

- Luiss Open - <https://open.luiss.it/en> -

## Text and Image Mining Methods for Business Research and Education

---

**Many consumer-firm interactions have moved from offline to online**, transforming platforms and marketplaces such as Google, Amazon, Facebook into crucial touchpoints of the customer journey. Nearly 70% of consumers rely on social media to resolve customer service issues (Ahmed, 2017), and 58% of them read reviews before choosing a restaurant (Gatherup, 2018). In addition, firms actively connect with consumers by starting conversations (e.g., Tweeting about a live event) or responding to online consumer complaints. Image-driven social media platforms (e.g., Instagram) are gaining unprecedented relevance, such that 71% of businesses use them to promote services, and 75% of consumers engage with those promotions (Clarke, 2019). Along with the **increase of digital interactions**, an unprecedented amount of unstructured data, mainly text and images, has disrupted business research.

Unlike structured data (i.e., numbers), which has been the main source of information that business analyzes for the past decades, unstructured data comes through a pathway of customers **telling stories to interconnect with other customers** and/or firms, rather than from solicited customer feedback surveys, predefined by organizations. Unveiling insights such as brand sentiment and audience segmentation from these text and image driven stories, requires the use of methods and tools from natural language processing, computer vision and machine learning. The present article describes the state-of-the-art of the text and image mining fields, discusses the key objectives and methods for business research, and explains the implications for business education.

Over the last five years, the increase in usability and **popularity of text and image analytics** (aka, text and image mining) methods has grown exponentially. In fact, recent business literature has dedicated several step-by-step guides in marketing (Berger et al. 2020), management (McKenny et al. 2018), retailing (deKimpe 2020) and **business in general** (Schwenzow et al. 2020), that should help researchers to use unstructured data for business insight. In addition, the expected growth rate for the global text mining market is 18.1% from 4.75 billion in 2019 to 16.85 billion in 2027 (Reports and Data 2020) and the **global image recognition** (aka, image mining) market size was valued at USD 27.3 billion in 2019 and is expected to register growth of 18.8% from 2020 to 2027 (Reports and Data 2020).

To keep pace with these changes, it is important understanding the key objectives and methods to implement text and image mining. We can identify 3 main objectives for text and image mining (Villarroel Ordenes and Zhang 2020). First, the operationalization of an observed or predefined construct. In this case, a **manager might be interested in the consumer sentiment or trust** derived from words and images in social media. Second, the identification of unobserved constructs or clusters. An example is when managers want to learn the most relevant topics that are discussed

in online reviews (e.g., hotel features), or when they want to identify different types of service experience (e.g., tangible vs. experiential) by clustering consumer images in social media. The third objective is the identification of **relationships amongst features in text or images**. Researchers might be interested in the words that most frequently occur with a determined brand (e.g., cars), or uncovering relations between the text and images that brands are using in social media.

Each of the aforementioned objectives can be accomplished by one or a combination of methods. Measuring consumer sentiment can be achieved by methods such as lexicon-based, machine learning, deep learning, ensembles, and transformer (Heitmann et al. 2020). There are also several methods for **the identification of topics** out of customer reviews such as Latent Dirichlet Allocation, Correlated Topic Models and Structured Topic models (Grewal et al. 2020), each taking different assumptions that might enhance their model fit in a business context. Identifying entity relations offers a broad range of methods to assess relationships between constructs, as well. For example, in **evaluating the similarity between brands**, researchers could use the Cosine or Word Embedding distances measures between the words' used to describe brands (Netzer et al. 2012). In image mining, applications such as Amazon Rekognition can identify objects or actions in images with high accuracy (e.g., humans, smiles, logos). In addition, advances in deep learning and neuronal nets have contributed to the development of customized algorithms to identify image types or motives that a researcher might be interested in (e.g., rugged brands; Liu, Dzyabura, and Mizik 2020).

The **aforementioned developments in text and image mining methods** resulted in a pressing need to cover them in business analytics curriculum. These courses face the challenge of balancing advanced technical capabilities in programming (e.g., R, Python, Knime) and statistics, with empirical demonstrations based on **real business data**. Cumbersome in achieving this balance is the implementation of applied projects that stimulate students to find the right business questions (exploratory, causal, or predictive?), suitable text and image data to answer these questions (e.g., social media, brand forums, or online reviews?), and the most efficient set of methods (e.g., supervised, unsupervised learning, or both?) and visualizations (e.g., which type of graph or figure?) to provide business insight. Without a curriculum that promotes student development in these three areas, it is likely that future managers will struggle in cutting the clutter from Big Data.

---

Article printed from Luiss Open: <https://open.luiss.it/en>

URL to article: <https://open.luiss.it/en/2020/12/18/text-and-image-mining-methods-for-business-research-and-education/>

Copyright © 2020 Luiss Open. All rights reserved.