

Decryptare codici: dai geroglifici alle notizie di mercato

Pre-print version of the following publication: | Versione pre-print della seguente pubblicazione:

Original Citation/Citazione:

Barone, Emilio; Barone, Gaia. (2025). Decryptare codici: dai geroglifici alle notizie di mercato.
<https://rdoi.net/10.13140/RG.2.2.12955.58407>

Availability/Disponibilità:

This version is available at: [11385/258718](https://doi.org/10.13140/RG.2.2.12955.58407) since: 2026-02-18T22:03:29Z - Questa versione è disponibile alla pagina:
[11385/258718](https://doi.org/10.13140/RG.2.2.12955.58407) dal: 2026-02-18T22:03:29Z

Publisher/Casa editrice:

Published version/Pubblicato:

DOI: <https://dx.doi.org/10.13140/RG.2.2.12955.58407>

License/Licenza:

Tutti i diritti riservati

Availability/Termini d'uso:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. Works made available under a Creative Commons license can be used according to the terms and conditions of said license. For all terms of use and more information see the publisher's website. | I termini e le condizioni relativi al riutilizzo della presente versione della pubblicazione sono disciplinati dalla politica editoriale. Le opere messe a disposizione con licenze Creative Commons possono essere utilizzate conformemente ai termini e alle condizioni previste da tali licenze. Per l'insieme delle condizioni di utilizzo e per ulteriori informazioni si rinvia al sito web dell'editore.

This item was downloaded from IRIS Luiss (<https://iris.luiss.it/>). When citing, please refer to the published version. | Questo documento è stato scaricato da IRIS Luiss (<https://iris.luiss.it/>). Per la citazione, fare riferimento alla versione pubblicata sul sito dell'editore.

(Article begins on next page | Il contributo inizia nella pagina successiva)

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/394479391>

Decriptare codici: dai geroglifici alle notizie di mercato Unscrambling Codes: From Hieroglyphs to Market News

Preprint · August 2025

DOI: 10.13140/RG.2.2.12955.58407

CITATIONS

0

2 authors, including:



Emilio Barone

LUISS Guido Carli, Libera Università Internazionale degli Studi Sociali

79 PUBLICATIONS 576 CITATIONS

SEE PROFILE

Decriptare codici: dai geroglifici alle notizie di mercato

Unscrambling Codes: From Hieroglyphs to Market News

Emilio Barone, Luiss Guido Carli

Gaia Barone, National College of Ireland

Keywords

Natural Language Processing, Hedge Funds, Cryptography

Jel codes

C38, C58, G23

Quest'articolo esamina alcuni eventi che hanno aperto la strada allo sviluppo della *sentiment analysis* (o *opinion mining*), una tecnica apparentemente utilizzata dall'*hedge fund* Medallion di Jim Simons per ottenere una performance "impossibile": un tasso di rendimento medio annuo del 66% nei 31 anni tra il 1988 e il 2018. La *sentiment analysis* è un potente strumento che utilizza l'«elaborazione del linguaggio naturale» (*natural language processing* - NLP), o linguistica computazionale, per determinare se una notizia avrà un impatto positivo, negativo o nullo sui titoli di una società e, in ultima analisi, per scoprire le sistematicità presenti nei prezzi delle azioni. Qui ne facciamo risalire le origini ad alcuni strumenti inventati nei secoli passati per riconoscere il significato dei simboli usati dagli esseri umani per comunicare tra loro in modo chiaro o criptato: i geroglifici egiziani, il cifrario di Giulio Cesare, le abbreviazioni di Fibonacci, il codice di Maria Stuarda, e così via.

This paper reviews some steps that paved the way for the development of sentiment analysis (or opinion mining), a technique apparently used by Jim Simons' Medallion fund for scoring an 'impossible' performance: a 66% annual average rate of return in the 31 years between 1988 and 2018. Sentiment analysis is a powerful tool that uses natural language processing (NLP), or computational linguistics, to determine whether a text is positive, negative or neutral about a company and, in ultimate analysis, to discover stock prices patterns. Here we trace back its origins to some tools invented during past centuries to recognize the meaning of symbols used by humans to communicate, plainly or secretly: Egyptians' hieroglyphs, Julius Caesar's cipher, Fibonacci's abbreviations, Mary Stuart's code, etc.

I. Introduzione

Cos'hanno in comune Jean François Champollion, archeologo ed egittologo francese, e James (Jim) Harris Simons, matematico e gestore di *hedge funds* (Figura 1)? Sono stati entrambi decifratore di codici. Champollion è riuscito a decifrare i messaggi crittografati presenti nella stele di Rosetta. Aveva studiato la lingua copta, che derivava dall'antico egiziano, e questa conoscenza gli permise di comprendere la sezione demotica della stele di Rosetta. Jim Simons si è laureato in matematica al MIT, ha un dottorato di ricerca in matematica a UC Berkeley ed è stato ricercatore ad Harvard. Poi si è messo a decifrare codici russi presso l'Institute for Defense Analysis (IDA), durante la Guerra Fredda.

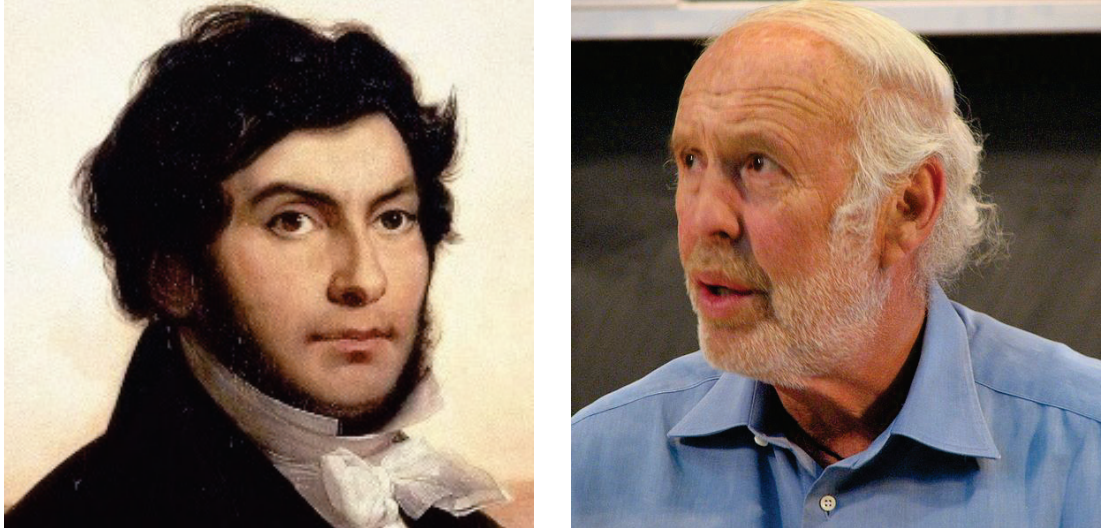
John Hull (2021) cita Jim Simons nel capitolo sul Natural Language Processing (NLP) del suo libro sul Machine Learning in Business:

In ogni momento, diventano disponibili nuove fonti di dati. Una possibilità è quella di cercare di stare un passo avanti rispetto alla maggior parte degli altri nello sfruttare queste nuove fonti. Un'altra possibilità è quella di sviluppare modelli migliori di quelli utilizzati dagli altri e mantenerli riservati. Renaissance Technologies è un hedge fund che ha seguito il secondo tipo di approccio. Ha sviluppato modelli sofisticati per comprendere il comportamento dei prezzi delle azioni e ha avuto un incredibile successo. Altri hedge funds non sono stati in grado di replicarne la performance. Nel periodo 1988-2018, il tasso di rendimento lordo di Medallion, il fondo di punta di Renaissance Technologies, è stato pari al 66% annuo. In particolare, nel 2008, quando lo S&P 500

ha perso il 38,5%, il tasso di rendimento di Medallion è stato all'incirca pari al 100%. Dopo che il fondatore di Renaissance Technologies, Jim Simons, si è ritirato a vita privata, nel 2009, il suo posto è stato preso da Robert Mercer e Peter Brown, entrambi esperti di linguistica computazionale. (p. 189)

Figura 1

Jean-François Champollion and James (Jim) Harris Simons



Fonte: Wikipedia.

2. Medallion Fund

Come riferisce Gregory Zuckerman (2019) nel suo *bestseller* «*The man who solved the market: How Jim Simons launched the quant revolution*», Simons lasciò Harvard nel 1964 per lavorare all'Institute for Defense Analysis (IDA) «un istituto di ricerca *d'élite* che assumeva matematici dalle migliori università per assistere la National Security Agency – l'agenzia di *intelligence* più grande e segreta degli Stati Uniti – nello scoprire e attaccare codici e cifrari russi. ... L'IDA ha insegnato a Simons come sviluppare modelli matematici per discernere e interpretare le sistematicità presenti in dati apparentemente privi di significato». (pp. 23-4).

Quando era all'IDA, Simons aveva scritto un articolo “a più mani” sui modelli probabilistici volti a prevedere il comportamento del mercato azionario. Rilevare le sistematicità presenti nei prezzi delle azioni divenne uno dei suoi obiettivi. Tuttavia, a quanto pare, non gli interessavano le interpretazioni:

“Non so perché i pianeti orbitano intorno al sole”, ha detto una volta Simons a un collega, suggerendogli di non dedicare troppo tempo a capire perché esistono sistematicità nel mercato. “Questo non vuol dire che non posso prevederle.” [Zuckerman (2019), p. 151]

Quali che siano le tecniche utilizzate da Jim Simons, i risultati ottenuti sono sorprendenti (Appendice A):

C'erano ragioni importanti che mi hanno indotto a raccontare la storia di Simons. Ex professore di matematica, Simons è probabilmente il trader di maggior successo nella storia della finanza moderna. Dal 1988, l'hedge fund di punta di Renaissance, il Medallion, ha generato rendimenti medi annui del 66%, accumulando profitti per oltre 100 miliardi di dollari (si veda l'Appendice 1 per come arrivo a queste cifre). Nessuno, nel mondo degli investimenti, si avvicina a questi risultati. Quelli ottenuti da Warren Buffett, George Soros, Peter Lynch, Steve Cohen e Ray Dalio sono tutti inferiori (si veda l'Appendice 2). [Zuckerman (2019), p. xvi]

I «tassi di rendimento logaritmici» (*log returns*) lordi coerenti con i dati dell'Appendice A sono riportati nella Tavola 1, insieme ai dati corrispondenti per lo S&P 500 Total Return Index (che tiene conto del reinvestimento dei dividendi). Il Medallion Fund (MF) ha sovra-performato lo S&P 500 in 28 anni su 31 e non ha mai registrato perdite. Un dollaro investito nel Medallion Fund alla fine del 1987 si è trasformato in 3.995.683 dollari dopo 31 anni!

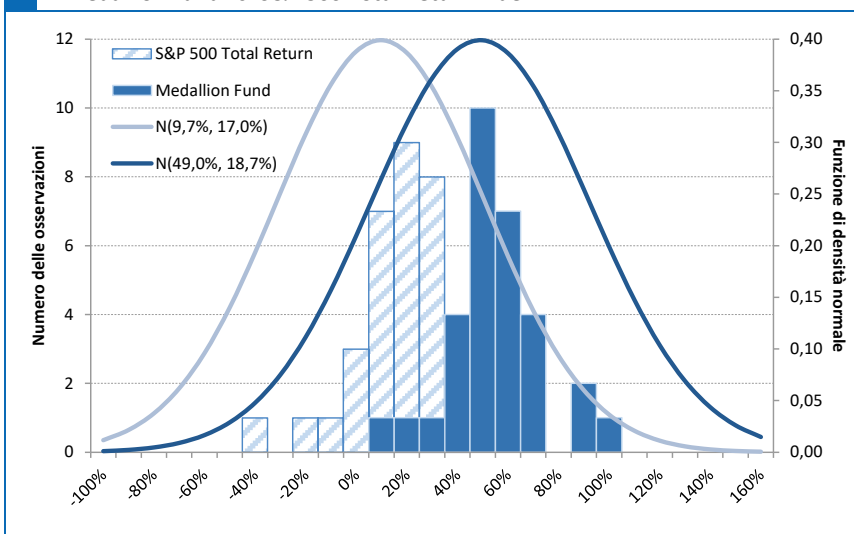
Tavola 1

Medallion Fund vs. S&P 500 Total Return Index: tassi di rendimento logaritmici lordi

| Fine anno | Medallion Fund | S&P 500 Total Return | Fine anno | Medallion Fund | S&P 500 Total Return |
|-----------|----------------|----------------------|-----------|----------------|----------------------|
| 1988 | 15,1% | 15,4% | 2005 | 45,6% | 4,8% |
| 1989 | 1,0% | 27,5% | 2006 | 61,0% | 14,7% |
| 1990 | 57,5% | -3,2% | 2007 | 86,1% | 5,3% |
| 1991 | 43,4% | 26,6% | 2008 | 92,5% | -46,2% |
| 1992 | 38,5% | 7,3% | 2009 | 55,7% | 23,5% |
| 1993 | 43,1% | 9,6% | 2010 | 45,4% | 14,0% |
| 1994 | 66,0% | 1,3% | 2011 | 53,7% | 2,1% |
| 1995 | 42,5% | 31,9% | 2012 | 45,0% | 14,8% |
| 1996 | 36,7% | 20,7% | 2013 | 63,6% | 28,1% |
| 1997 | 27,4% | 28,8% | 2014 | 56,0% | 12,8% |
| 1998 | 45,2% | 25,1% | 2015 | 52,7% | 1,4% |
| 1999 | 30,5% | 19,1% | 2016 | 52,2% | 11,3% |
| 2000 | 82,5% | -9,5% | 2017 | 61,7% | 19,7% |
| 2001 | 44,9% | -12,7% | 2018 | 56,8% | -4,5% |
| 2002 | 41,3% | -25,0% | | | |
| 2003 | 36,5% | 25,2% | μ | 49,0% | 9,7% |
| 2004 | 40,2% | 10,3% | σ | 18,7% | 17,0% |

Nota: le stime OLS della regressione dei *log-returns* del Medallion Fund vs. S&P 500 sono $\alpha = 0$ ($t = 0$) e $\beta = -0,5$ ($t = -3,3$).

Figura 2

Medallion Fund vs. S&P 500 Total Return Index

Fonte: Zuckerman (2019, Appendice 1) e Yahoo Finance (^SP500TR).

La Figura 2 riporta le distribuzioni dei tassi di rendimento annui di entrambi gli investimenti (MF e S&P 500). La media, μ , dei tassi di rendimento logaritmici del Medallion Fund (49,0%) è 5 volte maggiore della media dei tassi di rendimento logaritmici dello S&P 500 (9,7%), mentre le deviazioni standard, σ , sono simili (18,7% e 17,0%, rispettivamente). Sarà davvero impossibile battere quest'incredibile performance!

2.1 L'impossibile

“The Impossible” è il titolo di un film basato sullo tsunami che ha colpito l'Oceano Indiano nel 2004, un'osservazione estrema tra i disastri naturali a bassa frequenza e ad alta gravità. A volte, i fenomeni molto rari come i crolli dello S&P 500 del 19 ottobre 1987 (-20,5%; da 282,70 a 224,84) e del 16 marzo 2020 (-12,0%; da 2.711,02 a 2.386,13) sono chiamati eventi a “6 sigma” ossia a “6 deviazione standard”. Se X è una variabile normale standardizzata, la probabilità che risulti $X > 6$ [= $\text{INV.NORM.S}(1-1\text{E}-10)$ in Excel™] è pari a 0,0000000001, ovvero 1 su 10 miliardi.

Qual è la probabilità di ottenere un tasso di rendimento logaritmico medio annuo del 49,0% per 31 anni consecutivi senza mai registrare una perdita?

2.2 Leva finanziaria e ruota della roulette

Performance stellari non sono possibili senza che si usi la «leva finanziaria» (*leverage*). Un brillante esempio dell'effetto-*leverage* sulla performance è stato offerto da Mark Rubinstein (1999):

Ai fini di questa discussione, supponiamo che sulla ruota della roulette siano presenti solo i numeri da uno a 36 e che non ci sia lo zero o il doppio zero. Se mettiamo una fiche su ognuno dei numeri da 1 a 35, il più delle volte vinceremo una piccola somma. Supponiamo ora di aver preso in prestito la maggior parte del denaro necessario per acquistare le 35 fiches che vogliamo scommettere. Il più delle volte il rendimento sul capitale impiegato sarà molto alto ma, in media, una volta su 36, perderemo tutto. ... Il problema generale del tipo “uno-35” e “36” si pone quando valutiamo le strategie di investimento. Una lezione che possiamo trarre è che quando vediamo la performance di un gestore, dobbiamo chiederci: la sua è una strategia che vince 35 volte su 36, ma fallisce miseramente la 36-esima volta? Per via di questo problema, è facile lasciarsi ingannare dal track record di un gestore. [Rubinstein (1999, pp. 3-4)]

2.3 Natural Language Processing (NLP)

Il nostro *a priori* è che c'è molto di più della “buona sorte” dietro la performance stellare del Medallion Fund: (i) il miglior capitale umano; (ii) il massiccio utilizzo di computer; (iii) l'«elaborazione del linguaggio naturale» (*natural language processing* - NLP). Robert Mercer e Peter Brown, che hanno iniziato a collaborare con Jim Simons nel 1993, lavoravano entrambi presso l'IBM “alla ricerca di metodi per far sì che i computer migliorassero le trascrizioni del linguaggio parlato e le traduzioni delle lingue straniere” [Zuckerman (2019), p. 4]. L'obiettivo finale era la ricerca di un linguaggio comune, ossia la binarizzazione. Probabilmente, il *natural language processing*, ossia la linguistica computazionale, è stata la svolta-chiave che aiuta a spiegare la performance di MF. Gran parte dei *big data* disponibili sono sotto forma di parole scritte o parlate. Le quotazioni e le quantità scambiate sono statistiche di base che riassumono una gran quantità di informazioni, ma le notizie di mercato sono altrettanto importanti. Le regole del linguaggio sono difficili da comunicare a una macchina e le parole possono avere diversi significati. “Decriptare il codice del mercato” non è stato facile.

Nel resto di questo lavoro descriveremo alcuni passaggi del lungo viaggio compiuto dagli esseri umani per comunicare tra loro, in modo chiaro o criptato. Passeremo in rassegna i geroglifici egiziani (§3), il cifrario di Giulio Cesare (§4), il Liber Abaci di Fibonacci (§5), la scrittura speculare di Leonardo da Vinci (§6), il Codice di Maria Stuarda (§6), la decodifica Monte Carlo (§ 8), Shazam e Google Lens (§9), Kaggle (§10). E infine trarremo le conclusioni (§11).

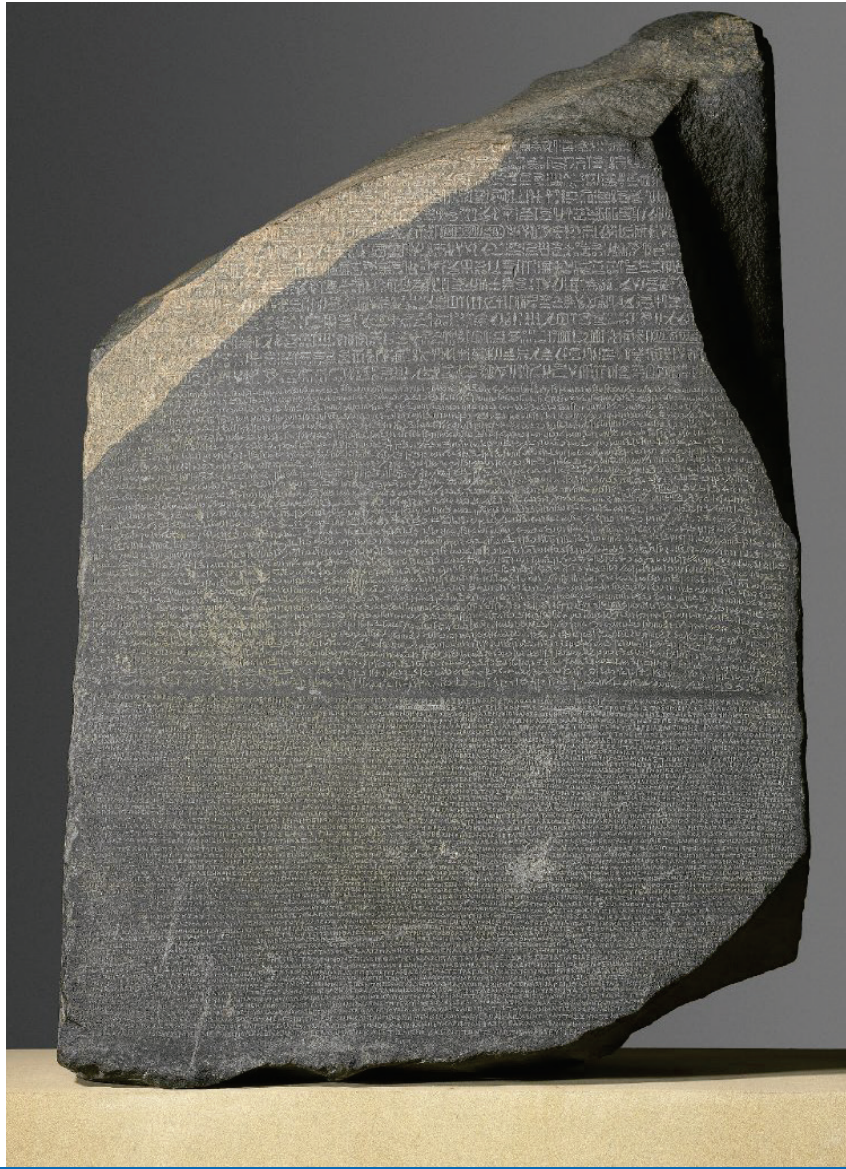
3. Geroglifici Egiziani

3.1 La Stele di Rosetta

La Stele di Rosetta è una stele di granito nero su cui sono incise tre versioni di un decreto emesso a Menfi, in Egitto, nel 196 a.C., per conto di Tolomeo V Epifane. Fu scoperta nel luglio 1799 dall'ufficiale francese Pierre-François Bouchard durante la campagna napoleonica in Egitto, mentre

scavava le fondamenta di una struttura volta a rafforzare una roccaforte nel delta del Nilo, nei pressi di el-Rashid (italianizzato in Rosetta). È stata esposta al British Museum fin dal 1802. Ha una forma irregolare, è lunga 114 cm e larga 72 cm (Figura 3).

Figura 3

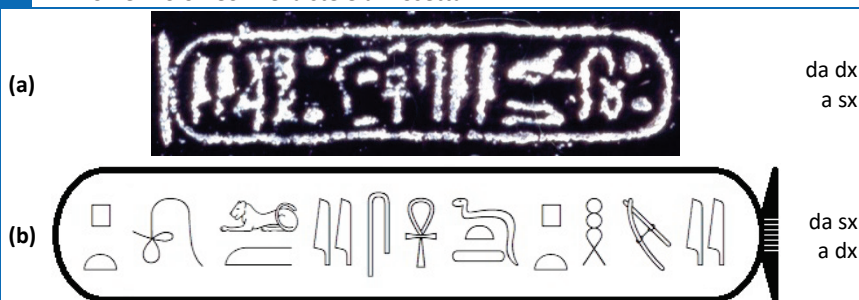
Stele di Rosetta

Fonte: © The Trustees of the British Museum.

La Stele di Rosetta riporta un decreto sacerdotale emanato dal faraone Tolomeo V Epifane. Il testo è ripetuto in tre blocchi: il geroglifico (il “linguaggio degli dei”, 14 righe), il demotico (il “linguaggio degli atti”, 32 righe) e il greco antico (53 righe). La traduzione del greco antico era relativamente facile. La traduzione del demotico (l’antica scrittura egiziana che precede il copto) era più difficile, ma i nomi propri (Tolomeo, Alessandro e Alessandria) furono rapidamente decifrati. La traduzione del geroglifico era ancora più difficile, ma presto si capì che all’interno degli ovali allungati (“cartigli”) erano stati riportati i nomi di re o faraoni.

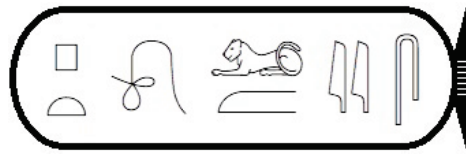
Nella Stele di Rosetta ci sono sei cartigli. In base alla traduzione del testo in greco antico, i sei «cartigli» (*cartouches*) dovevano chiaramente contenere il nome Tolomeo (Ptolemaios, in greco). Tre di essi apparivano come nella Figura 4, dove l’originale (a) va letto da destra a sinistra mentre la trascrizione (b) va letta da sinistra a destra (è a sinistra che guardano il leone e il cobra). Questi tre cartigli presentano il nome Tolomeo insieme a un titolo onorifico egiziano: “Tolomeo, vivente per sempre, amato da Ptah” [Robinson (2018), p. 125].

Figura 4

Il nome “Tolomeo” nella stele di Rosetta

Fonte: © The Trustees of the British Museum (originale).

Gli altri tre cartigli apparivano così:



Geroglifico



Letture di Champollion

P T O L M E S

























I geroglifici dell'antico Egitto rappresentano una delle prime forme di scrittura. Possono essere letti da destra a sinistra o da sinistra a destra. Si può distinguere la direzione in cui va letto il testo perché le figure degli esseri umani o degli animali sono sempre rivolte verso l'inizio del testo. Inoltre, i simboli in alto vengono letti prima di quelli in basso. Non ci sono spazi tra le parole, interruzioni di righe o punteggiature.

3.2 Traduttore di Google

Il 15 luglio 2020 (esattamente 221 anni dopo la scoperta della Stele di Rosetta) Google ha lanciato un traduttore di geroglifici che utilizza l'«apprendimento automatico» (*machine learning*) per decodificare l'antica lingua egiziana: <https://artsexperiments.withgoogle.com/fabricsius/en> (Tavola 2).

Tavola 2

Geroglifici dell'Antico Egitto

| A E O | A | B | CH | D | F PH V |
|---|--|--|---|--|--|
|  avvoltoio 1313F |  avambraccio 1309D |  piede 130C0 |  cappio 1337F |  mano 130A7 |  vipera cornuta 13191 |
| G | H | H | I Y | J (G) | K (C) |
|  vaso 133BC |  corda 1339B |  rifugio 13254 |  foglia di canna 131CB |  cobra 13193 |  cestino 133A1 |
| K (C) | L R | M | N | O U W | P |
|  collina 133D8 |  bocca aperta 1308B |  gufo 13153 |  acqua 13216 |  pulcino di quaglia 13171 |  sgabello 132AA |
| S (C) | SH | T | TH | TH | Z |
|  panno piegato 132F4 |  lago 13219 |  pagnotta 133CF |  (sconosciuto) 1340D |  pancia di mucca 13121 |  chiavistello 13283 |

Nota: 1.079 geroglifici in <https://fonts.google.com/noto/specimen/Noto+Sans+Egyptian+Hieroglyphs>.

4. Il Cifrario di Giulio Cesare

Giulio Cesare (12 luglio 100 a.C. - 15 marzo 44 a.C.) era molto in anticipo sui tempi. In un'epoca in cui l'analfabetismo era la norma, preferiva comunicare con i suoi generali mediante un codice segreto, il cifrario di Cesare. È Svetonio, segretario privato dell'imperatore Adriano, a raccontarci come Giulio Cesare usasse, per la sua corrispondenza confidenziale, un cifrario monoalfabetico in cui ogni lettera era sostituita dalla lettera che la seguiva dopo tre posizioni.

Extant et ad Ciceronem item ad familiares domesticis de rebus in quibus si qua occultius perferenda erant per notas scripsit id est sic structo litterarum ordine ut nullum verbum effici posset: quae si quis investigare et persequi volet quartam elementorum litteram id est D pro A et perinde reliquas commutet.

Ci sono anche sue lettere a Cicerone, nonché ai suoi familiari su questioni private, e in queste ultime, se aveva cose confidenziali da dire, le scriveva in codice, cambiando l'ordine delle lettere dell'alfabeto in modo che nessuna parola potesse essere compresa. Se qualcuno voleva decifrare e coglierne il significato, doveva sostituire la A con la quarta lettera dell'alfabeto, cioè la D, e così via.

Fonte: Suetonius (2008), *De vita Caesarum*, Ch. 1, §56.

Un cifrario simile fu utilizzato dall'imperatore Augusto (23 settembre 63 a.C. - 19 agosto 14 d.C.).

Orthographiam, id est formulam rationemque scribendi a grammaticis institutam, non adeo custodit ac videtur eorum potius sequi opinionem, qui perinde scribendum ac loquamur existiment. Nam quod saepe non litteras modo sed syllabas aut permutat aut praeterit, communis hominum error est.

Nec ego id notarem, nisi mihi mirum videtur tradidisse aliquos, legato eum consulari successorem dedisse ut rudi et indocto, cuius manu "ixi" pro "ipsi" scriptum animadverterit.

Quotiens autem per notas scribit, B pro A, C pro B ac deinceps eadem ratione sequentis litteras ponit; pro X autem duplex A.

Non rispettava rigorosamente l'ortografia, cioè le regole stabilite dai grammatici, e sembra piuttosto che condividesse l'opinione di chi crede che si debba scrivere così come si parla. Spesso invertiva le lettere o le sillabe, o le saltava, ma questi sono errori che commettono un po' tutti.

E non sarei qui a sottolinearli se non fossi rimasto sorpreso nel leggere che egli fece sostituire un governatore consolare, giudicandolo persona incolta e ignorante, perché aveva scritto «ixi» invece di «ipsi».

Ogni volta che scriveva in codice, sostituiva la A con la B, la B con la C e così via per le lettere successive; al posto della X scriveva AA.

Fonte: Suetonius (2008), *De vita Caesarum*, Ch. 2, §88.

Figura 5

Disco cifrante



Fonte: Wikipedia.

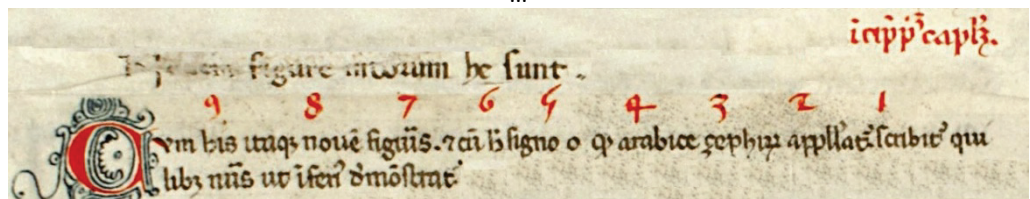
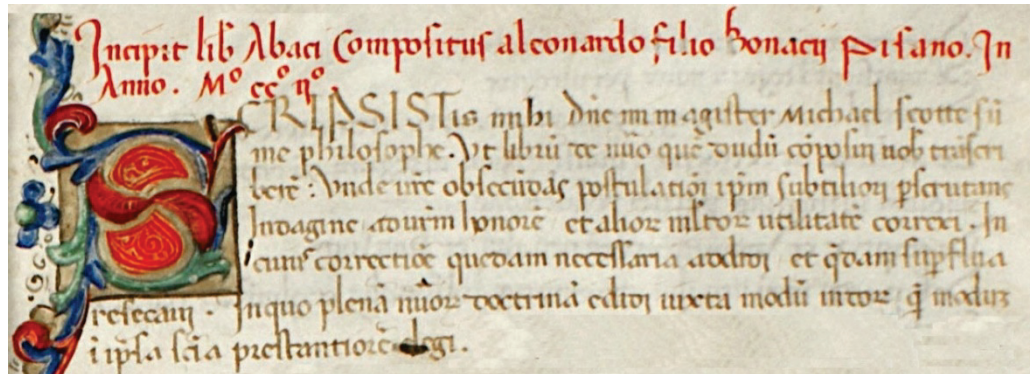
Il primo strumento di cifratura (e decifrazione) polialfabetica è stato descritto da Leon Battista Alberti nel suo *De Compendis Cyfris* intorno al 1467. Il trattato, che venne dato alle stampe solo un secolo dopo la morte dell'autore, è all'origine della moderna crittografia. La Figura 5 mostra un disco cifrante simile a quello progettato da Leon Battista Alberti. Il dispositivo è costituito da due piastre di rame, di diverso diametro, fissate su un perno e montate una sopra l'altra, in modo che possano ruotare in modo indipendente. L'utilizzo richiede che le parti interessate concordino la "chiave cifrante" composta da due lettere, una presente nella piastra esterna e una in quella interna.

5. Il Liber Abaci di Fibonacci

Prima che Gutenberg inventasse la stampa a caratteri mobili, intorno al 1440, i testi venivano scritti a mano. È il caso del Liber Abaci, il libro scritto nel 1202 da Leonardo Pisano, noto oggi ai matematici e agli scienziati di tutto il mondo con il nome di Fibonacci (Tavola 3). Anche se non esiste un codice segreto da decrittare, a volte è difficile “tradurre” i caratteri in un testo che sia leggibile dalle macchine.

Tavola 3

Il Liber Abaci di Fibonacci



Manoscritto | Liber Abaci, Leonardo Pisano, 1202 | Codice Magliabechiano, C. I, 2616, Badia Fiorentina, n. 73.

Incipit liber Abaci Compositus a Leonardo filio Bonacij Pisano In Anno. M^o cc^o ij^o.

SCRIPSI TIS mihi domine mi magister Michael Scotte, summe philosophe, vt librum de numero, quem dudum composui, vobis transcriberem: vnde uestrae obsecundans postulationi, ipsum subtiliori perscrutans Indagine ad uestrum honorem et aliorum multorum utilitatem correxii. In cuius correctione quedam necessaria addidj, et quedam superflua resecaui. In quo plenam numerorum doctrinam edidj, iuxta modum indorum, quem modum in ipsa scientia prestantiorem elegi.

Incipit primum capitulum. Nouem figure indorum he sunt

9 8 7 6 5 4 3 2 1 .

Cvm his itaque nouem figuris, et cum hoc signo 0, quod arabice zephirum appellatur, scribitur quilibet numerus, ut inferius demonstratur.

Latino “decriptato” | Boncompagni, B. (1857), *Il Liber Abaci di Leonardo Pisano*, Roma.

Here begins the Book of Calculation Composed by Leonardo Pisano, Family Bonaci, In the Year 1202.

You, my Master Michael Scott, most great philosopher, wrote to my Lord about the book on numbers which some time ago I composed and transcribed to you; whence complying with your criticism, your more subtle examining circumspection, to the honor of you and many others I with advantage corrected this work. In this rectification I added certain necessities, and I deleted certain superfluities. In it I presented a full instruction on numbers close to the method of the Indians, whose outstanding method I chose for this science.

Here Begins the First Chapter. The nine Indian figures are:

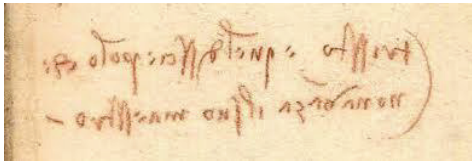
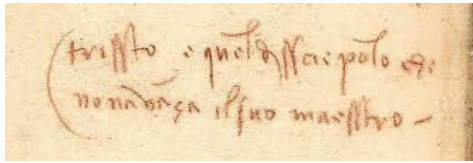
9 8 7 6 5 4 3 2 1.

With these nine figures, and with the sign 0 which the Arabs call zephir any number whatsoever is written, as is demonstrated below.

Traduzione in inglese | Sigler, L. E. (2002), *Fibonacci's Liber Abaci*, pp. 15 e 17.

6. La Scrittura a Specchio di Leonardo da Vinci

Leonardo da Vinci (15 aprile 1452 – 2 maggio 1519) scriveva al contrario, da destra a sinistra, la maggior parte dei suoi appunti personali, tentando così di mantenerli illeggibili da chiunque non fosse lui. Ad esempio, ha scritto:

| | | |
|--|--|--|
| [Testo in Italiano, criptato] ehc olopecsid leuq 'e otsirt) _ ortseam ous li aznava non) | [Testo in Italiano, decriptato] (tristo e' quel discepolo che non avanza il suo maestro _ | |
|  | ruotato in |  |
| [Testo in inglese, criptato] taht lipup eht si roop) _ retsam sih ssaprus ton seod) | [Testo in inglese, decriptato] (poor is the pupil that does not surpass his master _ | |


Fonte: 1493, Forster III, South Kensington Museum III 24b, <http://t.co/ON3xNdrOhb>.

7. Il Codice di Maria Stuarda

L'8 febbraio 1587, Maria Stuarda, Regina di Scozia, fu decapitata per alto tradimento, dopo diciannove anni di prigionia, in Inghilterra. Aveva 44 anni. Dalla sua prigionia, Mary Stuart (= Maria Stuarda) aveva scritto ai suoi sostenitori e alle cancellerie europee centinaia di lettere, di cui 50 ritrovate di recente nella Biblioteca nazionale francese. La corrispondenza, in codice, era stata decriptata da uno dei massimi esperti dell'epoca, Thomas Phelippes [Singh, (1999)]. Mary Stuart fu giudicata colpevole per il suo coinvolgimento nel complotto Babington, una cospirazione organizzata da Anthony Babington per assassinare Elisabetta I. L'alfabeto utilizzato da Mary Stuart è mostrato nella Tavola 4.

| Tavola 4 | | | | | | | | | | |
|----------------------------|-----|----|-----|------|-------|-------|-------|------|---------|------|
| Il codice di Maria Stuarda | | | | | | | | | | |
| A | B | C | D | E | F | G | H | I | J | K |
| | | | | | | | | | | |
| L | M | N | O | P | Q | R | S | T | U | V |
| | | | | | | | | | | |
| W | X | Y | Z | and | for | but | with | that | if | as |
| | | | | | | | | | | |
| of | the | by | so | not | when | from | this | is | in | say |
| | | | | | | | | | | |
| me | my | I | you | what | where | which | there | send | receive | pray |
| | | | | | | | | | | |

Nota: l'alfabeto contiene 5 caratteri "vuoti" da utilizzare per separare le parole (e rendere più complessa l'analisi frequenziale).



Fonte: <https://www.dcode.fr/mary-stuart-code>.

8. Decodificare con il Metodo Monte Carlo

Il primo passo per decifrare un codice è quello di capire qual è la lingua in cui è stato scritto. È poi possibile utilizzare un'analisi frequenziale. L'attuale versione del traduttore di Google, lanciata nell'aprile 2006, può riconoscere 108 lingue, dall'afrikaans allo xhosa, allo yiddish, allo yoruba e allo zulu [Wu, Y., et al. (2016)].

La distribuzione di frequenza delle lettere in lingua inglese, derivate da alcuni libri classici, è riportata nella Tavola 5.

Tavola 5

Lingua inglese: distribuzione di frequenza delle lettere.

| | | | | | | | |
|---|-----------|---|----------|---|----------|-------|------------|
| a | 6,51738% | h | 4,92888% | o | 5,96302% | v | 0,82903% |
| b | 1,24248% | i | 5,58094% | p | 1,37645% | w | 1,71272% |
| c | 2,17339% | j | 0,09033% | q | 0,08606% | x | 0,13692% |
| d | 3,49835% | k | 0,50529% | r | 4,97563% | y | 1,45984% |
| e | 10,41442% | l | 3,31490% | s | 5,15760% | z | 0,07836% |
| f | 1,97881% | m | 2,02124% | t | 7,29357% | Space | 19,18182% |
| g | 1,58610% | n | 5,64513% | u | 2,25134% | | 100,00000% |

Fonte: Art Owen (2008), Stat 362 (Monte Carlo Methods), Stanford University.

Nota: le frequenze riportate sopra si basano su questi classici (disponibili su www.literature.org): Origin of Species (Charles Darwin), The Voyage of the Beagle (Charles Darwin), Jane Eyre (Charlotte Bronte), Wuthering Heights (Emily Bronte), Tarzan of the Apes (Edgar Rice Burroughs), The Return of Tarzan (Edgar Rice Burroughs), Paradise Lost (John Milton). Tutte le lettere maiuscole sono state trasformate in minuscole. Tutti i numeri e tutti i caratteri speciali sono stati rimossi. I "ritorni a capo" sono stati considerati come spazi. Il numero dei caratteri presenti in questi files è pari a 5.086.936.

Il metodo Monte Carlo richiede che siano note anche le probabilità delle lettere adiacenti, cioè le probabilità di transizione del primo ordine:

$$\text{Prob} \equiv \text{Prob}(x_{i+1} = j \mid x_i = i) \text{ per } 1 \leq i, j \leq 27.$$

Un esempio reale di decodifica con il metodo Monte Carlo è stato mostrato da Persi Diaconis [(2009), pp. 1-3]:

Un giorno, uno psicologo del sistema carcerario statale mi si è presentato con una raccolta di messaggi in codice. ... Il problema era quello di decodificare i messaggi. Marc [Coram] ha ipotizzato che il codice fosse un semplice cifrario a sostituzione, in cui ogni simbolo rappresenta una lettera, un numero, un segno di punteggiatura o uno spazio. ... Mi piace quest'esempio perché a) è reale, b) non c'è dubbio che l'algoritmo [il metodo Monte Carlo] abbia trovato la risposta corretta e c) la procedura funziona anche se le ipotesi sottostanti non sono plausibili. In effetti, il messaggio è un mix di inglese, spagnolo e gergo carcerario. La plausibilità si basa solo sulle transizioni del primo ordine. Un tentativo preliminare basato solo sulle frequenze delle singole lettere non è riuscito a decrittare i messaggi.

9. Shazam e Google Lens

La binarizzazione è la tecnica di pre-elaborazione utilizzata per trasformare caratteri, suoni e immagini in cifre binarie (0, 1). Gran parte di tutte le informazioni disponibili è stata binarizzata. Il «riconoscimento ottico dei caratteri» (*optical character recognition* - OCR) ha reso possibile la conversione elettronica di testi «digitati» (*typed*), «scritti a mano» (*handwritten*) o «stampati» (*printed*) in testi codificati automaticamente. Il «riconoscimento ottico della musica» (*optical music recognition* - OMR) ha lo stesso scopo, ma si basa sugli «spartiti» (*printed sheet music*). Shazam, di proprietà di Apple, utilizza il microfono del dispositivo per identificare musica, film, pubblicità e programmi televisivi, mentre Google Lens identifica le immagini in pochi istanti (Tavola 6).

Tavola 6

Riconoscimento di musica e immagini: Shazam e Google Lens.

| | | | | | |
|---|---|---|--|---|---|
|  |  |  |  |  |  |
| Marilyn Monroe Arthur Miller | Romy Schneider Peter O'Toole | Romy Schneider Alain Delon | Michel Piccoli Romy Schneider | Humphrey Bogart Lauren Bacall | Jane Birkin Serge Gainsbourg |
|  |  |  |  |  |  |
| Riconosciuto | Alain Delon Nathalie Delon | Steve McQueen Neile Adams | Frank Sinatra Ava Gardner | Non riconosciuto | Grace Kelly Ray Milland |
|  |  |  |  |  |  |
| Gena Rowlands John Cassavetes | Maggie Gyllenhaal Peter Saarsgard | Ethan Hawke Winona Ryder | Riconosciuto | Non riconosciuto | Non riconosciuto |
|  |  |  |  |  |  |
| George Harrison Pattie Boyd | Jane Asher Paul McCartney | Mel Ferrer Audrey Hepburn | Non riconosciuto | Jacqueline Kennedy John Fitzgerald Kennedy | Jim Morrison Pamela Courson |
|  |  |  |  |  |  |
| Errol Flynn Olivia de Havilland | Isabella Rossellini David Lynch | Jacques Charrier Brigitte Bardot | Riconosciuto | Non riconosciuto | François Truffaut Catherine Deneuve |
|  |  |  |  |  |  |
| Gregory Peck Ingrid Bergman | Marcello Mastroianni Anna Karina | Richard Burton Elizabeth Taylor | Audrey Hepburn Mel Ferrer | Audrey Hepburn Mel Ferrer | Gad Elmaleh Audrey Tautou |
|  |  |  |  |  |  |
| Neile Adams Steve McQueen | Natalie Wood Robert Redford | Jack Nicholson Anjelica Huston | Anna Karina Jean-Luc Godard | Romy Schneider Alain Delon | Monica Vitti Michelangelo Antonioni |
|  |  |  |  |  |  |
| Jean-Paul Belmondo Jean Seberg | James Dean Anna Maria Pierangeli | Non riconosciuto | Non riconosciuto | Non riconosciuto | Riconosciuto |
|  |  |  |  |  |  |
| Marilyn Monroe Eli Wallach | Alain Delon Romy Schneider | Recognized | Alain Delon Nathalie Delon | Claude Mann Jeanne Moreau | Cary Grant Katharine Hepburn |
|  |  |  | | | |
| Monica Bellucci Vincent Cassel | Chiara Mastroianni Benicio del Toro | Märta Torén Dana Andrews | | | |

Fonte: YouTube (<https://youtu.be/ck57LnYScNQ>), Shazam (Burt Bacharach, Live in Sidney, 2008) e Google Lens.

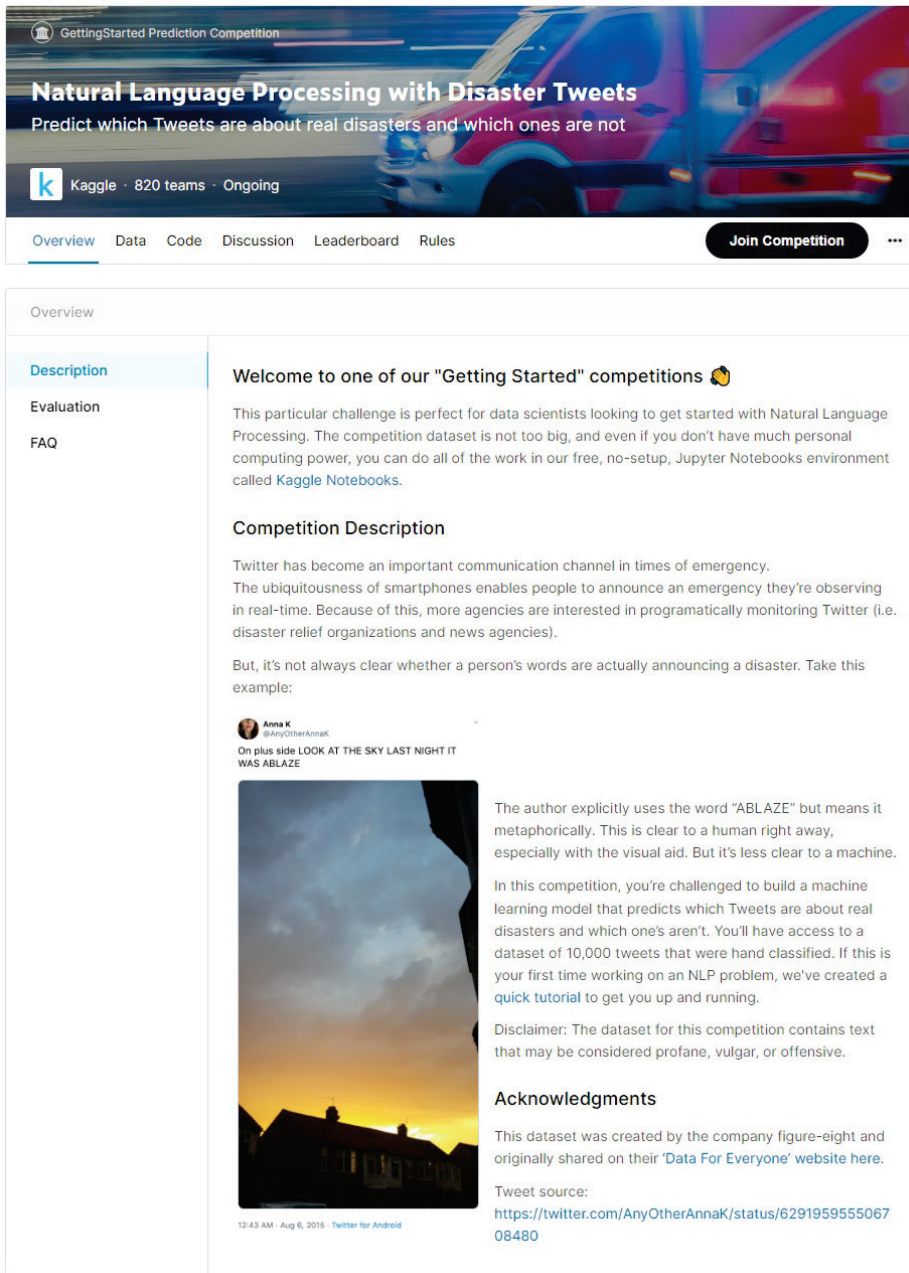
10. Kaggle

Per scoprire se una certa notizia, disponibile sui *social media* o altrove, avrà un impatto positivo, negativo o nullo su una certa azienda è necessario elaborare un'enorme quantità di informazioni. Un buon punto di partenza per imparare il *natural language processing* (NLP) è quello di partecipare a una delle competizioni organizzate da Kaggle, un'azienda di Google. Una di queste si intitola «Elaborazione del linguaggio naturale con *tweet* di disastri» (*Natural Language Processing with Disaster Tweets*). Si tratta di identificare i *tweet* che riguardano disastri "reali", separandoli dagli altri. Tutto il lavoro può essere svolto nell'ambiente Jupyter Notebooks di Kaggle (che è gratuito e non richiede alcuna configurazione), dove è possibile eseguire i propri codici in Python.

La competizione è in corso. La tavola 7 mostra alcuni passi del codice scritto dall'attuale *leader*.

Tavola 7

Kaggle contest: elaborazione del linguaggio naturale con tweet di disastri



The screenshot shows the Kaggle competition page for "Natural Language Processing with Disaster Tweets". The page title is "Natural Language Processing with Disaster Tweets" and the subtitle is "Predict which Tweets are about real disasters and which ones are not". The competition is ongoing, with 820 teams participating. The page includes a navigation menu with "Overview", "Data", "Code", "Discussion", "Leaderboard", and "Rules". A "Join Competition" button is visible. The "Overview" section is expanded, showing a "Description" tab. The description includes a welcome message, a competition description, and a tweet example. The tweet example shows a tweet from Anna K (@AnyOtherAnnaK) with the text "On plus side LOOK AT THE SKY LAST NIGHT IT WAS ABLAZE" and a photo of a sunset over a building. The description explains that the author uses the word "ABLAZE" metaphorically and that the competition is a machine learning challenge to predict real disasters from tweets. It also includes a disclaimer and acknowledgments.

Alcuni passi del codice utilizzato dall'attuale leader (Shahules)

Importazione delle librerie richieste.

Utilizzo del *backend* di TensorFlow. [TensorFlow: una piattaforma di *machine learning* (*open source*) sviluppata da Google]

Caricamento dei dati e acquisizione dell'idea base [Training set: 7613 righe e 5 colonne; Test set: 3263 righe e 4 colonne]

Distribuzione delle classi [Ci sono più *tweets* di classe 0 (nessun disastro) che di classe 1 (disastri reali)]

Analisi esplorativa dei tweets [a livello di carattere, di parola e di frase]

Numero di caratteri nei *tweets*, numero di parole in un *tweet*, lunghezza media delle parole in un *tweet*, *stopwords* comuni nei *tweets* [*stopwords*: parole così comunemente usate da contenere pochissime informazioni utili (*the, in, of, a, to, and, on, for, is, at, ...*)], analisi della punteggiatura [= > - : ? . | + # !] ' ; / = @ ~ * ...], analisi degli *Ngrams* [Ngram: una sequenza di *N* parole]

Pulizia dei dati [Rimozione di URLs, HTML *tags*, emoticons, punteggiatura. Correzione ortografica]

GloVe for Vectorization [GloVe: algoritmo di *machine learning* per ottenere rappresentazioni vettoriali delle parole.]

Numero di parole univoche: 20342

Modello base

11. Considerazioni Conclusive

Il tasso di rendimento medio annuo del Medallion Fund, l'*hedge fund* della Renaissance Technologies di Jim Simons, è stato pari al 66% nei 31 anni dal 1988 al 2018. Questa performance stellare è in forte contrasto con l'«ipotesi dei mercati efficienti» (*efficient market hypothesis* - EMH). Se i mercati fossero efficienti dal punto di vista informativo, nessuno potrebbe realizzare una performance pari a quella del Medallion Fund (MF). Qual è stata la chiave del successo di MF? Non è sufficiente elaborare dati numerici, come prezzi e quantità scambiate. Occorre analizzare una quantità enorme di notizie, disponibili sui *social media* e altrove, e scoprire se avranno un impatto positivo, negativo o nullo sui titoli delle società quotate. La chiave del successo di MF sembra essere l'acquisizione di vantaggi informativi attraverso l'«elaborazione del linguaggio naturale» (*natural language processing* - NPL).

“Ci sono sistematicità nel mercato”, ha detto Simons a un collega, “So che possiamo trovarle”. [Zuckerman (2019), p. 5].

Secondo le stime di Forbes, il patrimonio personale di Jim Simons è ora pari a \$30,7 miliardi (www.forbes.com/profile/jim-simons).

Bibliografia

Boncompagni, B. (1857), *Il Liber Abbaci di Leonardo Pisano*, Roma.

Diaconis, P. (2009), “The Markov Chain Monte Carlo Revolution,” *Bulletin of the American Mathematical Society*, 46(2), 179-205.

Hull, J. C. (2021), *Machine Learning in Business: Un Introduzione alla Scienza dei Dati*, 3^a ed., KDP.

Leonardo da Pisa (1202), *Liber Abaci*. Si veda **Sigler, L. E.**, *Fibonacci's Liber Abaci - A Translation into Modern English of Leonardo Pisano's Book of Calculation*, Springer, 2002.

Owen, A., (2008), “Monte Carlo theory, methods and examples,” <https://artowen.su.domains/mc/>.

Robinson, A. (2018), “Cracking the Egyptian Code: The Revolutionary Life of Jean-François Champollion”, Thames & Hudson.

Rubinstein, M. (1999), “The World According to Mark Rubinstein: Interview”, *Derivatives Strategy Magazine*.

Singh, S. (1999), *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*, Fourth Estate.

Suetonius, G. (2008), *The Twelve Caesars: The Lives of the Roman Emperors*, J. C. Rolfe (Trans.). St. Petersburg, FL: Red and Black Publications.

Wu, Y., et al (2016). “Google’s neural machine translation system: Bridging the gap between human and machine translation,” <https://arxiv.org/pdf/1609.08144.pdf>.

Zuckerman, G. (2019), *The man who solved the market: How Jim Simons launched the quant revolution*, Penguin Random House.