once benefits start showing up, they tend to share and expand that information. These meta-organizations play a vital role in creating a business ecosystem and developing innovation among the main actors of the ecosystem. It is beneficial though challenging at the same time to develop mutually organized ecosystems with a win-win attitude. With the wave of globalization and information technology, startups and SMEs play a significant part in modern business ecosystems. They also expect to get equal benefits as established corporates. Hence, ideal ecosystems provide equal opportunity for all actors within a given network to excel. A network of firms where market leaders would not play a challenging game as dominant players by setting all terms of collaboration and turning competition in their favor is considered an ideal ecosystem.

Overall, this chapter suggests that interactions and collaboration among networks are highly beneficial for suppliers and customers. However, these aspects can also create complex situations, and hence, these ecosystems must be analyzed at multiple levels. These levels could be in the hierarchy to understand the connection and information flow between distinct networks' components in contemporary business.

References

Adner, R., 2017. Ecosystem as structure: An actionable construct for strategy. Journal of Management, 43(1), pp.39-58.

Adner, R., 2012. The wide lens: A new strategy for innovation. Penguin Uk.

Adner, R, Kapoor, R, 2010. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. Strategic Management Journal 31, 306-333.

Kapoor, R. and Adner, R., 2007. Technology interdependence and the evolution of semiconductor lithography. Solid State Technology, 50(11), pp.51-55.

Adner, R., 2006. Match your innovation strategy to your innovation ecosystem—Harvard business review, 84(4), p.98.

Alexy O, George G, Salter AJ. 2013. Cui bono? The selective revealing of knowledge and its implications for innovative activity. Academy of Management Review 38: 270–291.

Agerfalk, P.l, Fitzgerald, B., 2008. Outsourcing to an unknown workforce: Exploring open sourcing as a global sourcing strategy. MIS Quarterly 32 (2), 385-409.

Ansari, S., Garud, R. and Kumaraswamy, A., 2016. The disruptor's dilemma: TiVo and the US television ecosystem. Strategic Management Journal, 37(9), pp.1829-1853.

Aarikka-Stenroos, L., & Ritala, P., 2017. Network management in the era of ecosystems: Systematic review and management framework. Industrial Marketing Management, 67, 23-36.

Bahrami, H., Evans, S., 1995. Flexible Re-Cycling and High-Technology Entrepreneurship. California Management Review 37 (3), 62-89.

Basole, RC., 2009. Visualization of interfirm relations in a converging mobile ecosystem. Journal of Information Technology 24, 144-159.

Basole, R. C., Russell, M. G., Huhtamäki, J., Rubens, N., Still, K., & Park, H., 2015. Understanding business ecosystem dynamics: A data-driven approach. ACM Transactions on Management Information Systems (TMIS), 6(2), 1-32.

Brusoni, S., and Prencipe, A. 2013. The organization of innovation in ecosystems: Problem framing, problem-solving, and patterns of coupling. In collaboration and competition in business ecosystems. Emerald Group Publishing Limited.

Carlile, P. R. 2002. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. Organization Science, 13(4), 442-455.

Christensen, C. M., & Rosenbloom, R. S. 1995. Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network. Research Policy, 24(2), 233-257.

Clarysse, B., Wright, M., & Bruneel, J. 2014. Creating Value in Ecosystems: Crossing the Chasm between Knowledge and Business Ecosystems. Research Policy, 43(7): 1164–1176.

Cosh, A., Hughes, A., 2010. Never mind the quality feel the width: University-industry links and government financial support for innovation in small high-technology businesses in the UK and the USA. Journal of Technology Transfer 35, 66-91.

Cusumano, M.A., Gawer, A., 2002. The elements of platform leadership. MIT Sloan Management Review Spring, 51-58.

Dalley, J. and Hamilton, B., 2000. Knowledge, context, and learning in the small business. International Small Business Journal, 18(3), pp.51-59.

Dattée, B., Alexy, O., & Autio, E. 2018. Maneuvering in poor visibility: How firms play the ecosystem game when uncertainty is high. Academy of Management Journal, 61(2), 466-498.

Desrochers, P., 2010. Industrial ecology and the rediscovery of inter-firm recycling linkages: Historical evidence and policy implications. Industrial and Corporate Change 11(5), 1031-1057.

Demsetz, H. 1991. The Theory of the Firm Revisited. in 0. E. Williamson and S. Winter (Eds.), The Nature of the Firm, New York: Oxford University Press, 159-178.

de Vasconcelos Gomes, L. A., Facin, A. L. F., Salerno, M. S., & Ikenami, R. K. 2018. Unpacking the innovation ecosystem construct Evolution, gaps, and trends. Technological Forecasting and Social Change, 136, 30-48.

Frels, lK., Shervani, T., Srivastava, RK., 2003. The Integrated Networks Model: Explaining Resource Allocations in Network Markets. Journal of Marketing 67, 29-45.

Frankort HTW. 2013. Open innovation norms and knowledge transfer in interfirm technology alliances: evidence from information technology, 1980-1999. Advances in Strategic Management 30: 239–282.

Gawer, A., & Cusumano, M. A. 2014. Industry platforms and ecosystem innovation. Journal of product innovation management, 31(3), 417-433.

Gawer, A., & Cusumano, M. A. 2002. Platform leadership: How Intel, Microsoft, and Cisco drive industry innovation (pp. 29-30). Boston: Harvard Business School Press.

Gulati, R., Puranam, P., & Tushman, M. 2012. Meta organization design: Rethinking design in inter-organizational and community contexts. Strategic Management Journal, 33(6), 571-586.

Grant, R. M. 1996. Prospering in dynamically competitive environments: Organizational capability as knowledge integration. Organization Science, 7(4), 375-387.

Garnsey, E., Leong, Y.Y., 2008. Combining resource-based and evolutionary theory to explain the genesis of bionetworks. Industry and Innovation 15 (6), 669-686.

Garnsey, E., Lorenzoni, G. and Ferriani, S., 2008. Speciation through entrepreneurial spin-off: The Acorn-ARM story. Research Policy, 37(2), pp.210-224.

Habbershon, T.G., 2006. Commentary: A framework for managing the familiness and agency advantages in family firms. Entrepreneurship theory and practice, 30(6), pp.879-886.

Hagel, J., & Brown, J. S. 2005. The only sustainable edge: Why business strategy depends on productive friction and dynamic specialization. Harvard Business Press.

Hecker, A. 2012. Knowledge beyond the individual? Making sense of a notion of collective knowledge in organization theory. Organization Studies, 33(3), 423-445.

Iansiti, M. and Levien, R., 2004. Strategy as ecology. Harvard business review, 82(3), pp.68-78.

Ilinitch, A. Y., D'Aveni, R. A., & Lewin, A. Y. 1996. New organizational forms and strategies for managing in hypercompetitive environments. Organization Science, 7(3), 211-220.

Jacobides, M. G., Cennamo, C., & Gawer, A. 2018. Towards a theory of ecosystems. Strategic Management Journal, 39(8), 2255-2276.

Javalgi, RG., Todd, P.R, Scherer, RF., 2005. The dynamics of global e-commerce: An organizational ecology perspective. International Marketing Review 22 (4), 420-435.

Järvi, H., Kähkönen, A. K., & Torvinen, H. 2018. When value co-creation fails: Reasons that lead to value co-destruction. Scandinavian Journal of Management, 34(1), 63-77.

Iansiti, M., & Levien, R. 2004. The keystone advantage: What the new dynamics of business ecosystems mean for strategy, innovation, and sustainability. Harvard Business School Press. Boston.

Isckia, T., 2009. Amazon's evolving ecosystem: A cyber-bookstore and Application Service Provider. Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 26(4), pp.332-343.

Iyer B, Lee CH, Venkatraman N. 2006. Managing in a small world ecosystem: some lessons from the software sector. California Management Review 48(3): 28-47.

Iyer, B. and Davenport, T.H., 2008. Reverse engineering Google's innovation machine. Harvard Business Review, 86(4), pp.58-68.

Kapoor, R. and Adner, R., 2007. Technology interdependence and the evolution of semiconductor lithography. Solid State Technology, 50(11), pp.51-55.

Kapoor, R., & Lee, J. M. 2013. Coordinating and competing in ecosystems: How organizational forms shape new technology investments. Strategic management journal, 34(3), 274-296.

Kikoski, C.K. and Kikoski, J.F., 2004. The inquiring organization: Tacit knowledge, conversation, and knowledge creation: Skills for 21st-century organizations. Greenwood Publishing Group.

Li, Y.R., 2009. The technological roadmap of Cisco's business ecosystem. Technovation, 29(5), pp.379-386.

Lusch, RF., 2010. Reframing Supply Chain Management: A Service-Dominant Logic Perspective. Journal of Supply Chain Management 47 (1), 14-18.

Mäkinen, S.J. and Dedehayir, O., 2012, June. Business ecosystem evolution and strategic considerations: A literature review. In 2012 18th International ICE Conference on Engineering, Technology, and Innovation (pp. 1-10). IEEE.

March, J.G. and Simon, H.A., 1958. Organizations John Wiley & Sons. New York.

McCarthy, I., Ridgway, K., Leseure, M. and Fieller, N., 2000. Organizational diversity, evolution, and cladistic classifications. Omega, 28(1), pp.77-95.

Meyer, A.D., Gaba, V., Colwell, K.A., 2005. Organizing Far from Equilibrium: Nonlinear Change in Organizational Fields. Organization Science 16 (5), 456-473.

Moore, J.F., 1993. Predators and prey: a new ecology of competition. Harvard business review, 71(3), pp.75-86.

Nehf, lP., 2007. Shopping for privacy on the internet. The Journal of Consumer Affairs 41 (2), 351-375.

Nicolini, Davide, Silvia Gherardi, and Dvora Yanow, eds. 2003. Knowing in Organizations: A Practice-Based Approach. Armonk, NY: M. E. Sharpe.

Nonaka, I., Toyama, R., & Konno, N. 2000. SECI, Ba, and leadership: a unified model of dynamic knowledge creation. Long-range planning, 33(1), 5-34.

Pentland, B.T., 1992. Organizing moves in software support hot lines. Administrative Science Quarterly, pp.527-548.

Pentland, B. T., & Rueter, H. H. 1994. Organizational routines as grammars of action. Administrative Science Quarterly, 484-510.

Pierce, L., 2009. Big losses in ecosystem niches: How core firm decisions drive complementary product shakeouts. Strategic management journal, 30(3), pp.323-347.

Ponomarov, S. Y., Holcomb, M.e., 2009. Understanding the concept of supply chain resilience. The International Journal of Logistics Management 20 (1), 124-143.

Polanyi, M., 1966. The tacit dimension. New York: Anchor Day. The Tacit Dimension1966.

Polanyi, M.1969.The logic of tacit inference. In Grene, M.(Ed.), Knowing and Being, Routledge & Keagan Paul, London.

Porter ME. 1980. Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press: New York.

Rohrbeck, R, H6lzle, K., Gemilnden, HG., 2009. Opening up for competitive advantage: How Deutsche Telekom creates an open innovation ecosystem. R&D Management 39 (4), 420-430.

Ritala, P., Agouridas, V., Assimakopoulos, D., & Gies, O. 2013. Value creation and capture mechanisms in innovation ecosystems: a comparative case study. International Journal of Technology Management, 63(3-4), 244-267.

Samila, S. and Sorenson, O., 2010. Venture capital as a catalyst to commercialization. Research Policy, 39(10), pp.1348-1360.

Santos, F.M., Eisenhardt, K.M., 2005. Organizational Boundaries and Theories of Organization. Organization Science 16 (5), 491-508.

Sharma, S. and Henriques, I., 2005. Stakeholder influences on sustainability practices in the Canadian forest products industry. Strategic management journal, 26(2), pp.159-180.

Schau, H. J., M. F. Smith, and P. I. Schau. 2005. The Healthcare Network Economy: The Role of Internet Information Transfer and Implications for Pricing. Ind

Sugai, P., 2005. Mapping the mind of the mobile consumer across borders. International Marketing Review 22 (6), 641-657.

Seidler-de Alwis, R. and Hartmann, E., 2008. The use of tacit knowledge within innovative companies: knowledge management in innovative enterprises. Journal of knowledge Management.

Tushman, M. L., & Nadler, D. A. 1978. Information processing as an integrating concept in organizational design. Academy of management review, 3(3), 613-624.

Tassey, G., 2010. Rationales and mechanisms for revitalizing US manufacturing R&D strategies. The Journal of Technology Transfer, 35(3), pp.283-333.

Teece, D.J., 2007. Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. Strategic management journal, 28(13), pp.1319-1350.

Tiwana, A., Konsynski, B., Bush, A.A., 2010. Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. Information Systems Research 21 (4), 675-687.

Tsatsou, P., Elaluf-Calderwood, S. and Liebenau, J., 2010. Towards a taxonomy for regulatory issues in a digital business ecosystem in the EU. Journal of Information Technology, 25(3), pp.288-307.

Tsujimoto, M., Kajikawa, Y., Tomita, J., & Matsumoto, Y. 2018. A review of the ecosystem concept—Towards coherent ecosystem design. Technological Forecasting and Social Change, 136, 49-58.

Thompson, J. D. 1967. Organizations in action: Social science bases of administrative theory. Transaction publishers.

Thomas, L. D. W., Autio, E., & Gann, D. M. 2014. Architectural Leverage: Putting Platforms in Context. Academy of Management Perspectives, 28(2): 198–219.

Vidgen, R. and Wang, X., 2006. From business process management to business process ecosystem. Journal of Information Technology, 21(4), pp.262-271.

Volberda, H. W. 1996. Toward the flexible form: How to remain vital in hypercompetitive environments. Organization Science, 7(4), 359-374.

Watanabe, C., Kondo, R., Ouchi, N. and Wei, H., 2004. A substitution orbit model of competitive innovations. Technological Forecasting and Social Change, 71(4), pp.365-390.

Wareham, J., Fox, P. B., & Cano Giner, J. L. 2014. Technology ecosystem governance. Organization Science, 25(4), 1195-1215.

West, J. and Wood, D., 2014. Evolving an open ecosystem: The rise and fall of the Symbian platform. In Collaboration and competition in business ecosystems. Emerald Group Publishing Limited.

Weber, E. P., & Khademian, A. M. 2008. Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. Public administration review, 68(2), 334-349.

Whitley, E.A. and Darking, M., 2009. Object lessons and invisible technologies. In Bricolage, Care and Information (pp. 348-366). Palgrave Macmillan, London.

Williamson, P. J., & De Meyer, A. 2012. Ecosystem advantage: How to successfully harness the power of partners. California management review, 55(1), 24-46.

Zacharakis, A.L., Shepherd, D.A. and Coombs, J.E., 2003. The development of venture-capital-backed internet companies: An ecosystem perspective. Journal of Business Venturing, 18(2), pp.217-231.

List of Tables and Figures

Figure 1: A schema of an ecosystem from the view of business management (Adner & Kapoor, 2010)

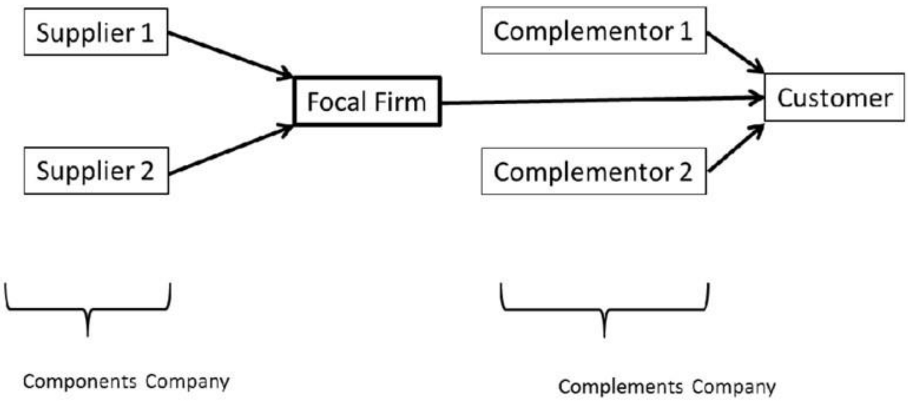


Figure 2: Ecosystem view of innovation management (Jacobides et al., 2018)

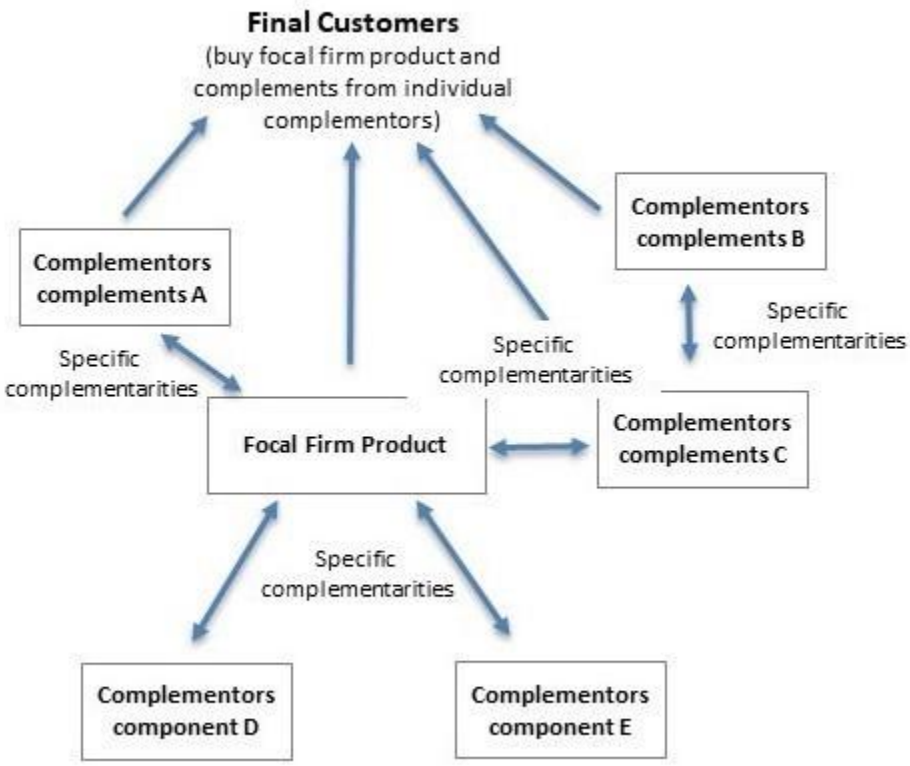
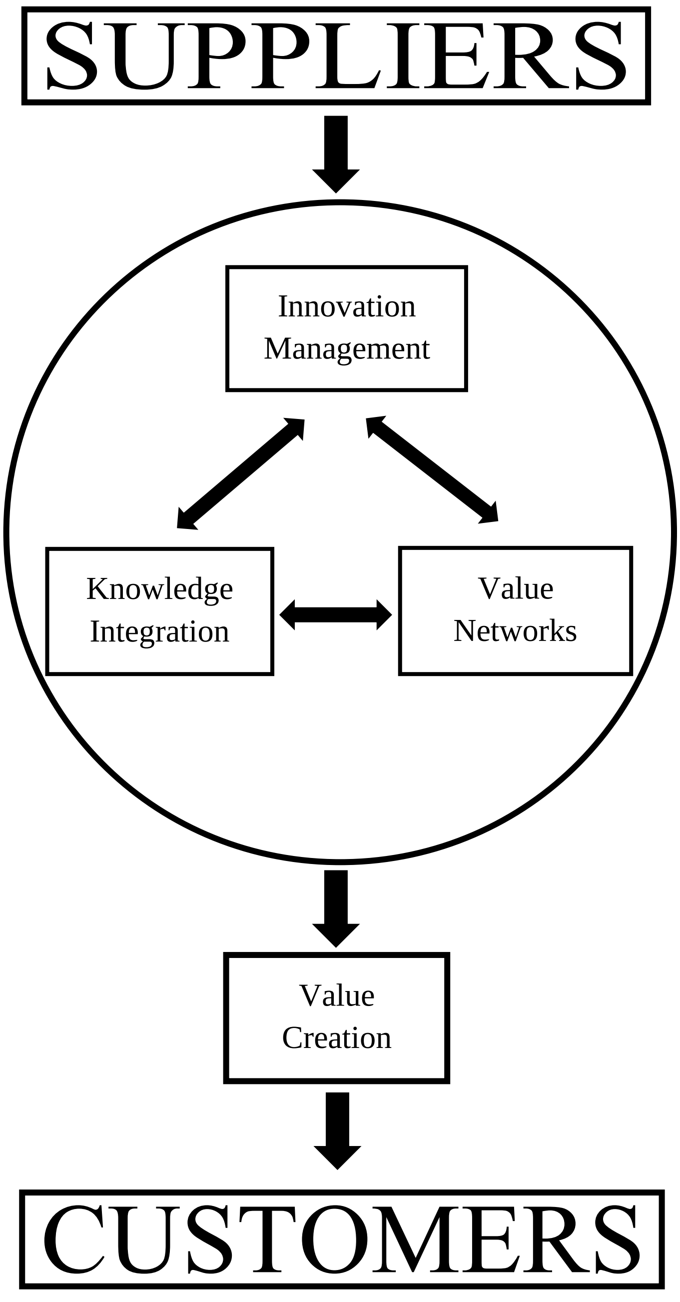


Figure 3: A Conceptual framework for business ecosystem



Chapter 2

Exploring Emerging Technological Trends in Automotive Ecosystem: A Patent Analysis

Abstract

The patents are becoming a source of critical information related to various technologies. This trend has led organizations to consider patent analysis as an essential element in their analysis and as a methodology for research and development. Indeed, patent-analysis could greatly help understand technological development, which is difficult to study otherwise, given its complexities. Exploring technological trends is vital for any effective technology policy, given the strategic significance of related opportunities and technological growth challenges. While previous research scholarship primarily relies on firm-level analysis, the value of patent-based approaches for exploring technological trends for an overall ecosystem has been underestimated. In this paper, we apply patent-analysis to study the technological development in the automotive ecosystem rather than on a firm-level.

To resolve this void in the literature, this study applies the topic modeling method using patents related to the automotive ecosystem registered at Worldwide Patent Statistical Database EPO's PATSTAT. We use the automotive ecosystem as a case study. The suggested approach is a generative model based on latent Dirichlet allocation (LDA), allowing identifying core topics related to the automotive sector. We further investigate the topics related to the automotive ecosystem on both technology-level and firm-level. The proposed approach offers implications both for academia and managers for the use of patents to explore technology development trends. This technological change occurs both inside and outside the organizational boundary creating an ecosystem. This diverse ecosystem consists of a focal firm, suppliers, complemetors such as information and communication service providers, and key actors related to smart charging infrastructure. In addition, we believe that our research on technical topics and their developments will significantly understand the technological environment in a holistic ecosystem rather being limited to a specific sector.

Keywords: Innovation ecosystems, technological trajectory, technology trend exploration, patent analysis, natural language processing, topic modeling, electric vehicles.

Introduction

Sustainability in the automotive ecosystem has become a strategic challenge and a gateway to businesses developing competitive advantages along with environmental issues and the adoption of developed technologies (Miller et al., 2014; Seitz et al., 2006; Vaz et al., 2017; Zanchi et al., 2018). One of the critical factors making sustainability a differentiation source is technology (Bernal-Conesa et al., 2017). With recent developments in the automotive sector, this role is becoming more relevant and diverse (Lin et al., 2019; Yun et al., 2018). It is established that the proper implementation of technology can improve fundamental sustainability functions (Planko et al., 2016; Son et al., 2016), as well as create new business models for its stakeholders (Aguilar-Fernández & Otegi-Olaso, 2018; Cillo et al., 2019; Mont et al., 2019).

Automotive companies are immersed in the innovation ecosystem. The ecosystem is defined as "the network of organizations including suppliers, distributors, customers, competitors, and government agencies involved in the delivery of a specific product or service for which the effect of technological introduction/improvements are not limited to an industry" (Moore, 1993). For all its main stakeholders, such an ecosystem adds value and eliminates the volatility inherent in dynamic value chains and rapidly evolving markets (Iansiti & Levien, 2004a). Growing and varying technological advancement choices can also give companies in the automotive ecosystem more opportunities to use advanced technology for their business (Borgstedt et al., 2016). The use of ecosystem-related technical capabilities has therefore become one of the best guarantees of sustainability in rapidly evolving competitive conditions and enables businesses to achieve sustainable competitiveness (Crabb & Johnson, 2010; Van den Hoed, 2007).

Simultaneously, these ecosystem-related technological changes posit automotive firms with significant challenges to overcome (Athanasopoulou et al., 2016; Pierce, 2009; Johansson & Deniz, 2014, Riasanow et al., 2017). Such technologies transform value chains, thereby expanding and blurring traditional organizational boundaries (Huber & Puto, 1983). Incumbent firms need to deal with fresh, unpredictable challengers, often prepared with the latest technical skills and plausible business models (Abdelkafi et al., 2013; Wells, 2013). The primary ecosystem stakeholders are not merely technology adopters, unlike conventional networks and alliances; instead, they are more likely to build innovations and establish sustainable competitive advantage (Iansiti & Levien, 2004a; Moore, 1993). As a result of this trend, firms might face greater uncertainty, especially those lacking experience of using technology for innovation (Teece, 2007).

Recognizing and realizing the current state of technology growth in this situation becomes a necessary prerequisite for considering the possibilities and avoiding the threats of technology (Basberg, 1984; Teichert & Mittermayer, 2002). Organizations are also highly involved in exploring external technology and controlling their ecosystem's internal technologies. Based on the expert's experience and insights, an expert-driven strategy is a traditional technology exploration technique (Lichtenthaler, 2004). As the technical reach and topics to be discussed have become broader and have continued to evolve, this strategy is becoming time-consuming, costly, and invalid. Therefore, organizations need a more organized way of reducing the strain of researching technical trends and collecting quality knowledge that businesses need to facilitate their creativity and tactical planning (Porter, & Cunningham, 2005).

As for the response, a large part of the literature on technology management has tried to establish methods and processes to promote the discovery and monitoring of technology trends. Significant advances have been made with the growing volume of technology-related data, such as research papers and patent applications, based on extracting valuable knowledge from such a wealth of information (Kerr et al., 2006). Scientific publications and patent documents are unstructured text data. For this purpose, text mining has become a commonly used tool for the complete or semi-automated analysis of textual data (Hashimi et al., 2015).

Patents have been generally taken as a proxy for technology exploration with the following benefits. First of all, proprietary inventions are worth analyzing because the regulatory process will ensure that they are novel and valuable (Worldwide Patents Database: PATSTAT, 2018). Since the patenting process is expensive and may take years, the patented inventions are likely to have high technical and economic importance (Basberg, 1987). Patent applications also contain bibliographic details such as dates of filing and issuance, assignees, and prior art, allowing researchers to analyze technology characteristics from different viewpoints, such as time, holders, and technology relationships. Patent papers can sometimes be viewed freely (Daim et al., 2006), and online patent databases are accessible these days. Patent data may also resolve the challenge of accessing objective technical development knowledge outside the confinement of the organization. Therefore, a suitable analysis of patent data will allow organizations and their networks to gather information on technological opportunities and threats(Porter & Detampel, 1995).

However, recent years have seen a significant increase in studies that investigate technological trajectories using patent data. These empirical studies investigated the adoption of technology (David & Olsen, 1992) and its effect on the performance of foal firms in the automotive sector (Popp, 2005; Huang et al., 2010). Few studies have taken a holistic methodological approach to investigate overall ecosystems' technological competencies (Adner & Kapoor, 2010; Adner & Kapoor, 2016; Jacobides et al., 2015). Using patent analysis to analyze the relationship between individual companies and the business networks surrounding them, this research impediment starts to take a step in maturing business ecosystems.

This research introduces a topic-based modeling approach to patent analysis that examines technological developments in the sustainable automotive ecosystem. Instead of processing metadata, such as patent classification codes (CPC/IPC), Latent Dirichlet allocation (LDA) is used to remove hidden topics behind patent records. In terms of patenting activity patterns and significant assignments for each issue, the topics defined by LDA are further investigated. This different initiative allows us to consider the automotive ecosystem's technological environment, both at the technical level and at the business level. The suggested approach is expected to offer patents to explore technology trends in a sustainable automotive ecosystem for the focal firm, suppliers, and complementary firms' perspective. The research question to be resolved by this study is:

How does technological trajectory evolve among key players within the automotive ecosystem?

The technical environment in the automotive ecosystem can be further explored on the basis of its creative competence at both technological and organizational levels.

Theoretical background

We will start this section by explaining the theoretical concept of ecosystem literature. We shall provide reasoning for using patent documents for technology exploration and how the ecosystem could be incorporated with the technological inter-dependence of key actors. Later, the topic modeling method is explained to find the most relevant topic in the automotive ecosystem. We will use the Latent Dirichlet Allocation algorithm and provide relevance of methodology with our ecosystem's theoretical framework for such analysis.

Literature in Ecosystem

For scholars, the analogy has long been a source of enlightenment to explain the phenomenon they are investigating. A biological ecosystem model has started to be adopted by business analysts to evaluate business relationships and strategic decision-making (Iansiti and Levien, 2004a). It is interesting to study used analogy of ecosystems (Moore, 1993) in business literature to study affiliated organizations, understand the dynamics resulting from it, and explain the observed phenomena. The importance of the dynamic network of business relationships inside and across industries is being recognized by managers and scholars. (Harte et al., 2001).

As for definition, "The business ecosystem describes the network of firms, which collectively produce a holistic, integrated technological system that creates value for customers" (Moore, 1993). According to Basole (2009), these businesses "co-evolve innovation capabilities" by working cooperatively as well as competitively to develop products and services (Lusch, 2010; Teece, 2007; Agerfalk and Fitzgerald, 2008).

Iansiti and Levien (2004b) extend Moore's concepts by defining the role of different actors in such networks and the relationship of these roles with their ecosystem's collective resources. These roles are described as the keystone, dominator, and niche player in the business ecosystem domain. Iansiti and Levien (2004a) show that business networks are rarely homogenous, and Some members have different and unequal positions. Iansiti and Levien (2004a; 2004b) define, “As a loosely coupled system, the network needs just interoperability and extensibility based on only adequate interaction satisfaction and protocol leverage. They understand that today's increasingly dispersed and networked industry structure is a relatively recent development that needs a new framework of thinking about health in the industry and what constitutes an industry in the first place”. Several different entities are involved in providing the customer with a commodity, which makes them share a shared fate that could be related to the product's fate.

From an ecosystem as perspective, researchers have stressed that various aspects of an ecosystem depend on the unit of analysis. In reviewing the literature, three broad groups of research domains have been identified: a “business ecosystem” stream, which centers on a firm and its environment; a “platform ecosystem” stream, which considers how actors organize around a platform and an “innovation ecosystem” stream, focused around a particular innovation or new value proposition and the constellation of actors that support it (Jacobites et al., 2018).

The first stream focuses on an individual business and considers the environment as a “community of organizations, institutions, and individuals impacting the firm's customers and supplies” (Teece, 2007). In such a scenario, the ecosystem is conceived as an economic construct of interacting actors that, through their activities, all influence each other, considering all related actors outside the borders of a single industry. According to Teece (2007), the ecosystem is the environment to be monitored and responded to by the organization, affecting its complex capacities and its capacity to establish a sustainable competitive advantage. (Jacobites et al., 2018).

The platform ecosystem stream focuses on a specific class of technologies platforms and the interdependence between platform sponsors and their complementary components (Parker et al., 2016). In this view, the ecosystem includes the sponsor of the platform among all vendors of the complementary portion of the ecosystem (Ceccagnoli et al., 2012; Gawer & Cusumano, 2008) that make the platform more appealing to customers. Complementors can not only produce complementary innovation by connecting to the platform but also reach the customers of the platform, in a direct or indirect form, as in the examples of independent software vendors affiliated to SAP (Ceccagnoli et al., 2012) or developers making video games for particular consoles (Cennamo & Santaló, 2013).

The third line of research focuses on focal innovation and (upstream) elements and (downstream) complements that facilitate it (Kapoor & Lee, 2013). It describes the ecosystem as "the collaborative arrangements by which companies integrate their products into a coherent, customer-facing solution" (Adner, 2006). The focus is on understanding how interdependent players collaborate to build and sell technologies that support the end-user (Jacobides et al., 2018).

It is essential to notice that the innovation method enables clients to use the end product rather than the focal firm. Therefore, the definition of the ecosystem seeks to capture the relation between a core product, its components, and its complementary products/services, i.e., complement. For users, these complementary products/services collectively bring value. Companies that generate focal novation might or might not collaborate to complementor (Adner & Kapoor, 2010). The degree to which organizations align with various agreements can impact their ability to generate value for the end customer (Adner, 2017). The ecosystem casts a net around an interdependent network that offers focal and complementary innovations (Brusoni & Prencipe, 2013). Ecosystem research has investigated how the link between the innovator and its supplements influences the ability of both organizations to coordinate investments in new technology and marketing (Leten et al., 2013). A key factor involved is sharing information among key actors that affect relationships within the business ecosystem(Frankort, 2013). Such collective agreements impact the ecosystem's wellbeing and survival (Leten et al., 2013).

The underlying mechanism in an ecosystem is rendered by Peltoniemimi to understand the conceptual framework (2004), yet the basis for such a theoretical framework is at the base level creating a simulation model for network collaboration. Several empirical studies have also been performed based on the developed framework by Den Hartigh and Van Asseldonk (2004) and Quaadgras (2005). Such studies take innovation as a focus of research and use network approaches to examine ecosystems, such as the effect of network structure on the performance of the organization and network. Scholars like Moore (2006), Foer (2006), and Gundlach (2006) would later address a different focus of the study by using this idea to discuss problems in antitrust cases. Den Hartigh et al. (2006) uses network theory in the Dutch IT ecosystem to establish ecosystem health measures and empirically evaluates added value. However, these studies neglect how innovation is explored among the network of companies working in an ecosystem with diverse skills. Other studies explored the ecosystem stream further.

Quaadgras (2005) uses network theory to empirically understand the actions of big, diverse companies about network joining based on the absorptive power and exploration/exploitation model. In order to evaluate firm output in the network, this model can also be further developed to some degree. Firm efficiency, however, does not reflect how the environment is connected to technology or even goods. Jimenez (2007), who uses network theory to analyze network health in a less complicated manner, has made an advance at the conceptual level. As described by Iansiti and Levien (2004b), he approaches two of the three metrics, i.e., efficiency and robustness, by using a structural approach in which the emphasis of the study is on attributes of relations between actors. His research, however, did not touch on the third measure described by Iansiti and Levien (2002), i.e., the formation of niche actors within the ecosystem.

Although Moore (2006) believes that the idea of the ecosystem related to business dimensions could solve the shortcomings of previous structures such as strategic alliances and virtual organizations (Moore, 1996), it is vital to resolve some critical issues. For the innovation research community, the competence-dependent network approach offered by the innovation ecosystem (Adner and Kapoor, 2010) is essential for technology exploration. There has been a shortage of analytical instruments that provide practitioners with value (Adamovicius et al., 2006). To address this limitation to empirical study in the ecosystem, we use patents for technology exploration in the automotive ecosystem.

Patent Analysis for investing technological trajectory

For technology analysis within the ecosystem construct, we need more empirical studies to understand the practical implications of the ecosystem phenomenon to analyze the value created by an ecosystem. Adner and Kapoor (2010) use firm performance as the target variable. This measurement is valuable, yet it fails to capture technology exploration among connected actors within the ecosystem. We consider the ecosystem framework presented by Adner and Kapoor (2010) and apply patent analysis in the automotive ecosystem context.

Patents are considered as a standard proxy for measuring innovation (Schmookler, 1966). The interpretation of patent statistics as a measure of inventive operation has a long history (Scherer, 1965). It has become the most common way of measuring the creative output of businesses, industries, and even nations (Aghion et al., 2016). The literature (Aghion et al., 2016; Braun et al., 2011; Griliches, 1990) discusses many general benefits of using patent data for scientific work. First, in a similar technical area, there is the possibility to differentiate inventions (Duch-Brown & Costa-Campi, 2015). Other data, e.g., R&D investments, are not available on such a disaggregated level (ICEV, HEV, BEV, and FCEV[4]). Second, this study aims at exploring an innovation ecosystem that specializes in one of the four technologies. These specialized firms, such as data aggregators, are usually not in the focus of management research. Patent data makes an exploratory approach as each player in an ecosystem with at least one patent in one of the technologies, regardless of company size, market, and other variables, are part of the sample. Third, due to the availability of annual numbers, patent data allows a time-series study (Basberg, 1987). As a consequence, for each year during the measurement period, patterns in each technology advancement can be identified.

[4] ICEV=Internal Combustion Engine Vehicles, BEV= Battery Electric Vehicles, FCEV= Fuel Cells Electric Vehicles, HEV=Hybrid Electric Vehicles

Several items in a patent document may be split into structured and unstructured data, depending on semantics and format (Tseng et al., 2007). Depending on the types of data, the analytical methods do differ (Ernst, 1997). Structured data, on the one hand, includes objects that are standardized in both semantics and formats. Patent numbers, filing dates, registration dates, assignees, patent classification codes, and citations are representative examples. For decades, research has focused on structured data analysis (Archibugi & Planta, 1996; Ernst, 1997; Tseng et al., 2007). Bibliometric analysis has been a common method for analyzing structured datasets (Narin, 1994). Depending on the topic of a technology inquiry, the basic approaches are different. The statistical analysis of bibliometric knowledge to determine technical innovations (Narin et al., 1987) and competitiveness are some representative examples, Study by co-assignees to consider technical cooperation (Lei et al., 2013), co-classification analysis for exposing technological convergence patterns (Choi et al., 2015; Yun et al., 2016), and examination of citations to recognize significant patents or areas of technology (Cho & Shih, 2011).

In patent data, data in the form of texts with variable lengths and diverse contexts are classified as unstructured data. This data includes the title, abstracts, claims, and descriptions of patent documents' inventions. Since the methods and algorithms for dealing with textual data, such as text mining and natural language processing, have been developed, comprehensive technological innovation research has used textual data in patent documents (Hyun et al., 2020; Lei et al., 2019; Lupu, 2017; Trappey et al., 2020). Although patent texts for technology exploration are not used in a particular way, a common task is to define technology topics from text data and examine their characteristics. The particular approaches for defining technology topics differ from the clustering of related patent documents based on keywords. (Kim et al., 2016; Yoon & Park, 2004) to the topic modeling methods such as latent Dirichlet allocation (Lee & Sohn, 2017; Chen et al., 2017; Jeong & Yoon, 2017; Lee et al., 2005). By combining other methods like bibliometric analysis (Basberg, 1987; Lei et al., 2013), network analysis (Yoon & Park, 2004), and topic modeling method (Blei, Ng, & Jordan, 2003), more beneficial insights can be gained.

To summarize, the advantages of patents as a data source for innovation research and technology management and organized data analytics approaches such as bibliometric analysis have become the standard for technology exploration. There is some recent research with unstructured data in different contexts such as health care (Ghosh & Guha, 2013), logistics

(Choi & Song, 2018), energy (Benites-Lazaro et al., 2018); however, these studies are only focused on focal firm or supply chain. There is a lack of content analyses in the field of ecosystems. Our research would focus on the innovation ecosystem, including focal firms, suppliers, and complementors (Adner and Kapoor, 2010), and would investigate how key technologies can be understood among each actor.

Topic modeling approach in identifying patent topics.

Topic models are a series of mathematical algorithms that expose the key themes of a comprehensive collection of unstructured documents known as the corpus (Blei, 2012). They are considered ‘‘generative’’ models (Blei, Ng, & Jordan, 2003; Zhang, 2012) as they assume a specific probabilistic document generation process exists. Topic models also assume that;

1) Each document is a combination of topics

2) Each topic has its probability distribution over words.

The aim of the topic modeling algorithm is therefore to estimate the parameters of this process of probabilistic document generation, i.e., the distribution of the topic per document and the distribution of words per topic, by observing the words used in actual records in the corpus (Blei, Ng, & Jordan, 2003).

From both algorithmic and functional viewpoints, topic models have many strengths. First of all, they have mathematical foundations that help us understand the document generation process. Second, without the assistance of human experts, they do not need any prior marking of documents to be examined (Blei et al., 2003). Finally, they are immediately able to arrange and summarize documents (Steyvers & Griffiths, 2004). Because of these advantages, topic models have recently gained considerable attention and have been successfully applied to a wide variety of tasks in text mining (Amado et al., 2018; Isoaho et al., 2019; Shi et al., 2018; Yan, 2014; Zha & Li).

LDA, introduced by Blei et al. (2003), is the most commonly used algorithm for topic modeling and has some functional advantages over other algorithms for topic modeling. The mixture of existing topics can be calculated in new documents without updating the current model. In addition, since it has a fixed number of parameters, regardless of the size of the corpus, it can accommodate high volumes of documents (Blei, 2012). LDA can therefore avoid overfitting, allowing it to be easily applied to tasks of text mining in which large amounts of text documents are continually produced.

The fundamental concept of LDA is that documents are expressed over latent topics as random mixtures, each of which is described by a distribution over phrases. Figure 1 demonstrates the method of LDA document generation with an illustrative example. In Figure 1, $\omega_{d,i}$ is the *i*th word in the d_{th} document, and $z_{d,\ i}$ is the topic assigned to the phrase $\omega_{d,\ i}$. θ_d is the topic proportions for the d_{th} document, whereas φ_k is the word distribution for the *k*th topic. α and β are the Dirichlet hyperparameters for θ_d and φ_k, respectively. N, D, and K denote the number of words in a document, the number of documents in a corpus, and the number of topics across the corpus, respectively. Nodes are random variables, and edges indicate dependence. Furthermore, only the shaded node is visible, while the others are concealed or latent. Finally, replicated variables are shown by plates.

In LDA, each word in a document is created through the procedure described. First, for the entire corpus, per-topic word distributions, i.e., the frequency of use of each word within a topic, are calculated. Then, for each document, a per-document topic proportion, which indicates each topic's prevalence in the target document, is determined. Finally, by taking into account the assigned topic and its distribution through words, each word is chosen.

LDA training is equivalent to inferring the latent variables, i.e., the per-topic word distributions φ_k, per-document topic proportions θ_d, and per-word topic assignment $z_{d,\ i.}$ A well-trained LDA model is one where a document's selected terms are very close to the words that appear in that document. Therefore, the LDA parameters are tuned on the basis of each document's observed terms to increase the probability of individual word appearances predicted by the model.

--- Insert Figure 1 about here ---

Once the LDA inference is completed for *K* topics, a *D* by *K* per-document topic proportion matrix θ and a V by *K* per-topic word distribution matrix φ is obtained as shown in Table1a and Table 2b. Note that each row's sum is 1 for θ, whereas the sum of each column is 1 for φ. In addition, note that if a one-dimensional row vector is averaged, it can be understood as the distribution of the corpus-level topic (Yan,2014). As per Steyvers & Griffiths (2004), subsequent text analysis tasks can be performed with these inferred distributions; φ can be used to decrease the dimensionality for document categorization or classification, while φ can be used to exploit the relationship between latent topics, track the shift in topic distributions over time, and identify topics that are increasing and falling in popularity.

--- Insert Table 1a about here ---

--- Insert Table 1b about here ---

Recently, many attempts have been made to use LDA to classify research topics in the related academic field. Some of its research development has been discussed in the next section.

LDA (Latent Dirichlet allocation) model for technology identification

In light of the rising amount of text data and advanced computational research, generative modeling methods, such as topic modeling, have gained prominence by automatically discovering latent patterns in documents. Individual algorithms vary from each other, but a group of keywords that often occur together is usually believed to contain topics (Blei, Ng & Jordan, 2003). Texts with different topics, in other words, may use different vocabulary (Shi et al., 2018). Topic models are built on the basis of this premise as probabilistic models that aim to find a set of words from the collection of documents that better describe the themes or topics in the text document (Steyvers & Griffiths, 2004). Topic models enable us to classify related documents through the use of words (i.e., topics) by finding latent associations between words and topics across the range of documents and to annotate documents according to these topics (Alghamdi & Alfalqi, 2015). Therefore, the topic-based approach has offered a convenient way to organize, understand and summarize a complex array of textual knowledge in different areas (Blei, 2012).

Among several topic modeling methods, Latent Dirichlet allocation (LDA) is a Bayesian topic model and has recently been widely used (Blei, Ng & Jordan, 2003; Blei, 2012). LDA is a technique that assumes that writing is a result of word choice (Steyvers & Griffiths, 2004; Blei, 2012). Because of this premise, LDA models display documents with k topics per document as a mixture distribution and each record/document as a multinomial distribution over phrases (Lee and Kang, 2018).

The issue with defining the topics of documents is that we can only examine documents (i.e., the keywords contained in those documents). The composition of the topic is hidden, such as topics on their own, topic distributions on a document, term distributions for each topic, and topic allocations for each word (Choi & Song, 2018). LDA attempts to solve this challenge by looking backward from the distribution of words observed in particular documents to infer a hidden nature of the topic (Blei et al., 2003). Inference means estimating the conditional distribution of hidden variables in Bayesian methods, given all the documents, called posterior distribution. Unlike other traditional methods such as the mixture of single-word models (Nigam et al., 2010), Latent Dirichlet Allocation (Grün & Hornik, 2011) enables a more realistic topic model by using a mixed-membership model where the documents are distributed to different degrees at the same time to multiple topics, and the topic distributions differ over documents. It is also a complete probabilistic generative model that enables a robust estimation of the latent factor structure. Probabilistic Latent Semantic Indexing (pLSI), which Hofmann (2001) suggested earlier than LDA, shares the basic concept of the writing process and portrays a single text as a mixture of topics. Some approaches do not generate probabilistic models at the document level (Blei, Ng & Jordan, 2003). How pLSI allocates probability to such unseen documents is unclear, so it is difficult to apply it to new documents.

Moreover, the number of parameters to be determined increases linearly with the size of the document collection, making it difficult to overfit (Lee et al., 2010). LDA overcomes the problems of LSA by adding hidden random variables for document subject mixture weights (Blei, Ng & Jordan, 2003). LDA is a generative algorithm that efficiently deals with new documents. LDA also avoids the overfitting issue and the increase in parameter numbers with the size of data.

The LDA model is used in the study of technical developments found in scientific papers. Any potential examples are as follows. Lee et al. (2005) forecast emerging topics by reviewing the patent literature with Relation Prediction and LDA modeling. Chen et al. (2017) implement latent semantic analysis to analyze patent claim topics, revealing shifts in the subject over time. Lee and Sohn (2017) examine financial banking patents through LDA and identified emerging themes in the document. However, these case studies failed to represent the importance of key entities in patenting research and development.

Research Methodology

As shown in Figure 2, our presented scheme includes four steps: data collection, pre-processing of data, extraction of key topics, and exploration of topics. Step one is to collect patent documents relating to the automotive ecosystem and then prepare those documents for review by a topic-modeling algorithm. For the third stage, we generate a list of automobile-related patent topics. Finally, in the fourth stage, the established activity patterns are further explored.

--- Insert Figure 2 about here ---

Data Collection

The patent documents related to the automotive ecosystem are collected from PATSTAT's online database. There are certain advantages of using PATSTAT that are of great significance for this study. PATSTAT is an initiative by the European Patent Office for the benefit of families. The database includes over seventy-seven million patent applications and forty-five million issued patents ensuring data consistency and accuracy (Worldwide Patents Database: PATSTAT, 2018). Research may be done to determine when the patent application was filed, how the procedure has progressed, who the inventors are, and whether and when it is granted. The database of patents is a suitable source for investigating technological developments because it offers a representative selection of patents and a complete image of the most advanced technologies (Kim & Lee, 2015).

In previous evaluations of how different companies operate, patent classifications are used to find related patents. (Lee & Sohn, 2017). However, industry-level patent analysis does not meet the challenge of inter-connected network dependencies highlighted by the ecosystem framework (Adner and Kapoor, 2010). We have proposed an ecosystem level analysis where not only automobile manufacturing firms will be considered but, inter-dependent actors like components (e.g., battery providers) and complementors (data centers and information systems) will also be part of the analysis.

When searching for patents on ecosystem-based, it is critical to utilize the online database's correct search method. In most cases, technology classifications (Aghion et al., 2016; Golembiewski et al., 2015; Lanzi et al., 2011), relevant keywords (Oltra and Saint Jean, 2009; Sick et al., 2016; van den Hoed, 2005), or a combination of all of the above (Braun et al., 2011; Karvonen et al., 2016; Wang et al., 2012; Borgsted, Neyer & Schewe, 2017) are used. For the purpose of identifying patent classifications that focus on ICEV, BEV, FCEV, and HEV technologies, a combined search query of patent classifications and keywords is effective for several reasons (Borgsted, Neyer & Schewe, 2017).

A patent can be assigned to one or more distinct technical classes, as the patent examiner gives the patent to at least one technology class. The International Patent Classification is a system used by patent attorneys to classify patent applications based on their different patents (Duch-Brown and Costa-Campi, 2015; World Intellectual Property Organization, 2015). A prerequisite for learning about similar technologies is that they, as a group, are found in one or several of these classes. The problem is that the classes provide many patents that are correctly assigned to the given technological fields but partly suffer from a lack of automotive ecosystem usability (Borgsted, Neyer & Schewe, 2017). The H01M-008 (Fuel cells in terms of Manufacture) includes patents related to fuel cells beyond their automotive uses. The keywords to find within the title and abstract of the further patent specifies the data necessary for an effective automotive ecosystem. By requiring the terms "vehicle," "car," or "automotive" to be inside the patent, this strategy aims to ensure patentability. Second, such keywords are used in green technology to exclude any patents granted to one of the technologies. These words are found to be relevant to the study's subject (Oltra and Saint Jean, 2009; van den Hoed, 2005; Borgsted, Neyer & Schewe, 2017). Table 2 shows the results of the search.

--- Insert Table 2 about here ---

In this search query, IP signifies a classification system, the first set of keywords fulfills automotive utilization, whereas the second keyword set is the satisfying dimension of the technological ecosystem. For example, battery electric vehicle (BEV) related patents contain automotive features such as ‘vehicle,’ ‘car,’ or ’automobile’ and have ecosystem-related features such as ‘electric car,’’ battery,’ ‘charging infrastructure.’ The query for BEV-related patents would enlist technologies that might belong to focal automotive companies, battery charging systems that belong to suppliers as components, or complementors such as charging infrastructure, which belongs to data aggregators, energy providers, and ICT-related companies.

The observation period for this research is from 1990 to 2017, depending on the quality date of the patent, the first time the patent has been submitted to the patenting authority anywhere in the world (Nagaoka et al., 2010; OECD, 2009). The California Air Resources Board's legal criteria for reducing greenhouse emissions in 1990 have been used as a pivot point for the advancement of alternative technologies (van den Hoed, 2007). Selecting this year as the starting point for the analysis follows several studies (Berggren et al., 2009; Oltra and Saint

Jean, 2009; van den Hoed, 2005; Braun et al., 2011; Mueller et al., 2015). A patent application's publication is usually delayed by about 18 months, starting with the priority date. The review requires patents up to the year 2017 because of this shortcoming.

Using the document parsing techniques (Dridan, 2013), the relevant information is extracted from documents such as the document title, abstract, keywords, authors, assignee, class code, and reference citation. The output is then written into a relational database table (CSV file). Among these objects, the abstract is used as the input to LDA to define topics since it is usually restricted to a single technical solution. In total, for all four technologies, the dataset consists of 68,762 patents.

Borgsted, Neyer & Schewe (2017) propose a manual validity method to verify the results. We check for a total of 3,685 patents (around 5 percent of the dataset), distributed proportionally over all four technologies. The dataset's total validity is 92%, which is a decent result. If the content includes the given type of technology and the probability of automotive use, a patent is valid.

Data Preprocessing

We use an abstract feature from our relational dataset that indicates a detailed description of the invention for our case study. Thus, the abstract text will be used as an input to our model. As the abstract given in the patent is in an unstructured text format, it should be pre-processed and transformed into a structured format using the text mining technique for further analyses (Kim et al., 2016). For preprocessing, abstract in a free-text format requires processing tasks with natural language processing (NLP) pipeline, including tokenization, n-gram model, lemmatization, stop-word removal vector-space representation, as shown in Figure 3.

--- Insert Figure 3 about here ---

First, every sentence of the abstract is split into words, and these are then used as tokens. (Feldman & Sanger, 2007). It is essential to split these tokens into a sequence of 1-token, i.e., unigram, 2-token sequence, i.e., bigram, 3-token sequence, i.e., trigram. These token sequences add meaning to our text data. For example, electric vehicle technology; if this is taken as a unigram, then “electric,” “vehicle,” and “technology” have meaning on their own.

Similarly, combining these in bigram will also add context to our corpus, such as “electric vehicle” and “vehicle technology.” The resulting n-gram tokens go through the lemmatization process that reduces variants of a term to the root of its structural components. By representing a group of inflected forms as a single object, the analysis can be more accurate (Hotho et al., 2005). Then, the stop words that seldom contribute to the semantic representation of the documents are eliminated. There is no universal consensus on which words are and are not stop words, but function words are usually known to stop words. In addition, the common words of the language are not included in the stop-word list since those words seldom express details in the documents (Feldman & Sanger, 2007).

Finally, the patent abstract text is taken as matrix representation using the bag-of-words (BOW) model and term frequency-inverse document frequency (TF-IDF) vector model. We shall use both these vectorization methods and later compare the results of both to our LDA model. A bag-of-words model (Zhang et al., 2010) converts arbitrary text into vectors by counting how many times each word occurs. Meanwhile, the TF-IDF weighting method has been used to quantify the value of a given term in a text (Kuang & Xu, 2010; Zhang et al., 2010). The following formula determines the TF-IDF keyword frequency value;

$$\omega_{i,j} = TF\mathrm{i}, j \times log\left(N/\mathrm{d}f_j\right)$$

$TF_{i,j}$ = relative frequency of term j in the patent abstract i,

N = the total number of patent documents/abstract, and

df_j = the number of patent abstracts that include the term j.

This equation implies that if a word appears regularly in a document and infrequently in other documents/records, the term is likely to describe the document and differentiate it from other documents.

Finally, we filter out the terms that do not appear in more than 95% of the abstract document to exclude terms that rarely occurred. By incorporating the count vectors and TF-IDF vector for every patent abstract, we generate matrices of dimension (68762, 12630), the LDA input explained in the next step.

Topic Identification

This study performs LDA to classify k major subjects in related patents to identify topics from text data, i.e., patent abstract. The LDA algorithm is used for the patent abstracts in both the bag of words and the TF-IDF-matrix. LDA on the patent abstract is supposed to produce two outputs: similarities between patents and categories and distributions of topics. Each patent is assigned to one of the k topics based on the likelihood of each subject appearing in patent documents for the patent. In addition to grouping patents with related topics, LDA also offers a way to perform latent topic analysis by computing a topic distribution over each topic. The top n terms that are most widely used in describing the k topics are listed for use on the automotive patents. For evaluation of the LDA model, we use Perplexity (Zhao et al., 2015) and coherence score (Mimno et al., 2011).

Topic Evaluation

Next, the established topics are further analyzed for patterns over time and the most involved assignees within topics. To this point, each field has been explored through a holistic interpretation of patents that reveal the topic. To gauge patents at the firm level, the number of patents is the most basic kind of detail to look at. Other patent indices for technology relevance and firms' technology capacity are focused on it (Ernst, 2003). In order to perform the analysis, researchers use patent application-based indexes for both topic levels and firm levels.

On the one hand, to capture the patenting activity patterns at the topic level, each topic's current weightage in all patents relevant to the automotive ecosystem and the shift in the patent weightage over time are examined. The number of patent applications in a specific field is determined as the total number of patent applications divided by the total number of patent applications in that area. The shift in the topic's patent weight and the compound annual growth are calculated from time to time (CAGR). Based on the new indices, a classification system that includes the established topics within the patenting landscape as suggested by (Choi & Song, 2018). The parameters capture two distinct aspects of a company's patenting operation, i.e., current position and growth rate. From these indices, the proposed structure defines four distinct types of topics specific to the automotive ecosystem.

- The dominant topic with a large patent weightage and positive CAGR of patent weightage,

- The emerging topic with a small patent weightage and negative CAGR of patent weightage,
- The saturated topic with a large patent weightage and negative CAGR of patent weightage, and
- The declining topic with a small patent weightage and the positive CAGR of patent weightage.

On the other hand, to find out which automotive ecosystem key player has placed patents on similar technologies, the assignee has revealed technology advantage (RTA) is used. In this case, an assignee firm could be a focal firm, component, or complement (Iansiti, M., & Levien, R. 2004a). A firm's revealed technology advantage in a particular topic is the share of technology in an assignee's overall patents, divided by the global share of this technology in all patents (Ernst, 2003). RTA is obtained by normalizing technology share to the value between 1 and 0; it enables us to gauge a firm's competitive status relative to other firms in a specific technical domain. (Ernst, 2003). Precisely, Assignees with a relative technology share greater than 0.5 are examined in each topic as major assignees.

It is reasonable that technology leaders in an ecosystem are expected to perform much patent activity and possess a higher competitive technological advantage (Fabry et al., 2006). More substantial evidence of the positive relationship between the number of patent applications and the innovation portfolio of companies has been found in several previous empirical studies (Deng et al., 1999; Ernst, 1995; Hall & Jaffe, 2000; Lerner, 1994; Shane, 2001).

Analysis and Results

Employing the query search shown in Table 1, we find 68762 patent applications from 1990-2017. We collect them in HTML format by web crawling for further analyses each date of 03 October 2018). In order to extract the relevant material, including title, abstract, assignees, filing date, register year, classification code, citation, the patent documents are parsed. Described in a relational format such as comma-separated values (CSV), these patent objects are stored in a database.

All 68762 patents' abstracts are pre-processed by tokenization, lemmatization, stop-word removal, and vector-space representation. These pre-processing tasks on the patents 'abstract are performed with a Python IDE called Juypter Notebook. Tokenization is performed using regular expression(re) and genism module, lemmatization and stop-word removal are performed using the NLTK (Natural Language Toolkit) modules provided in the Python environment. We have also applied unigram, bigram, and trigram for having a full semantic overview of the patent abstract. The patent abstracts were then translated into a feature vector using the related Python function. There are two ways to look at what is important in a document; the bag of words model and the TF-IDF matrix. (Kuang & Xu, 2010; Zhang et al., 2010). The vector representation is employed to construct the count vector and TF-IDF matrix. We omitted words that do not appear in at least 95% of abstracts to select words commonly used in our data collection. Inputting the remaining terms into the LDA model, we create matrices of dimension (68762, 12630).

Essential Features of patent related to the automotive ecosystem

Before we discuss the results of topic identification by LDA, we will begin by explaining a few basic features of our data, which will help explain the technological progress in the automotive ecosystem in terms of overall patenting operation. As depicted in figure 4, in a linear trend, the number of similar patent applications rises uniformly from 1990 to 2005.

Since 2005, a considerable amount of patent applications have gone up. This upward trend may be in line with the consensus that ICTs have an increasingly influential position on the change and innovation of the automotive industry. (Peters et al., 2015). Our data's peak point is of the year 2011 with 5,737 applications, after which there is a constant decrease in applications. A steep decline in patent applications in 2015 is due to the patent filling phase at the EPO, which entails a considerable period between the filing and disclosure or registration of applications, varying from several months to over two years.

--- Insert Figure 4 about here ---

Secondly, the information helps us recognize major automobile firms and organizations active in the patenting process. As a result of the consolidation process that cleans out misspellings and variants of the assignees' names, 13,437 separate assignees are listed. Ten top assignees on the patenting activity are presented in Table 3. It is noteworthy that automotive companies and companies from various industries are part of the automotive ecosystem, such as BYD and Hitachi. Other organizations are not in the top ten patent applications, such as Siemens, Tesla, Samsung, and LG. The application counts for these organizations are not as high as patents from auto manufacturing companies yet represent the traditionally considered organizations outside the automotive sector.

--- Insert Table 3 about here ---

Technological topics in ecosystem

Based on the matrix representation discussed previously for patent abstracts, this study has performed LDA to identify important topics in automotive ecosystem-related patents using Python's genism module. Before implementation, we need to take care of two issues with the Latent Dirichlet Allocation model. One is related to choosing a better matrix representation model, and the other is related to the optimal number of topics selection. Both approaches have been of key interest in topic modeling research (Cao & Fei-Fei, 2007; Ignatenko, 2019; Kim & Gil, 2019; Syed & Spruit, 2017; Wallach: 2006).

We take matrix representation options for matrix representation, i.e., a bag of words and TF-IDF, both with dimensions of 68762 ×12630. We use perplexity and coherence score metrics to evaluate our unsupervised machine learning model (Zhao et al., 2015; Mimno et al., 2011). The model with a lower perplexity score and higher coherence score is considered to be better. With the LDA model bag, we get perplexity of -6.08 and a coherence score of 0.50 whereas, the LDA model with TF-IDF representation gives a -7.92-perplexity value of 0.57 coherence score. Therefore, we use the LDA model with TF-IDF metrics representation. As for choosing the optimal number of topics, Figure 5 outlines the coherence score, C, for the number of topics across the validation set.

--- Insert Figure 5 about here ---

The coherence score keeps increasing with the number of topics; it may make better sense to pick the model that gives the highest score before flattening out or a significant drop. In this case, we pick 8 for the number of topics with the highest coherence value of 0.557. Using Gibbs sampling, this 8-topic LDA model is designed to estimate the posterior distribution of latent topic structures suggested by Phan et al. (2008). We use pyLDAvis (Python module) for interactive topic model visualization, as shown in Figures 6a to 6h. These Figures show the identification of a topic based on the 30 most frequent terms occurrence, inter-topic distance map via multidimensional scaling, overall term frequency, estimated term frequency within the selected topic. Saliency[5] and relevance[6] have been calculated to determine word distributions with their latent topics and words' relevancy to the given corpus, respectively.

--- Insert Figure 6a, b, c, d, e, f, g, h about here ---

Identification of innovative ecosystem roles among topics

Any patent document goes through one of the eight most probable topics. The top 10 words in each subject are found to be the most commonly used words in that topic, as shown in Table 4. Finally, using these terms, the eight identified topics in the automotive ecosystem-related patents are labeled. We have provided domain-specific labels for each topic in order to provide a competence set for each topic. Additionally, each topic labeled competence is identified with its role (focal, component, complement) in the innovation ecosystem.

--- Insert Table 4 about here ---

The topics include some conventional automotive sector competencies in topics 1, 5, 6, 7, 8. Interestingly, all of these topics belong to technologies innovated at the focal firm-level in the automotive ecosystem. Topic 2 and topic 3 are associated with battery charging systems, with topic2 is related to cell battery methods that could be used in all four technologies. In contrast, topic3 is related to battery charging management developed for battery electric vehicles (BEV), hybrid electric vehicles (HEV) only. These two topics belong to technologies innovated at supplier firms known as 'component' in the ecosystem. Finally, topic 4 belongs to vehicle charging infrastructure that involves ICT and data aggregators, making it complement technology in ecosystem networks.

Trends in patenting activities in the innovation ecosystem

As of the yearly trend, the number of patent applications tends to increase consistently from 1990- 2004. There is a noticeable increase after 2005 in topics 1 to 6; however, we can find lower numbers in topics 7 and 8 (see Figure 7). This pattern of Topics from 1 to 6 is consistent with Figure 5. There is a constant rise from 1990-2004 with further increase in 2011. From

[5] saliency (term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

[6] relevance (term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Figure 7, it is clear that it is not possible to generalize the overall patent distribution trend to topic distribution because each topic has its distribution. It is also essential to understand how some focal technologies have increased patent applications over time, and others have declined. Unlike the traditional patent analysis approach, it is more useful to examine the relative weight of patents by topics to recognize the modern technical emphasis. It is interesting to find out patent weight, i.e., share in each topic, has changed over time and how this changing trend can be seen in ecosystem networks.

--- Insert Figure 7 about here ---

Since Patent Application disclosure or register opening takes some time, there is a time lag (in both pronouncements) within the patent filing phase. So, the number of patent applications in the last few years is generally not a very accurate patent study variable. The research studies like the one shown in the diagram show that in 2014, the number of patent applications among colleges with the largest declines. Since a 2004 patent weighs more, the base year for measurement used is 2011, not 2004. Meanwhile, the time frame for determining the CAGR (Compound annual growth rate) of the patent share is calculated as the last five years between 2009 and 2013. Looking at all financial years, there is a continuous rise and fall in the sum of patents as of Figure 3, implying that patenting activities in the automotive ecosystem have become earnest. Table 5 summarizes the weightage of patent applications in the year 2011 and the CAGR of patent applications from 2009-2013 by automotive ecosystem topics.

--- Insert Table 5 about here ---

Each technological topic is classified as one of four types of dominant, emerging, saturated, and declining topics, according to its current patent weightage and patent application CAGR. The average patent weightage in 2011 by topic was 8.3%. The patent weightage of a particular topic is calculated by referring to a previously collected patent counting system as part of statistical analysis. The reference value of the CGAR of the patent share is set as 0. The result of this condition is shown in Table 5.

Torque coordinated control in-vehicle systems (Topic 1), Battery, battery management, and battery charging in automobiles. (Topic 3), The Fuel injection system of ICEV (Topic5) is classified as the dominant topic, having a current patent weight larger than the average (8.3%)

and a positive CAGR of the patent application as per Table 5. The current key areas where technological progress is actively accomplished and where potential growth is also anticipated can be considered to be such technologies across diverse actors in the ecosystem. In the ecosystem network, topic 1 represents a focal-technology related to all four types of vehicles; Topic 3 shows component-technology, which is mainly associated with battery charging in BEV and HEV, Topic 5 depicts focal-technology that is only relevant to ICEV.

--- Insert Table 6 about here ---

We have methods for battery cell charging systems (Topic 2) for saturated topics, which can be used for all vehicles' technologies in ecosystems. Interestingly, it is interesting to notice how topic 3, which is related to lithium-ion charging batteries mainly used in BEV and HEV, lies in dominating technology. In contrast, topic 2 related to car batteries has become saturated. However, both technologies belong to component technology in the ecosystem network[7].

Coming to declining topics, we have technologies that all are generic and can be used for all types; however, their primary candidate is ICEV. It reflects that there is not much innovation occurring in ICEV technology. It is worth noticing how these technologies are related to focal technology, i.e., automobile manufacturing. We will discuss organizations and their dependence on these technological competencies. Topic 4 is related to charging infrastructure and data aggregation, which belong to complementary technologies. We find that these technologies are emerging with a small patent weight, yet potential growth is expected. One limitation for this case is that recent development in charging and data aggregation is through open-source community.

For this reason, we might expect a biased result as we may not get a proper innovation pattern for topic 4. However, it is clear from other studies that the charging infrastructure is at the stage of growth (Andrews et al., 2020; Chen et al., 2020; Pagani, 2019; Zhang, 2019). We may also argue that it is correctly listed as an emerging technology.

[7] Notice our study shows saturation in terms of innovation only- this technology may still be in demand in commercial sense.

Major assignee and their topic association in the innovation ecosystem

According to the ecosystem classification of three key technologies, we organize major assignees, i.e., focal technology, component technology, and complementary technology in Figure 8. It enables us to examine the major assignee's profile of automotive ecosystem-related patented technology. For instance, Toyota Motor Corp. and Ford Global technologies possess all three automotive ecosystem classes, i.e., control the full value chain network. Such organizations are called vertical integrated with management literature since they own their suppliers, distributors, or retail locations to retain their value or supply chain. In ecosystem literature, these organizations also add value for their end customers by owning complementary roles. It is interesting to note that most automobile manufacturers have appeared in "focal technology" such as Nissan, Honda, BMW. These organizations seem to innovate only at focal technologies and rely on other organizations to fulfill component and complementary innovations. It is important to note that focal technology consists of 36,918 out of 68,762 patent applications making it 53.69% of its contribution towards the patent dataset.

--- Insert Figure 8 about here ---

As for component technologies, we can find names like BYD, Panasonic, Siemens, Wanxiang. We can see that these are not the first to strike mind when we think of the automotive sector; however, they play an integral role in the automotive ecosystem. These innovations actively play a role in producing technologies that would control supplier end for their focal innovations. Component class type is composed of 22197 (32.29%) of our patent dataset. Finally, the complementary innovation type only contains 9647 (13.34%) patent applications where we find few major organizations in complementary technology part such as Toyota, Hyundai, Ford, Hitachi. Complementary innovations are also contributed by organizations having competence other than the traditional automotive domain, such as Proterra, Qualcomm, RWE, NextEV. These organizations add value for end customers bundled with focal innovation and component innovation within the ecosystem network. An organization such as Tesla appears only in complementary part with any significant patent application count. Though they perform all their activities within the organization, i.e., taking care of the whole ecosystem, their innovation profile only highlights the complementary role.

In Figure 9, we have shown our innovation classification in terms of emerging, dominant, saturated, and declining technologies. For understanding the assignee role, we have taken revealed technological advantage (RTA)[8] into account that provides an index to measure an assignee's relative specialization in technology. RTA for patent applications is standardized to fit into a [0, 1] interval, where 0 to 0.5 reflects no specialization and 0.5 to 1 indicates a revealed technological advantage in a class. Due to the inclusion of RTA, we do not find firms with a higher patent count but firms with a specialized contribution towards its ecosystem. We find that some service companies such as Tesla, Qualcomm, and Siemens are innovating effectively in emerging technologies and are creating enough revealed technological advantage. Highly integrated companies such as Ford and Toyota are not in emerging technologies as their RTA was not found in the range of 0.5 to 0.1.

For dominating technologies, Toyota, Ford, Hitachi, BYD, Continental automotive GmbH , and BMW are the assignees with a specialized contribution. Notably, the dominant technology class consists of OEM (Original equipment manufacturer) and component (battery management) organizations. For a traditional automotive supply chain, the collaboration of OEM and supplier end is crucial – a result that is consistent with the studies from (Borgstedt et al., 2017).

For saturated technologies, we have automobile companies all from Japan except the Peugeot Citroen group. Toyota is investing in both dominant and saturated technologies because some organizations keep a strong position within the ecosystem (Gulati et al., 2012). From an ecosystem perspective, automobile companies are targeting their innovative efforts in technologies that are rendered saturated. It is crucial to notice that these saturated technologies may still be significant for commercial and strategic purposes; therefore, further study of the ecosystem from a management perspective would be interesting.

Finally, declining technologies contain all automobile assignees. As per Figure 9, these technologies belong to topics 6, 7, and 8 related to focal innovation, and their current patent weightage is very small. As for the innovation ecosystem, these technologies are still adopted by organizations that would miss the mark. Their innovation portfolio is not targeting upcoming or in-demand trends in the automotive domain. These organizations are only focused

[8] The RTA is defined as the share of a technology in an assignee's overall patents, divided by the global share of this technology in all patents(OECD, 2009)

on the current dominant design (ICEV), and if rendered obsolete, these organizations would also lose their competitive advantage (Afuah, 2004).

--- Insert Figure 9 about here ---

Discussion

The motivation for investigating innovation trends in the automotive ecosystem

This study employs an automotive ecosystem framework to study innovation patterns using patents' data not restricted to classification-based patent analysis (Aghion et al., 2016; Golembiewski et al., 2015; Lanzi et al., 2011) or firm-based approaches (Cohen et al., 2000; Griliches, 1990; Oltra and Saint Jean, 2009; van den Hoed, 2005). The results of the study show that the assigned patents in the automotive ecosystem are from various industrial fields, and there is an interdisciplinary trend of innovation. For example, we could find companies from the automobile, battery management, ICT, and energy sectors, making it an incredibly diverse and innovative ecosystem. Additionally, no single IPC classification is found dominant, thereby confirming the advantage of our approach to targeting patent data for investigating innovation trends.

It is also worth noting that this study does not restrict the target data to the business ecosystem because of the importance of generating innovation among various actors in the automotive ecosystem. Previous studies analyzing ecosystems to identify key actors' strategic value have targeted their study to the alliance and network interaction (Basole, 2009; Harte et al., 2001; Lusch, 2010; Agerfalk and Fitzgerald, 2008). The exact relation and interdependence between business ecosystems and networks are not clear so far (Wulf, A., & Butel, 2017). It remains unclear where one actor's role in the business ecosystem begins and ends (Jacobides et al., 2018). By studying knowledge creation and technological trends in the ecosystem using patens, each actor's distinct role.

We have used patents as a standard criterion for innovation. However, each actor in the ecosystem has a different competence profile; therefore, inspecting assignees as their citation number or patent size is not appropriate. We instead investigate how actors are segmented into focal, component, and complement by their innovation activity. This study's focus is not limited to investigate how focal firms may achieve competitive advantage over others due to strong strategic alliances with their component and complement. Instead, it highlights how innovation bundled in focal technology, component technology, or complementary technology might add value to the end customer. Therefore, we study how innovation activities within the automotive ecosystem could promise end customers' technological value.

The benefit of ecosystem framework for Technology Trend Exploration

Our study shows that innovation distribution among key actors is not uniform, focusing on the patent dataset 53.69%. Automobile companies are still at the heart of technology development, whereas component and complement provide the puzzle's remaining pieces. We find that there is s shift from focal firm-based innovation towards component and complementary related innovation, as shown in Figure10. Thus, understanding the technological landscape in the automotive ecosystem is vital in their technological competencies. For example, complementary technology such as data aggregators with a lower patent profile and being at the initial business development stage can leverage the cooperation with firms with larger firm size and more robust innovation portfolios to facilitate knowledge and technology transfer within the ecosystem (Gao et al., 2019). This kind of collaboration can even enable complementary technology to generate their niche yet competitive innovation profile (Autio, E., & Thomas, 2014; Jacobides et al., 2018). It has been recognized empirically in the organizational management literature that a big organization is favorable to technology sensing for information acquisition and awareness of new technological growth (Garrison, 2009). Thus, the ecosystem as a framework for technological trend exploration, as suggested in this study, would help emerging complementary companies playing a pivotal role in the future of the automotive ecosystem.

On the other side, focal firms could take advantage of domain knowledge other than their core competencies. The significant link in the literature between company size and technology sensing is based on the idea that technology sensing relies on the expertise of individuals (Clarysse et al., 2014). For example, an OEM would have better knowledge about the design of the automotive ecosystem. Service-based solutions may not be the core competence of that organization, complementary service-providers in the automotive ecosystem may add value without disturbing focal firms' competence set. Being informed of technical developments outside of organizational competence is a prerequisite for making this strategy successful. In this case, the suggested approach of this study will help automotive organizations identify candidate partnering organizations by finding knowledge on who in their respective ecosystem has competitive advantages.

Patent for the Exploration of Technology-Driven Innovation ecosystems

Only novel and non-obvious scientific inventions may be patented as per patent definition. (Worldwide Patents Database: PATSTAT, 2018). The technological invention can also be useful for the patent, which means that it can be technically applied for industry use. (Borgstedt et al., 2017; Worldwide Patents Database: PATSTAT, 2018). The patent-based approach, therefore, encourages practitioners and scientists to pursue technical progress that has a high potential for industrial applicability (Porter & Detampel, 1995). In addition, the availability of patent data helps a company to access external innovation (Daim et al., 2006). Patents are considered reliable for this paper because of the above-mentioned characteristics. They are well-aligned with their primary focus: creating an intelligent tool to promote the discovery of technology developments in the automotive ecosystem.

In terms of the necessary technical advancement types, incremental and radical innovations are a common classification scheme. Innovation's radicalness is typically defined within a continuum of innovation's novelty and effect (Green et al., 1995). Because of the criteria for patents for creativity, non-obviousness, and utility, patent research is more likely to include innovation data that poses progressive steps forward or entirely different approaches to problems, rather than incremental improvements that branch off and expand on previous ideas (Worldwide Patents Database: PATSTAT, 2018). On the patent-based calculation of how radical innovation is, there are two different opinions. The first view suggests that individual patents have a similar degree of influence and novelty within a category and timeline. This method, therefore, tests the radicality of a patent category at a specific time by using the construction of an S-curve technology life cycle (Foster, 1985). Specifically, over time, patent counts in a category are adapted to an S-curve, divided into phases of introduction, development, maturity, and decline. The earlier a patent category is drawn upon this step graph, the more radical the technical breakthrough associated with it is (Lee & Su, 2017).

In this study, the suggested solution relates to the first view. Patent counts do not directly match the s-curve, but the four topic categories refer to the four stages of the life cycle of technology. The two parameters, i.e., the current weighting of the patent and the CAGR of the patent application, indicate the topic's success and growth rate, respectively. The emerging topic with a small patent weightage and negative CAGR of patent application could correspond to the introduction phase. The dominant topic with a large patent weightage and positive CAGR of a patent could be matched with the growth phase. The saturated topic with a large patent weightage and negative CAGR of a patent can be regarded as the mature phase (Choi and Song,2018). As shown in Table 5 in the previous section, the decreasing topic with a small patent weighting and the positive CAGR of a patent application can be considered the declined phase.

The second view towards radical innovation posits that individual patents should have different levels of novelty and impact, even if applied simultaneously. The methodology focused on this point of view tests the radicalness of patent-level innovation, primarily through citation-based indices and invention content (Dahlin & Behrens, 2005). Since this paper focuses on identifying groups of similar patents based on their innovation content (i.e., technical topics and the investigation of group-level features of automotive ecosystem-related technologies), it is beyond the reach of this paper to classify which patents are more radical within a topic. The incorporation of the review of individual patents, however, may be useful in certain situations, such as technology benchmarking.

Topic Identification and Exploration in the ecosystem

Depending on the time window of topic patents, the topics behind automotive ecosystem-related patents discovered by LDA could be different. For instance, If the time window is constrained to the most current patents, only patents relating to the new automotive technologies are deemed to be included. Therefore, if the emphasis is on catching cutting-edge technology at the time of research, better results could be obtained by using the latest patents as base data for the identification of the topic. In comparison, the entire collection of collected patents is used in this analysis. However, for this analysis, the effect of the time window on the recognition of the topic will not be significant since the patents filed after 2005 shows a substantial count.

It is essential to mention that the limited time window provides less information about technological topics over time. It is not necessary to use the latest patents because this study seeks to show what technical issues are in the automotive ecosystem and how related technological topics have been modified. On the other hand, a large time window may include the risk of considering and allocating undue importance to obsolete or outdated topics. For this reason, the patent data were narrowed down from 1990-2017 topic exploration. The starting period is defined due to the California Air Resources Board’s legal requirement in the year 1990 concerning reducing greenhouse gases. Therefore, this research will provide information on historical developments in technical topics in the automotive environment and the current status of the topics by using different time windows for the subject recognition and exploration.

Contributions

This research contributes to mitigating gaps in the lack of empirical studies directly on innovation ecosystems. Autio & Thomas, 2014: Jacobites et al. (2018) use of patent data for the innovation study in general (Kache & Seuring, 2017). We analyzed patent documents' textual data, which have rarely been used in research despite their value. Also, the automotive-related patents were defined based on a combination of the keywords and patent classifications. This study could investigate technological trends in the automotive ecosystem more thoroughly and provide specific information about patents' topics. From the perspective of research methods, this research also contributes to ecosystem innovation research by proposing a topic-modeling approach to patent data with LDA, an advanced algorithm for textual data analysis. Given the rise in automotive-related patents, the usefulness of our topic-modeling-based approach has increased because it can effectively handle a large amount of unstructured text content.

Furthermore, LDA helps to group and examine patent documents that display similar patterns of words called topics. The result is close to clustering algorithms in certain aspects; clustering algorithms are based on a black-box mechanism. On the other hand, LDA offers a way to view patent documents grouped into a subject by specifically describing topics using word probability distributions over topics (Blei, 2012).

We also made considerable practical implications for managers. Our proposed approach would improve the understanding of what technological fields are relevant in the automotive ecosystem, who are the leading players in each sector, and how they have changed over time. The eight most optimized topics in automotive-related patents included focal technology activity, such as design and manufacturing of vehicles, and interdisciplinary technological areas, such as battery management that signify component role and charging infrastructure that refers to complementary activity. This study tried to examine dynamic technological topics in the automotive ecosystem and the future potential innovative activities.

Four distinct subject types were developed by the classification scheme using the current patent weightage and the CAGR as criteria: dominant, evolving, saturated, and declining. As a result, the classification scheme helps us to consider the various technologies that have led to innovation in the holistic automotive ecosystem when evaluating a technological subject. As a result of this study, traditional technological areas were classified as saturated and declining topics that belong to focal technology. On the other side, emerging topics were in complementary technology roles. It implies that the automotive ecosystem is also affected by rapid technological development, and more innovation is taking place on complementary technology than focal technology. It also allows for the future to predict the importance of a particular technical field.

This study shows that many of the major assigned patents related to the automotive ecosystem come from different sectors. For instance, OEMs are technological leaders among the ecosystem, but their technology competency tends to be limited to interdisciplinary technological areas such as data aggregation, communication, and ICT. More IT, energy, and data-related companies seem likely to be expanding their influence in the automotive ecosystem. It reflects convergence trends in the automotive ecosystem and implies the increasing possibility of competition and cooperation from unexpected industries and technology sectors. The research on investigating the automotive ecosystem and the resulting innovations using patent data is still in the infant stage. Our study exhibits converging features among different players in the ecosystem.

In terms of patent studies, researchers investigating the classification of patents related to the automotive ecosystem should start with the keyword and IPC classes associated with the topics described in this report. This holistic ecosystem approach to explore technological trajectories using patent data can be more valuable than firm-based cases. The major assignees' analysis can also help researchers and practitioners discover and monitor the innovative leaders and their relationship with niche players within their ecosystem.

Limitations

There are drawbacks to this analysis that require further studies. For example, this research can be strengthened by changing the policy of data collection. This study took a keyword and classification combination approach to define automotive ecosystem patents, recognizing the limitations of classification-based and firm-based approaches. However, by its design, this approach to identifying target patents suffers from the limitation that the quality and quantity of the data depends largely and predictably on search keywords and IPC classifications. This study used this approach to determine automotive ecosystem-related patents, but the resulting patents are not likely to represent the overall ecosystem. For example, only 212 patents filed by BMW are included in the data collected in this report. However, we have been able to find reports in recent years on BMW's growing patent application activities in related areas (BMW Group Investor Relations Presentation Q2, 2019). It suggests a strong probability that certain patents do not contain keywords or IPC classes, even though their proprietary innovations are relevant to the automotive industry. The data collection policy, therefore, needs to be strengthened. On the other side, the data collected in this study could be utilized as a good starting point. In defining search keywords for automotive ecosystem-related patents, the phrases that often occur with automotive companies in our data may be contenders. Another issue is that most ICT and data-related companies use open-source technologies that are patent agnostic; therefore, those technologies' contributions could not be realized in this study.

Second, this study may have shortcomings from the analysis aspect. The technical issues described help to illuminate the technological environment to some extent in the automotive ecosystem. However, by subdividing the listed subjects, there is still room for progress. Some topics include a larger number of patent documents, such as topic 1 has 20742 patent documents. Therefore, in terms of using and utilizing technological trajectories, the additional division may cover various technologies. It can also provide richer data to break down the topics into detailed sub-topics and examine patterns at the sub-topic level. The subtopic-level analysis is more suitable for investigating the technological landscape according to the business model components or value chain components in the ecosystem framework, enabling identifying opportunities for technology-based innovation.

References

Adner, R. and Kapoor, R., 2016. Innovation ecosystems and the pace of substitution: Re-examining technology S-curves. Strategic management journal, 37(4), pp.625-648.

Adner, R. and Kapoor, R., 2010. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. Strategic management journal, 31(3), pp.306-333.

Adner, R., 2006. Match your innovation strategy to your innovation ecosystem—Harvard business review, 84(4), p.98.

Autio, E. and Thomas, L., 2014. Innovation ecosystems. The Oxford handbook of innovation management, pp.204-288.

Amado, A., Cortez, P., Rita, P. and Moro, S., 2018. Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. European Research on Management and Business Economics, 24(1), pp.1-7.

Anggraeni, E., Den Hartigh, E. and Zegveld, M., 2007, October. Business ecosystem as a perspective for studying the relations between firms and their business networks. In ECCON 2007 Annual meeting (pp. 1-28).

Andrews, D.J., Polmateer, T.L., Wheeler, J.P., Slutzky, D.L., and Lambert, J.H., 2020. Enterprise risk and resilience of electric-vehicle charging infrastructure and the future mobile power grid. Current Sustainable/Renewable Energy Reports, pp.1-7.

Afuah, A. 2004. Does a focal firm's technology entry timing depend on the impact of the technology on co-opetitors? Research Policy, 33: 1231-1246.

Afuah, A., 2000. How much do your co-opetitors' capabilities matter in the face of technological change?. Strategic management journal, 21(3), pp.397-404.

Adomavicius, G., Bockstedt, J., Gupta, A. and Kauffman, R.J., 2006, January. Understanding patterns of technology evolution: An ecosystem perspective. In Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06) (Vol. 8, pp. 189a-189a). IEEE.

Archibugi, D. and Planta, M., 1996. Measuring technological change through patents and innovation surveys. Technovation, 16(9), pp.451-519.

Athanasopoulou, A., Bouwman, W.A.G.A., Nikayin, F.A. and de Reuver, G.A., 2016. The disruptive impact of digitalization on the automotive ecosystem: a research agenda on business models, platforms, and consumer issues. Proceedings of the 29th Bled eConference.

Aguilar-Fernández, M.E. and Otegi-Olaso, J.R., 2018. Firm size and the business model for sustainable innovation. Sustainability, 10(12), p.4785.

Abdelkafi, N., Makhotin, S. and Posselt, T., 2013. Business model innovations for electric mobility—what can be learned from existing business model patterns?. International Journal of Innovation Management, 17(01), p.1340003.

Aghion, P., Dechezleprêtre, A., Hemous, D., Martin, R., and Van Reenen, J., 2016. Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. Journal of Political Economy, 124(1), pp.1-51.

Alghamdi, R. and Alfalqi, K., 2015. A survey of topic modeling in text mining. Int. J. Adv. Comput. Sci. Appl.(IJACSA), 6(1).

Basberg, B.L., 1987. Patents and the measurement of technological change: a survey of the literature. Research Policy, 16(2-4), pp.131-141.

Braun, F.G., Hooper, E., Wand, R. and Zloczysti, P., 2011. Holding a candle to innovation in concentrating solar power technologies: A study drawing on patent data. Energy Policy, 39(5), pp.2441-2456.

Basberg, B.L., 1987. Patents and the measurement of technological change: a survey of the literature. Research Policy, 16(2-4), pp.131-141.

Basberg, B.L., 1984. Patent statistics and the measurement of technological change—an assessment of the Norwegian patent data, 1840–1980. World Patent Information, 6(4), pp.158-164.

Blei, D.M., 2012. Probabilistic topic models. Communications of the ACM, 55(4), pp.77-84.

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent Dirichlet allocation. Journal of Machine Learning Research, 3(Jan), pp.993-1022.

Borgstedt, P., Neyer, B. and Schewe, G., 2017. Paving the road to electric vehicles–A patent analysis of the automotive supply industry. Journal of cleaner production, 167, pp.75-87.

Borgstedt, P., Kirschke, F., Neyer, B. and Schewe, G., 2016, September. The impact of competing powertrain technologies on the formation of automotive innovation networks. In 2016 Portland International Conference on Management of Engineering and Technology (PICMET) (pp. 2141-2151). IEEE.

Benites-Lazaro, L.L., Giatti, L. and Giarolla, A., 2018. Topic modeling method for analyzing social actor discourses on climate change, energy, and food security. Energy research & social science, 45, pp.318-330.

Bernal-Conesa, J.A., de Nieves Nieto, C. and Briones-Peñalver, A.J., 2017. CSR strategy in technology companies: Its influence on performance, competitiveness, and sustainability. Corporate social responsibility and environmental management, 24(2), pp.96-107.

BMW Group Investor Relations Presentation Q2.,2019. available at:

https://www.bmwgroup.com/content/dam/grpw/websites/bmwgroup_com/ir/downloads/en/2019/Investor_Presentation/BMW_Investor_Presentation_2019.pdf

Cao, L., & Fei-Fei, L.,2007. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. In 2007 IEEE 11th International Conference on Computer Vision (pp. 1-8). IEEE.

Clarysse, B., Wright, M., Bruneel, J. and Mahajan, A., 2014. Creating value in ecosystems: Crossing the chasm between knowledge and business ecosystems. Research Policy, 43(7), pp.1164-1176.

Cho, T.S. and Shih, H.Y., 2011. Patent citation network analysis of core and emerging technologies in Taiwan: 1997–2008. Scientometrics, 89(3), pp.795-811.

Choi, J.Y., Jeong, S. and Kim, K., 2015. A study on diffusion pattern of technology convergence: Patent analysis for Korea. Sustainability, 7(9), pp.11546-11569.

Choi, D. and Song, B., 2018. Exploring technological trends in logistics: Topic modeling-based patent analysis. Sustainability, 10(8), p.2810.

Chen, T., Zhang, X.P., Wang, J., Li, J., Wu, C., Hu, M., and Bian, H., 2020. A Review on Electric Vehicle Charging Infrastructure Development in the UK. Journal of Modern Power Systems and Clean Energy, 8(2), pp.193-205.

Chen, H., Zhang, G., Zhu, D. and Lu, J., 2017. Topic-based technological forecasting based on patent data: A case study of Australian patents from 2000 to 2014. Technological Forecasting and Social Change, 119, pp.39-52.

Chuang, J., Manning, C.D. and Heer, J., 2012, May. Termite: Visualization techniques for assessing textual topic models. In Proceedings of the international working conference on advanced visual interfaces (pp. 74-77).

Cillo, V., Petruzzelli, A. M., Ardito, L., & Del Giudice, M.,2019. Understanding sustainable innovation: A systematic literature review. Corporate Social Responsibility and Environmental Management, 26(5), 1012-1025.

Crabb, J. M., & Johnson, D. K.,2010. Fueling innovation: the impact of oil prices and CAFE standards on energy-efficient automotive technology. The Energy Journal, 31(1).

David, P. A., & Olsen, T. E.,1992. Technology adoption, learning spillovers, and the optimal duration of patent-based monopolies. International Journal of Industrial Organization, 10(4), 517-543.

Daim, T.U., Rueda, G., Martin, H. and Gerdsri, P., 2006. Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technological Forecasting and Social Change, 73(8), pp.981-1012.

Dahlin, K.B. and Behrens, D.M., 2005. When is an invention really radical?: Defining and measuring technological radicalness. Research Policy, 34(5), pp.717-737.

Deng, Z., Lev, B. and Narin, F., 1999. Science and technology as predictors of stock performance. Financial Analysts Journal, 55(3), pp.20-32.

Duch-Brown, N. and Costa-Campi, M.T., 2015. The diffusion of patented oil and gas technology with environmental uses: a forward patent citation analysis. Energy Policy, 83, pp.267-276.

Ernst, H., 2003. Patent information for strategic technology management. World patent information, 25(3), pp.233-242.

Ernst, H., 1997. The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry. Small business economics, 9(4), pp.361-381.

Ernst, H., 1995. Patenting strategies in the German mechanical engineering industry and their relationship to company performance. Technovation, 15(4), pp.225-240.

Fabry, B., Ernst, H., Langholz, J. and Köster, M., 2006. Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry. World Patent Information, 28(3), pp.215-225.

Feldman, R. and Sanger, J., 2007. The text mining handbook: advanced approaches in analyzing unstructured data. Cambridge university press.

Foster, R.N., 1985. Timing technological transitions. Technology in Society, 7(2-3), pp.127-141.

Gawer, A., & Cusumano, M. A. 2002. Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation. Boston, MA: Harvard Business School Press.

Gawer, A., & Cusumano, M. A. 2008. How companies become platform leaders. MIT Sloan Management Review, 49: 28.

Gao, Y., Liu, X. and Ma, X., 2019. How do firms meet the challenge of technological change by redesigning the innovation ecosystem? A case study of IBM. International Journal of Technology Management, 80(3-4), pp.241-265.

Garrison, G., 2009. An assessment of organizational size and sense and response capability on the early adoption of disruptive technology. Computers in Human Behavior, 25(2), pp.444-449.

Green, S.G., Gavin, M.B. and Aiman-Smith, L., 1995. Assessing a multidimensional measure of radical technological innovation. IEEE transactions on engineering management, 42(3), pp.203-214.

Griliches, Z., 1990. Patent statistics as economic indicators: a survey part 2. NBER.

Gulati, R., Puranam, P., & Tushman, M. L. 2012. Meta-organization design: Rethinking design in inter-organizational and community contexts. Strategic Management Journal, 33(6): 571-586.

Gulati, R., Nohria, N., & Zaheer, A. 2000. Strategic networks. Strategic Management Journal, 21(3): 203-215.

Ghosh, D. and Guha, R., 2013. What are we ‘tweeting about obesity? Mapping tweets with topic modeling and Geographic Information System. Cartography and geographic information science, 40(2), pp.90-102.

Gerken, J.M. and Moehrle, M.G., 2012. A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. Scientometrics, 91(3), pp.645-670.

Hashimi, H., Hafez, A. and Mathkour, H., 2015. Selection criteria for text mining approaches. Computers in Human Behavior, 51, pp.729-733.

Hall, B.H., Jaffe, A.B., and Trajtenberg, M., 2000. Market value and patent citations: A first look (No. w7741). National bureau of economic research.

Hashimi, H.; Hafez, A.; Mathkour, H. Selection criteria for text mining approaches. Comput. Hum. Behav. 2015, 51, 729–733.

Hannah, D.P. and Eisenhardt, K.M., 2018. How firms navigate cooperation and competition in nascent ecosystems. Strategic Management Journal, 39(12), pp.3163-3192.

Huang, C., Arundel, A. and Hollanders, H., 2010. How firms innovate: R&D, non-R&D, and technology adoption.

Hornik, K. and Grün, B., 2011. topic models: An R package for fitting topic models. Journal of statistical software, 40(13), pp.1-30.

Hofmann, T., 2001. Unsupervised learning by probabilistic latent semantic analysis. Machine learning, 42(1-2), pp.177-196.

Hotho, A., Nürnberger, A. and Paaß, G., 2005, May. A brief survey of text mining. In Ldv Forum (Vol. 20, No. 1, pp. 19-62).

Hyun, Y.G., Han, J.H., Chae, U., Lee, G.H., and Lee, J.Y., 2020. A Study On Technical Trend Analysis Related to Semantic Analysis of NLP Through Domestic/Foreign Patent Data. Journal of Digital Convergence, 18(1), pp.137-146.

Huber, J. and Puto, C., 1983. Market boundaries and product choice: Illustrating attraction and substitution effects. Journal of consumer research, 10(1), pp.31-44.

Iansiti, M., & Levien, R. 2004a. The Keystone Advantage: What the New Dynamics of Business Ecosystems Mean for Strategy, Innovation, and Sustainability: Harvard Business School Press.

Iansiti, M., & Levien, R. 2004b. Strategy as ecology. Harvard Business Review, 82: 68-78.

Iyer B, Lee CH, Venkatraman N. 2006. Managing in a small world ecosystem: some lessons from the software sector. California Management Review 48(3): 28-47.

Ili, S., Albers, A. and Miller, S., 2010. Open Innovation in the automotive industry. R&d Management, 40(3), pp.246-255.

Ignatenko, V., Koltcov, S., Staab, S. and Boukhers, Z., 2019, February. Fractal approach for determining the optimal number of topics in the field of topic modeling. In Journal of Physics: Conference Series (Vol. 1163, No. 1, p. 012025). IOP Publishing.

Isoaho, K., Gritsenko, D. and Mäkelä, E., 2019. Topic modeling and text analysis for qualitative policy research. Policy Studies Journal.

Jacobides, M.G., Cennamo, C. and Gawer, A., 2018. Towards a theory of ecosystems. Strategic Management Journal, 39(8), pp.2255-2276.

Jacobides, M., Cennamo, C. and Gawer, A., 2015. Industries, ecosystems, platforms, and architectures: Rethinking our strategy constructs at the aggregate level. Unpublished manuscript.

Jacobides, M.G., Knudsen, T. and Augier, M., 2006. Benefiting from innovation: Value creation, value appropriation and the role of industry architectures. Research Policy, 35(8), pp.1200-1221.

Johansson, M. and Deniz, S., 2014. Ecosystem Changes in the Automotive Industry: Electric Vehicles and the Increased Importance of Networked Business Models.

Jaffe, A.B. and Trajtenberg, M., 2002. Patents, citations, and innovations: A window on the knowledge economy. MIT press.

Jeong, B. and Yoon, J., 2017. Competitive intelligence analysis of augmented reality technology using patent information. Sustainability, 9(4), p.497.

Kapoor, R., 2014. Collaborating with Complementors: What Do Firms Do?', Collaboration and Competition in Business Ecosystems (Advances in Strategic Management, Volume 30).

Kerr, C.I., Mortara, L., Phaal, R. and Probert, D.R., 2006. A conceptual model for technology intelligence. International Journal of Technology Intelligence and Planning, 2(1), pp.73-93.

Kache, F. and Seuring, S., 2017. Challenges and opportunities of digital information at the intersection of Big Data Analytics and supply chain management. International Journal of Operations & Production Management.

Kuang, Q., and Xu, X., 2010, August. Improvement and application of TF• IDF method based on text classification. In 2010 International Conference on Internet Technology and Applications (pp. 1-4). IEEE.

Kim, G., Lee, J., Jang, D. and Park, S., 2016. Technology clusters exploration for patent portfolio through patent abstract analysis. Sustainability, 8(12), p.1252.

Kim, S.W. and Gil, J.M., 2019. Research paper classification systems based on TF-IDF and LDA schemes. Human-centric Computing and Information Sciences, 9(1), p.30.

Lei, L., Qi, J. and Zheng, K., 2019. Patent analytics based on feature vector space model: A case of IoT. IEEE Access, 7, pp.45705-45715.

Lee, W.S. and Sohn, S.Y., 2017. Identifying emerging trends of financial business method patents. Sustainability, 9(9), p.1670.

Lee, S., Song, J. and Kim, Y., 2010. An empirical comparison of four text mining methods. Journal of Computer Information Systems, 51(1), pp.1-10.

Lee, H. and Kang, P., 2018. Identifying core topics in technology and innovation management studies: A topic model approach. The Journal of Technology Transfer, 43(5), pp.1291-1317.

Lerner, J., 1994. The importance of patent scope: an empirical analysis. The RAND Journal of Economics, pp.319-333.

Lichtenthaler, E., 2004. Technological change and the technology intelligence process: a case study. Journal of Engineering and Technology Management, 21(4), pp.331-348.

Lei, X.P., Zhao, Z.Y., Zhang, X., Chen, D.Z., Huang, M.H., Zheng, J., Liu, R.S., Zhang, J. and Zhao, Y.H., 2013. Technological collaboration patterns in the solar cell industry based on patent inventors and assignees analysis. Scientometrics, 96(2), pp.427-441.

Lin, W.L., Cheah, J.H., Azali, M., Ho, J.A. and Yip, N., 2019. Does firm size matter? Evidence on the impact of the green innovation strategy on corporate financial performance in the automotive sector. Journal of Cleaner Production, 229, pp.974-988.

Lupu, M., 2017. Information retrieval, machine learning, and Natural Language Processing for intellectual property information.

MacMullan, H.,2018, August 15. Worldwide Patents Database: PATSTAT - Research-IT. Retrieved from https://research-it.wharton.upenn.edu/news/worldwide-patents-database-patstat/

Miller, L., Soulliere, K., Sawyer-Beaulieu, S., Tseng, S., & Tam, E.,2014. Challenges and alternatives to plastics recycling in the automotive sector. Materials, 7(8), 5883-5902.

Mont, O., Whalen, K., & Nussholz, J.,2019. Sustainable innovation in business models: celebrated but not interrogated. In Handbook of Sustainable Innovation. Edward Elgar Publishing.

Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A.,2011. Optimizing semantic coherence in topic models. In Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (pp. 262-272).

Moore, J. F. 1993. Predators and prey: a new ecology of competition. Harvard Business Review, 71: 75-86.

Moore, J. F. 1996. The Death of Competition: Leadership and strategy in the age of business ecosystems. New York, NY: HarperBusiness.

Narin, F., 1994. Patent bibliometrics. Scientometrics, 30(1), pp.147-155.

Narin, F., Noma, E. and Perry, R., 1987. Patents as indicators of corporate technological strength. Research Policy, 16(2-4), pp.143-155.

Nigam, K., McCallum, A.K., Thrun, S. and Mitchell, T., 2000. Text classification from labeled and unlabeled documents using EM. Machine learning, 39(2-3), pp.103-134.

OECD, 2009. Patents as statistical indicators of science and technology. In: OECD Patent Statistics Manual. OECD Publishing, Paris. http://dx.doi.org/10.1787/ 9789264056442-3-en.

Oltra, V., Saint Jean, M., 2009. Variety of technological trajectories in low emission vehicles (LEVs): a patent data analysis. J. Clean. Prod. 17 (2), 201-213.

Orsato, R.J., and Wells, P., 2007. U-turn: the rise and demise of the automobile industry. Journal of Cleaner Production, 15(11-12), pp.994-1006.

Parker, G., Van Alstyne, M.W. and Jiang, X., 2016. Platform ecosystems: How developers invert the firm. Boston University Questrom School of Business Research Paper, (2861574).

Pagani, M., Korosec, W., Chokani, N. and Abhari, R.S., 2019. User behavior and electric vehicle charging infrastructure: An agent-based model assessment. Applied Energy, 254, p.113680.

Phan, X.H., Nguyen, L.M. and Horiguchi, S., 2008, April. Learning to classify short and sparse text & web with hidden topics from large-scale data collections. In Proceedings of the 17th international conference on World Wide Web (pp. 91-100).

Planko, J., Cramer, J.M., Chappin, M.M. and Hekkert, M.P., 2016. Strategic collective system building to commercialize sustainability innovations. Journal of Cleaner Production, 112, pp.2328-2341.

Pierce, L., 2009. Big losses in ecosystem niches: How core firm decisions drive complementary product shakeouts. Strategic management journal, 30(3), pp.323-347.

Porter, A.L. and Detampel, M.J., 1995. Technology opportunities analysis. Technological Forecasting and Social Change, 49(3), pp.237-255.

Popp, D., 2005. Lessons from patents: Using patents to measure the technological change in environmental models. Ecological Economics, 54(2-3), pp.209-226.

Porter, A.L. and Cunningham, S.W., 2005. Tech mining. Competitive Intelligence Magazine, 8(1), pp.30-36.

Peters, S., Chun, J.H. and Lanza, G., 2015. Digitalization of automotive industry–scenarios for future manufacturing.

Riasanow, T., Galic, G. and Böhm, M., 2017. Digital transformation in the automotive industry: Towards a generic value network.

Scherer, F.M., 1965. Firm size, market structure, opportunity, and the output of patented inventions. The American economic review, 55(5), pp.1097-1125.

Seitz, M.A., and Wells, P.E., 2006. Challenging the implementation of corporate sustainability: The case of automotive engine remanufacturing. Business Process Management Journal, 12(6), pp.822-836.

Son, I., Kim, J., Park, G. and Kim, S., 2018. The Impact of Innovative Technology Exploration on Firm Value Sustainability: The Case of Part Supplier Management. Sustainability, 10(10), p.3632.

Shi, T., Kang, K., Choo, J. and Reddy, C.K., 2018, April. Short-text topic modeling via non-negative matrix factorization enriched with local word-context correlations. In Proceedings of the 2018 World Wide Web Conference (pp. 1105-1114).

Sievert, C. and Shirley, K., 2014, June. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Srinivasan, R., Lilien, G.L. and Rangaswamy, A., 2002. Technological opportunism and radical technology adoption: An application to e-business. Journal of Marketing, 66(3), pp.47-60.

Shane, S., 2001. Technological opportunities and new firm creation. Management Science, 47(2), pp.205-220.

Syed, S. and Spruit, M., 2017, October. Full-text or abstract? Examining topic coherence scores using latent Dirichlet allocation. In 2017 IEEE International conference on data science and advanced analytics (DSAA) (pp. 165-174). IEEE.

Steyvers, M. and Griffiths, T., 2007. Probabilistic topic models. Handbook of latent semantic analysis, 427(7), pp.424-440.

Schmookler, J.A.C.O.B., 1966. Invention and Economic Growth (Cambridge: Harvard).

Tseng, Y.H., Lin, C.J. and Lin, Y.I., 2007. Text mining techniques for patent analysis. Information processing & management, 43(5), pp.1216-1247.

Trappey, A.J., Trappey, C.V., Wu, J.L. and Wang, J.W., 2020. Intelligent compilation of patent summaries using machine learning and natural language processing techniques. Advanced Engineering Informatics, 43, p.101027.

Teichert, T. and Mittermayer, M.A., 2002, August. Text mining for technology monitoring. In IEEE International Engineering Management Conference (Vol. 2, pp. 596-601). IEEE.

Vaz, C.R., Rauen, T.R.S. and Lezana, Á.G.R., 2017. Sustainability and innovation in the automotive sector: A structured content analysis. Sustainability, 9(6), p.880.

Van den Hoed, R., 2007. Sources of radical technological innovation: the emergence of fuel cell technology in the automotive industry. Journal of Cleaner Production, 15(11-12), pp.1014-1021.

Wallach, H.M., 2006, June. Topic modeling: beyond bag-of-words. In Proceedings of the 23rd international conference on Machine learning (pp. 977-984).

Wareham, J., Fox, P.B. and Cano Giner, J.L., 2014. Technology ecosystem governance. Organization Science, 25(4), pp.1195-1215.

Wulf, A. and Butel, L., 2017. Knowledge sharing and collaborative relationships in business ecosystems and networks. Industrial Management & Data Systems.

Wells, P., 2013. Sustainable business models and the automotive industry: A commentary. IIMB Management Review, 25(4), pp.228-239.

Worldwide Patents Database: PATSTAT, 2018. Available online:

https://www.epo.org/searching-for-patents/business/patstat.html (accessed on 7 March 2018).

Yun, J. J., Jeong, E., Lee, Y., & Kim, K.,2018. The effect of open innovation on technology value and technology transfer: A comparative analysis of the automotive, robotics, and aviation industries of Korea. Sustainability, 10(7), 2459.

Yun, J.J., Jeong, E. and Park, J., 2016. Network analysis of open innovation. Sustainability, 8(8), p.729.

Yoon, B. and Park, Y., 2004. A text-mining-based patent network: Analytical tool for high-technology trend. The Journal of High Technology Management Research, 15(1), pp.37-50.

Zanchi, L., Delogu, M., Zamagni, A., & Pierini, M.,2018. Analysis of the main elements affecting social LCA applications: challenges for the automotive sector. The International Journal of Life Cycle Assessment, 23(3), 519-535.

Zapata, C., & Nieuwenhuis, P.,2010. Exploring innovation in the automotive industry: new technologies for cleaner cars. Journal of Cleaner Production, 18(1), 14-20.

Zha, D., & Li, C., 2019. Multi-label dataless text classification with topic modeling. Knowledge and Information Systems, 61(1), 137-160.

Zhao, W., Chen, J.J., Perkins, R., Liu, Z., Ge, W., Ding, Y. and Zou, W., 2015, December. A heuristic approach to determine an appropriate number of topics in topic modeling. In BMC bioinformatics (Vol. 16, No. 13, p. S8). BioMed Central.

Zhang, Y., Jin, R., & Zhou, Z. H.,2010. Understanding bag-of-words model: a statistical framework. International Journal of Machine Learning and Cybernetics, 1(1-4), 43-52.

Zhang, W., Yoshida, T., and Tang, X., 2011. A comparative study of TF* IDF, LSI, and multi-words for text classification. Expert Systems with Applications, 38(3), pp.2758-2765.

Zhang, L., Zhao, Z., & Kan, Z.,2019. Private-sector partner selection for public-private partnership projects of electric vehicle charging infrastructure. Energy Science & Engineering, 7(5), 1469-1484.

List of Tables and Figures

Table 1a: LDA output example for Per-document topic proportion (θ_d) Lee& Kang, (2018).

	Topic 1	Topic 2	Topic 3	....	Topic K	Sum
Doc 1	0.20	0.50	0.10	....	0.10	1
Doc 2	0.50	0.02	0.01	....	0.40	1
Doc 3	0.05	0.12	0.48	....	0.15	1
…….	....	....	....	....	....	....
Doc N	0.14	0.25	0.33	....	0.14	1

Table 1b: LDA output examples for Per-topic word distribution (φ_k) Lee& Kang, (2018).

	Topic 1	Topic 2	Topic 3	…….	Topic K
Word 1	0.01	0.05	0.05	…….	0.10
Word 2	0.02	0.02	0.01	…….	0.03
Word 3	0.05	0.12	0.08	…….	0.02
…….	…….	…….	…….	…….	…….
Word V	0.04	0.01	0.03	…….	0.07
Sum	1	1	1	1	1

Table 2: Search queries for Automotive ecosystem

Tech	Search query
ICEV	IP=(F02D41/00 OR F02F7/00 OR F02D13/00 OR F02B2275/00 OR F02F1/00 OR F01M11/00 OR F02B1/00 OR Y02T10/00 OR Y02T10/10 OR Y02T90/50 OR Y02T10/70) AND Kywd1= (vehicle* OR car OR automobile*) AND Kywd2= ("internal combustion engine" OR "ic engine" OR "diesel engine")
HEV	IP=(B60W20/00 OR B60K6/00 OR B60W10/00 OR B60W2510/00 OR B60W2710/00 OR B60Y2200/00 OR Y10S903/00 OR B60W30/00 OR Y02T10/00 OR Y02T90/169 OR B60Y2200/92 OR Y02T90/168 OR Y04S30/10 OR Y02E60/721) AND Kywd1= (vehicle* OR car OR automobile*) AND Kywd2= ("hybrid vehicle*" OR "hybrid electric vehicle*" OR "hybrid propulsion") AND ("battery") AND ("charging stations")
BEV	IP=(H01M10/00 OR H01M2/00 OR H02J7/00 OR H01M2220/00 OR Y02T10/00 OR B60K6/00 OR B60K1/00 OR B60L2240/00 OR B60W20/00OR B60L2240/72 OR Y02T10/7005 OR Y02T90/10 OR Y02T90/12 OR Y02T90/16 OR B60L2240/70 OR Y02T 90/128 OR B60L2240/60 ORB60L53/00) AND Kywd1= (vehicle* OR car OR automobile*) AND Kywd2= (("electric vehicle*" OR "electric car" OR "electric automobile*") AND ("battery") AND ("charging infrastructure")
FCEV	IP=(B60W20/00 OR B60K6/00 OR B60W10/00 OR Y02T10/00 OR B60W2510/00 OR B60W2710/00 OR B60W30/00 OR H01M8/00 OR B60W2520/00 OR Y02T90/30 OR Y02T90/32 OR Y02T90/34 OR Y02T90/42) AND Kywd1= (vehicle* OR car OR automobile*) AND Kywd2= ("fuel cell*")

Table3: Top listed assignees for the ecosystem-related patents

Rank	Assignee	# of patents	Percentage
1	Toyota Motor Corp	8125	11.82%
2	Honda Motor Co Ltd	3977	5.82%
3	Nissan Motor No Ltd	2885	4.22%
4	Bosch Corp	2770	4.05%
5	Hyundai Motor Co Ltd	2597	3.80%
6	Ford Global Technologies Inc	1813	2.65%
7	BYD	1412	2.0%
8	Denso Corp	1097	1.60%
9	Renault SAS	859	1.25%
10	Hitachi Ltd	759	1.11%

Table 4: List of topics generated with their given labels

Topic #	Key terms/words	Competence Labels	Ecosystem
1	"engine", "torque", "clutch", "mode", "speed", "motor", "gear", "drive", "hybrid", "transmission".	Torque coordinated control in-vehicle systems	Focal technology
2	"voltage", "battery", "circuit", "pile", "module", "dc", "charge", "converter", "current", "cell".	Methods for battery cell charging system	Component technology
3	"charge", " lithium ", "automobile" + "utility", "power", "body", "vehicle", "battery", "box", "connect".	Lithium-ion battery charging in BEV and HEV.	Component technology
4	"charge", "information", "wireless", "data", "communication", "battery", "unit", "station", "power ", "management".	Charging infrastructure for BEV and HEV.	Complement technology
5	"valve", "combustion", "fuel", "oil", "cylinder", "pressure", "gas", "exhaust", "engine", "internal".	Fuel injection system of ICEV	Focal technology
6	"cell", "cool", "fuel", "air", "stack", "heat", "water", "fill", "temperature", "coolant".	Temperature controller development for water-cooled PEM fuel cell systems	Focal technology
7	"layer", "electrode", "catalyst", "interaction", "man", "ion", "carbon", "material", "evaluation", "node".	Catalytic oxidation in control of vehicle exhaust emissions	Focal technology
8	"powertrain", "photovoltaic", "signal", "pure", "event", "gradient", "vibration", "downhill", "slope", "current".	Automotive Control Systems for engine or vehicle	Focal technology

Table 5: Weightage of patent applications as per automotive ecosystem related topics

Topics number	Weightage of patent	CAGR of patent applications
Topic 1	34.30%	6.51%
Topic 2	16.52%	-0.77%
Topic 3	15.17%	1.53%
Topic 4	5.49%	1.71%
Topic 5	20.11%	2.58%
Topic 6	3.31%	-6.28%
Topic 7	3.93%	-8.26%
Topic 8	1.13%	-3.71%

Table 6: Classification of technological topics in the automotive ecosystem

Emerging technology topic (small patent weightage, positive CAGR) Topic 4	*Dominant Tech topic* (Large Patent weightage, positive CAGR) Topic 1, Topic 3, Topic 5
Declining technology topic (small patent weightage, negative CAGR) Topic 6, Topic 7, Topic 8	*Saturated technology topic* (Large Patent weightage, negative CAGR) Topic 2

Figure 1: Topic generation process using Latent Dirichlet Allocation as per Blei et al.(2003).

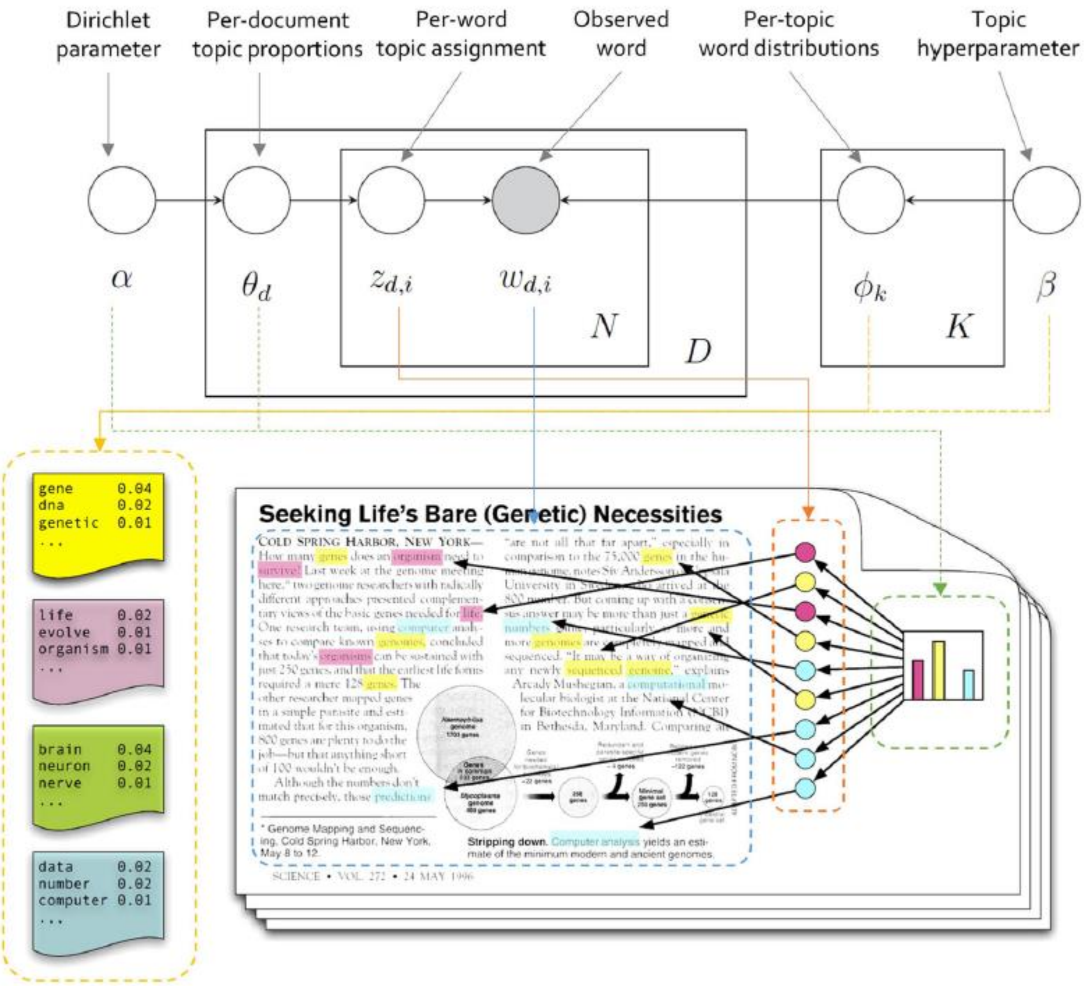


Figure 2: The overall framework for the proposed pipeline

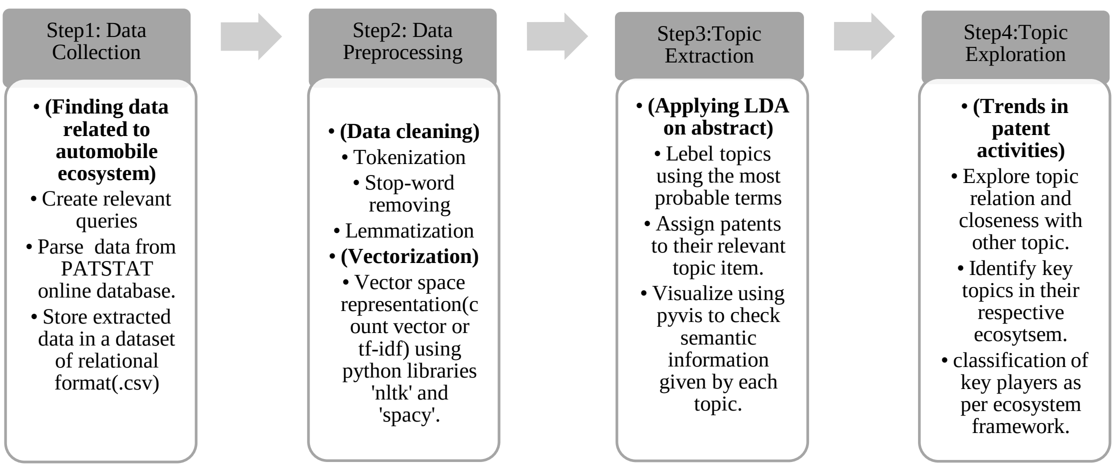


Figure 3: Data Preprocessing

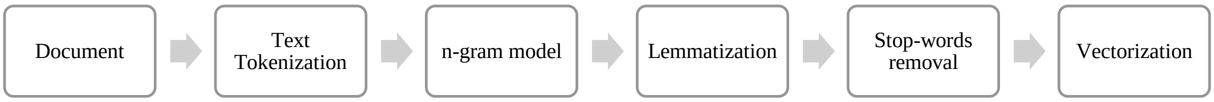


Figure 4: Distribution of patent applications over years

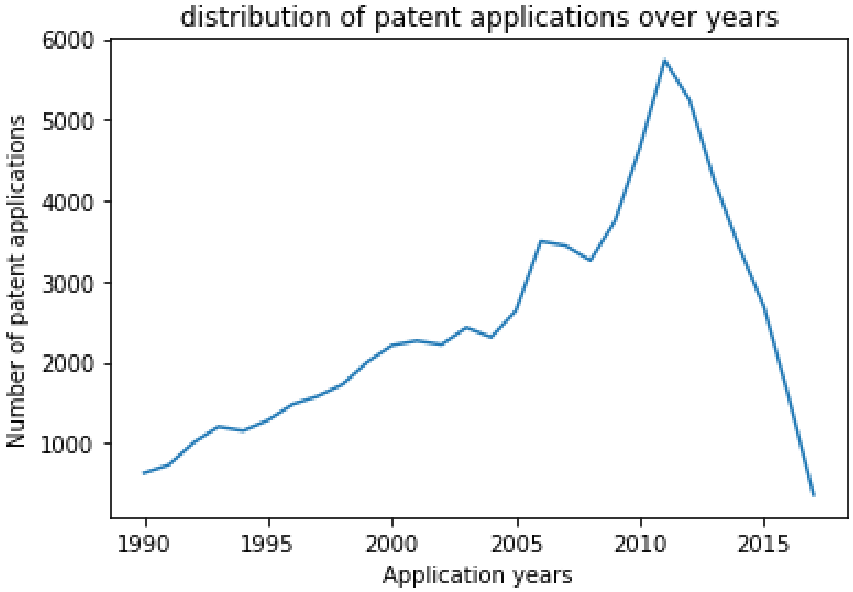


Figure 5: Calculating the maximum number of topics using the coherence score

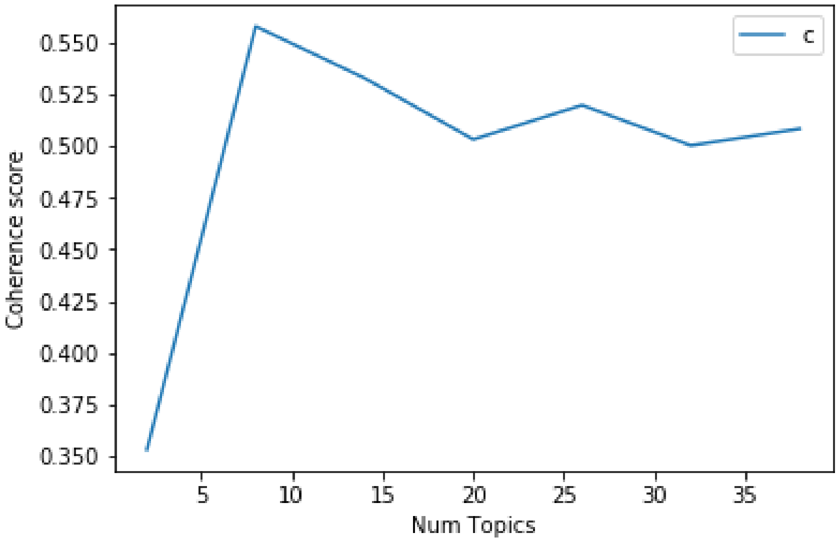


Figure 6a: Identification of Topic1 with most relevant terms

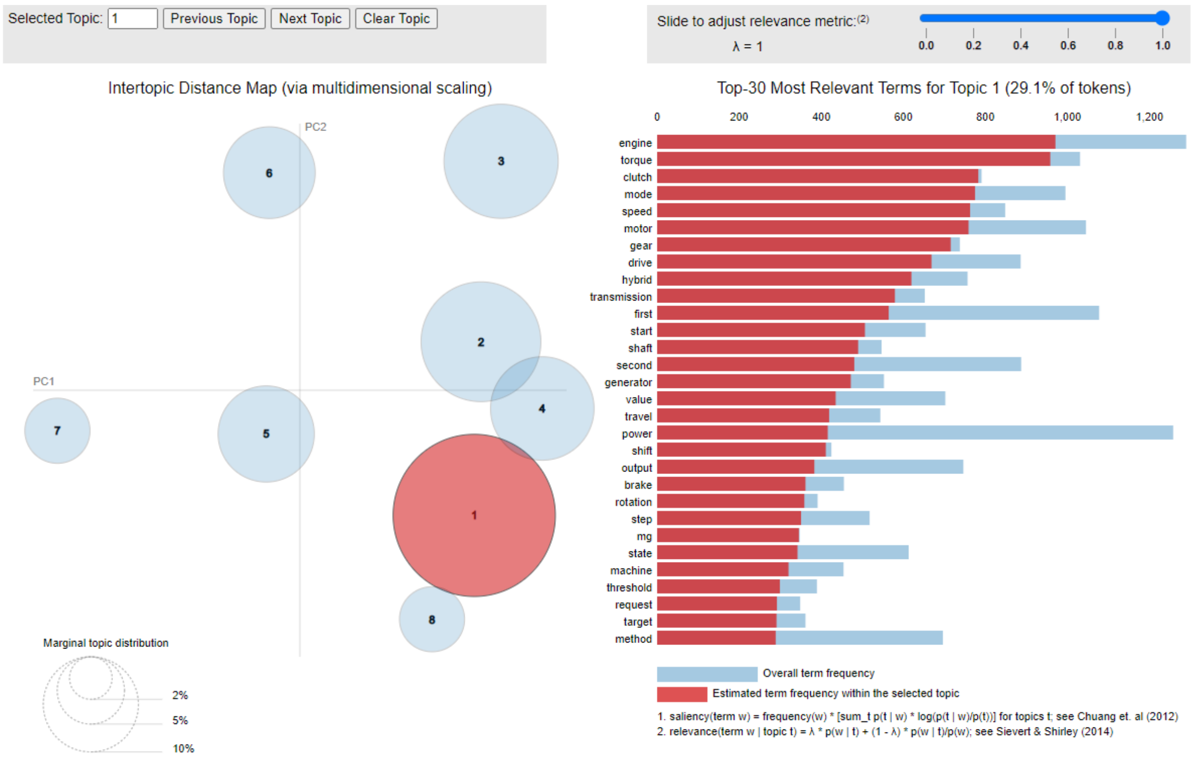


Figure 6b: Identification of Topic2 with most relevant terms

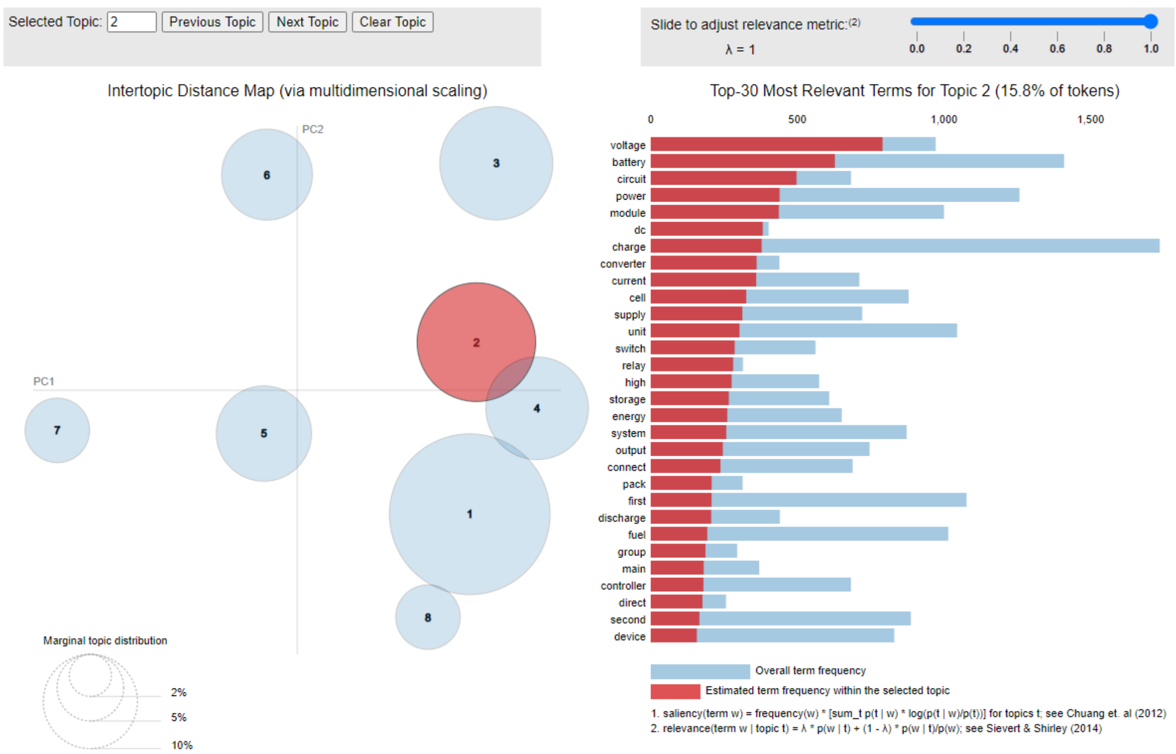


Figure 6c: Identification of Topic3 with most relevant terms

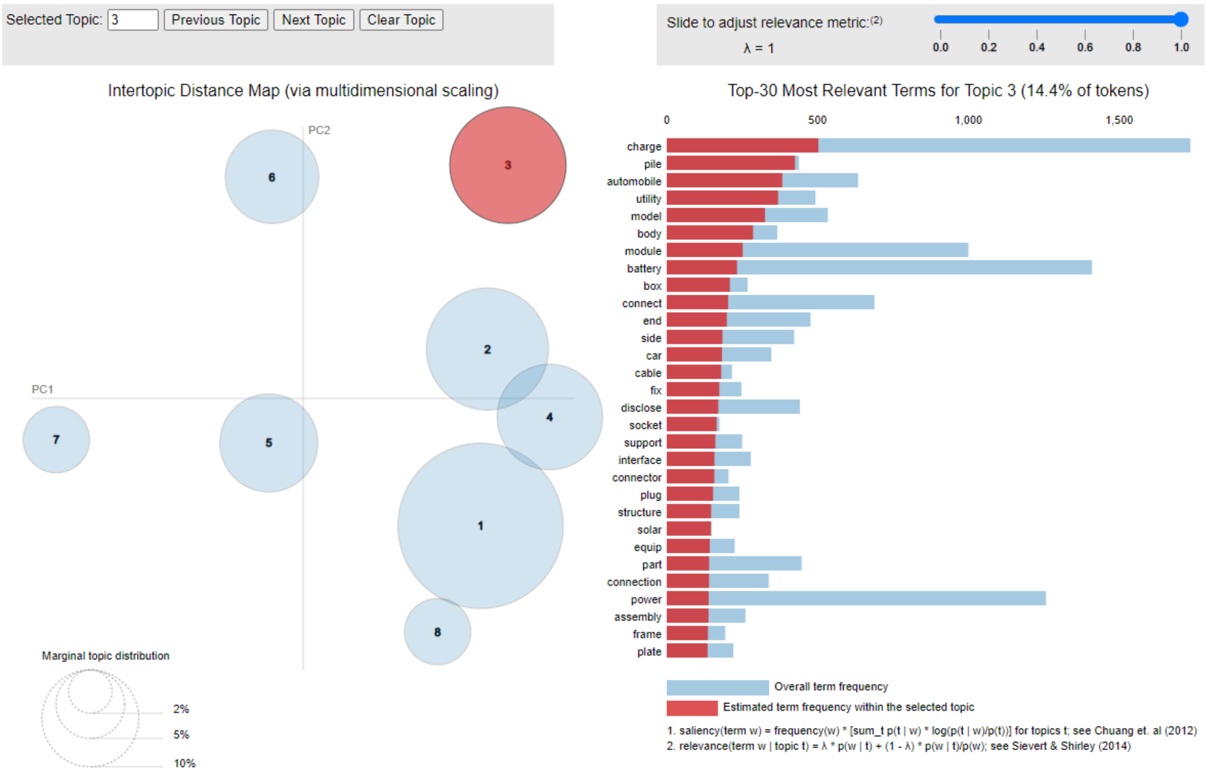


Figure 6d: Identification of Topic4 with most relevant terms

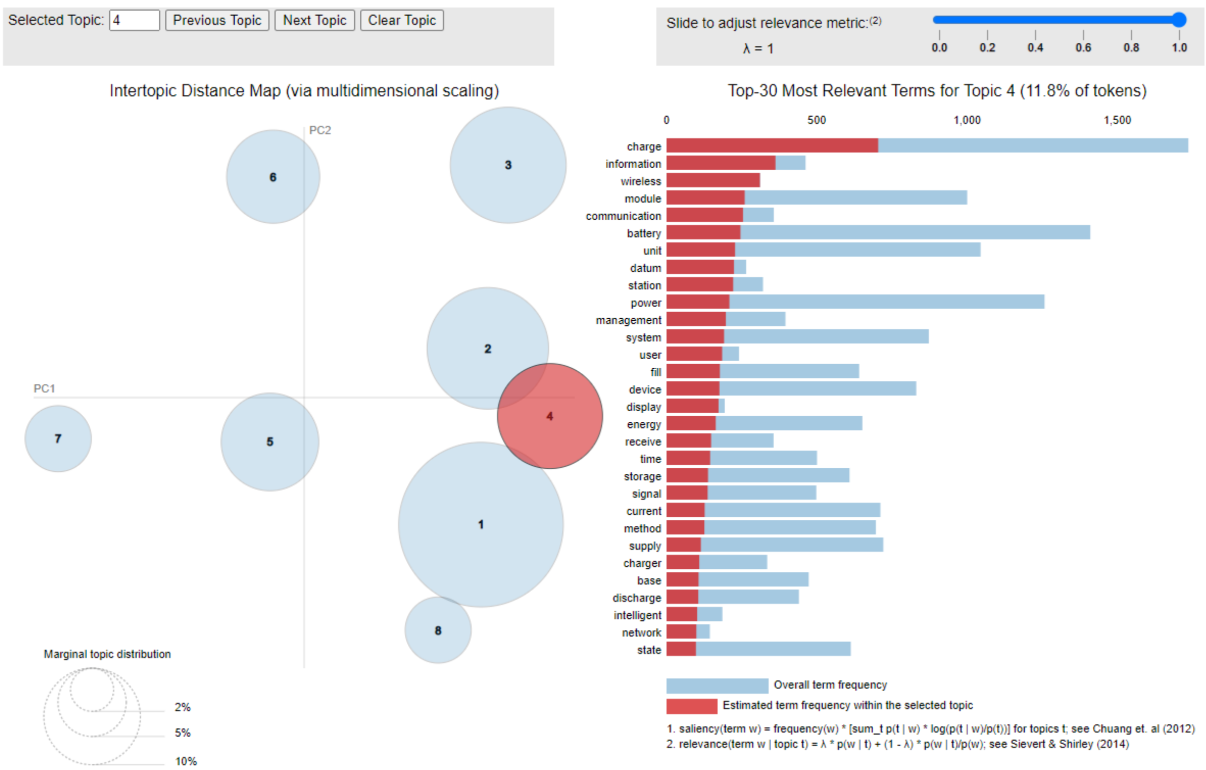


Figure 6e: Identification of Topic5 with most relevant terms

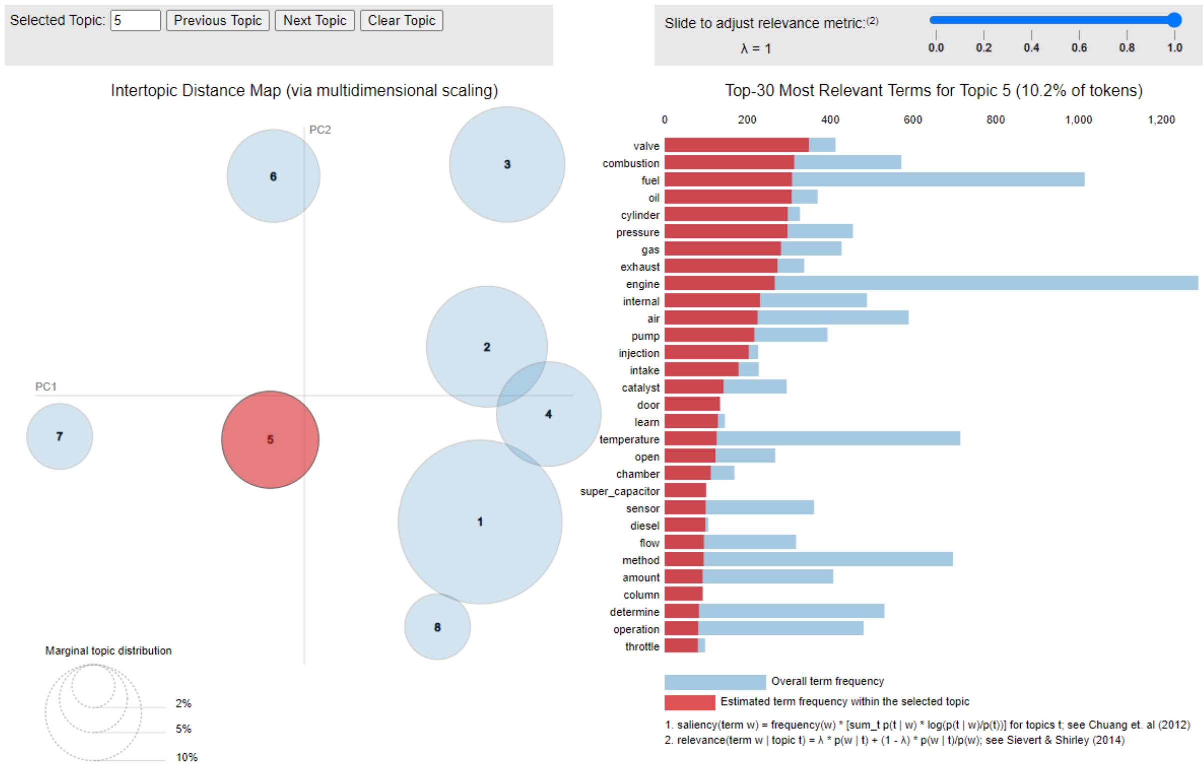


Figure 6f: Identification of Topic6 with most relevant terms

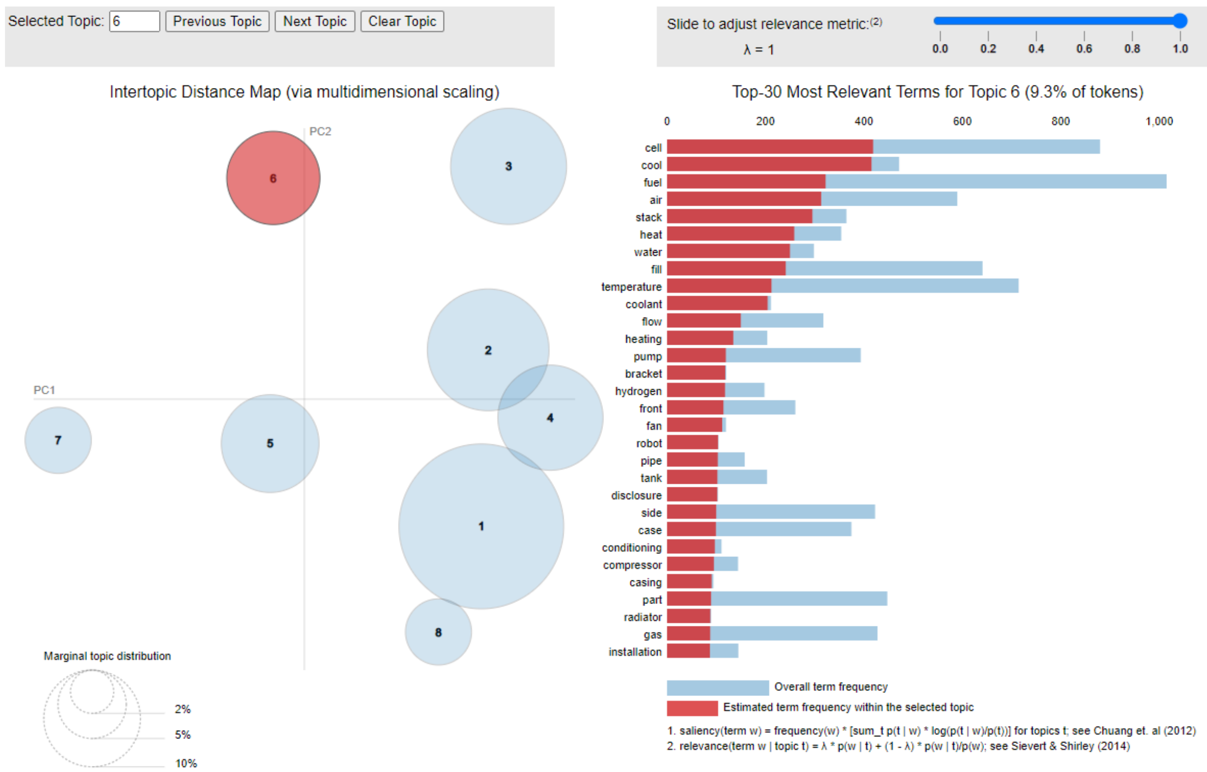


Figure 6g: Identification of Topic7 with most relevant terms

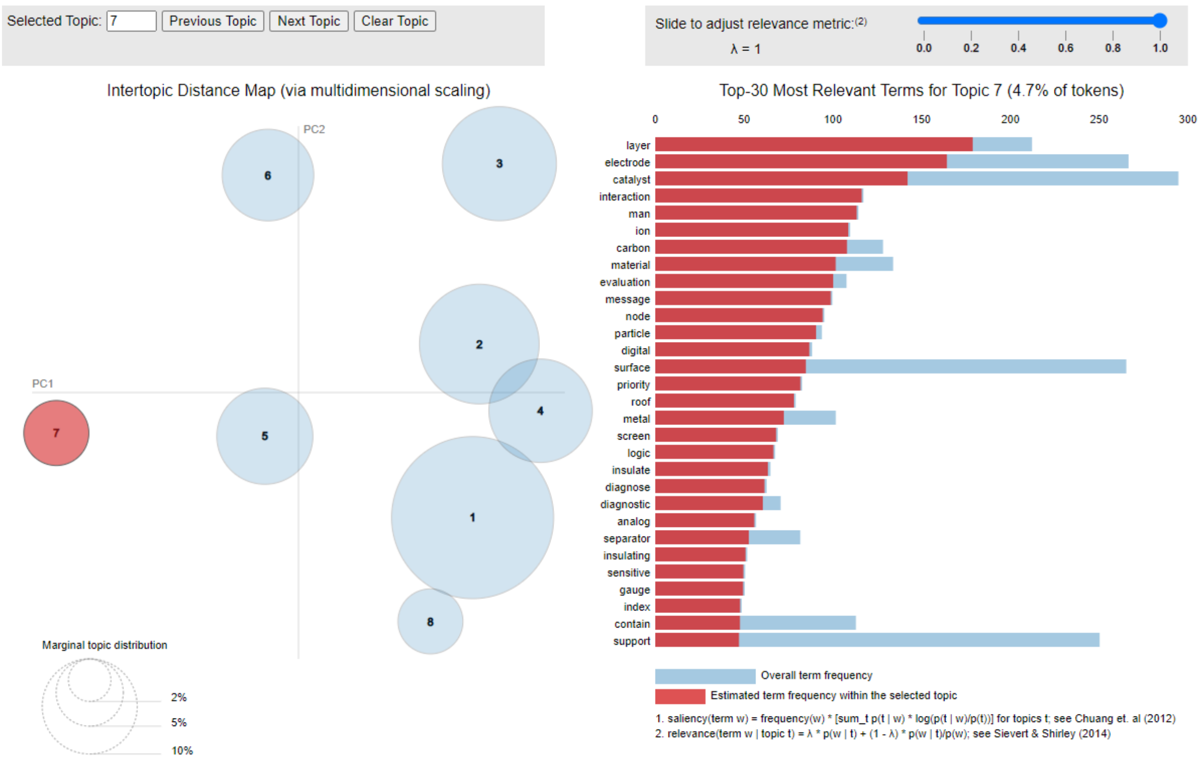


Figure 6h: Identification of Topic8 with most relevant terms

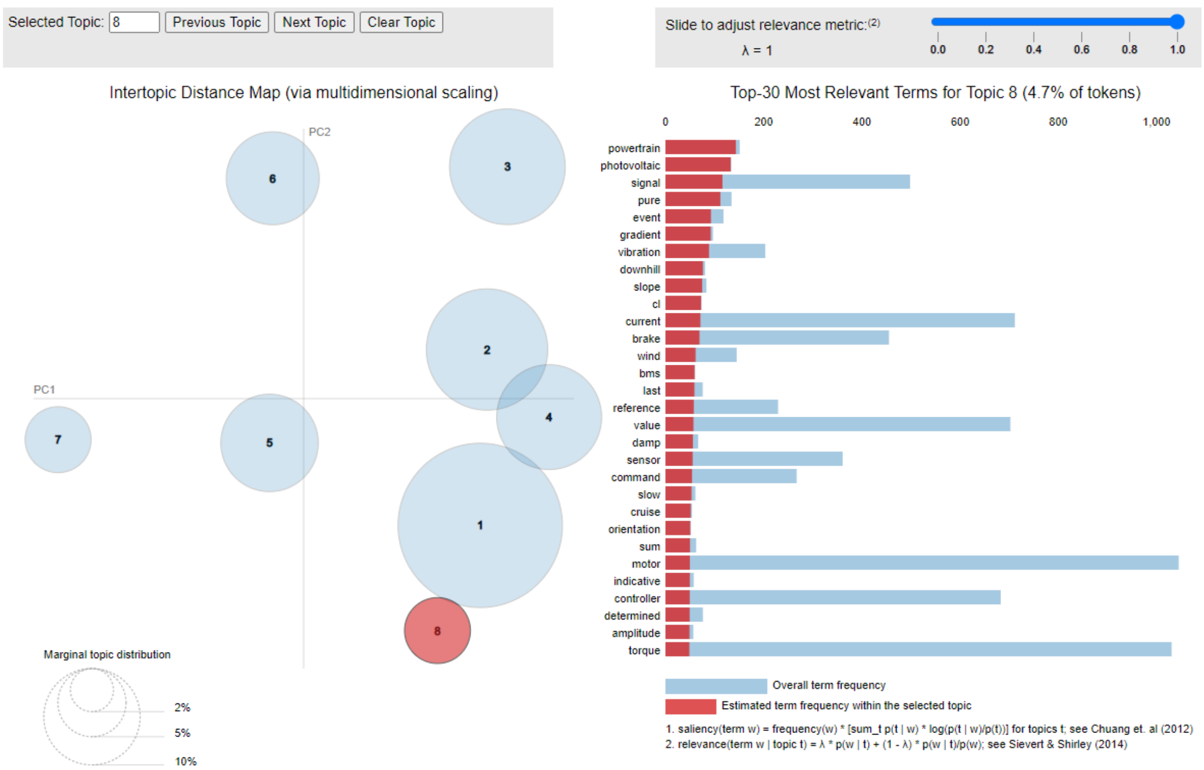


Figure 7: Distribution of Topics from 1990-2017

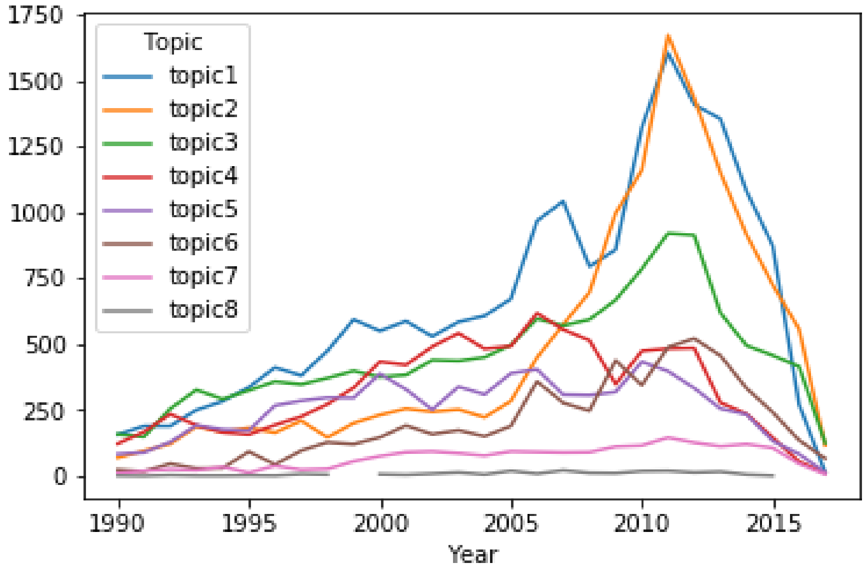
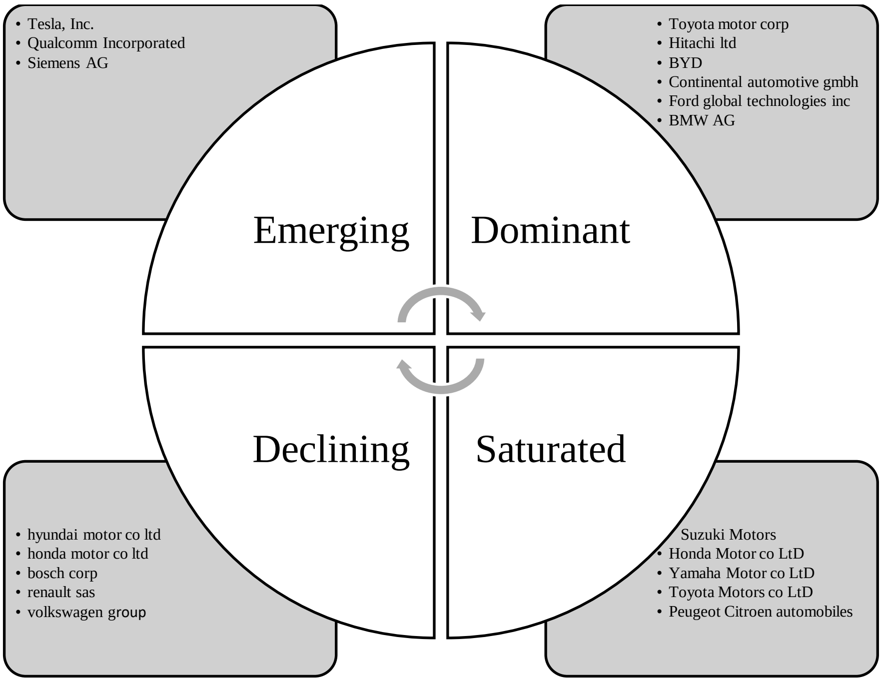


Figure 8: Major assignee in terms of the automotive ecosystem

Focal	Component	Complementors
•Toyota Motor Corp.	•BYD	•Toyota Motor Corp
•Honda Motor co. LtD	•Bosch Corp	•Hyundai Motor
•Nissan Motor co. LtD	•Panasonic	•Hitachi Automotive Systems
•Ford Global Technolgies	•Toyota Motor Corp	•Ford Global Technolgies
•Mitsubishi Motors Corp.	•Hitachi Automotive Systems	•Tesla, Inc.
•Bosch Corp	•Siemens AG	•Qualcomm Incorporated
•Denso Corp	•Ford Global Technolgies	•RWE AG
•GM Global Technologies	•Wanxiang	•Proterra, Inc.
•BMW	•Daimler AG	•Chargepoint Inc.
•Suzuki Motor Corp	•Continental AG	•NextEV USA Inc.

Figure 9: Major assignee in terms of the automotive ecosystem



Chapter 3

Analyzing the technological adoption through customers' perceived value: A case study for Autonomous Vehicle (AV)

Abstract

In the early stages of new technologies, there is always some ambiguity as to their potential effectiveness. Firms within the ecosystem may reduce this ambiguity by assessing the customer's sentiments when consumers ultimately decide to embrace or reject modern technology. We are identifying the key factors that affect customers' opinions is highly important for companies. Taking the ecosystem for Autonomous Vehicles, a representative case of technological innovations, in this paper, we are trying to recognize the risk and benefit perceptions that lead to adopting new technologies. We take a machine learning method for text classification and use extensive twitter data (455,727 tweets from June 2016 to January 2019) as our methodology.

We will discuss two performance determinants: risk rate and benefit rate, which allow estimating risk and benefit perceptions on Twitter. Our results suggest that organizations should take customers' perception of risk and benefit into account. Perception has a critical role to play in the success of the adoption of emerging technologies and particularly AV. The risk and benefit perception advocate further research on the adoption of autonomous vehicles (AV). In addition, we illustrate some of the most pivotal aspects of public understanding of AV and direct the management of such new technologies in order to increase their probability of adoption.

Keywords

Technology adoption, customers' sentiment, benefit perception, risk perception, autonomous vehicles, natural language processing, recurrent neural networks.

Introduction

The automotive sector research has been affected by machine learning over the years due to consistent technological evaluation in automation. It has seen multiple trials in automation over its life cycle till the very recent artificial intelligence surge. It is vital to have a vehicle in development with keeping efficiency and safety as the key (Kyriakidis et al., 2015). In the last decade, we can identify rich research that is focused on the technology of autonomous vehicles

(Kyriakidis et al., 2015). Automotive firms, suppliers, data aggregation, and related IT industries comprise the autonomous vehicle ecosystem actively pursuing this scholarship. Autonomous vehicles are also sought by highly qualified IT companies such as Tesla, Google, and Apple (Fagnant and Kockelman, 2015; Spinrad, 2014). However, an emerging technology's significant key factor in technology adoption by its customers (Davis et al., 1989; Venkatesh et al., 2007).

In recent years, autonomous vehicles (AV) have become a controversial topic. Multiple media outlets such as newspapers, television, and social media have discussed the coming trend in the autonomous vehicles’ ecosystem. Concerns associated with regulations of driving (Zmud and Sener, 2017), centered on who drives and assumes responsibility for accidents, are issues of heated debates. Nevertheless, many sectors are not limited to the transport and automotive industries, exploring autonomous vehicles' possibility (Gogoll and Müller, 2017). These sectors seem convinced that autonomous vehicles (AV) will be the future of mobility. In such technologies, concerns coming from the public should be of key importance; however, we have seen a trend of innovation and investment in technology while public perception being somewhat underestimated (Piao et al., 2016). It is also important to point out that opinions coming from experts of a domain may differ from the public (Blake, 1995). Conover (1994) argues that risk and benefit perceptions could be very different from the real implications of technologies using the first intelligent vehicle handling system. Research on other technological innovations also indicates that perceptions of risk and benefit are essential factors for public adoption of technology (Butakov and Ioannou, 2015; Gogoll and Müller, 2017; Pendleton et al., 2015). Therefore, public perceptions eventually evaluate whether autonomous vehicles (AV) would dominate the upcoming transport landscape. Thus, Public opinion is a crucial factor to be considered, especially for the initial adoption of technological innovations (Bansal et al., 2016). Studies discussing the public understanding of technological innovations and their adoption, such as autonomous vehicles (AV) technology, remain limited across several countries and may shift over time (Butakov and Ioannou, 2015). Initially, We address this lack of study by outlining the outcomes of previous studies on public sentiment and perception of autonomous vehicles. Next, we build an approach to quantify and respond to public perceptions that facilitate customers' sentiment. This strategy enables us to use the large quantities of knowledge publicly accessible through social media to predict the adoption of emerging technologies. In particular, this paper seeks to answer the following important and yet open questions:

- To predict technology adoption, how could we assess public expectations towards autonomous vehicle (AV) ecosystems?
- How do incidents impact the end consumers' perception of the autonomous vehicles ecosystem among the public?

We create an approach to automatically recognize and track the perceived perception, i.e., risks and benefits of new technologies. Investigation for consumers' perception is conducted from short 280-character text messages posted on the social media site called 'Twitter' to find persuasive answers to our proposed questions. We build on scientific literature and text mining methods (McCorkindale, 2010; Tan, 1999) that enable the extraction of information from text documents. We use Twitter to collect a stream of opinions about autonomous vehicles, one instance of currently emerging technology. Based on autonomous vehicles' risk and benefit perception, we consider major advances and obstacles in the future adoption of this new technology and the direct emerging technologies.

Literature background

This section provides an overview of current literature on the significance of adoption towards autonomous vehicles (AV) from both innovation ecosystem (Adner, 2016; Jacobides et al., 2018) and public adoption perspectives (Butakov and Ioannou, 2015; Gogoll and Müller, 2017; Kohl et al.,2018; Pendleton et al., 2015). We present an introduction to technology adoption literature, the current scientific knowledge of the autonomous vehicle ecosystem, and studies assessing autonomous vehicles' adoption (AV). By summarizing the theoretical context for our analysis through propositions, we may conclude this section.

Theoretical background for Technology adoption

A significant aim of the innovation literature is to explain the dynamics of technology adoption. The standard embodiments of the adoption of technology have been analyzed by technology lifecycle, i.e., S-curves (Foster, 1986). The approach to the technology life cycle holds that the degree of performance change in a given technology for a fixed-time unit is relatively limited during the early stages of production. The pace of improvement increases until the stage of maturity as the technology is clearly represented, at which point the technology hits its limits and the output effect of additional sought to improve the technology is subject to diminishing returns.

We will combine technology adoption from both innovation literature and the information system research stream. In this paper, we will focus on end customers instead of firm-based view (Bharati and Chaudhury,2006; Bruque and Moyano, 207; Dasgupta et al., 1999; Woiceshyn and Daellenbach, 2005). Therefore, the technology adoption in our study is close to the technology acceptance model (Venkatesh et al., 2007) and is a vital source of research endeavors (Davis et al., 1989). The adoption model for technology describes and forecasts when and why technology will be used by individuals based on their consumption behavior (Fishbein and Ajzen, 1975), perceived ease of use(Lancelot Miltgen et al., 2013), and perceived usefulness (Davis et al., 1989). We may describe perceived usefulness or benefit as the likelihood of a particular technology system improving the efficiency of the consumer for a given task. The perceived ease of use represents the effort a user requires to solve a given task while using the technology framework. The principle of the acceptance model of technology (Davis, 1989) in the same spirit of technology adoption (Butakov and Ioannou, 2015) is that perceived ease of use determines the strength of the customers' behavioral intention to use a specific technology against a given problem. The behavioral intention then leads to actual usage, as defined in the reasoned action theory by Fishbein and Ajzen (1975). The technology adoption model (Davis et al., 1989) has its roots in the theory of reasoned action, which is considered an influential theory within social psychology scholarship.

Several researchers have expanded the technology adoption model to take into account the value of risk perception for user adoption (Davis, 1989; Venkatesh et al., 2003; Zmud et al., 2016). Martins et al. (2014), on the other hand, inspect the adoption of internet banking (Lancelot Miltgen et al.,2013) and conclude that the perception of risk is crucial for organizations. They find that privacy risk and the risk of being subject to internet banking fraud are essential for accepting Internet banking services. Lancelot Miltgen et al. (2013) study the end-user adoption of biometrics and find that privacy risk is vital for accepting biometrics. These studies demonstrate that determining the perception of risk involves domain expertise in order to recognize the related risks to a specific technology. However, it is not included in any focal technology, despite some studies considering risk perception as an additional factor (Venkatesh et al., 2003). The understanding of risk depends on the new technology itself and is therefore tricky to de-escalate (Sharma et al., 2009).

The core feature of risk perception for end-users 'adoption is recognized by public adoption studies. Prior research indicates that people have rejected many innovations because of cultural controversies, creating adverse commercialization effects (Gupta et al., 2012). Rejection of autonomous vehicles' ecosystem may have severe consequences for organizations since there is a lot on stake. We can find recent projects with considerable investments in R&D of this technology (Hohenberger et al., 2016). The events and reported accidents with autonomous vehicles, such as the recent Google autonomous vehicles' accident (Salon, 2020), could lead to apprehension and hesitation to adopt this new technology (Hohenberger et al., 2016). Even modern ICT companies and conventional automotive companies are investing in autonomous vehicle research and development, and they may suffer severe financial harm. If this innovative technology fails to find wide adoption, they might not get expected returns on their investments.

There is a unique model to explore technology adoption in the customers'' acceptance proximity introduced by Siegrist (2000) and studies in the information system scholarship. It primarily focuses on the relationship between risk and benefit expectations, confidence, and public adoption. It has been discovered that trust affects expectations of risks and benefits, which directly affects the adoption of technology (Siegrist, 2000). It is crucial to distinguish between real risks and benefits and their expectations in order to understand the conceptual basis for the technology acceptance model. Slovic's (1987) seminal work explains the understanding of threats associated with new technologies unknown to most individuals and incomprehensible. To analyze the risks more critically, instead of personal testing or technologically sophisticated study, people rely on emotional judgments based on media reporting (Griffin and Hauser, 1993). Such rulings may also be vulnerable to heuristic biases that may not lead to optimal or logical choices (Kahneman and Tversky, 1979).

However, according to Alhakami and Slovic(1994), the definition of risks and benefits is unclear, implying that when analyzing emerging technologies, people do not distinguish between risks and benefits. Thus, when predicting their actions, we should not expect individuals to make reasonable decisions based on evidence but rather consider their perceptions. Trust, which decreases cognitive complexity when assessing emerging technology, often affects these attitudes (Earle and Cvetkovich, 1995). Instead of evaluating them, individuals trust other organizations to correctly analyze and implement new technologies. As far as autonomous vehicle technology is concerned, it is crucial that regulations and governance bodies are trusted by individuals. This faith will ensure that autonomous vehicles are safe for drivers and the public on the roads (Choi and Ji, 2015).

The risks and benefits of an evolving technology viewed by non-professional personnel can differ dramatically from the risks and benefits of technology experts (Bongaerts et al., 2016). Due to non-professional perceived risks, emerging technologies and products can cause excessive anxiety and resistance to their use (Zmud et al., 2016). We have an example where innovative technologies such as nanotechnology and genetically modified food have failed to be embraced in the public domain. Simultaneously, the advantages outweigh the risks from a scientific point of view due to subjective perceptions (Gupta et al.,2015). In an early stage of product growth, recognizing potential risks and benefits, as is currently the case with autonomous vehicles, enables businesses to behave ambidextrously (Duncan, 1976; Christensen,1997); that is, to change their main innovations to comply with existing market demand.

Both innovation studies and information system literature depend primarily on questionnaire surveys to predict and clarify technology adoption (Sharma et al., 2009; Venkatesh et al., 2003). For each construct, a survey typically consists of several components that have been tested in prior research and tailored to the application domain (Venkatesh et al., 2016). The survey-based questionnaire is then administered to a population sample and analyzed after collecting data using econometrics methods such as regression, structural equation modeling (SEM). With such restrictions, a time-consuming and laborious outcome is achieved. The questionnaire, for instance, is vulnerable to common process bias. Artifactual covariance, a major validity threat for survey-based social sciences research fields, is introduced by this process bias (Sharma et al., 2009).

Theoretical background for Autonomous Vehicles (AV) Ecosystem

An innovation ecosystem is comprised of focal technology, component technology, and complementary technology (Adner, 2010). Currently, all key stakeholder of the autonomous vehicles' ecosystem is vying to be the first on the market. They see the tremendous potential and the technical challenges of this new technology, but they weaken the adoption of customers (Rogers, 2003). As in the case of autonomous vehicles (AV) technology, driving automation is categorized into different automation levels. Three widely used descriptions of these levels are present. Five levels of driving automation (Gasser and Westhoff 2012; NHTSA 2013) are defined by the German Federal Highway Research Institute (BASt) and the US National Highway Traffic Safety Administration (NHTSA), while six levels of automation are defined by the Society of Automotive Engineers (SAE International, 2014). The concepts are identical in addition to the distinct number of degrees of automation. A comparison of the three descriptions is given by Kyriakidis et al. (2015). In order to exemplify the various levels of automation, we use the BASt concept for our analysis. All definitions aim to differentiate between driving automation systems providing driver only, assisted, partially automated, highly automated, and fully automated. Table 1 summarizes driving automation stages.

--- Insert Table 1 about here ---

Modern automation systems for driving, such as the Tesla Automatic Autopilot, require drivers to monitor the vehicle at any time, irrespective of current conditions. Therefore, they need to be considered partially automated, which, according to BASt definitions, only offers level 3 driving automation. The term autonomous vehicle is, however, generally affiliated with these vehicles. Note that we use the word 'fully automated' after Gasser and Westhoff (2012) when referring to level 5 automation. According to Yadron and Tynan(2016), The major downside to the use of a quasi-autonomous Autopilot driving automation system at level 3 is that the driver must be in a position to regulate driving at all times. As per Krok(2015), by leaving the driver's seat when driving on public roads using a level 3 driving automation system, drivers often disregard this requirement. The driver finds it difficult to get back into the loop and react appropriately to such traffic circumstances. Previous studies (Gold et al., 2013; Kork, 2015; Körber et al., 2016) show that unrealistic benefit perceptions can negatively impact the safety of individuals and, therefore, indirectly affect negatively to public adoption.

With 1,533 respondents, a public opinion survey on autonomous vehicles in the U.S., the U.K., and Australia reveal that 56 percent of individuals have positive views on autonomous vehicles. In contrast, 13.8 percent share negative reservations, and 29.4 percent are neutral about the subject (Schoettle and Sivak, 2014). The Consumer Technology Association said that 70 percent of U.S. drivers show a desire to test autonomous vehicles. Around 60 percent of car drivers show a willingness to replace their cars with fully autonomous vehicles (Markwalter, 2015). Payre et al. (2014) conduct a survey of 421 French drivers and have found that 68.1% are prepared to use autonomous vehicles (Markwalter, 2015; Payre et al., 2014). Supporters argue that autonomous vehicles will reduce car accidents by a significant amount because 93 percent of car accidents are due to human errors (Treat et al., 1977; Fagnant and Kockelman, 2015). However, adversaries of this view argue that autonomous vehicles could pose new and currently unknown risks, such as device failures or behavioral compensation. Schoettle and Sivak (2014) conclude that autonomous vehicles might be no safer than an average driver.

Thus, if self-driven and human-driven vehicles use the same roads, AV can increase the number of total vehicle accidents.

Generally, surveys show that individuals understand autonomous vehicles, although they only know a bit about them (Fraedrich et al., 2016). Previous research shows that the benefit perception positively affects the adoption of technology (Hohenberger et al., 2017; Kohl et al., 2018). However, focusing solely on the advantages of autonomous vehicles may not be a viable strategy to increase their initial adoption. If autonomous vehicles are commonly available when used in active cruise control, individuals can begin to identify potential safety problems and risks. With the introduction of active cruise control in production cars, the public began to understand their lack of control, resulting in a lack of adoption (Eckoldt et al., 2012). Therefore, the threats and disadvantages of autonomous vehicles need to be conveyed to car manufacturers. There is a need to prevent or even counteract myths regarding both threats and advantages.

Implementing questionnaires, which may not be appropriate for innovative technologies such as autonomous vehicles, is the prevalent tool for assessing adoption. In such scenarios, Fraedrich and Lenz (2014) suggest a questionnaire respondent who possibly has neither detailed knowledge nor experiences of autonomous vehicles that lead to biased outcomes. Therefore, rather than using defined questionnaires, the conduction of exploratory research (König and Neumayr, 2017) is recommended in this field at an early stage.

Research Propositions

The customers' perception followed through sentiment provides fascinating research possibilities for innovative technologies (Griffin and Hauser, 1993). According to Ward et al. (2017), the expectations of risk and benefit of potential customers are likely to play a key role in the adoption of autonomous vehicles. In order to recognize any emerging technology problems or public adoption, risk and profit perceptions should be closely monitored. Prior research has performed exploratory analysis for text data on the expectations of risk and benefit of (Bazilinskyy et al., 2015) and has shown that this is a promising method (Fraedrich and Lenz, 2014; Kohl et al., 2018).

Recent studies have placed considerable effort into the manual coding of all data but have struggled with this very long and time-consuming strategy. (Bazilinskyy et al., 2015), which could still have biases. Kauer et al. (2012) recommend researching expectations over time, as they are likely to evolve as the public becomes more aware of this technology. Thus, the effect of critical accidents involving autonomous vehicles on public perceptions will be an interesting theme to research. We present the following proposals to answer our research questions in order to address these results and recommendations of previous research:

1) Social media trends associated with benefits of autonomous vehicles such as relaxing driving experience, increased safety, and reduced mobility costs, increase benefit perception of autonomous vehicles.

2) Social media news concerning the risks of autonomous vehicles such as fatal accidents, security concerns increases the risk perception of autonomous vehicles.

Research Methodology

We use a new method to define benefits and risks, as Kohl et al. (2018) suggested. We get exciting results by analyzing the vast amount of existing data about the autonomous vehicle on Twitter. The theoretical foundation from Neuendorf (2016) using quantitative content analysis, which facilitates qualitative evidence-based analysis and enhances previous qualitative approaches (Bazilinskyy et al., 2015). Although unstructured social media knowledge has traditionally used methods of content analysis (McCorkindale, 2010), we use machine learning to automate much of the coding process. In marketing research, this approach is similar to sentiment analysis (Okazaki et al., 2014). It has the advantage that by using the supervised text classification method, only a small portion of the data requires manual coding. Using this approach, we avoid survey-related issues such as common method variance (Sharma et al., 2009), as discussed earlier.

For sentiment analysis, we adopt the analysis process suggested by Okazaki et al. (2014). It includes data extraction, data planning, preprocessing of data, implementation of the model, and validation of the model. Our risk and benefit perception analysis methodology is close to sentiment analysis, which helps us to adopt a traditional method of automating customers' sentiment. We have some variations in the process of sentiment analysis b integrating the steps of model generation and validation into one phase as per (Feldman, 2013). Our analysis process is given as follows:

First, we get tweets relevant to autonomous vehicle technology using the Twitter Search API. This stage includes data extraction, data planning, and data annotation. Second, we process the text data in tweets to improve data quality, decrease data noise, reduce dimensionality, and avoid misclassification. This stage is referred to as data preprocessing. Next, we construct the deep learning model (recurrent neural network) and then test it using model validation methods. Fourth, we predict out of sample data using our developed model to classify the out of sample tweets. In order to answer our research questions, we then evaluate the classified tweets qualitatively and quantitatively.

Data extraction

Twitter has also proved to be a reliable source of knowledge for the prediction and tracking of various phenomena and case studies (Tumasjan et al., 2010), from political election to outbreaks of disease (St Louis and Zorlu, 2012). Twitter users have a per-message limit of 280 characters. Despite this restriction, tweets contain valuable natural language encoded details (Pak and Paroubek, 2010). The extraction of this information from the overwhelming amount of noise present on Twitter is an enormous effort. To extract information from tweets, we draw on previous results from sentiment analysis (Medhat et al., 2014; Okazaki et al., 2014; Pak and Paroubek, 2010) and text data classification (Kohl et al.,2018; McCorkindale, 2010; Neuendorf, 2016). As sentiment analysis is not explicitly applicable to the extraction of perceptions of risk and benefit, we need to expand previous approaches. Traditionally, it only attributes to a given comment a polarity (Medhat et al., 2014), i.e., a positive or negative sentiment.

New Twitter technologies involve tweet bots (Haustein et al., 2016) that are difficult to distinguish from real individuals and devices that interact through Twitter through the Internet of Things (Kranz et al., 2010). The findings of Twitter analyses, therefore, require careful consideration. In particular, Twitter bots are becoming increasingly good at mimicking human speech and writing patterns. Researchers are worried about the large-scale penetration of "social bots" that can hardly be discerned by humans (St Louis and Zorlu, 2012). A quantitative analysis is carried out by social bots, such as the analysis of tweet counts, not just from Twitter but also from other social media such as Facebook (Haustein et al., 2016; Boshmaf et al., 2011) is becoming very challenging to differentiate. We shall compare our analysis results in-line with previous studies (Kohl et al.,2018; Okazaki et al.,2014; Tumasjan et al., 2010) to detect any possible distortions of the Twitter data.

Another problem with tweets is that they are not available directly to researchers. In this research, we only collect data samples that the Twitter Search API returns, calculated by optimization algorithms, and are not a representative sample of the total tweets (Ruths and Pfeffer, 2014). It is noteworthy that the user-base over Twitter is not a representative population sample. It could have been a concern for data validity but, social media platforms, including Twitter, have a large and diverse audience from various social and interest groups (Pak and Paroubek, 2010) and hence, is a valuable tool for assessing the opinions of people. We anticipate Twitter users to be more open to emerging innovations such as autonomous vehicles, which, based on representative population samples, could lead to slightly more favorable results than previous surveys (Twitter, 2020).

Despite the drawbacks of Twitter mining, our methodology helps millions of active Twitter users to reach the feelings of customers. This results in substantially more claims about the adoption of autonomous vehicles than in previous studies. We use a deep learning algorithm to automate the classification of tweets related to autonomous vehicle technology to deal with the extensive amount of data from Twitter. We thus avoid the painstaking manual coding of qualitative information, thus making this study possible.

Data Planning

Tweets written in English concerning autonomous vehicles collected using the Twitter Search API are part of our dataset. (Twitter, 2020). For our data, we choose those organizations that are leading in autonomous vehicle technology. Additionally, we take the country and region of those organizations into consideration because there are very few countries with proper infrastructure for autonomous vehicle testing on the road. We use beautiful soup API to scrap tweets not older than one week (Twitter, 2020). A realistic longitudinal study, however, involves the ability to collect tweets for longer periods through the regular collection of tweets and storage in a database. For the last seven days, a free edition of tweeter accounts helps us to fetch tweets, so we have collected tweets regularly. We use a comma-separated values (CSV) data file to save the tweets returned by the Twitter Scraping API. This dataset includes a unique tweet identifier, timestamp of given tweets, quarter details of those time periods, date of creation of tweets, user ID, username, screen name, and the message body, including text. The tweets are then moved to an in-memory database to process them effectively. We use IPython Jupyer Notebook integrated development environment (IDE) to process data using python modules such as request and pandas to read the twitter-based CSV dataset. We start data collection for this analysis on June 6, 2016, with the last tweets on January 18, 2019. We use the following query combining key technology and company-related keywords for our Twitter Search API requests inspired by a previous study on autonomous vehicles (Kohl et al., 2018):

Self-drive OR driverless OR autonomous OR tesla OR apple OR ford OR google OR Waymo OR Opel OR gm OR general motors OR Volkswagen OR VW OR Daimler OR Mercedes OR Benz OR BMW OR Audi OR Porsche OR Nissan OR Toyota OR Honda OR Suzuki.

Until data collection, we correct the above search queries. They consist of a mix of keywords related to the subject, including names of US-based companies actively working on autonomous vehicles and U.S. car manufacturers and active German and Japanese leading car manufacturers. For our case, company-related queries result in many tweets that are not concerned with autonomous vehicles.

We then used regular expressions to filter out unrelated data samples. To maintain the tweets containing our necessary words, such as driverless, self-driving, autonomous driving, automated driving, autonomous vehicle, and automated car, we add a regular expression module. We also include a slight variation of the term with the help of regular expressions, such as “driverless” or “driver-less.” With the given queries and filtering method, we eventually get 455,727 relevant tweets.

We select a subset of tweets for training a text classifier model. For this purpose, we use a dataset of 15,000 tweets labeled as beneficial, risky, and neutral. Top tweets are popular tweet trends that many other Twitter users have engaged with (Twitter, 2016c). Analyzing such popular tweets relating to autonomous vehicles, we get an overview of the discussion of this topic on Twitter, which helps to acquire knowledge of the latest trends related to our case study. We refrain from evaluating these tweets because they represent just a tiny fraction of the actual tweets released from June 2016 to January 2019 and are probably highly skewed by Twitter's proprietary selection algorithms. We only use them as training data for machine learning classification. As shown in Table 2, In 43.4% of the tweets in the training dataset (i.e., 6513 tweets), no information about risk and benefit perceptions is present and is, thus, categorized as “neutral.” Tweet class.

--- Insert Table 2 about here ---

Data annotation

This labeling text data process is significant to build an accurate training dataset for text classifiers. A machine learning algorithm is an iterative operation, and it is, therefore, usual to review and change labels until a reasonable level of label accuracy is achieved. In practice, To review and update data labels, researchers construct custom annotation systems as correctly labeled data is crucial for model quality. The text classifier cannot learn the ground truth efficiently if there are problems with the names, leading to incorrect predictions. One approach that researchers have used to enhance their label data accuracy (Zhang & Yang, 2015) is through audit workflows. Audit workflows allow a group of reviewers to check the accuracy of labels and adjust them if required. Instead, we use a built-in audit workflow applying the Amazon SageMaker Ground Truth tool. It performs both label verification and label adjustment for semantic segmentation. This approach makes sure that labeling instructions and rules are consistent among all reviewers. This method is also valuable for reviewing jobs made by other reviewers involved in the labeling process.

--- Insert Figure 1 about here ---

Model Implementation

We use a natural language processing pipeline in-line with (Graliński et al., 2013). We analyze this process starts with unstructured text data as input for identifying significant patterns. There might be special characters, new phases, or punctuation symbols that do not contribute to the context. Such unwanted entities create noise. Hence, we clean the data by removing these patterns. Then, the data is converted into features to verify the cleaning procedure. After the feature extraction phase, the data is transformed into matrices keeping syntactic and lexical meaning using ‘word embedding.’ We feed this data in the form of matrices to our machine learning model. We use LSTM (Long short-term memory) algorithm to predict our tweets’ perception state.

--- Insert Figure 2 about here ---

Data Preprocessing

We clean the twitter data by removing noises to create consistent input for our model and reducing dimensionality to avoid misclassification. Data cleaning for text analysis is standard in machine learning scholarship (Okazaki et al., 2014). The text is easily polluted by different types of noise, so analyzing it without pre-processing could yield erroneous results. The process of data preprocessing consists of two steps, i.e., data cleaning and vectorization. The data cleaning procedure is about noise removal, lexicon normalization, and object standardization. We refer to vectorization as mapping words or phrases from vocabulary to a corresponding vector of real numbers. We feed these numeric data points to our machine learning algorithm for training. The text cleaning process consists of three steps:

Noise Removal: We disregard something in the data that is not explicitly applicable to the context. Noise removal is about removing characters, digits, and chunks of text that can interfere with given text analysis (Li et al., 2018).

Text Normalization: Another type of textual noise is about the multiple representations exhibited by a single word. For example – “work,” “worker,” “worked,” “works,” and “working” are the different variations of the word ‘work.’ We have used a technique called “lemmatization” (Zhang et al., 2019). The main advantage of lemmatization is that it considers the word's context to determine which is the intended meaning the user is searching. The preprocessing technique improves efficiency, thus lowering the noise.

Text Standardization: Special terms and phrases used in the text, such as acronyms, hashtags with attached words, and colloquial slang, are all part of the text (Gupta & Joshi, 2017). This process involves a transformation of text into a canonical (standard) form. For example, we can transform the word “2moro”, “tomrw” and “2rrw” as just “tomorrow.”

--- Insert Figure 3 about here ---

Tables 3a and 3b show text data before and after the cleanup show some of the most frequently occurring words known as “tokens” in our data vocabulary. Unigram is a sequence of one word, bigram is a sequence of two words, and trigram is a sequence of three words present in our dataset.

--- Insert Table 3a and 3b about here ---

Feature Extraction

To verify our data cleaning step, we conduct a comparative analysis for original raw data versus cleaned text data. We extract additional features to check the difference and quality of our cleaning procedure. We show the results of the last five samples from the dataset in Figure 4a. Here, we get text data per tweet with the total number of words in each tweet, total characters in a given tweet, number of stop-words, number of hashtag symbols (shows trends in tweets), number of @ symbols (shows any mentioned entity), numeric digits, and upper-case letters. All of these features measure data cleaning activity.

--- Insert Figure 4a about here ---

--- Insert Figure 4b about here ---

Figure 4b shows the last five samples of cleaned text in our dataset with the same features mentioned above. It indicates that we reduce the total number count of each tweet. We remove punctuation, digits, and hyperlinks. Punctuation will not be used to evaluate classification as we will not examine grammar. Next, we remove English stop words as provided by the ‘Natural Language Toolkit NLTK’ library. Stop words are expressions that do not provide specific text classification details (e.g., "to," "by," and "a"), so we do not use these phrases in further analysis. In addition to removing the English language stop words, we remove the Twitter-specific stop words such as “via” and “rt.” We find a reduced noise pattern in character count, number of stop-words, number of hashtag symbols, number of @ symbols, digits, and upper-case letters. Hence, we managed to reduce data noise and extra tokens considerably. Through this text cleaning procedure, the machine learning algorithm may compute more efficiently.

Word Embedding

Our processed tweets now contain words that are in reduced form for the text classification model. However, machine learning models cannot process words. Thus, we need to convert it into a given number of matrices. Traditionally, we implement the bag of words (BOW) approach assigns a unique integer to a unique word, and then keeps counting how many times that specific word has occurred in our corpus (Zhao & Mao, 2017). We are referring to terms as sparse vector representations since there are lots of zeros in the vector representation of words. This representational approach has a few disadvantages. First of all, no relationship between words is recorded. We need a mathematical representation of words to carry meaning rather than integer values representing words for our analysis. Secondly, the sparse representation of words needs ample vector space as our vocabulary size grows, so it is not an efficient approach in our case study (Li et al., 2017).

We deal with sparse matrix representation through the word embedding process (Ren et al., 2016). In a coordination method, we interpret words as a collection of matrices in which related words are put closer together in the form of vector space, based on a corpus of relationships. This approach provides us a matrix with context. This method is not computationally costly since the word similarity is captured by a dense matrix representation, using closeness between two vectors (Ren et al., 2016; Li et al., 2017).

Model building

The fundamental principle of text classification with the supervised machine learning technique is to automatically assign a certain number of target classes to documents using a much smaller collection of training data than the total number of classification documents. Our training data with the size of 15,000 samples contains automated labeling of target classes, i.e., risk-related tweets, benefit-related tweets, or neutral tweets. Based on these classifications against their relevant text document, the machine learning algorithm creates a model that determines how to classify new documents. We may use various machine learning algorithms for our text classification problem such as Naïve Bayes (Kim et al., 2006), Maximum Entropy Classification (Nigam, 1999), or Support Vector Machines (Pang et al., 2002; Kohl et al., 2018, Joachims et al., 2001). These models are easy to implement; however, these algorithms are not suitable for our case. As text data is in the form of a sequence, we need to implement a model that computes the probability of several words in a sequence. For this reason, we use Recurrent Neural Networks (RNN).

An RNN uses a partially linked neural network that involves a few of its layers in a loop. That loop is typically an iteration of two inputs adding or concatenating, multiplication of representative matrix, and a non-linear function forming recurrent layers. These layers let the training model maintain information in 'memory' over time. However, our dataset contains tweets with variable lengths varying from few characters to a maximum of 280 limits. Therefore, it can be challenging to train standard RNNs because the loss function's gradient decays exponentially with time. We refer to this issue in the machine learning domain as a vanishing gradient (Chung et al., 2014).

To solve the vanishing gradient problem, our training model requires learning long-term temporal dependencies. Long Short-Term Memory, i.e., LSTM networks (Hochreiter and Schmidhuber, 1997; Gers et al., 2000), deal with vanishing gradient by introducing new gates. These gates are input and forget gates that allow better control over the gradient flow and let "long-range dependencies" to be better maintained. For our case, we would follow a text classification model architecture that generally consists of the following components connected in sequence, as shown in Figure 5.

--- Insert Figure 5 about here ---

For model validation, we use the train-test split method using the scikit-learn module. We divide our train set into 11250 samples and the test set into 3750 samples. As an input, the LSTM network gets fed with a sequence of vectors representing the word embeddings. Our data can learn from word embeddings while training a neural network on the classification problem. Additionally, we add padding to make all the vectors of the same length. As for the model, we deploy a standard one hidden-layer LSTM, trained for predicting three-class perception, i.e., neutral, risk-related, and benefit-related class. A step-by-step procedure for input data is following:

- By converting each text into either a sequence of integers or a vector.
- Set a maximum number of words in every perception to 280 (maximum allowed by Twitter).
- Truncate and pad the modeling input sequences such that these inputs are all of the same lengths.

We provide LSTM (Long Short-Term Memory) network structure below:

- The first layer uses vectors with a length of 100 to represent each word in the embedded layer.
- In natural language processing models, SpatialDropout1D performs variational dropout. Dropout is useful in the prevention of overfitting in neural network models.
- The LSTM dense layer with 100 memory units is the next.
- The output layer is made up of 3 output values, one per class.

- Softmax is the activation function for multi-class classification.
- We use Adam as an optimization algorithm to take care of our neural network's weights and learning rate so that loss would not begin to diverge after decrease to a point.
- We use categorical cross-entropy as the loss function because it is a multi-class classification problem.

Model evaluation

On three-class perception prediction (risk, neutral, benefit) of complete sentences, the model achieves 81.1% weighted average accuracy. We also compute each class's accuracy based on the confusion matrix with actual versus predicted results comparison (Figure 6).

--- Insert Figure 6 about here ---

We can see that the accuracy value is good as for the sequence model. We encode our classes using the one-hot encoding method. We encode Benefit related tweets as ‘0’, neutral sentiments as ‘1’, and risk-related sentiments as ‘2’. For evaluating individual classes, accuracy is not a useful metric as it only provides how the overall model predicts and not for each class at the individual level (Valverde-Albacete, 2013). Table 4 offers other metrics of model evaluation that also provide very consistent results.

--- Insert Table 4 about here ---

Results

Tweet analysis in terms of benefit and risk ratio

With an overall total of 455,727 tweets, we obtain 177,489 (38.95%) neutral tweets, 182,533 (40.05%) benefit related tweets, and 95,705 (21.01%) risk-related tweets about autonomous vehicles. We then calculate risk ratio (RR) and benefit ratio (BR) as given below:

RR= (RT)/(RT+BT) = (95705) / (95705+182533) =0.343
BR= (BT)/(BT+RT) = (182533) / (182533+95705) =0.657
where,
RT = risk-related tweets, BT= benefit related tweets

The RR defines the ratio of AV-related risk tweets to the total number of tweets except for neutral tweets that do not provide risk or benefit details. BR is similarly defined and can also be calculated as 1-RR if RR is already referred to as RR + BR equal to one.

--- Insert Table 5 about here ---

In 2016, we do identify 66,864 (39.3%) tweets as neutral, 62,723(36.9%) as benefits, and 40,369(23.7%) as related to risks about autonomous vehicles, as shown in Table 5. It is interesting to see the pattern of occurrence of each sentiment class each year. In our training data, we have higher classes of neutral sentiments of consumers; however, this is not the overall dataset. This symmetry of classes indicates that our neural network does not fall in the class imbalance trap (Huang et al., 2006).

In 2016, neutral tweets are of the highest frequency in our dataset; however, we can see this pattern varies over time. There is a drop in neutral tweets observed in the year 2018 reducing from 65,635 to 44,901 tweets. It is the pivotal point where the polarity pattern shifts in our dataset, tilting towards benefit-related tweets being 63,779. Risk-related tweets are still lower than benefit and neutral tweets in an overall dataset and also year vice. In 2019, benefit tweets remain the highest in terms of occurrence. We only collect data for one-month of 2019, i.e., January, which is why this occurrence pattern is not very significant. Overall, we see a reliable perception class pattern over the years in our analysis. We find the distribution of sentiments from 2016-2019 in Figure 7.

--- Insert Figure 7 about here ---

There is a low occurrence of risk-related tweets in our dataset. It seems that the pattern of risk tweets is consistently low. For this reason, it is crucial to see how risk-oriented tweets vary over months. A further in-depth analysis considering monthly trends shows interesting trends, as shown in Figure 9. We can find that risk tweets do not change very frequently except for three points. These points are January 2016, May 2016, and March 2018. These anomalies are the points that require further inspection from real-world events. We can also find a very sharp decrease in January 2019. Since tweets are collected only for the first week of January 2019, this dip does not significantly affect our results.

--- Insert Figure 8 about here ---

In Table 4, RR in 2016 is equal to 0.391, and the BR ratio value is 0.609. These values show some deviation from overall RR and BR values of 0.343 and 0.657, respectively. This ratio remains very consistent on year vice trends. The year 2019, where we had very few samples, shows consistent RR and BR ratio scores. We can conclude that our calculated RR and BR ratios remain consistent with overall BR and RR ratios' values. We have shown the distribution of benefit ratios and risk ratios from 2016-2019 in Figure 9.

--- Insert Figure 9 about here ---

Over the years, we can find a stable BR and RR ratio, which means that our model of neural network classifier behaves consistently over time. The classifier does not fail to distinguish risk and benefit tweets among the vast number of neutral tweets as time passes, and topics shift. In addition, as they become more known to the public, the conversation about autonomous vehicles may also concentrate more on the technology's risks and benefits. The nature of RR and BR did not shift dramatically from 2016 to 2019, suggesting that RR and BR could be a rigorous measure to quantify the perception of risk and benefit. Closer RR inspection reveals that between the months, it did have some outlier trends.

Figure 9 shows BR and RR ratios where the benefit ratio is predominately higher than RR. It shows that the results of the year vice BR and RR ratio make sense. However, the relation between BR and RR changes at three instances. January 2016, May 2016, and March 2018 are the milestones for these cases. We can monitor how Twitter users respond to specific news stories, announcements, or other events by examining the BR and RR ratio pattern of tweets in detail and whether changes in perceived risks or benefits are affected by these events. We discover that there is a correlation between events related to autonomous vehicles and the content of tweets. This finding enables us to state that tweets about autonomous vehicles, while subjective, are linked to evidence and events in the real world. We will provide some examples below.

The driver of a Tesla Model S in Handan, China, was killed in January 2016 when his car crashed into a stationary truck. In our data, we could detect this event as it was often stated in tweets with a ratio value of 0.65 in RR. In total, we point out 9337 risk tweets about autonomous vehicles in our dataset compared to 5021 benefit tweets. RR ratio remained lower for three consecutive months. It seems that one accident from one specific company did not hurt the overall business ecosystem of autonomous vehicles since public perception remained positive.

On May 7, 2016, in Williston, Florida, the first recorded fatal accident involving a Tesla involved in Autopilot mode took place. The driver got killed in a crash with an 18-wheel tractor-trailer. We see a rise with the RR ratio above the BR ratio value at 0.62 in our results, indicating the impact of this accident. We could observe tweets with risk orientation when spikes again reach 8824 risk tweets versus 5399 benefit-related tweets. It is worth noticing how the pattern goes back towards a ‘normal’ high BR ratio against a low RR ratio next month. By late June 2016, the National Highway Traffic Safety Administration of U.S. An opened an investigation into the fatal autonomous accident. Even in this instance, aggregate benefit tweets show higher distribution as per Figure 10 than risk-related tweets.

--- Insert Figure 10 about here ---

It is vital to notice that organizations also reacted quickly towards subsequent accidents to maintain their public image. Tesla’s official statement in late June 2016 explains “neither autopilot nor the driver noticed the white side of the tractor-trailer against a brightly lit sky, so the brake was not applied” gave a perspective of “not all machine error” (Tesla, 2016). Another promising example is Tesla's announcement of a $2.6 billion contract with Solar City to combine solar power with autonomous vehicles in August 2016. These positive initiatives create a positive perception among end consumers, reflecting that the benefit ratio again became consistent and higher than the risk ratio.

The ratio of benefit remained higher than the risk ratio for 2017 in our dataset. This high point does not mean that there were no risk-oriented tweets; instead, they overcame benefit-related tweets. In March 2018, this pattern was broken. Elaine Herzberg's death was the first recorded case of a pedestrian fatality following a collision that took place late on the evening of 18 March 2018 involving an autonomous vehicle. We see a jumpy in risk tweets to 7160, which are higher than 5209 benefit tweets. The risk ratio goes to 0.57, and another spike of debate against self-drive cars and artificial intelligence, in general, erupted the media. By this time, other leading car manufactures like BMW, Ford, Volkswagen, and Toyota have joined the race, and therefore, Tesla’s one negative incident did not stop the momentum. We can see that public perception comes back to the benefit mindset in April 2018, as shown in Figure 10. We recognize these three critical incidents and their public perception in Figure 8 and Figure 10, where risk tweets show a rise in an otherwise consistent trend.

There is another unusual event of January 2019, which is due to smaller data samples. Even though smaller yet not insignificant, KIA, in early January 2019, announced considering the interactive cabin as a focal point. As observed in Figure 10, this news might promote a positive perception on social media regarding autonomous vehicles.

Tweet analysis concerning leading organizations within the AV ecosystem

We recognized that the car manufacturers dominate the discussion about autonomous vehicles by analyzing the data. This domination pattern shows that there are significant initiatives made by car manufacturers in recent years due to which public interest has raised towards them. Figure 11 shows the overall number of tweets for each car manufacturer in the dataset. It is interesting to see German car manufacturers' dominance (the top three car companies are from Germany). Furthermore, it is remarkable that IT companies such as Google and Apple are linked that firmly to this topic. We must consider, however, that the dataset is biased since we queried Twitter only for English tweets leaving out the Chinese market. Despite this limitation, we have got few tweets that refer to Baidu in their discussion.

--- Insert Figure 11 about here ---

In Figure 12, we can see some of the critical words occurring in risk tweets and organization names such as Tesla, Volkswagen, General Motors have occurred frequently. Multiple autonomous vehicle incidents, as mentioned above, have affected the reputation of Tesla in public. Volkswagen, which has gone through much bad press due to the carbon emission scandal, has its image affected even in the autonomous vehicles’ domain.

--- Insert Figure 12 about here ---

We have analyzed the RR to the leading organizations in the AV ecosystem for a detailed analysis of the driverless car ecosystem. Figure 13 shows that all companies mentioned in our dataset have a benefit ratio higher than the risk ratio indicating that end-customer perception towards autonomous vehicles and leading organizations have been positive. However, closer inspection of each company may provide insightful findings.

Google and Waymo are at the top, having RR scores of 0.49 and 0.48. this score is not higher than the threshold points of 0.5 but, it is very close to the risk domain. The users seem to be concerned about not being able to drive anymore since Waymo had rider services, i.e., no human safety driver at the wheel (Forbes, 2019). Furthermore, the Google Car's minor accidents have received much negative public attention taking its RR close to 0.5. On the other hand, most car manufacturing companies such as Porsche, Audi, and BMW are among the lowest RR scores, suggesting that public trust is higher for car manufacturers than service and IT companies. We see that, despite the incidents arising from the abuse of Tesla's Autopilot, Tesla is getting close to conventional car manufacturers. Daimler-current Benz's driving automation system, Intelligent Drive, is even more stringent than Tesla's system. The abuse of the driving automation system is thus much less likely with the Tesla system. However, the additional restrictions may be seen by users as another lack of control, which is one of the key concerns about autonomous vehicles (Rödel et al., 2014; Langdon et al., 2018). It is also interesting that German auto manufacturers are considered the least risky by end consumers. We need to understand that as the number of tweets varies significantly between the carmakers, the findings may be biased. There are 13314 tweets about BMW and just 889 about Apple in our dataset. The RR makes the findings comparable as per figure 13, but before drawing any immediate conclusions, audiences should keep this disparity in mind.

--- Insert Figure 13 about here ---

Discussion and key findings

We discuss validity before we discuss and interpret our outcomes. First, our results are compared with previous studies. In a study conducted in Germany, Fraedrich et al. (2016) found that 46 percent of its 1163 respondents had a favorable autonomous vehicle connotation. For their views on autonomous vehicles, Schoettle and Sivak (2014) surveyed 1533 candidates from the U.S., U.K., and Australia and found that 56.8 percent of respondents had at least a marginally positive opinion. Kohl et al. (2018) implemented tweet analysis and found that 55% of their samples were with positive context. We collected data for a more extended period compared to given studies. Additionally, we did not limit our analysis to tweets' class distribution being positive, negative, or neutral. Instead, we measured BR and RR, and our RR score is 65 percent, which is 10 percent higher than Kohl et al. (2018)'s calculated rate of positive opinions score. Therefore, our data remains in-line with the research domain relevant to the public perceptions of risk and benefit perception.

We find solid performance for our methodology that machine learning methods on Twitter can evaluate the risk and benefit perceptions of the public about autonomous vehicles. Our training data have imbalanced classes, with neutral being higher in proportion; however, our dataset has found a higher proportion of benefit tweets than neutral tweets. The prediction result also verifies that the machine learning approach is better for large volume and unstructured data. Our studies' rise in a positive trend might be due to our search for companies that belong to those countries that have developed infrastructure and advanced automotive industry, i.e., USA, Japan, and Germany.

Second, as our in-depth review of tweets in the previous section has shown, our analysis indicates that news and events have been expressed in tweets. We have therefore found support for our research questions as well. We further address our results in the following and compare them to previous studies on the adoption of autonomous vehicles.

Given the results of the previous study and the RR and BR values determined in this study, we infer that individuals have an overall positive view of autonomous vehicles; however, such unique autonomous vehicle-related reservations are also noted. For instance, if we assume that autonomous vehicles were released today, technology adoption could not be fully achieved due to underlying apprehension. Before the public is faced with a growing number of increasingly autonomous cars or even completely automated autonomous vehicle technology, it is essential to address public concerns. In order to evaluate the tweets over time, we measured the BR and RR values at different points of time and found a marginal decrease in RR from 39.1 percent to 36.3 between 2016 and 2019. This decrease could mean that the anxiety of people has not been addressed.

In addition, there could be a prejudice created by a social reinforcement of risk perceptions (Kasperson et al., 1988), i.e., individuals appear to speak more about adverse developments than positive growth. While it is a critical finding that social amplification of risk perceptions might exist in social media, the results are still severe. We see in many cases that social media contributes more and more to inflated expectations of risk that lead to irrational behaviors. The pessimistic results depict that the adoption of autonomous vehicles by both manufacturers and suppliers within the autonomous vehicle ecosystem could be significantly diminished by a single accident.

We noticed several tweets that showed a distorted view of a hazard, such as, "If I wish to die of quick & painful death, then those horrible looking self-driving cars are not that bad after all.". However, autonomous vehicles may minimize them dramatically instead of raising the likelihood of fatal accidents (Fagnant and Kockelman, 2015). Other research has also found skewed views that could alter as individuals become more acquainted with autonomous vehicles (Woisetschläger, 2016; Bansal et al., 2016).

People also shared mistrust and communicated their passion for driving towards the autonomous vehicle manufacturing firms. For example: "Sorry @Google, we do not trust your technology not going to buy a self-driving car I like driving my own." In this situation, the understanding of advantages is skewed. While in some circumstances driving can be pleasant, we are also faced with less enjoyable driving elements such as traffic jams, long boring highways with speed limits, or in increasingly crowded cities hunting for a parking space. In marketing research, we find some issues that predict a lack of emotional connection because of the loss of driving pleasure when a car drives itself (Olson, 2017). Few researchers have discussed the autonomous vehicle experience of end-users (Rödel et al., 2014; Pettersson and Karlsson, 2015; Niculescu et al., 2017). To mitigate this issue, other research suggests adapting the driving style of autonomous vehicles to the driving style of their users (Kraus et al., 2009; Butakov and Ioannou, 2015; Kuderer et al., 2015). Previous research also suggests allowing drivers to take back control of their cars if they want in order to reduce confidence problems and maximize driving pleasure (Yap et al., 2016).

People also showed concern for their privacy and protection; for example, "What if #driverless #cars be hacked, and hackers get to drive my car?" expresses the fear of hackers that could take control of your vehicle. As autonomous vehicles complement IT services, hackers can implement viruses that could be spread from vehicle system to vehicle system. This security risk could prove real as hacker activities have been noted on current autonomous vehicle systems. The safety problem is due to the greater connectivity of autonomous vehicles (Lee et al., 2016). These hacking attacks may cause car passengers and other road users financial and physical harm and even injury, which is more dangerous than getting a hacked computer. Autonomous vehicle (AV) manufacturers and service providers need to be aware of hackers and deploy techniques to prevent hackers from targeting their vehicles effectively. Previous research reveals that one of the main fears about autonomous vehicles has been security risks and cyber-attacks (Zmud and Sener, 2017).

Drivers are constantly exposed to higher levels of electromagnetic field radiation with all the positive onboard components such as GPS, remote controls, power accessories, Bluetooth, Wi-Fi, audio, and radio components. As it is shown in," Just knew about radiation in self-drive cars. Artificial Intelligence cannot do anything right. #selfdriving". Electronic radiation exposure can cause a variety of severe health concerns. There is more focus towards on-road safety measures, and health cost has not been discussed much in media. As per Fleetwood (2017), Companies should open up about such health concerns before becoming a serious problem.

Many people liked to save time using autonomous vehicles with regard to the tweets listed as having benefits of autonomous vehicles. For example: "I'm excited for autonomous vehicles because I like to be driven around until I fall asleep, but you can't really ask someone to do that after you turn 4." Another case: "Just saw some man taking a selfie while driving and I shook but then I realized he was in his self-drive Tesla. Amazing time to join sleepy club in car". Such tweets are a case of a misguided perception of benefit since only complete automation allows sleeping while driving. The existing level of autonomous cars is level 3, i.e., partly automated, and it is possible that it will take many years for us to achieve level 4 or even level 5, i.e., full automation.

Meanwhile, drivers misuse current autonomous vehicles by leaving the driver's seat while driving on a public road using the Autopilot feature of a Tesla Model S (Krok, 2015). Intentionally or inadvertently, they endanger the lives of their own and others and possibly impact the public's adoption of autonomous vehicles, as seen by the fatal self-drive Tesla accidents back in 2016. However, manufacturers should be conscious that individuals continue to use self-drive vehicles for pleasure purposes, such as sleeping or taking selfies while in an autonomous vehicle. Other studies (Cosh et al., 2017) confirm that sleeping is usual practice when being in an autonomous vehicle.

In general, the technological innovation put into autonomous vehicles (AV) impresses individuals. For example: "Spotted the first #Tesla #Model3 on the road in the Netherlands, test car at a @Fastned station with CCS connector. "Most benefit tweets reflected that people were excited to find something new on the road. For example:" Lukas just spotted the Google car."

People also show a high level of interest in the leadership role and pragmatic approach (Ulrich, 2007). We earlier discussed three major fetal accidents from 2016-2019, all with Tesla self-drive cars. Elon Musk-CEO of Tesla Motors, replied for raising concerns: “It is super messed up that Tesla crash resulting in a broken ankle is front-page news and the 40,000 people who died in US auto accidents alone in past year get almost no coverage” (Elon Musk, 10:54 pm, 14 May 2018). This tweet was interpreted positively at large with sympathetic interpretation such as “Keep it up, autopilot is better than humans,” “The media just hate Tesla, whatever bad news happens it is on the front page,” and “This is getting out of hand…. Journalism has lost its integrity” (Jappy, 2019). These supporting tweets show that public perception towards the adoption of technology is also related to leadership (Ulrich, 2007)

Autonomous vehicles (AV) pioneers have found that people are enthusiastic about this new technology and its advantages. As a result, they are investing in the production of autonomous vehicles and have already promised features that will only be introduced in several years. This anticipation could trigger a misunderstanding of the potential benefits and inflated risk perceptions of autonomous vehicles if communication strategies are not changed. According to Nees(2016), it may also have detrimental effects on public adoption of autonomous vehicles by concentrating solely on the benefits and even creating unrealistic benefit expectations.

Implications of our study

Based on our findings, we recognized the need for developers, vehicle manufacturers, and complementary stakeholders to listen to the feelings of future potential customers. The objective best solution or superior development of new technology will fail if it does not gain or generate public support from the end consumer. Active public acceptance management is therefore mandatory to reduce the risk of the failure of emerging technology. Companies need to reconsider their communication strategies for autonomous vehicles (AV) to resolve skewed views of the benefit and risks of autonomous vehicles (Kasperson and Kasperson, 1996). It is already evident with the first available level 3 automated cars as per the German Federal Highway Research Institute (BASt). An exaggeration of benefits could lead to misuse of autonomous vehicles, initial user dissatisfaction, and catastrophic results. The public's overstatement of risks could lead to opposition to autonomous vehicles (AV) before they are even widely available. (Kleijnen et al., 2009; König and Neumayr, 2017).

In addition, practitioners can ensure that the full potential of autonomous vehicles (AV) is utilized by incorporating the benefits mentioned in social media, as stated in this study. Initial field experiments and case studies show that people embrace more autonomous vehicles after using prototypes (Alessandrini et al., 2011; Pendleton et al., 2015; Christie et al., 2016; Portouli et al., 2017; Madigan et al., 2017). Initial personal experience with prototypes leads to less susceptibility to skewed perceptions of autonomous vehicles and should therefore be made more publicly accessible through, for instance, autonomous vehicle events by manufacturers, establishing additional model regions and test tracks, or creating autonomous vehicle driving experience centers. We also found that in the next generation of Level 3 and Level 4 driving automation systems, user interface design will play a key role, as Tesla Autopilot's understanding of risk and benefit perception demonstrates. Rather than increased protection due to device limitations, users are mainly aware of the increased loss of control induced by them, decreasing adoption.

Limitations

This research has some limitations. It could further enhance the use of machine learning algorithms other than recurrent neural networks, such as neural networks with stacked layers (Cao et al., 2017) or transformers (Groenwold et al., 2020). However, we do not expect significant improvements since our analysis of the RNN model performance already showed promising results. Recurrent neural networks are usually among the strongest performers for sequential data, such as text classification problems (Chung et al., 2014). Additionally, data collection was limited to the companies that have developed an autonomous vehicle (AV) ecosystem. This limitation did not include companies and complementors from the small-medium size hierarchy.

Instead, more studies should concentrate on more specific categories of risk and benefits. Slovic (1987) identifies risks as a combination of two factors: unknown risks that are not measurable and unknown to those exposed, and dreaded risks that can be devastating and fatal globally. Also, Hohenberger et al. (2017) split the advantages of autonomous vehicles into the monetary outcome of innovation, time spent on development, and safety benefits for end-customers. Using sub-categories of risks and benefits, future research may examine risk and benefit expectations or identify new categories influencing the acceptance of technology based on social media data.

Future work

In order to help the in-depth qualitative study of tweets after classification, future work may be useful to use topic modeling (Blei, 2012: Debortoli et al., 2016). The next step in research could involve other outlets of social media, such as Facebook, Reddit, or blogs, in the study. Although our Twitter analysis appears to be comparable to current survey samples, widening the study to other digital channels may explore discrepancies in the explanation and distribution of perceptions across social media sites. It will, however, entail widening the technological platform, as the frameworks of other social media sites vary greatly from Twitter. Besides analyzing opinions expressed in written text, customers' sentiments can also be extracted from other media such as recorded speech and videos, as Brown (2017) mentioned, who analyzed YouTube videos about driving automation experiences. A more comprehensive overview of opinions and perceptions will contribute to the inclusion of multiple outlets and enhance our understanding of the different facets of new technology important to potential customers and society.

Conclusion

Technology adoption in management is considered one of the most valued research streams. We extended this scholarship from sole technology to a holistic ecosystem. We did not focus on the focal firm as a unit of analysis but, we instead bring customers' perception as key to the emerging technology ecosystem. By this assumption, we made technology adoption analysis based on modern natural language processing methods by using tweets from social media.

We have shown that using machine learning to automatically identify social media is a promising approach to examine the adoption of new technologies such as autonomous vehicles by analyzing 455,727 tweets. Although Twitter data is vulnerable to certain biases, our findings are consistent with previous studies. Our methodology mitigates some of the analytical limitations in data collection and biases of online surveys for technological innovations and time-consuming manual coding. In addition, our approach allows the influence of such incidents on the public understanding of new technology to be calculated. The perceived risks and benefits model in our study can be integrated with conventional adoption models for questionnaire-based survey studies. We found quantitatively and qualitatively that customers' risk and benefit perception is the key determinant for technology adoption within an emerging ecosystem. Exaggerated expectations of risk or benefit may lead to irrational behaviors, as seen in some tweets. On the other hand, the organization may take notice of customers' sentiments for the adoption of an autonomous vehicle ecosystem.

References

Adner, R. and Kapoor, R., 2016. Innovation ecosystems and the pace of substitution: Re-examining technology S-curves. Strategic management journal, 37(4), pp.625-648.

Adner, R. and Kapoor, R., 2010. Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. Strategic management journal, 31(3), pp.306-333.

Adner, R., 2006. Match your innovation strategy to your innovation ecosystem—Harvard business review, 84(4), p.98.

Alessandrini, A., Holguin, C., and Parent, M., 2011, October. Advanced transport systems showcased in La Rochelle. In 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC) (pp. 896-900). IEEE.

Alhakami, A.S. and Slovic, P., 1994. A psychological study of the inverse relationship between perceived risk and perceived benefit. Risk analysis, 14(6), pp.1085-1096.

Bansal, P., Kockelman, K.M. and Singh, A., 2016. Assessing public opinions of and interest in new vehicle technologies: An Austin perspective. Transportation Research Part C: Emerging Technologies, 67, pp.1-14.

Bazilinskyy, P., Kyriakidis, M. and de Winter, J., 2015. An international crowdsourcing study into people's statements on fully automated driving. Procedia Manufacturing, 3, pp.2534-2542.

Blake, E.R., 1995. Understanding outrage: how scientists can help bridge the risk perception gap. Environmental Health Perspectives, 103(suppl 6), pp.123-125.

Bharati, P. and Chaudhury, A., 2006. Studying the current status of technology adoption. Communications of the ACM, 49(10), pp.88-93.

Bruque, S. and Moyano, J., 2007. Organizational determinants of information technology adoption and implementation in SMEs: The case of family and cooperative firms. Technovation, 27(5), pp.241-253.

Bongaerts, R., Kwiatkowski, M. and König, T., 2017. Disruption technology in mobility: customer acceptance and examples. In Phantom ex Machina (pp. 119-135). Springer, Cham.

Boshmaf, Y., Muslukhov, I., Beznosov, K. and Ripeanu, M., 2011, December. The socialbot network: when bots socialize for fame and money. In Proceedings of the 27th annual computer security applications conference (pp. 93-102).

Brown, B., 2017. The social life of autonomous cars. Computer, 50(2), pp.92-96.

Burns, L.D., 2013. A vision of our transport future. Nature, 497(7448), pp.181-182.

Butakov, V. and Ioannou, P., 2015, September. Driving autopilot with personalization feature for improved safety and comfort. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 387-393). IEEE.

Blei, D.M., 2012. Probabilistic topic models. Communications of the ACM, 55(4), pp.77-84.

Cao, W., Song, A., and Hu, J., 2017. Stacked residual recurrent neural network with word weight for text classification. IAENG International Journal of Computer Science, 44(3), pp.277-284.

Chih-Chung, C. and Chih-Jen, L., 2004. LIBSVM: a library for support vector machines [EB/OL].

Chung, J., Gulcehre, C., Cho, K. and Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

Choi, J.K., and Ji, Y.G., 2015. Investigating the importance of trust on adopting an autonomous vehicle. International Journal of Human-Computer Interaction, 31(10), pp.692-702.

Christie, D., Koymans, A., Chanard, T., Lasgouttes, J.M. and Kaufmann, V., 2016. Pioneering driverless electric vehicles in Europe: the city automated transport system (CATS).

Conover, G.D., 1994. The eleven commandments for IVHS. In Proceedings of VNIS'94-1994 Vehicle Navigation and Information Systems Conference (pp. 503-506). IEEE.

Cosh, K., Wordingham, S. and Ramingwong, S., 2018. Investigating public opinion regarding autonomous vehicles: a perspective from Chiang Mai, Thailand. In IT Convergence and Security 2017 (pp. 3-10). Springer, Singapore.

Davis, F.D., 1985. A technology acceptance model for empirically testing new end-user information systems: Theory and results (Doctoral dissertation, Massachusetts Institute of Technology).

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, pp.319-340.

Davis, F.D., Bagozzi, R.P. and Warshaw, P.R., 1989. User acceptance of computer technology: a comparison of two theoretical models. Management Science, 35(8), pp.982-1003.

Debortoli, S., Müller, O., Junglas, I. and vom Brocke, J., 2016. Text mining for information systems researchers: An annotated topic modeling tutorial. Communications of the Association for Information Systems, 39(1), p.7.

Dasgupta, S., Agarwal, D., Ioannidis, A. and Gopalakrishnan, S., 1999. Determinants of information technology adoption: An extension of existing models to firms in a developing country. Journal of Global Information Management (JGIM), 7(3), pp.30-40.

Duncan, R.B., 1976. The ambidextrous organization: Designing dual structures for innovation. The management of the organization, 1(1), pp.167-188.

Earle, T.C. and Cvetkovich, G., 1995. Social trust: Toward a cosmopolitan society. Greenwood Publishing Group.

Fagnant, D.J. and Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice, 77, pp.167-181.

Fleetwood, J., 2017. Public health, ethics, and autonomous vehicles. American journal of public health, 107(4), pp.532-537.

Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), pp.82-89.

Fishbein, M. and Ajzen, I., 1977. Belief, attitude, intention, and behavior: An introduction to theory and research.

Fraedrich, E. and Lenz, B., 2014. Automated driving: Individual and societal aspects. Transportation research record, 2416(1), pp.64-72.

Fraedrich, E., Cyganski, R., Wolf, I. and Lenz, B., 2016. User perspectives on autonomous driving: A use-case-driven study in Germany.

Forbes. 2019. Waymo Says More Of Its Self-Driving Cars Operating 'Rider Only' With No One At Wheel. [ONLINE] Available at https://www.forbes.com/sites/alanohnsman/2019/10/28/waymos-autonomous-car-definition-if-you-need-a-drivers-license-its-not-self-driving/#3a19062f6478. [Accessed 13 July 2020].

Gasser, T.M. and Westhoff, D., 2012, July. BASt-study: Definitions of automation and legal issues in Germany. In Proceedings of the 2012 road vehicle automation workshop. Automation Workshop.

Gogoll, J. and Müller, J.F., 2017. Autonomous cars: in favor of a mandatory ethics setting. Science and engineering ethics, 23(3), pp.681-700.

Gold, C., Damböck, D., Lorenz, L. and Bengler, K., 2013, September. "Take over!" How long does it take to get the driver back into the loop?. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 57, No. 1, pp. 1938-1942). Sage CA: Los Angeles, CA: Sage Publications.

Griffin, A. and Hauser, J.R., 1993. The voice of the customer. Marketing Science, 12(1), pp.1-27.

Gupta, I. and Joshi, N., 2017, December. Tweet normalization: A knowledge-based approach. In 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS) (pp. 157-162). IEEE.

Gupta, N., Fischer, A.R.H. and Frewer, L.J., 2015. Ethics, risk, and benefits associated with different applications of nanotechnology: a comparison of expert and consumer perceptions of drivers of societal acceptance. NanoEthics, 9(2), pp.93-108.

Gupta, N., Fischer, A.R. and Frewer, L.J., 2012. Socio-psychological determinants of public acceptance of technologies: a review. Public Understanding of Science, 21(7), pp.782-795.

Graliński, F., Jassem, K. and Junczys-Dowmunt, M., 2013. PSI-toolkit: A natural language processing pipeline. In Computational Linguistics (pp. 27-39). Springer, Berlin, Heidelberg.

Groenwold, S., Honnavalli, S., Ou, L., Parekh, A., Levy, S., Mirza, D., and Wang, W.Y., 2020. Evaluating the Role of Language Typology in Transformer-Based Multilingual Text Classification. arXiv preprint arXiv:2004.13939.

Haboucha, C.J., Ishaq, R. and Shiftan, Y., 2017. User preferences regarding autonomous vehicles. Transportation Research Part C: Emerging Technologies, 78, pp.37-49.

Haustein, S., Bowman, T.D., Holmberg, K., Tsou, A., Sugimoto, C.R. and Larivière, V., 2016. Tweets as impact indicators: Examining the implications of automated “bot” accounts on Twitter. Journal of the Association for Information Science and Technology, 67(1), pp.232-238.

Hohenberger, C., Spörrle, M. and Welpe, I.M., 2016. How and why do men and women differ in their willingness to use automated cars? The influence of emotions across different age groups. Transportation Research Part A: Policy and Practice, 94, pp.374-385.

Hohenberger, C., Spörrle, M. and Welpe, I.M., 2017. Not fearless, but self-enhanced: The effects of anxiety on the willingness to use autonomous cars depend on individual levels of self-enhancement. Technological Forecasting and Social Change, 116, pp.40-52.

Hopkins, D.J. and King, G., 2010. A method of automated nonparametric content analysis for social science. American Journal of Political Science, 54(1), pp.229-247.

Huang, Y.M., Hung, C.M. and Jiau, H.C., 2006. Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. Nonlinear Analysis: Real World Applications, 7(4), pp.720-747.

Jappy, T. ed., 2019. The Bloomsbury Companion to Contemporary Peircean Semiotics. Bloomsbury Publishing.

Jacobides, M.G., Cennamo, C. and Gawer, A., 2018. Towards a theory of ecosystems. Strategic Management Journal, 39(8), pp.2255-2276.

Joachims, T., 2001, September. A statistical learning model of text classification for support vector machines. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 128-136).

Kahneman, D., 1979. Prospect theory: An analysis of decisions under risk. Econometrica, 47, p.278.

Kasperson, R.E., and Kasperson, J.X., 1996. The social amplification and attenuation of risk. The annals of the American academy of political and social science, 545(1), pp.95-105.

Kasperson, R.E., Renn, O., Slovic, P., Brown, H.S., Emel, J., Goble, R., Kasperson, J.X. and Ratick, S., 1988. The social amplification of risk: A conceptual framework. Risk analysis, 8(2), pp.177-187.

Kauer, M., Franz, B., Schreiber, M., Bruder, R. and Geyer, S., 2012. User acceptance of cooperative maneuver based driving–a summary of three studies. Work, 41(Supplement 1), pp.4258-4264.

Kleijnen, M., Lee, N. and Wetzels, M., 2009. An exploration of consumer resistance to innovation and its antecedents. Journal of economic psychology, 30(3), pp.344-357.

König, M. and Neumayr, L., 2017. Users' resistance towards radical innovations: The case of the self-driving car. Transportation research part F: traffic psychology and behavior, 44, pp.42-52.

Kohl, C., Knigge, M., Baader, G., Böhm, M., and Krcmar, H., 2018. Anticipating acceptance of emerging technologies using Twitter: the case of self-driving cars. Journal of Business Economics, 88(5), pp.617-642.

Körber, M., Gold, C., Lechner, D. and Bengler, K., 2016. The influence of age on the take-over of vehicle control in highly automated driving. Transportation research part F: traffic psychology and behavior, 39, pp.19-32.

Kranz, M., Roalter, L. and Michahelles, F., 2010, May. Things that twitter social networks and the internet of things. In What can the Internet of Things do for the Citizen (CIoT) Workshop at The Eighth International Conference on Pervasive Computing (Pervasive 2010) (pp. 1-10).

Kraus, S., Althoff, M., Heißing, B. and Buss, M., 2009, June. Cognition and Emotion in autonomous cars. In 2009 IEEE Intelligent Vehicles Symposium (pp. 635-640). IEEE.

Kim, S.B., Han, K.S., Rim, H.C. and Myaeng, S.H., 2006. Some effective techniques for naive Bayes text classification. IEEE transactions on knowledge and data engineering, 18(11), pp.1457-1466.

Krok, A., 2015. This is the stupidest misuse of Tesla's autopilot yet.

Kuderer, M., Gulati, S. and Burgard, W., 2015, May. Learning driving styles for autonomous vehicles from the demonstration. In 2015 IEEE International Conference on Robotics and Automation (ICRA) (pp. 2641-2646). IEEE.

Kuhn, M., 2008. Building predictive models in R using the caret package. Journal of statistical software, 28(5), pp.1-26.

Kyriakidis, M., Happee, R. and de Winter, J.C., 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation research part F: traffic psychology and behavior, 32, pp.127-140.

Langdon, P., Politis, I., Bradley, M., Skrypchuk, L., Mouzakitis, A. and Clarkson, J., 2017, July. Obtaining design requirements from the public understanding of driverless technology. In International Conference on Applied Human Factors and Ergonomics (pp. 749-759). Springer, Cham.

Lee, E.K., Gerla, M., Pau, G., Lee, U. and Lim, J.H., 2016. Internet of Vehicles: From intelligent grid to autonomous cars and vehicular fogs. International Journal of Distributed Sensor Networks, 12(9), p.1550147716665500.

Li, X., Wang, Y., Zhang, A., Li, C., Chi, J. and Ouyang, J., 2018. Filtering out the noise in short text topic modeling. Information Sciences, 456, pp.83-96.

Li, Y., Pan, Q., Yang, T., Wang, S., Tang, J. and Cambria, E., 2017. Learning word representations for sentiment analysis. Cognitive Computation, 9(6), pp.843-851.

Madigan, R., Louw, T., Wilbrink, M., Schieben, A., and Merat, N., 2017. What influences the decision to use automated public transport? Using UTAUT to understand public acceptance of automated road transport systems. Transportation research part F: traffic psychology and behavior, 50, pp.55-64.

Markwalter, B., 2017. The path to driverless cars [CTA insights]. IEEE Consumer Electronics Magazine, 6(2), pp.125-126.

Martins, C., Oliveira, T. and Popovič, A., 2014. Understanding the Internet banking adoption: A unified theory of acceptance and use of technology and perceived risk application. International Journal of Information Management, 34(1), pp.1-13.

McCorkindale, T., 2010. Can you see the writing on my wall? A content analysis of the Fortune 50's Facebook social networking sites. Public relations journal, 4(3), pp.1-13.

Medhat, W., Hassan, A. and Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), pp.1093-1113.

Miltgen, C.L., Popovič, A. and Oliveira, T., 2013. Determinants of end-user adoption of biometrics: Integrating the “Big 3” of technology adoption with privacy context. Decision Support Systems, 56, pp.103-114.

Nees, M.A., 2016, September. Acceptance of self-driving cars: An examination of idealized versus realistic portrayals with a self-driving car adoption scale. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 60, No. 1, pp. 1449-1453). Sage CA: Los Angeles, CA: SAGE Publications.

Neuendorf, K.A., 2016. The content analysis guidebook. [H. Bakhshi., & V. Jalaeean Bakhshande, Persian trans]. Mashhad: Jahad-e-Daneshgahi Publication.

Niculescu, A.I., Dix, A. and Yeo, K.H., 2017, May. Are you ready for a drive? User perspectives on autonomous vehicles. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (pp. 2810-2817).

Nigam, Kamal, John Lafferty, and Andrew McCallum. "Using maximum entropy for text classification." In IJCAI-99 workshop on machine learning for information filtering, vol. 1, no. 1, pp. 61-67. 1999.

Okazaki, S., Dìaz-Martìn, A.M., Rozano, M. and Menendez-Benito, H.D., 2014. How to mine brand Tweets: Procedural guidelines and pretest. International Journal of Market Research, 56(4), pp.467-488.

Olson, E.L., 2017. Will songs be written about autonomous cars? The implications of self-driving vehicle technology on consumer brand equity and relationships. International Journal of Technology Marketing, 12(1), pp.23-41.

Pak, A. and Paroubek, P., 2010, May. Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

Pang, B. and Lee, L., 2009. Opinion mining and sentiment analysis. Comput. Linguist, 35(2), pp.311-312.

Pang, B., Lee, L. and Vaithyanathan, S., 2002. Thumbs up? Sentiment classification using machine learning techniques. arXiv preprint cs/0205070.

Payre, W., Cestac, J. and Delhomme, P., 2014. Intention to use a fully automated car: Attitudes and a priori acceptability. Transportation research part F: traffic psychology and behaviour, 27, pp.252-263.

Pendleton, S., Uthaicharoenpong, T., Chong, Z.J., Fu, G.M.J., Qin, B., Liu, W., Shen, X., Weng, Z., Kamin, C., Ang, M.A. and Kuwae, L.T., 2015, September. Autonomous golf cars for a public trial of mobility-on-demand service. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 1164-1171). IEEE.

Pettersson, I. and Karlsson, I.M., 2015. Setting the stage for autonomous cars: a pilot study of future autonomous driving experiences. IET intelligent transport systems, 9(7), pp.694-701.

Piao, J., McDonald, M., Hounsell, N., Graindorge, M., Graindorge, T., and Malhene, N., 2016. Public views towards implementation of automated vehicles in urban areas. Transportation research procedia, 14(0), pp.2168-2177.

Portouli, E., Karaseitanidis, G., Lytrivis, P., Amditis, A., Raptis, O. and Karaberi, C., 2017, June. Public attitudes towards autonomous minibusses operating in real conditions in a Hellenic city. In 2017 IEEE Intelligent Vehicles Symposium (IV) (pp. 571-576). IEEE.

Ramos, J., 2003, December. Using TF-IDF to determine word relevance in document queries. Proceedings of the first instructional conference on machine learning (Vol. 242, pp. 133-142).

Rödel, C., Stadler, S., Meschtscherjakov, A. and Tscheligi, M., 2014, September. Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. In Proceedings of the 6th international conference on automotive user interfaces and interactive vehicular applications (pp. 1-8).

Rogers, E.M., 2003. Diffusion of innovations/Everett m. rogers. NY: Simon and Schuster, 576.

Rubinkam, M., 2015. Driverless truck meant to improve safety in work zones.

Ren, Y., Wang, R. and Ji, D., 2016. A topic-enhanced word embedding for twitter sentiment classification. Information Sciences, 369, pp.188-198.

Ruths, D. and Pfeffer, J., 2014. Social media for large studies of behavior. Science, 346(6213), pp.1063-1064.

SAE International, 2014. Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.

Salon. 2020. Google engineer triggered self-driving car accident that went unreported | Salon.com. [ONLINE] Available at:

https://www.salon.com/2018/10/16/googles-self-driving-cars-involved-in-unreported-crashes/. [Accessed 12 February 2020].

Schoettle, B. and Sivak, M., 2014. A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia. University of Michigan, Ann Arbor, Transportation Research Institute.

Sharma, R., Yetton, P. and Crawford, J., 2009. Estimating the effect of common method variance: The method—method pair technique with an illustration from TAM Research. Mis Quarterly, pp.473-490.

Siegrist, M., 2000. The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. Risk analysis, 20(2), pp.195-204.

Slovic, P., 1987. Perception of risk. Science, 236(4799), pp.280-285.

Socher, R., Lin, C.C.Y., Ng, A.Y. and Manning, C.D., 2011, January. Parsing natural scenes and natural language with recursive neural networks. In ICML.

Spinrad, N., 2014. Google car takes the test. Nature, 514(7523), pp.528-528.

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H. and Demirbas, M., 2010, July. Short text classification in Twitter to improve information filtering. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (pp. 841-842).

St Louis, C. and Zorlu, G., 2012. Can Twitter predict disease outbreaks? Bmj, 344.

Tan, A.H., 1999, April. Text mining: The state of the art and the challenges. In Proceedings of the pakdd 1999 workshop on knowledge discovery from advanced databases (Vol. 8, pp. 65-70). sn.

Treat, J.R., Tumbas, N.S., McDonald, S.T., Shinar, D., Hume, R.D., Mayer, R.E., Stansifer, R.L. and Castellan, N.J., 1979. Tri-level study of the causes of traffic accidents: final report. Executive summary. Indiana University, Bloomington, Institute for Research in Public Safety.

Tumasjan, A., Sprenger, T.O., Sandner, P.G., and Welpe, I.M., 2010, May. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In the Fourth international AAAI conference on weblogs and social media.

Twitter (2016a) The search API. https://dev.twitter.com/rest/public/search. Accessed February 8, 2019

Twitter (2016b) Public API: GET search/tweets. https://dev.twitter.com/rest/reference/get/search/tweets. Accessed February 9, 2019

Twitter (2016c) Help center: the basics. https://support.twitter.com/articles/131209. Accessed February 9, 2019

Tesla (2016) A Tragic Loss. https://www.tesla.com/blog/tragic-loss.

Accessed 2 July 2020.

Ulrich, D. and Smallwood, N., 2007. Building a leadership brand. Harvard Business Review, 85(7/8), p.92.

Venkatesh, V., Thong, J.Y. and Xu, X., 2016. Unified theory of acceptance and use of technology: A synthesis and the road ahead. Journal of the Association for Information Systems, 17(5), pp.328-376.

Venkatesh, V., Davis, F. and Morris, M.G., 2007. Dead or alive? The development, trajectory, and future of technology acceptance research. Journal of the association for information systems, 8(4), p.1.

Venkatesh, V., Morris, M.G., Davis, G.B. and Davis, F.D., 2003. User acceptance of information technology: Toward a unified view. MIS quarterly, pp.425-478.

Valverde-Albacete, F.J., Carrillo-de-Albornoz, J. and Peláez-Moreno, C., 2013, September. A proposal for new evaluation metrics and result visualization technique for sentiment analysis tasks. In International Conference of the Cross-Language Evaluation Forum for European Languages (pp. 41-52). Springer, Berlin, Heidelberg.

Ward, C., Raue, M., Lee, C., D'Ambrosio, L. and Coughlin, J.F., 2017, July. Acceptance of automated driving across generations: The role of risk and benefit perception, knowledge, and trust. In International Conference on Human-Computer Interaction (pp. 254-266). Springer, Cham.

Woisetschläger, D.M., 2016. Consumer perceptions of automated driving technologies: an examination of use cases and branding strategies. In Autonomous Driving (pp. 687-706). Springer, Berlin, Heidelberg.

Woiceshyn, J. and Daellenbach, U., 2005. Integrative capability and technology adoption: evidence from oil firms. Industrial and Corporate Change, 14(2), pp.307-342

Yadron, D. and Tynan, D., 2016. Tesla driver dies in the first fatal crash while using autopilot mode. The Guardian, 1.

Yap, M.D., Correia, G. and Van Arem, B., 2016. Preferences of travelers for using automated vehicles as last-mile public transport of multimodal train trips. Transportation research part a: policy and practice, 94, pp.1-16.

Zhao, R. and Mao, K., 2017. Fuzzy bag-of-words model for document representation. IEEE transactions on fuzzy systems, 26(2), pp.794-804.

Zhang, H., Sproat, R., Ng, A.H., Stahlberg, F., Peng, X., Gorman, K. and Roark, B., 2019. Neural models of text normalization for speech applications. Computational Linguistics, 45(2), pp.293-337.

Zhang, J., Yang, X. and Appelbaum, D., 2015. Toward effective Big Data analysis in continuous auditing. Accounting Horizons, 29(2), pp.469-476.

Zmud, J. and Sener, I.N., 2019. Consumer Acceptance, Trust, and Future Use of Self-driving Vehicles.

Zmud, J.P. and Sener, I.N., 2017. Towards an understanding of the travel behavior impact of autonomous vehicles. Transportation research procedia, 25, pp.2500-2519.

List of Tables and Figures

Table 1: Automation Levels provided by Gasser & Westhoff (2012).

Level	Nomenclature	Description
1	Driver only	The driver continuously (throughout the complete trip) accomplishes longitudinal (accelerating/braking) and lateral(steering) control.
2	Assisted	The driver continuously accomplishes either lateral or longitudinal control. The remaining task is accomplished by the automating system to a certain level. The driver must permanently monitor the system. The driver must at any time be prepared to take over complete control of vehicle.
3	Partially automated	The system takes over lateral and longitudinal control (for a certain amount of time and/or in specific situation). Driver must permanently monitor the system. The driver must at any time be prepared to take over complete control of vehicle.
4	Highly automated	The system takes over lateral and longitudinal for a certain amount of time in specific situations. Driver does not need to permanently monitor the system if it is active. If necessary, driver is requested to take over control by the system with a certain time buffer. All system limits are detected by the system. The system is not capable of re-establishing the minimal risk condition from every initial state.
5	Fully automated	The system takes over lateral and longitudinal control completely within the individual specification of the application. The driver does not need to monitor the system. Before specified limits of the application are reached, the system requests the driver to take over with sufficient time buffer. In absence of a takeover, the system will return to the minimal risk condition by itself. All system limits are detected by the system; the system is capable to return to the minimum risk condition in all situations.

Table 2: Overview of statistics of the Twitter data sample

	Tweets related to risk class	**Tweets related to benefit class**	**Tweets related to neutral class**
Total Number (15000)	2575	5912	6513
In percentage	17.1%	39.4%	43.4%

Table 3a: Top ten Unigram, Bigram, Trigram before cleaning

Unigram		**Bigram**		**Trigram**	
Tokens	Frequency	Tokens	Frequency	Tokens	Frequency
com	11054	Twitter com	8425	pic twitter com	5653
the	9561	pic twitter	5653	autonomouscars	1453
Twitter	8663	self-driving	3638	http bit ly	929
to	5751	driving cars	1455	http ow ly	310
http	5191	General Motors	1211	for self-driving	223
in	4807	http bit	929	http buff it	209
driving	4204	http www	901	driving cars http	164
self	4007	Mercedes Benz	709	Porsche pic twitter	155
is	2686	on the	436	http fb me	127
BMW	2647	for the	435	autonomousvehicles	104

Table 3b: Top ten Unigram, Bigram, Trigram after cleaning

Unigram		Bigram		Trigram	
Tokens	Frequency	Tokens	Frequency	Tokens	Frequency
self-drive	2738	autonomousvehicles	1862	autonomouscars	295
BMW	2368	General Motors	1062	Ford Motor company	80
Twitter	8663	Tesla Motors	363	General Motor company	73
Porsche	2179	Porsche 911	230	Porsche 911 Turbo	35
cars	1756	Mercedes Benz	199	Google self-driving	32
Mercedes	1606	Ford Motors	145	Porsche 911 GT3	32
Audi	1459	motor company	88	Mercedes Benz Stadium	31
New	1150	new BMW	83	autonomouscars Bitly	28
general	1123	Elon Musk	72	Testing autonomouscars	25
Bitly	983	Google self-drive	47	autonomouscar project	20

Table 4: Model Performance with various evaluation matrices

Classification Classes	precision	recall	f1-score	support
Benefit related tweets	83%	84%	83%	1512
Risk related tweets	70%	59%	64%	632
Neutral tweets	83%	87%	85%	1606
Macro average	79%	77%	77%	3750
Weighted average	81%	81%	81%	3750

Table 5: Total tweets each year in terms of risk and benefit classes

Yearly trend	Tweets with neutral class	Tweets with benefit class	Tweets with risk class	Risk ratio	Benefit ratio
2016	66,864(39.3%)	62,723(36.9%)	40,369(23.7%)	0.392	0.608
2017	65,635(45.3%)	55,895(38.6%)	23,139(15.9%)	0.293	0.707
2018	44,901(31.8%)	63,779(45.2%)	32,120(22.8%)	0.335	0.665
2019	89(29.5%)	135(44.8%)	77(25.5%)	0.363	0.637
Overall	177,489(38.95%)	182,533(40.05%)	95,705(21.01%)	0.343	0.657

Figure 1: Data labeling using Amazon SageMaker Ground Truth

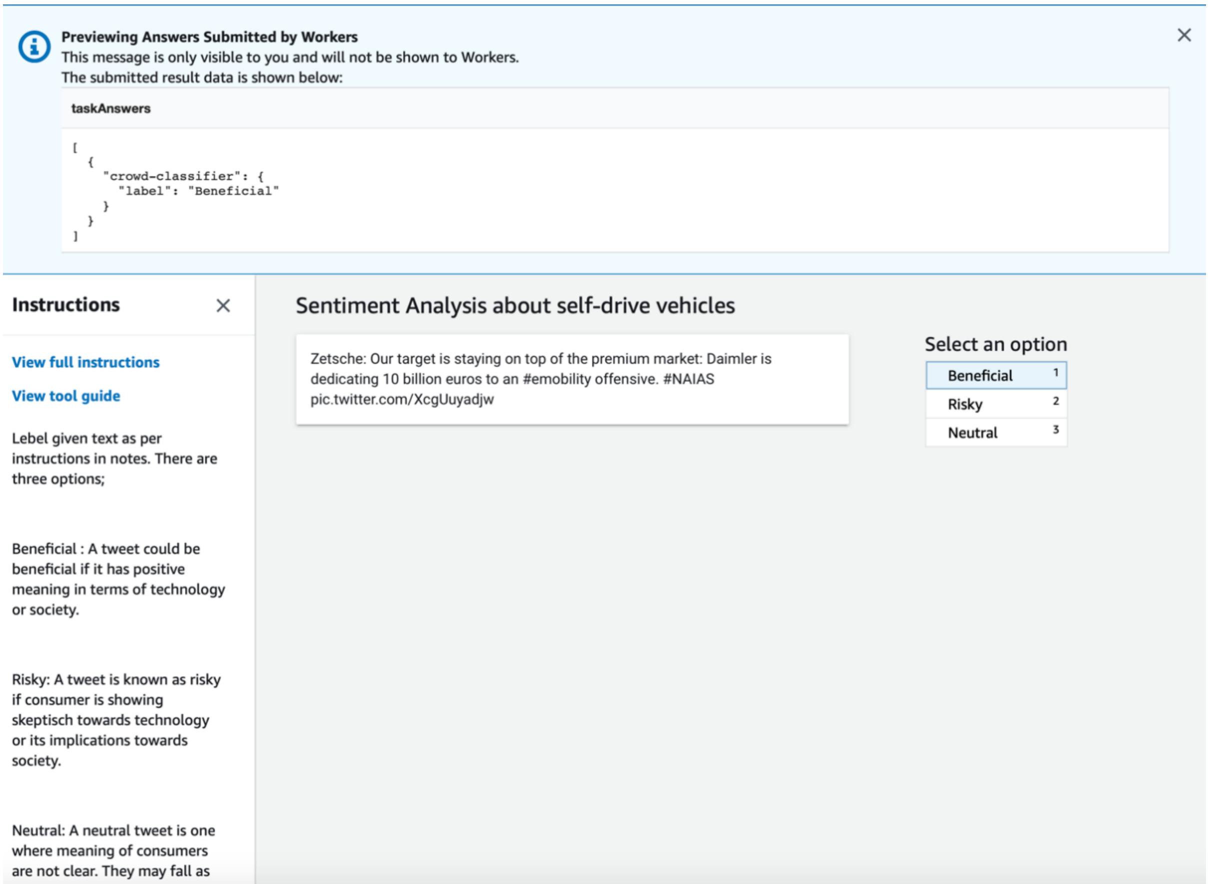


Figure 2: NLP data pipeline implemented for sentiment prediction.

Input (Text) → Data Clean → Feature Extraction → Word Embedding → ML Algorithm → Perception Prediction

Figure 3: Data cleaning pipeline

Raw Text → Noise Removal → Normalization → Standardization → Clean Text

Figure 4a: Data before applying NLP cleaning pipeline

In [49]:
```
untidy_data.tail()
```

Out[49]:

	TEXT	word_count	char_count	stopwords	hastags	@	numerics	upper
14995	Elon Musk: Google won't compete with Tesla on ...	16	127	5	0	0	0	0
14996	In which I am touched by the altruism of every...	17	241	5	0	0	0	1
14997	Intergalactic dream trio on the track #BMW #Mp...	14	162	2	2	6	0	1
14998	And its good morning all from australia pic.tw...	8	66	3	0	0	0	0
14999	http://caferacerpasion.com 1979 BMW R100RS #...	15	167	1	4	0	1	5

Figure 4b: Data after applying NLP cleaning pipeline

```
In [50]: tidy_data.tail()
```

Out[50]:

	TEXT	word_count	char_count	stopwords	hastags	@	numerics	upper
14995	elon musk google wont compete tesla selfdrivin...	9	57	0	0	0	0	0
14996	touched altruism everyday driver beta testing ...	10	98	0	0	0	0	0
14997	intergalactic dream trio track bmw mpower lien...	12	141	0	0	0	0	0
14998	good morning australia	3	22	0	0	0	0	0
14999	httpcaferacerpasioncom bmw streettracker bmw m...	10	93	0	0	0	0	0

Figure 5: Model Architecture

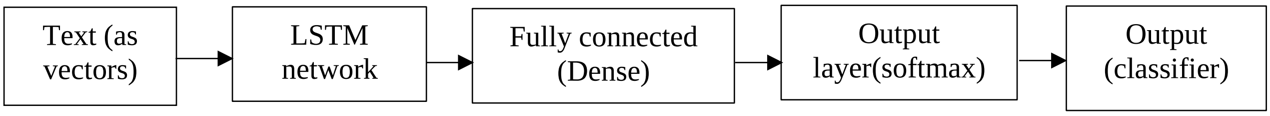


Figure 6: Confusion Matrix for multi-class classifier

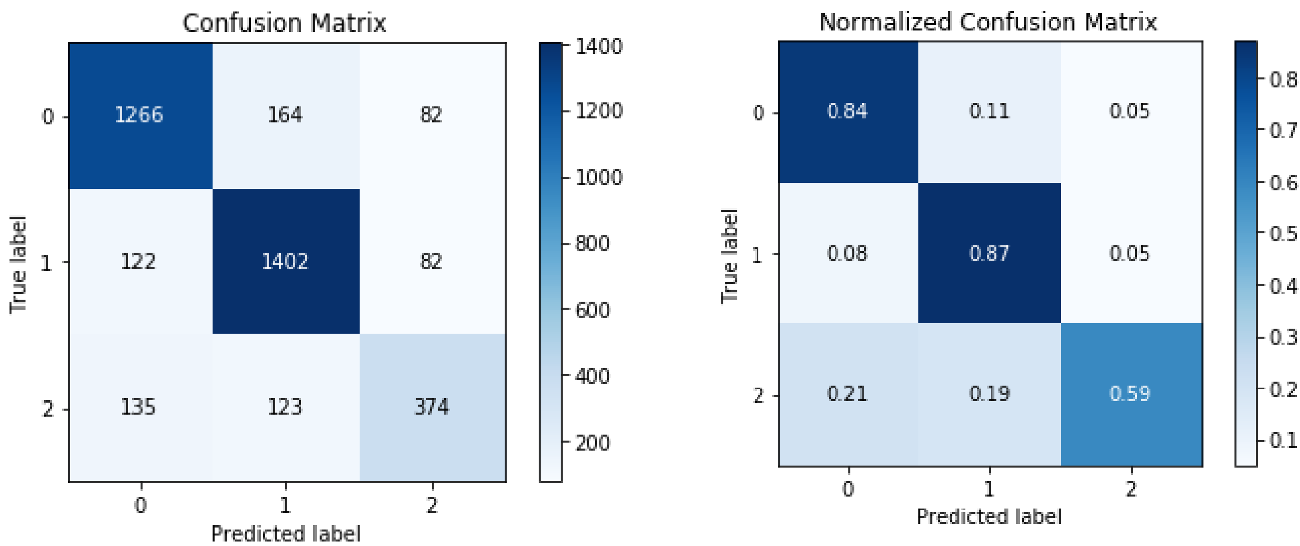


Figure 7: Distribution of sentiments from 2016-2019

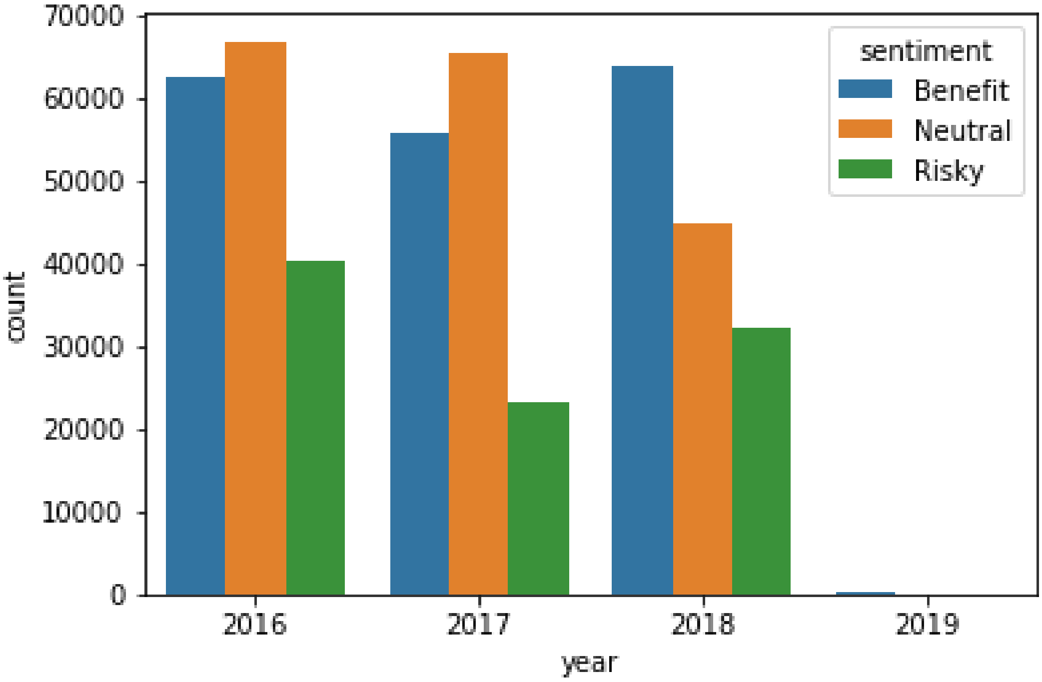


Figure 8: Distribution of risk-related tweets every month

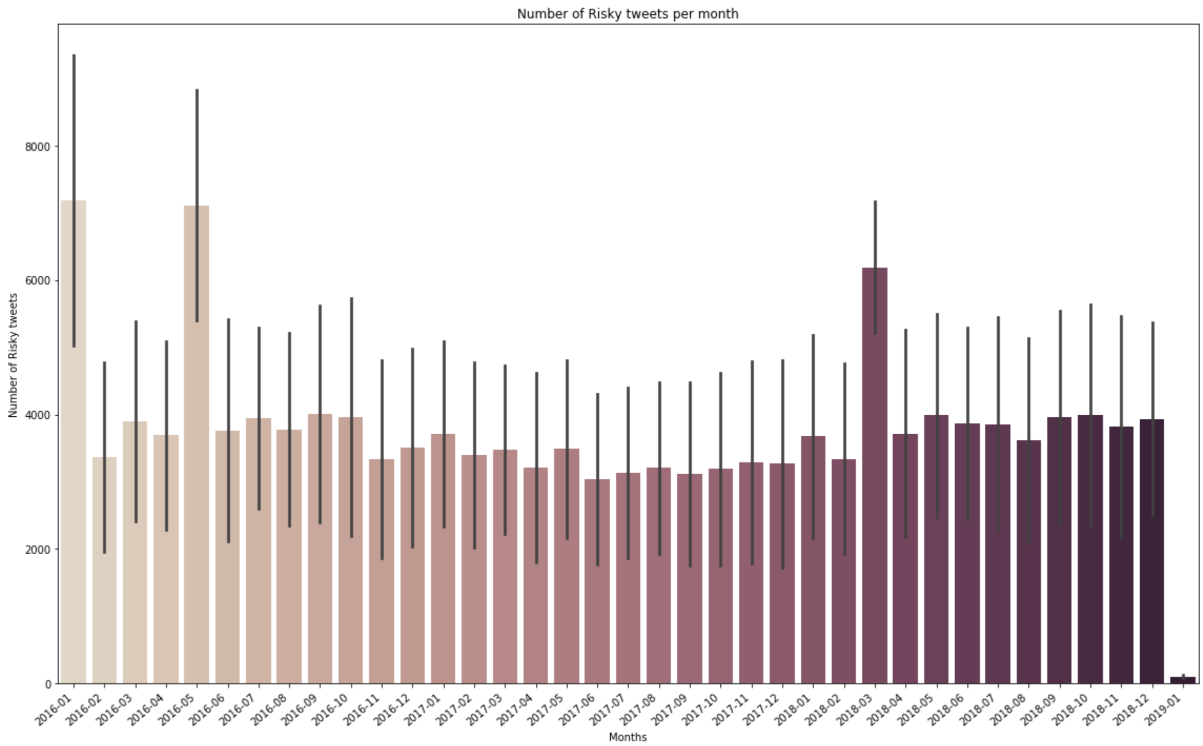


Figure 9: Distribution of BR & RR ratios from 2016-2019

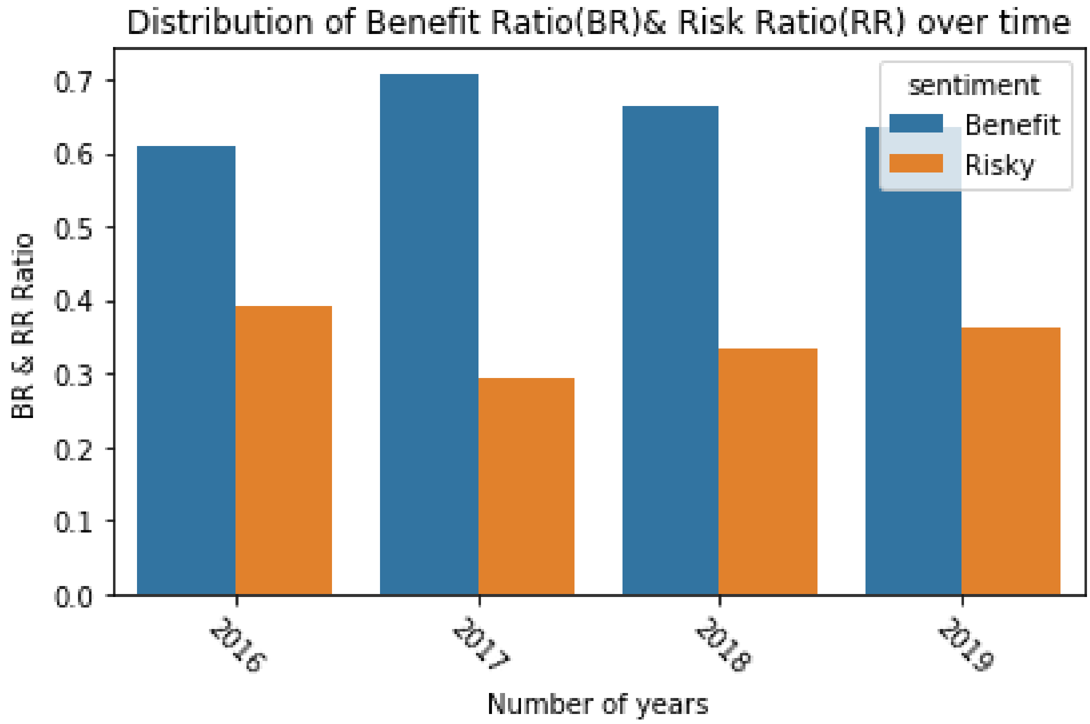


Figure 10: Distribution of BR & RR ratios every month

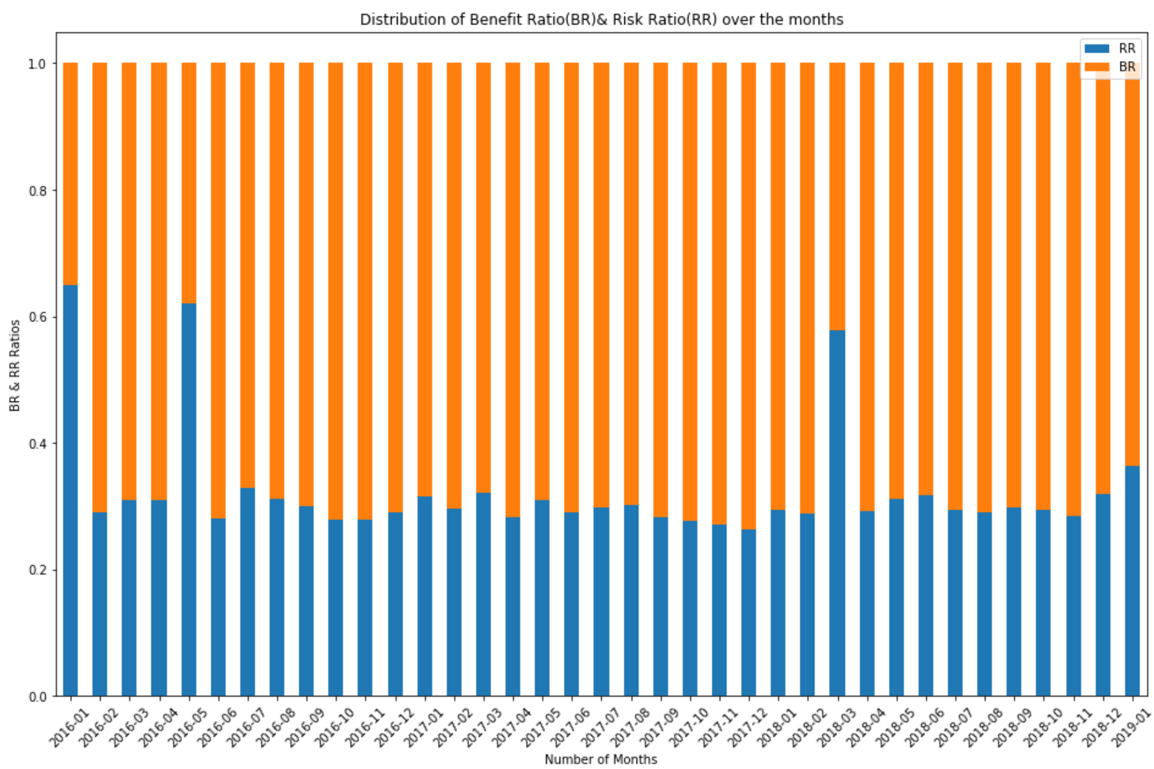


Figure 11: Total number of tweets for leading players in the ecosystem

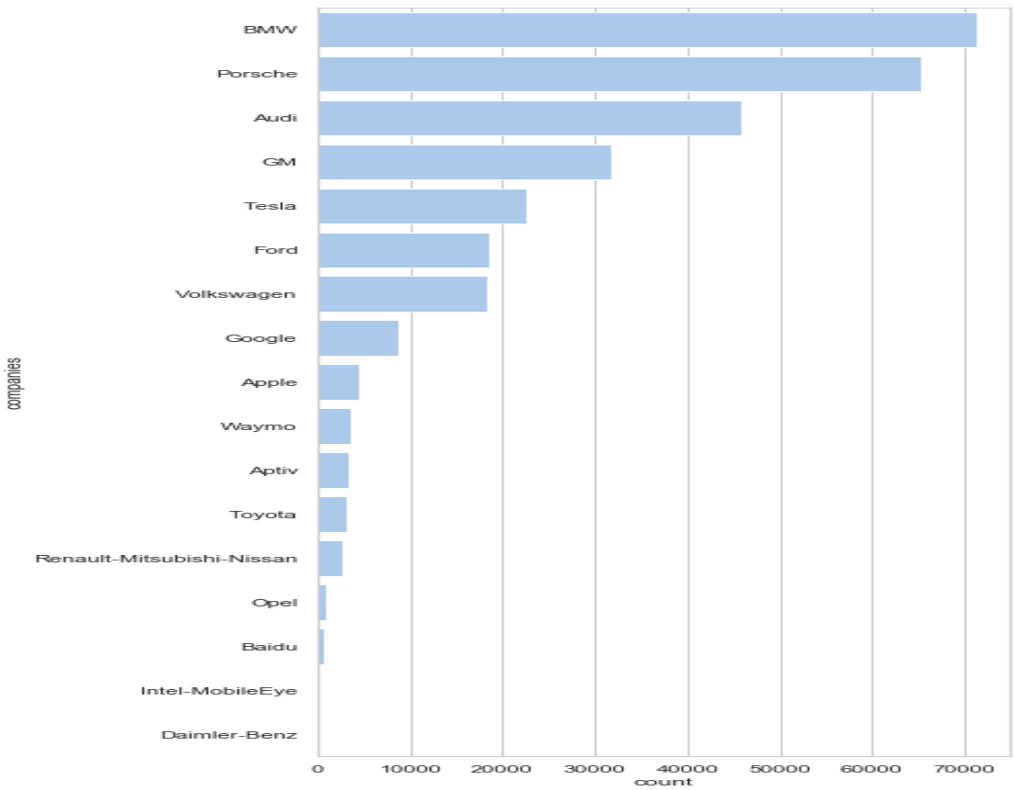


Figure 12: Word Cloud related to risk tweets about autonomous vehicles

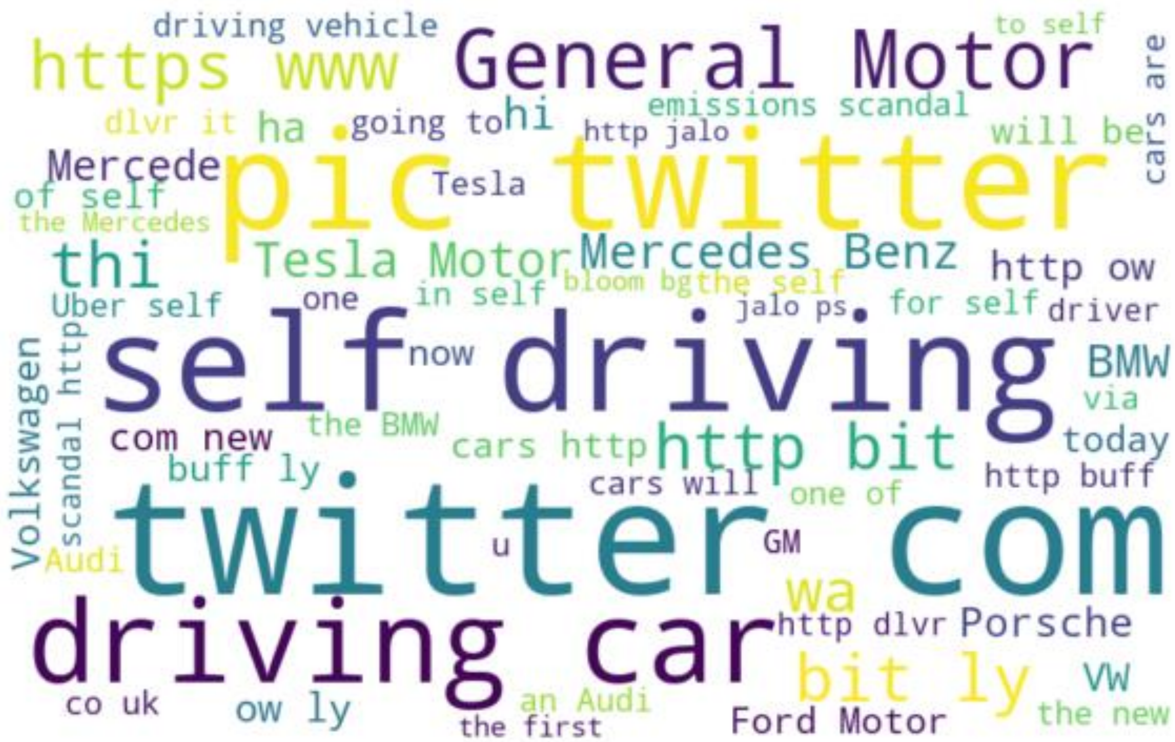
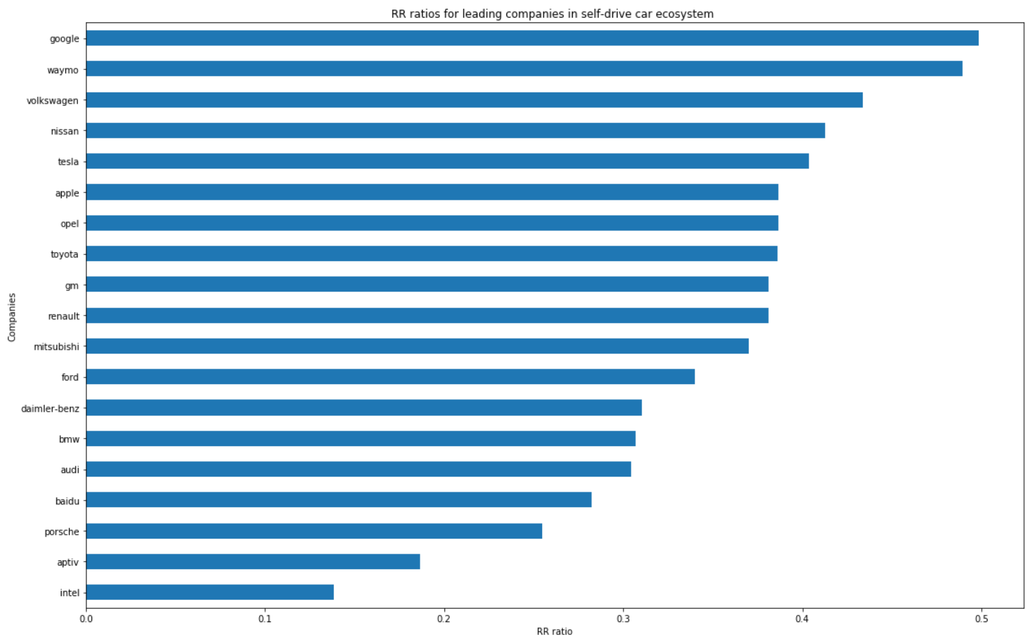


Figure 13: Risk ratio of leading players in self-drive vehicle ecosystem



Concluding Remarks to Thesis

In this study, we explore the concept of business ecosystems as a research perspective for studying the relationship between individual organizations and their respective business networks. Novel research avenues are needed to investigate the relationships between focal technology and its components and complementors. Such a scenario becomes vital since technical practices shift from a dominantly stand-alone perspective to a network perspective. The business ecosystem framework is an exciting place to start from such a view.

Several academic and managerial perspectives are provided by the ecosystem conceptualization offered in this study. Our research leads to the identification of key actors in the business ecosystem and their interrelationships and dependencies. A manager, thereby an organization, who can better visualize and interpret their business ecosystem, may have a competitive advantage in their industry. A thorough understanding of one's business ecosystem could help a firm move faster and more efficiently than its rivals while leveraging expertise from other key stakeholders. Understanding technology interactions within the ecosystem also have ramifications for technology proliferation(Adner and Kapoor, 2016), platform centralization(Ceccagnoli et al., 2012; Gawer & Cusumano, 2008; Parker et al., 2016), and regularization of technological ecosystems(Wareham et al., 2014).

This paper adds to existing research streams by proposing an ecosystem framework that explores innovation processes and offers an in-depth understanding of the interactions of various technological actors in such systems. We conceptualize an innovation ecosystem as a complex, dynamic system that includes the collective impact of each actor within the network. We analyze this impact in terms of the integration and interaction of the technology ecosystem, which offers a landscape of innovation ecosystems, and comprehensively assess their consequences. Furthermore, this approach explores the interaction between knowledge and business economies and hence, investigates the interplay between component, focal technology, and complementor. Since there is so much uncertainty about cutting-edge technological trajectories, such a perspective is particularly valuable for understanding technological advances.

A few limitations to the study presented in this article provide opportunities for further investigation. Since business ecosystems are likely to have a wide variety of key actors with different value chains, research in this field focuses on a shared syntax that will provide deep insight into better management of these ecosystems. Another limitation is that we only use a limited number of fundamental viewpoints to explore the business ecosystem's critical aspects. In this study, we use the value creation perspective(Adner and Kapoor, 2010) to examine technological trajectories. While we think this perspective will cover essential parts of the innovation ecosystem framework, it might well be worth extending the set of theories from information systems and strategic management literature.

Future research might concentrate on evaluating technical relationships in a business ecosystem. Similar to how a focal technology provides the platform in a single ecosystem, it's conceivable that a single actor's interrelationships have specific capabilities in that ecosystem. Researchers in technology management and information system could contribute to this area of research by studying the dynamics and capabilities of a focal firm, component, or complementor within ecosystems. Regarding technological capabilities, the interactions between key players in business ecosystems are currently unexplored territory. Another avenue for research would be to look at ecosystems from the standpoint of specific technologies. This stream would focus on findings managerially related issues that technology entrepreneurs might face. Another corresponding stream of research investigates how ecosystem relationships change over the course of the platform technology's lifecycle. A holistic view of the ecosystem becomes more critical as the interaction of complementor with its focal technology becomes central in value creation.