

State of the Art Review for Autonomous Fire Detection and Suppression

João Silva¹, Ricardo Sousa¹, João Santos¹, Nuno Costa¹,
José Araújo¹, Diogo Formosinho¹, and Francisco Cabrita¹

Instituto Superior de Engenharia do Porto
{1150425, 1160900, 1161023, 1171584,
1180943, 1210056, 1210058}@isep.ipp.pt

Abstract. This article presents an overview of the state of the art regarding the application of machine learning powered robotic apparatus and multi agent systems for fire detection and suppression in an industrial context.

Finish abstract

Keywords: Fire Detection · Fire Suppression · Marketing · Sentiment Analysis · Product Development.

1 Introduction

All industrial operations are founded on a production model, whereas each independent operation produces a given product [31]. Specifically, we see this model applied in natural resource extraction, where a given industrial operation extracts a raw material from the environment, creating value from said extraction.

One such example is the lumber industry, which focuses on the process of obtaining wood from trees. This process is usually destructive and requires special attention to sustain a stable raw material output. In order to do this, these operations usually undertake constructive actions to plant trees, in order to sustain a tree lifecycle that guarantees a steady availability of trees suitable for extraction.

These actions however require not only an initial effort but a continuous attention to the state of the environment.

With the advent of the Industry 4.0, this data gathering has increased significantly in both volume and scope, further extending its reach within society and market [12].

Regarding the quality control process, there exist already principles that include the clients for a given product within the quality control processes. Customer feedback is a vital component of quality improvement and assurance, allowing for the market to dictate product issues (and allowing the industrial operation to fix said issues) and guide new product and service development [9]. This customer feedback can be captured in a variety of ways, from direct means (through direct feedback and surveys) and indirect means (error and failure logging for internet connected products, retainment metrics and social media activity regarding a given product).

In general, this data collection usually occurs within social networks [2], and as such new methodologies and heuristics are required in order properly capture and analyse the available data. This article provides an overview of the state of the art for the areas of social network data mining and language processing, both essential within the problem space mentioned previously, taking into account the expect problems (and possible solutions) the stated problem may encounter.

1.1 Social Network Data Mining

Social media is undoubtedly where billions of people connect from all corners of the world, ages and nationalities to share opinions, experiences, feelings, pictures, videos and more. This opened the door for public and private sectors to promote, benefit, learn, analyse, and improve organizations based on data provided by social media [26].

The evolution of Information Technology has generated large amounts of databases and immense data in various areas. The research in databases has given rise to an approach to store and manipulate invaluable data for further decision making. Data mining is the process of extraction of information and discovery of patterns from massive amounts of data; this process is also referenced as knowledge discovery process; knowledge mining of data; knowledge extraction or data/pattern analysis. The goal of this technique is to discover patterns that were previously unrecognized. Once these uncharted patterns are found they can be utilized to make business decisions.

Social Media and big data combined forces and created a genuine field called social media mining, like data mining but circumscribed to the virtual social world of Twitter, Facebook, Instagram, and others like [26]. This is the process of representing, analysing, and extracting actionable patterns from social media data feeds. In other words, this mining occurs when a company collects data about users and analyses it with the objective to form conclusions about these users. Often this information is used to build marketing campaigns to specific market segments [26].

1.2 Semantic Data Processing

After obtaining the available data we must then process it. This processing is usually done using semantic techniques, including but not limited to sentence segmentation and tokenization. Additionally, we also researched into feature extraction from a given sentence, in order to obtain information regarding specific attributes mentioned.

Sentence segmentation is the process of determining the longer processing units consisting of one or more words. This task involves identifying sentence boundaries between words in different sentences. Since most written languages have punctuation marks which occur at sentence boundaries, sentence segmentation is frequently referred to as sentence boundary detection, sentence boundary disambiguation, or sentence boundary recognition. All these terms refer to the

same task: determining how a text should be divided into sentences for further processing [6].

Tokenisation is the process of breaking up the sequence of characters in a text by locating the word boundaries, the points where one word ends and another begins. For computational linguistics purposes, the words thus identified are frequently referred to as tokens. In written languages where no word boundaries are explicitly marked in the writing system, tokenisation is also known as word segmentation, and this term is frequently used synonymously with tokenisation [6].

In practice, sentence and word segmentation cannot be performed successfully independently from one another. For example, an essential subtask in both word and sentence segmentation for English is identifying abbreviations, because a period can be used in English to mark an abbreviation as well as to mark the end of a sentence. In the case of a period marking an abbreviation, the period is usually considered a part of the abbreviation token, whereas a period at the end of a sentence is usually considered a token in and of itself [6].

Finally, it is also important for a sentiment analysis system to be able to obtain the important features in a given sentence. With this, researchers can more easily understand what is mentioned alongside the keywords previously selected, providing additional insight into the opinion of the studied population on the selected topic.

1.3 Sentiment Analysis

Sentiment analysis is a field in Natural Language Processing, whose main objective is to identify the opinion of authors regarding specific entities [18]. With the growth of social networks, this field started to have an important role considering it allows the monitoring of opinions and sentiments regarding different topics for a given population at large.

The applications/libraries that implement sentiment analysis usually parse the sentences written by the authors into numerical scores (sentiment), and the lower that score is, the worst is the sentiment of that sentence. When we apply this analysis to a large number of people, we will get a good overview of the general opinion regarding the analysed topic. Commonly, the opinions fall into a category of “positive”, “neutral”, or “negative”, depending on the score of the sentences.

Additionally, efforts have been made in the sentiment analysis space to have an accurate analysis from emoji related sentiment. The main obstacle to this detection is related to the way emojis can be used. In some scenarios, emojis are used with a sarcastic or ironic sense in sentences, which can be a problem in an automatic detection scenario.

In 2015, Novak et al [18], introduced the first emoticon lexicon, which provides a sentiment table to each entry with a vector of values. This table introduces a positive, neutral, and negative value from 0 to 1 and an average value from -1 to 1.

1.4 Sentiment Analysis in Marketing and Quality Control

With Facebook daily users close to the three billion mark it is evident that almost everyone is on social media [7]. This leads to unmeasurable amounts of free information that can prove invaluable to the Marketing and Quality Control departments. In the following sections, we will explore some potential applications of Sentiment Analysis in both Marketing and Quality Control.

2 Areas of Interest

All of the aforementioned topics have areas of interest in terms of application techniques, field presence and research direction. We present the most common fields where these topics are applied and the research options for the topic, taking into account the most recent developments in usage and research.

2.1 Social Network Data Mining

Data mining has importance regarding the process of finding patterns and forecasting trends, discovering knowledge in many areas not to say all areas that in some way produce data related to users. Data that is very important and valuable to public or private organizations. There are many areas like surveillance, medicine, entertainment, where data mining techniques are explored. Audio, video, and image mining established its own separate research field [4].

Sentiment Analysis and Opinion Mining are two interesting subjects of study, social media, economics, financial, and insurance, topics might have distinct perspectives when looking to an opinion from the author and reader standing point. “The housing price has gone down, which is bad for the economy”; the authors point out the negative impact of the price drop to the economy. However, this sentence can be perceived in both ways by readers. Negative for sellers but not for buyers [23]. Researchers either ignore the issue or assume a standing point in their analysis – opinion holders assume to be the consumers of public unless stated [23]. As we see above this complex, subjective and advanced topics are important to study in data mining since a wrong analysis can build wrong opinion summarization [23].

FinTech is a hot topic these days and Data Mining is a core to financial stock market forecast using Data Mining techniques. To determine for example if it is possible to predict if the closing price of a given stock will increase or decrease on the following day. The combination of multiple Data Mining techniques allows to automatically generate predictions and test them. The results would show the significance level of the closing price and the portfolio being managed. Over the last three decades the amount of data has increased and its volume will keep growing considerably in the future [15]. This is another area of interest from which organization take great value – even if it is used to understand investors risk exposure and behaviours or minimize investment losses [15].

2.2 Semantic Data Processing

Until recently, the problem of robustness was rarely addressed by Nature Language Processing (NLP) systems, which normally could process only well-formed input conforming to their hand-built grammars. The increasing availability of large corpora in multiple languages that encompass a wide range of data types (e.g. newswire texts, email messages, closed captioning data, optical character recognition (OCR) data, multimedia documents) has required the development of robust NLP approaches, as these corpora frequently contain misspellings, erratic punctuation and spacing, and other irregular features. It has become increasingly clear that algorithms which rely on input texts to be well-formed are much less successful on these different types of texts.

Most existing segmentation algorithms for natural languages are both language-specific and corpus-dependent, developed to handle the predictable ambiguities in a well-formed text. Depending on the origin and purpose of a text, capitalization and punctuation rules may be followed very closely (as in most works of literature), erratically (as in various newspaper texts), or not at all (as in email messages or closed captioning).

For a language like English, where a great deal of the initial tokenisation can be done by relying on white space, tokenisation is considered to be majorly solved and is not rated as a sentence segmentation technique. However, for unsegmented languages, evaluation of word segmentation is critical for improving all aspects of NLP systems. A sentence segmentation algorithm's performance is usually reported in terms of a single score equal to the number of punctuation marks correctly classified divided by the total number of punctuation marks. It is also possible to evaluate sentence segmentation algorithms using recall and precision, given the numbers of false positives, punctuation marks erroneously labelled as a sentence boundary, and false negatives, actual sentence boundaries not labelled as such.

In cases such as written Thai where punctuation gives no reliable information about sentence boundaries, locating sentence boundaries is best treated as a special class of locating word boundaries. Even languages with relatively rich punctuation systems like English present surprising problems. Recognizing boundaries in such a written language involves determining the roles of all punctuation marks which can denote sentence boundaries: periods, question marks, exclamation points, and sometimes semicolons, colons, dashes, and commas. In large document collections, each of these punctuation marks can serve several different purposes in addition to marking sentence boundaries. A period, for example, can denote a decimal point or a thousands marker, an abbreviation, the end of a sentence, or even an abbreviation at the end of a sentence. Ellipsis (a series of periods (...)), can occur both within sentences and at sentence boundaries. Exclamation points and question marks can occur at the end of a sentence, but also within quotation marks or parentheses (really!) or even (albeit infrequently) within a word, such as in the band name Therapy?. However, conventions for the use of these two punctuation marks also vary by language; in Spanish, both

can be unambiguously recognized as sentence delimiters by the presence of ‘i’ or ‘j’ at the start of the sentence.

Regarding feature selection, we find a wide research field, where solutions based on both semantic and machine learning techniques are available and researched upon.

Usually feature selection algorithms take into account the frequency of a given word in the evaluated sentence or corpus, such is the case for bag of words (BoW) [27]. It may also take into account the statistical data from the data set, such as CHIR [20].

2.3 Sentiment Analysis

Sentiment Analysis can be applied in a large variety of contexts, such as:

- Social media monitoring
- Customer support ticket analysis
- Brand monitoring and reputation management
- Listen to voice of the customer (VoC)
- Listen to voice of the employee
- Product analysis
- Market research and competitive research

The usage of sentiment analysis in these fields allow companies to have a better automated feed on how customers feel about certain products and at the same time automatize the customer support services with a correct output to the doubt and comments given by the users.

3 Academic Production

Most of the topics mentioned in this topic have seen steady or increased interest, as measured by the number of research papers published. This allows us to envision a future with more and better options to solve problems that integrate these topics as available solution paths.

3.1 Social Network Data Mining

The analysis was not very conclusive – by looking into figure 1, it seems to us that Data Mining as a field of study gave in recent years place to a broader area called data science; although there are still many articles being produced recently but not as many as with the term “Data Science” as a technique in which the extraction and transformation of data is also present at its core.

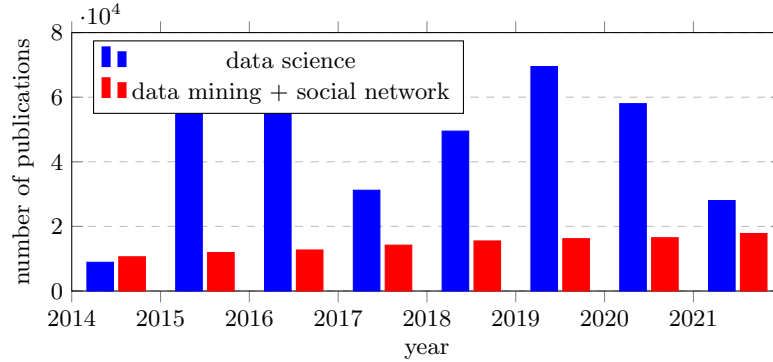


Fig. 1: Number of released papers with specific keywords per year

3.2 Semantic Data Processing

The topic of sentence segmentation and tokenization has been subject of increased attention for the past years. From the analysis presented in figure 2 we can steadily growing academic production, albeit with an overall small production volume.

We ascertain that this is due to the afterthought that sentence segmentation can be in more sophisticated machine learning system, which may reduce it's overall interest when compared with other language processing topics.

Regarding feature selection, we include an initial term in order to properly filter out ambiguity from other fields with similar terminology. With this, we also find that feature selection still plays an important and growing field in natural language processing, at a greater scale than sentence segmentation.

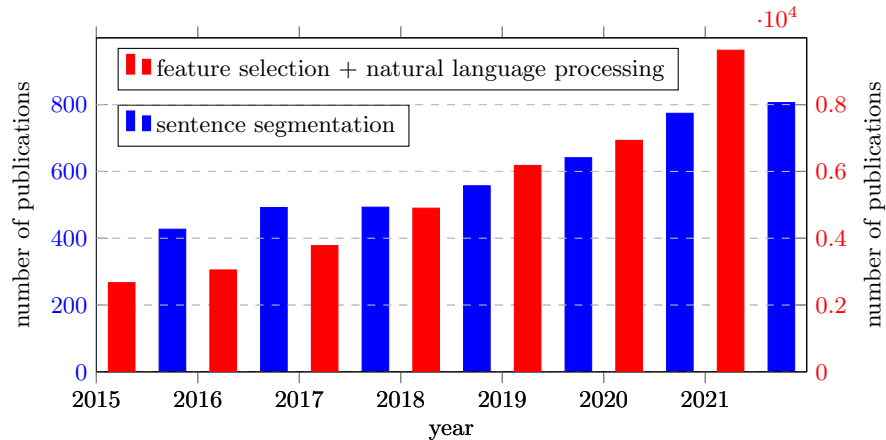


Fig. 2: Number of released papers with specific keywords per year

3.3 Sentiment Analysis

Sentiment analysis is a popular topic for research throughout the years, as we can see in figure 3. This graph show us that interest and research into this topic has been steady for the last decade, mostly due to the maturity of this field.

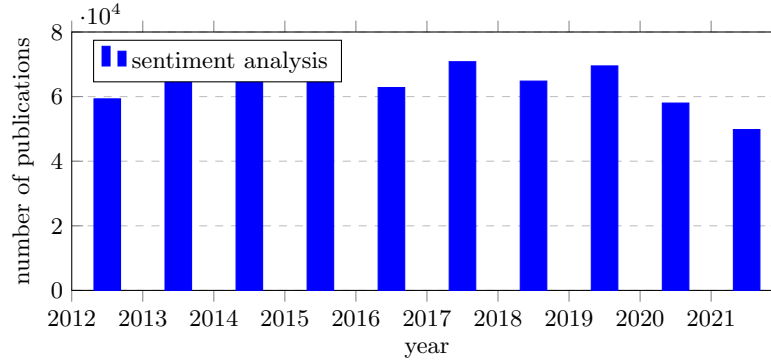


Fig. 3: Number of released papers with specific keywords per year

4 State of the Art

4.1 Social Network Data Mining

The size of the data at our disposal and its generation speed is too big to fit in structures of generic existing database architectures in such a way that must be different ways to oversee it [30].

Predictive tasks of data mining include Classification, Regression and Deviation Detection. Descriptive tasks include Clustering, Summarization, Connection Learning and Sequential Patterns. Clustering or unconfirmed learning is an important but difficult problem; the impartial of clustering is to partition a set of unlabelled objects into standardized groups or clusters.

Clustering is a process to automatically divide volumes of data into groups that are related to each other on a basis of common themes or properties – to find out what a collection contains. It is applied to identify properties of a set of data as date, cost, etc., and divide them into clusters. Clusters or subsets can be used to find hidden similarities [17].

Many clustering techniques are used by applications to establish or realizing structure in data, such as data mining information retrieval image subdivision and machine learning; as a result, clusters of data can be appeared with dissimilar figures, magnitudes, data scarceness and grades of split-up [30]. All these complexities introduce noise in the data and can mask the true nature of the structure of data.

Data mining and Natural Language Processing are interlinked in a way that one extends the analysis of data the extraordinary field of comprehension of data. Romero and Ventura established this relation that emerging text mining technology attempt to extract meaningful information from unstructured textual data. Text mining can be used to identify efficiently and systematically, extract, manage, integrate, and explore knowledge from texts [17] [28]. As an example, the most important difference between naïve Bayes and logistic regression is that logistic regression is a discriminative classifier while naïve Bayes is a generative classifier, meaning a generative model would have the goal of understanding what dogs look like and what cats look like – by contrast, a discriminative model is only trying to learn to distinguish the classes and perhaps without learning much about them. All the dogs in the training set wear collars and cats not [14].

During a detailed survey about data mining techniques used in social media, investigators of Western Ontario and Sharjah Universities performed a deep exploration, including the percentages of data mining types applied with artificial intelligence algorithms by researchers in social media. Technique list below.

- AdaBoost
- Artificial Neural Networks (ANN)
- Apriori
- Bayesian Networks (BN)
- Decision Trees (DT)
- Density Based Algorithm (DBA)
- Fuzzy
- Generic Algorithm (GA)
- Hierarchical Clustering (HC)
- K-Means
- K-nearest Neighbors (k-NN)
- Linear Discriminant Analysis (LDA)
- Linear-Regression (Lin-R)
- Logistic Regression (LR)
- Markov
- Maximum Entropy (ME)
- Novel
- Support Vector Machines (SVM)
- Wrapper

4.2 Semantic Data Processing

An extensive word list combined with an informed segmentation algorithm can help to achieve a certain degree of accuracy in word segmentation, but the greatest barrier to accurate word segmentation is in recognizing unknown words, words not in the lexicon of the segmenter. This problem is dependent both on the source of the lexicon as well as the correspondence (in vocabulary) between the text in question and the lexicon. Another obstacle to high-accuracy word

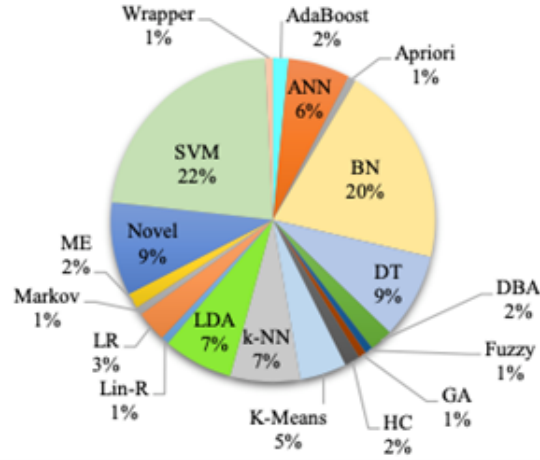


Fig. 4: Data Mining Techniques among selected papers

segmentation is the fact that there are no widely-accepted guidelines as to what constitutes a word, and there is therefore no agreement on how to “correctly” segment a text in an unsegmented language. Native speakers of a language do not always agree about the “correct” segmentation, and the same text could be segmented into several very different (and equally correct) sets of words by different native speakers. A simple word segmentation algorithm is to consider each character a distinct word. This is a practical for Chinese because the average word length is very short (usually between one and two characters, depending on the corpus) and actual words can be recognized with this algorithm. Although it does not assist in tasks such as parsing, part-of-speech tagging, or text-to-speech systems, the character-as-word segmentation algorithm has been used to obtain good performance in Chinese information retrieval, a task in which the words in a text, play a major role in indexing.

A very common approach to word segmentation is to use a variation of the maximum matching algorithm, frequently referred to as the greedy algorithm. The greedy algorithm starts at the first character in a text and, using a word list for the language being segmented, attempts to find the longest word in the list starting with that character. If a word is found, the maximum-matching algorithm marks a boundary at the end of the longest word, then begins the same longest match search starting at the character following the match. If no match is found in the word list, the greedy algorithm simply segments that character as a word (as in the character-as-word algorithm above) and begins the search starting at the next character.

Successful high-accuracy segmentation requires a thorough knowledge of the lexical and morphological features of the language. Languages like Japanese and Korean have writing systems that incorporate alphabetic, syllabic and logographic symbols. Modern Japanese texts, for example, frequently consist of many

different writing systems: Kanji (Chinese symbols), hiragana (a syllabary for grammatical markers and for words of Japanese origin), katakana (a syllabary for words of foreign origin), romanji (words written in the Roman alphabet), Arabic numerals, and various punctuation symbols. In some ways, the multiple character sets make tokenisation easier, as transitions between character sets give valuable information about word boundaries.

4.3 Sentiment Analysis

Valence Aware Dictionary for sEntiment Reasoning (VADER) An example of a state of art model that had impact in sentiment analysis field is the Valence Aware Dictionary for sEntiment Reasoning (VADER), which is a rule-based model for general sentiment analysis. VADER outperformed previous Sentiment Analysis models such as LIWC, ANEW, the General Inquirer and SentiWordNet [13], especially in text extracted from Social Networks. In contrast to the previous sentiment analysis models, VADER can interpret accurately the sentiment of some common slang and emojis, which increases the performance of sentiment analysis, especially in informal environments such as social networks.

VADER uses five different heuristics to measure the sentiment of sentences:

- Punctuation: the type and amount of punctuation has a direct impact in the sentiment of the sentence. For instance, “4.0 Industry is cool!!!” has a more positive sentiment than “4.0 Industry is cool”
- Capitalization: if a word is in capital letters, the sentiment is higher, therefore it should have a higher impact in the overall sentiment of the phrase. Example: “The traffic today was HORRIBLE” was a worst sentiment than having the word “horrible” in lower case
- Degree Modifiers (intensifiers or booster words): take in consideration the words before a word that has positive or negative impact in the sentence. For instance, if someone describes something as “incredibly good”, the sentiment is way more positive than having “kind of good”
- Contrastive conjunction “but”: usually when a sentence is divided by a “but” word, the dominant sentiment is in the second part of the sentence. VADER reduces up to 50% the value of sentiment before the “but”, while those after increase to 150% the value of the sentiment [5]. For instance: “I love the new iPhone but the battery is weak”
- Examining the tri-gram preceding a sentiment-laden lexical feature: heuristic used to find negated sentences that flip the polarity of the sentence. If a negation is captured, VADER multiplies the sentiment of the sentence by -0.74. Example of negated sentence that flips the polarity: “The food here isn’t really all that great” [13]

Using these heuristics, VADER was able to similar results than individual humans rating the sentiment of text tweets. In those tests, r (correlation to ground truth, mean of 20 human raters) for VADER was $= 0.881$ while the r for individual humans was 0.888 [13].

TextBlob is a textual processing Python library capable of doing different tasks such as:

- Noun phrase extraction
- Part-of-speech tagging
- Sentiment analysis
- Classification (Naive Bayes, Decision Tree)
- Tokenization (splitting text into words and sentences)
- Word and phrase frequencies
- Parsing
- n-grams
- Word inflection (pluralization and singularization) and lemmatization
- Spelling correction
- Add new models or languages through extensions
- WordNet integration

The sentiment property output given by TextBlob returns a tuple that contains both the polarity, which is the sentiment of the sentence, which ranges from -1 to 1 and subjectivity, which is the personal feelings, which range from 0 to 1, where 0 is very objective and 1 very subjective.

TextBlob was built based on NLTK and Pattern, also providing many of the native features [25].

5 Applications

5.1 Social Network Data Mining in Marketing and Quality Control

Data mining is a mature technology and currently widespread and used across many distinct industries in a regular basis. Retail stores, hospital, banks, and insurance companies are among the many using it to feed the mechanics of their decision processes. “Data mining has wide application domain almost in every industry where the data is generated that’s why data mining is considered one of the most important frontiers in database and information systems” [4], Organizations combine data mining with statistics, pattern recognition, and other important tools to find patterns and connections that would otherwise be difficult to find. Data mining allows them to learn more about their customers and users as well as their competitors in a way they can make smart marketing decisions.

Artificial intelligence applied to Data Mining introduced techniques that use information and knowledge intelligently; research categories like Neural networks; Algorithm architecture; Dynamic prediction based approaches; Analysis of systems architecture; Intelligence agent systems; Modelling and its applications; Knowledge-based systems; Systems optimizations; Information systems [21].

Practical examples of data mining applied to marketing but not exclusively:

- FBTO Dutch Insurance Company, challenged to reduce email costs; increase efficiency of marketing campaigns and increase cross-selling to existing customers – resulted in marketing team with new capabilities to predict the effectiveness of its campaigns; increased efficiency of marketing campaign creation, optimization, and execution. Decreased mailing costs by thirty five percent and increased conversion rates by forty percent.
- In 2017 the Journal of Advertising published a study about the utilization of social media mining techniques to gauge users’ perception of a variety of brands. Twitter feed API was used over a six-month period to examine tweets about different stores, telecoms, consumer electronics and footwear companies. The data was collected, compiled, and boil down to a general topic and sentiment with very specific results. Circa fifteen percent of tweets about fast-food restaurants were about promotions and sixty six percent of tweets expressing negative sentiment about Comcast company [26].
- The ability to continually change and provide new understanding is the principal advantage of Data Mining and will be the core of Data Mining applications in future [21]. This means that whenever possible, Data Mining will be applied to help decision makers take the knowledge one step further by learning with new data structures and patterns.

5.2 Sentiment Analysis in Marketing and Quality Control

It is important to consider consumers’ opinions on any product, however, surveys can lead to abstract ratings. A better way to obtain a consumer’s opinion is through sentiment analysis on social networks. There are various applications for Sentiment Analysis:

- Hotel reviews [16]
- Restaurant reviews [10] [19]
- Stock price prediction [24]

Sentiment analysis has been used to automatically classify consumer comments concerning marketing 4Cs aspects (customer, cost, convenience, communication) [22].

During the COVID-19 pandemic, some papers also suggested the use of Sentiment Analysis in order to better understand the pandemic effects on subjects such as vaccine rejection and false information spread. Some papers suggest a deeper study of the impact of the pandemic through Sentiment Analysis in order to be better prepared for the next one [1].

References

1. Alamoodi, A., Zaidan, B., Zaidan, A., Albahri, O., Mohammed, K., Malik, R., Almahdi, E., Chyad, M., Tareq, Z., Albahri, A., Hameed, H., Alaa, M.: Sentiment analysis and its applications in fighting covid-19 and infectious diseases: A systematic review. *Expert Systems with Applications* **167**, 114155 (2021). <https://doi.org/https://doi.org/10.1016/j.eswa.2020.114155>, <https://www.sciencedirect.com/science/article/pii/S0957417420308988>

2. Anderson, J., Narus, J.: Selectively pursuing more of your customer's business. *Sloan Management Review* **44**, 42 (03 2003)
3. Belch, G.E., Belch, M.A.: Advertising and promotion. Irwin/McGraw-Hill (2001)
4. Bharati, M., Ramageri, B.: Data mining techniques and applications. *Indian Journal of Computer Science and Engineering* **1** (12 2010)
5. Calderon, P.: Vader sentiment analysis explained - pio calderon. Medium (Mar 2018), <https://medium.com/@piocalderon/vader-sentiment-analysis-explained-flc4f9101cd9>
6. Dale, R., Moisl, H., Somers, H.: Handbook of Natural Language Processing. Taylor & Francis (2000), <https://books.google.pt/books?id=VoOLvxyX0BUC>
7. Department, S.R.: Most used social media 2021 (Mar 2022), <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
8. Evans, J.R., Evans, J., Lindsay, W.: The Management and Control of Quality. Principles of Management Series, South-Western College Pub. (1999), <https://books.google.pt/books?id=8ORCAQAIAAJ>
9. Fundin, A.P., Bergman, B.L.: Exploring the customer feedback process. *Measuring Business Excellence* **7**(2), 55–65 (Jan 2003). <https://doi.org/10.1108/13683040310477995>, <https://doi.org/10.1108/13683040310477995>
10. Gan, Q., Ferns, B.H., Yu, Y., Jin, L.: A text mining and multidimensional sentiment analysis of online restaurant reviews. *Journal of Quality Assurance in Hospitality & Tourism* **18**(4), 465–492 (2017). <https://doi.org/10.1080/1528008X.2016.1250243>, <https://doi.org/10.1080/1528008X.2016.1250243>
11. Hague, P.N., Hague, P., Hague, N., Morgan, C.A.: Market Research in Practice: How to Get Greater Insight from Your Market. Kogan Page Limited (2013)
12. Hood, J., Brady, A.: Industry 4.0 engages customers (Dec 2016), <https://www2.deloitte.com/us/en/insights/focus/industry-4-0/customer-engagement-strategies-digital-manufacturing-enterprise.html>
13. Hutto, C., Gilbert, E.: VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media* **8**(1), 216–225 (May 2014), <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
14. Injadat, M., Salo, F., Nassif, A.B.: Data mining techniques in social media: A survey. *Neurocomputing* **214**, 654–670 (2016). <https://doi.org/https://doi.org/10.1016/j.neucom.2016.06.045>, <https://www.sciencedirect.com/science/article/pii/S092523121630683X>
15. Kaliyaperumal, S., Sathik, M., Arumugam, P.: Financial stock market forecast using data mining techniques. *Lecture Notes in Engineering and Computer Science* **2180** (03 2010)
16. Kasper, W., Vela, M.: Sentiment analysis for hotel reviews (01 2011)
17. Kendi, C.: Social media mining methods used to improve business intelligence at safaricom public limited company (2018)
18. Kralj Novak, P., Smailović, J., Sluban, B., Mozetič, I.: Sentiment of emojis. *PLOS ONE* **10**(12), 1–22 (12 2015). <https://doi.org/10.1371/journal.pone.0144296>, <https://doi.org/10.1371/journal.pone.0144296>
19. Krishna, A., Akhilesh, V., Aich, A., Hegde, C.: Sentiment analysis of restaurant reviews using machine learning techniques. In: Sridhar, V., Padma, M., Rao, K.R. (eds.) *Emerging Research in Electronics, Computer Science and Technology*. pp. 687–696. Springer Singapore, Singapore (2019)

20. Li, Y., Luo, C., Chung, S.M.: Text clustering with feature selection by using statistical data. *IEEE Transactions on Knowledge and Data Engineering* **20**, 641–652 (2008)
21. Liao, S.H., Chu, P.H., Hsiao, p.y.: Review: Data mining techniques and applications - a decade review from 2000 to 2011. *Expert Systems with Applications: An International Journal* **39**, 11303–11311 (09 2012). <https://doi.org/10.1016/j.eswa.2012.02.063>
22. Lin, H.C.K., Wang, T.H., Lin, G.C., Cheng, S.C., Chen, H.R., Huang, Y.M.: Applying sentiment analysis to automatically classify consumer comments concerning marketing 4cs aspects. *Applied Soft Computing* **97**, 106755 (2020). <https://doi.org/https://doi.org/10.1016/j.asoc.2020.106755>, <https://www.sciencedirect.com/science/article/pii/S1568494620306931>
23. Liu, B.: Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* **5**(1), 1–167 (2012). <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>, <https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
24. Liu, J., Lu, Z., du, W.: Combining enterprise knowledge graph and news sentiment analysis for stock price prediction. In: *Proceedings of the Annual Hawaii International Conference on System Sciences*. Hawaii International Conference on System Sciences (2019), <http://dx.doi.org/10.24251/hicss.2019.153>
25. Loria, S.: textblob documentation. Release 0.15 **2** (2018)
26. McCourt, A.: Social media mining: The effects of big data in the age of social media (Mar 2018), <https://law.yale.edu/mfia/case-disclosed/social-media-mining-effects-big-data-age-social-media>
27. McTear, M., Callejas, Z., Griol, D.: *The Conversational Interface: Talking to Smart Devices*. Springer (May 2016)
28. Romero, C., Ventura, S.: Educational data mining: A review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *IEEE Transactions on* **40**, 601 – 618 (12 2010). <https://doi.org/10.1109/TSMCC.2010.2053532>
29. for Standardization, I.O.: ISO9001 - Quality Management. Standard, Geneva, CH (Sep 2015)
30. Tappe, C.G., Chouragade, P.M.: *Analysis and data mining techniques in social media* (2016)
31. Webster, M.: Definition of industry (Mar 2022), <https://www.merriam-webster.com/dictionary/industry>