UNIVERSITY OF PIRAEUS – DEPARTMENT OF INFORMATICS

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

MSc «Advanced Informatics and Computing Systems – Software Development and Artificial Intelligence

«Προηγμένα Συστήματα Πληροφορικής – Ανάπτυξη Λογισμικού και Τεχνητής Νοημοσύνης»

MSc Thesis

Μεταπτυχιακή Διατριβή

Thesis Title: Τίτλος Διατριβής:	**Sentiment Analysis on Twitter** Ανάλυση Συναισθήματος στο Twitter
Student's name-surname: Ονοματεπώνυμο Φοιτητή:	**Evangelos Gkouskos** Ευάγγελος Γκούσκος
Father's name: Πατρώνυμο:	**Dimitrios** Δημήτριος
Student's ID No: Αριθμός Μητρώου	**ΜΠΣΠ18008**
Supervisor Επιβλέπων	**Dionisios Sotiropoulos, Assistant Professor** Διονύσιος Σωτηρόπουλος, Επίκουρος Καθηγητής

Ημερομηνία Παράδοσης **December 2022 / Δεκέμβριος 2022**

3-Member Examination Committee

Τριμελής Εξεταστική Επιτροπή

Dionisios Sotiropoulos **Assistant Professor**	**Evangelos Sakkopoulos** **Associate Professor**	**George Tsihrintzis** **Professor**
Διονύσιος Σωτηρόπουλος Επίκουρος Καθηγητής	Ευάγγελος Σακκόπουλος Αναπληρωτής Καθηγητής	Γεώργιος Τσιχριντζής Καθηγητής

Acknowledgments

I want to thank first of all my parents for their encouragement all these years, despite any obstacles on the road. Special thanks to my advisor, who gave me this opportunity and for introducing me to this extraordinary field. Lastly, I am particularly grateful to my wife for all her supportiveness, without it, I would not be the same.

Περίληψη

Η άνθηση του Web 2.0 και η ευκολία που παρέχει σε συνδυασμό με τον Covid-19, έδωσε στους ανθρώπους τη δυνατότητα για αλληλεπίδραση διαδικτυακά μεταξύ τους περισσότερο από ποτέ. Η αλληλεπίδραση αυτή ποικίλει, από επικοινωνία μέσω Μέσων Κοινωνικής Δικτύωσης έως αγοραπωλησίες προϊόντων στο εξωτερικό. Η ικανότητα κάποιου ανθρώπου να κάνει κριτικές ή σχόλια για προϊόντα, αγαθά ή ακόμα και πολιτικά πρόσωπα έχει αλλάξει τη συμπεριφορά και τις ανάγκες των ανθρώπων ως προς τη κατανάλωση και το εμπόριο γενικότερα και έθεσε την επικοινωνία σε μια νέα εποχή.

Πλέον, όλο και περισσότεροι άνθρωποι κάνουν κριτικές και σχόλια στο διαδίκτυο. Αναμφισβήτητα, αυτές οι κριτικές τράβηξαν το ενδιαφέρον από διάφορους επιχειρηματικούς κλάδους και από την ακαδημαϊκή κοινότητα.

Η ανάλυση συναισθήματος, γνωστή και ως εξόρυξη γνώμης είναι το επιστημονικό πεδίο το οποίο αναλύει κριτικές, γνώμες ακόμα και αισθήματα και εξάγει το συναίσθημα που απορρέει.

Ο σκοπός της διπλωματικής αυτής εργασίας είναι η ανάλυση δεδομένων από το Κοινωνικό Δίκτυο Twitter, να ελέγξει το συναίσθημά τους χρησιμοποιώντας διαφορετικές μεθόδους και τελικώς να συγκρίνει τα αποτελέσματα. Οι μέθοδοι που θα χρησιμοποιηθούν αποτελούνται από αλγόριθμους μηχανικής και βαθιάς μάθησης. Ξεκινάμε με μια ιστορική αναδρομή της ανάλυσης συναισθήματος και έως και τη μορφή που βρίσκεται σήμερα. Στη συνέχεια, παρουσιάζουμε τους αλγορίθμους που χρησιμοποιήθηκαν για την συγγραφή της εργασίας αυτής, καθώς και της διαδικασίας της προετοιμασίας και επεξεργασίας των δεδομένων ώστε να διαμορφωθούν κατάλληλα για τη μελέτη τους. Τέλος, ολοκληρώνουμε την εργασία εξετάζοντας την ακρίβεια του εκάστοτε αλγορίθμου και παρουσιάζουμε τις νέες μεθόδους και το μέλλον που διαφαίνεται στο πεδίο της ανάλυσης συναισθήματος.

Abstract

The rise of Web 2.0 and the convenience it brought in conjunction with Covid-19, enabled people to collaborate with each other online more than ever. This collaboration varies, from communicating through Social Media to cross-boarding selling, or buying. The capability of someone to review or to make comments about products, goods, or even political characters changed people's behavior and needs regarding consumption and trade more generally and put communication into a new era.

Nowadays, more and more people make critics or comments on the internet. Undeniably, those reviews attracted attention from various industries in business and the academic community.

Sentiment analysis, also known as opinion mining is the scientific field that studies and analyzes reviews, critics, opinions, or even emotions and extracts the sentiment that derives (Liu, 2015).

The goal of this thesis is to analyze data from Twitter, determine their sentiment using different mechanisms and finally compare the results. Those mechanisms include machine learning algorithms and deep learning approaches. We begin with the evolution of sentiment analysis through time and in which condition we find it today. Later, we will present the algorithms that we used for this study, in addition to the preparation and preprocessing steps we did to form the data accordingly, for our task. Lastly, we conclude this study by investigating the accuracy of its algorithm and introduce new avenues and future work in the field of sentiment analysis.

Contents

List of Figures

List of Tables

Chapter 1

1.1 Introduction

Even though sentiment analysis (SA) is a scientific subfield of Linguistics and Natural Language Processing (NLP) that has a long history, it is in the eye of research for no longer than 12 years. Its flourish coincided with the rapid growth of Web 2.0, which allowed consumers to barrage social media and websites with opinions, reviews, discussions, and sentiments. This opinionated data turned out to be very valuable for the business industry and society generally.

More and more people nowadays leave their reviews and opinions online. This fact, could not be ignored by corporations worldwide. According to Forbes[1], the necessity for the inclusion of online review research must be highlighted in new business strategies. It is no coincidence, that 92.4% of consumers are influenced by online reviews for their purchasing decisions. Furthermore, the impact of the pandemic Covid-19 in people's everyday life and the limitations it brought, encouraged online trade and raised review interaction by 50% concerning the pre-pandemic levels[2]. Recent research states that irrespective of the average ratings, businesses with more reviews online increase their revenue by 54%[3].

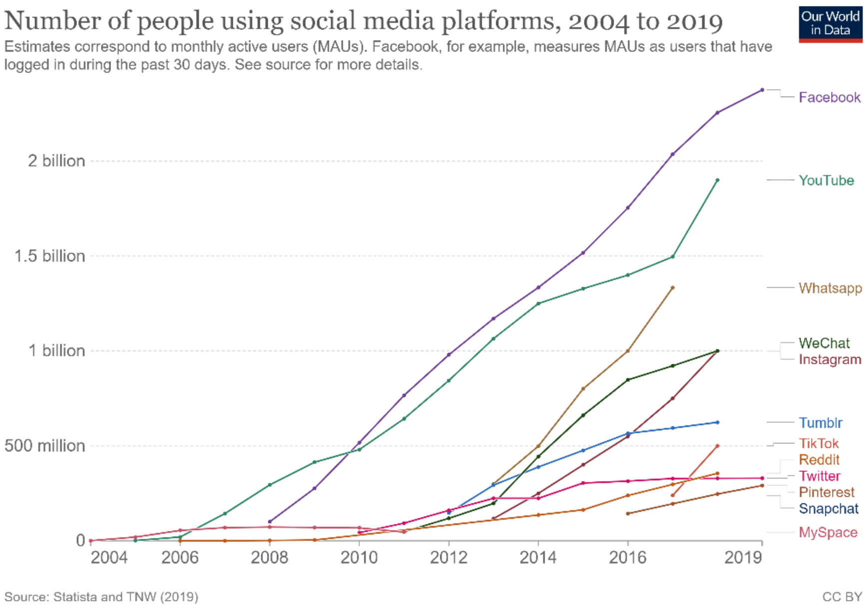

Figure 1. The number of people using social media platforms

Apart from the overwhelming number of reviews online over the last years for consumer goods, we observed the same increment in online comments, opinions, and experiences that people share mostly through Social Media (SM). As we can see in figure 1 from Our World in Data[4], 4.70

[1] Forbes, Article: Online Reviews: The Customer Reference That Never Sleeps, https://www.forbes.com/sites/forbestechcouncil/2022/08/16/online-reviews-the-customer-reference-that-never-sleeps/, last visit 15/11/2022
[2] Reviewtrackers, Article: Online Reviews Statistics and Trends: A 2022 Report by ReviewTrackers, https://www.reviewtrackers.com/reports/online-reviews-survey/, last visit 15/11/2022
[3] Searchengineland, Article: Review counts matter more to local business revenue than star ratings, according to study, https://searchengineland.com/review-counts-matter-more-to-local-business-revenue-than-star-ratings-according-to-study-320271 last visit 16/11/2022
[4] Our World in Data, Article: The rise of social media, https://ourworldindata.org/rise-of-social-media, last visit 17/11/2022

billion people around the world use social media, and 5.1% of those are brand new users within the last year.

Nowadays, people are accustomed to posting not only reviews and comments but also deeper thoughts and experiences. (Burbach, et al., 2020) elaborate on the power that SM gained in formatting opinions and how they influence viewers' personalities. One of the fundamental uses of SM today is to spread information and opinion formation. For example, there are not just a few occasions until today, where social networks were used as a tool for online radicalization[5] [6] and spreading political or even terrorist propaganda[7] [8]. (Kusen, et al., 2019) present a systematic study about the influence of emotional valence shifts on the messaging behavior of social network users on three large-scale platforms, Twitter, Facebook, and YouTube.

People no longer stay inactive when it comes to reading opinions about a subject of their interest, but they take part in the debate and participate vividly stating their thoughts. This new nature of Social Media platforms, stimulated organizations and the academic community to work further on the sentiment analysis problem.

The demand for automated systems that extract sentiments from texts and classify them is more than ever. The lack of time is usually an excuse for a company that does not respond to online reviews even though according to Trust Pulse[9] businesses that respond to reviews are 1.7 more plausible to be described as trustworthy than those which do not. Therefore, following this trend, we observed a significant increase (99%) in the proportion of published papers focusing on sentiment analysis during recent years concerning those which were issued 20 years ago (Mäntylä, Mika V.; Graziotin, Daniel; Kuutila, Miikka, 2018).

1.2 Evolution of Sentiment Analysis

As mentioned above, there was very little research on sentiment analysis before the 2000s. (Liu, 2020) pinpoints this fact, due to the lack of opinionated text in digital forms. Indeed, early in the 21st century, we have the first papers published, the most noted of which were by (Pang, et al., 2002) which presented the differentiation between typical machine learning algorithms in online movie reviews for sentiment analysis classification and (Turney, 2002) who elaborated on the same problem. The next year, it was the paper by (Nasukawa & Yi, 2003), in which perhaps the term sentiment analysis was first introduced in the world.

During the following years, the proliferation of online platforms where people could share thoughts and sentiments such as social media networks (e.g. Facebook[10]), blogs or microblogs (e.g. Twitter[11]), forums (e.g. Reddit[12]), or even review sites (e.g. Trip Advisor[13]) pushed up diverse data. Hence, we reach the point where (Pang & Lee, 2008), published probably one of the most influential papers, which was an excellent aggregation of the sentiment analysis problem until then.

[5] Vice, Article: An expert explains how social media can lead to the "Self-Radicalization" of terrorists, Article: https://www.vice.com/en/article/qbxnz5/we-asked-an-expert-how-social-media-can-help-radicalize-terrorists, last visit 27/11/2022

[6] FBI, Article: ISIL Online: Countering Terrorist Radicalization and Recruitment on the Internet and Social Media, https://www.fbi.gov/news/testimony/isil-online-countering-terrorist-radicalization-and-recruitment-on-the-internet-and-social-media-, last visit 27/11/2022

[7] New York Times, Article: A Good-Will Government Was Possible in Israel, https://www.nytimes.com/2022/11/27/opinion/a-good-will-government-was-possible-in-israel.html/, last visit: 27/11/2022

[8] The McGill International Review, Article: Platforms of Propaganda: How Social Media Sites Facilitate and Spread Disinformation, https://www.mironline.ca/platforms-of-propaganda-how-social-media-sites-facilitate-and-spread-disinformation/, last visit 27/11/2022

[9] Trust Pulse, Article: Online Review Statistics You Need to Know, https://trustpulse.com/online-review-statistics/, last visit 20/11/2022

[10] Facebook, site: https://www.facebook.com/

[11] Twitter, site: https://www.twitter.com/

[12] Reddit, site: https://www.reddit.com/

[13] Tripadvisor, site: https://www.tripadvisor.com/

After the 2010s, the increase of deep learning methodologies provided better sentiment accuracies than before. Undoubtedly, deep networks seem to outperform traditional techniques and provide state-of-the-art results (Goodfellow, et al., 2016), in various disciplines ranging from visual recognition to self-driving cars. So, there was no chance for deep learning to stay out of natural language processing area. As a result of this growth, sentiment analysis developed to be one of the most active research areas today (Zhang, et al., 2018).

1.3 Sentiment Analysis Applications

In recent years, an enormous wave of applications dominated the internet. In a similar trend, we have witnessed the growth of opinionated texts, comments, and thoughts shared on those applications. Organizations and companies are far more concerned with leveraging those opinions, for either building up their customer base or finding out opinions about their services and products. Likewise, thoughts and concerns could be useful also for political parties, and so forth. Thus, we have now reached the point at which sentiment analysis applications have spread to nearly every possible area from product reviews, health care, and stock market prediction to political strategies and elections (Liu, 2015).

1.3.1 Applications in Business

Customer satisfaction is a necessary condition that contributes to the viability of every company. Seeking ways to discover firstly, if their products have a positive or negative impact on the marketplace and secondly, ways to enhance products' reputation. Hence, organizations comprehend the fact that using a program that automatically aggregates people's comments about a product and extracts the overall sentiments of those comments not only, ensures time for the company but also relieves employees of the task of identifying sites that contain useful information and summarizing the opinions in them, which on most of the occasions are tricky to decipher (Zhang, et al., 2018).

While in the past, the traditional ways for a company to understand customers' opinions towards its brand, it should conduct surveys or opinion polls. Henceforth, it is no longer indispensable, since this information is spread publicly on the web, also known as external data, and also in the form of customer feedback which every company keeps as internal data (Liu, 2015). So, it is now very easy for a company to collect and automatically classify emotions from reviews, and social media comments. This helps in decision-making and customer approach strategies.

According to (Wankhade, et al., 2022), companies that use sentiment analysis outperform in three major tasks:

- Increment of customer happiness, through effective dealing with real-time problem tracking, improved products, and market uniqueness.
- Customer satisfaction analysis, by the abundance of opinions online about each product, leading to product optimization, to reverse negative attitudes, if any.
- Track problems in real-time and tackle them directly, since customers can disperse their experience on social media, from their first contact with the product.

Besides customer satisfaction detection, (Wankhade, et al., 2022) highlight another popular application, which is market research. In this task which belongs to business intelligence, sentiment analysis aims to give a comprehensive picture of consumer values, which contains for example where a company stands regarding the competition. This helps in developing marketing decisions and alternative strategies for their product campaigns. Usually, data for analysis is collected from online platforms such as social media, blogs and micro-blogs, online shopping and e-commerce platforms, and so forth.

1.3.2 Politics

(Norris, 2000) distinguishes three time periods in the context of political strategies for communication and campaigns:

- Pre-modern campaign period, characterized by strong bonds between political parties and their voters.
- Direct campaign period, which contains meetings and physical contacts.
- Modern campaign period, where face-to-face contacts are dwindling and ties are loosened. On the contrary, advertising campaigns are increased via the internet, television, and advertising.

With the advent of Social Network Sites, political parties seized upon their potential to fill the gap between politicians and voters, that the modern period caused. Furthermore, according to (Lilleker, et al., 2010) study, extremists and fringe or new parties, who were lacking public and media attention fell on fertile ground to diffuse their ideas. In addition to that, SNSs and generally Web 2.0 applications provided instant interaction and combined multiple features in such a way as to incite citizens' activities and enhance political participation[14] and mobilization (Vergeer, et al., 2013). Hence, we reach a point where the political arena differs from the past, in terms of increased campaign periods (Brown, 2002; Blumenthal, 1982) and social polarization.

With the passage of time, and especially during the latest years, it is observed that political parties intensify their presence on social network sites, mostly via Twitter, which is concerned to be the main tool for disseminating information and positions. Given the growing use of social media, for even more online representation (Woolley, 2020), it is encouraged the use of social bots in online political debates, and their main tasks include: first, influence for malicious intentions, second further polarizing political conversations[15], and third diffuse fake news and misinformation[16] (Guess & Lyons, 2020).

Alongside with abovementioned strategies, there are many more, however, analysis of these matters would be beyond the scope of this thesis, whose purposes are among others, opinion manipulation[17], stimulating social movements, and many more. So, the advance of sentiment analysis methods over political agendas was very crucial and being used for many years. Insights extracted from opinions and sentiments are valuable for every government. Such insights could help in economic decision-making, international relations, and even diplomacy (Liu, 2015).

Albeit the fact that there are many applications in politics, the most ubiquitous are those relating to social media monitoring (Aisopos, et al., 2011) or identifying the political sentiment of users (Khatua, et al., 2020). Likewise, very popular globally, are the applications that are used for election predictions (Xia, et al., 2021; Yaqub, et al., 2017). In other related works, we can find applications that quantify public sentiment and the impact that formed towards Brexit (Ilyas, et al., 2020) or how economical changes impact the citizens in South Korea (Aich, et al., 2017). Generally, the political area is an area where sentiment analysis has a wide range of implementations with a continuing interest that seems to evolve even more in the future.

1.3.3 Finance and Stock Market

The need to develop models for efficient financial forecasts is an important area for research for a very long time now. In the beginning, the lack of enough data was making research more

[14] FirstMonday, Article: Key differences between Web 1.0 and Web 2.0, https://firstmonday.org/article/view/2125/1972, last visit 28/11/2022
[15] IEEE Spectrum, Article: How political campaigns weaponize social media bots, https://spectrum.ieee.org/how-political-campaigns-weaponize-social-media-bots, last visit 28/11/2022
[16] FirstMonday, Article: Social bots distort the 2016 U.S. Presidential election online discussion, https://firstmonday.org/article/view/7090/5653, last visit 28/11/2022
[17] Medium, Article: The Rise of the Weaponized AI Propaganda Machine, https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b, last visit 28/11/2022

difficult than in later years. With the advent of Web 2.0, the abundance of information scattered throughout the internet facilitated the procedures and now made sentiment analysis applications in finance particularly effective. It is no coincidence, that during the later years, we have witnessed an exponential increase in research in natural language-based financial forecasting (Xing, et al., 2018).

In the early stages, the data sources were mostly opinions and comments from blogs and microblogs such as Twitter. The main goal was to predict stock market price movements identified by the classification of those opinions. Such work was done by (Das & Chen, 2007), in which they presented a methodology for sentiment classification from messages and opinions extracted from stock message boards using a "crawler" program. They conclude by analyzing the correlation between stock prices and sentiment. Another study presented by (Zhang, et al., 2011) shows how emotionally charged tweets become a predictor for some of the most popular stock markets such as Dow Jones or NASDAQ. In this respect, (Bollen, et al., 2011) elaborated on the public mood and the way it affects closing prices and correlates to the value of the Dow Jones Industrial Average periodically.

During the later years, even if the main resource for data mining remains social media, various papers exploit also financial news articles (Kalyani, et al., 2016). Along a similar line but utilizing an alternative data source, (Koukaras, et al., 2022) used two sentiment analysis tools, TextBlob and VADER for stock market prediction and seven different machine learning algorithms. Alongside Twitter, the data used for this study were also from StockTwits[18] (a social media platform especially for investing and trading), and financial data from Finance Yahoo[19].

1.4 Twitter

Twitter is one of the most popular social networking services, a micro-blogging platform that was introduced in 2006, where people can communicate through quick and short messages, called tweets. Apart from that, people can also post photos, videos, links, etc. Anyone can follow other users, groups, or even companies, which means that they can see their messages uploaded to their Twitter Home timeline. In addition, Twitter allows replies and responses to enable discussions and debates. It also provides an action called "retweet" where a user can forward tweets of other users to his followers. During the first years, Twitter allowed users to write messages up to 140 characters, but now this limit is extended to 280. Generally, a tweet can be seen by the followers of the uploader, but if the tweet is public it is visible to anyone, who searches Twitter with a keyword, also known as "hashtag", contained in the tweet. A property that those hashtags have (keywords with the sign "#" in the beginning) is that they create topics, that people all over the world can contribute to and discuss. Furthermore, like most other social media sites, if someone finds a useful or interesting tweet he can click the like button, represented by a small heart. Another interesting feature of Twitter is that it utilizes those hashtags, by gathering the tweets containing them and counting them in order to generate a list with the most popular tweets for that specific time, called "trends". Users can explore this list, either by the most popular trends near their place or by their interests.

From its beginning, Twitter gained popularity for various reasons, the first of which is because it is designed to be user-friendly. Twitter is extremely straightforward and thus manages to create users from all ages, cultures, and social classes. Its promptness in message uploading and its two-sided communication (Vergeer, et al., 2013), earned the attention of entrepreneurs, companies, politicians, and even celebrities, although it is not as popular as Facebook or Instagram. Knowing that such users, many of whom are an inspiration for others, are active on this platform, made Twitter a suitable tool for content sharing. Many businesses now, use Twitter

[18] https://stocktwits.com/
[19] https://finance.yahoo.com/

as a marketing tool to promote and advertise their products and services[20]. They also target groups of people and customize their marketing strategies to create community interest in their brand[21]. Apart from the aforementioned, Twitter is a substantial tool for media discourse and news reporting (Ahmad, 2010). More and more journalists use Twitter to share their thoughts and opinions about breaking news[22]. It is no coincidence that in a survey conducted by Muck Rack, with a participation of more than 2500 journalists, regarding the question of which is the most valuable social media platform, Twitter ranked first with 77% of the journalists preferring Twitter[23]. When it comes to Twitter users, it turns out that 48% of those get their news from Twitter[24].

In a nutshell, Twitter provides people with the convenience of a one-stop platform for multiple uses, for example from being informed about the news near them to searching for things of their interest. Twitter is among the 10 most favorite social media sites[25] with a penetration rate reaching 45%[26] among users in the UK. The total number of users globally stands at around 396.2 million, with at least 500 million tweets sent daily[27]. For those reasons among others, organizations and academic community place Twitter as their main data resource for sentiment analysis.

[20] Lifewire, Article: What is Twitter and How Does it Work? https://www.lifewire.com/what-exactly-is-twitter-2483331, last visit 30/11/2022
[21] Upwork, Article: An Introduction to Twitter Marketing Strategy: Basics and Examples for 2022, https://www.upwork.com/resources/twitter-marketing, last visit 30/11/2022
[22] Journalism, Article: Which Twitter features are most useful for Journalists?, https://www.journalism.co.uk/news/twitter-tries-to-reel-in-journalists-with-a-raft-of-new-features/s2/a962175/, last visit 30/11/2022
[23] Muck Rack, Article: New Muck Rack survey: The state of journalism 2022, https://muckrack.com/blog/2022/03/15/state-of-journalism-2022, last visit 30/11/2022
[24] Blog Twitter, Article: How many people come to Twitter for news? As it turns out, a LOT, https://blog.twitter.com/en_us/topics/insights/2022/how-many-people-come-twitter-for-news, last visit 30/11/2022
[25] SEJ, Article: The Top 10 Social Media Sites and Platforms 2022, https://www.searchenginejournal.com/social-media/biggest-social-media-sites/#close, last visit 30/11/2022
[26] CyberCrew, Article: Top 25 Surprising Twitter Statistics UK Edition [2022], https://cybercrew.uk/blog/twitter-statistics-uk/, last visit 30/11/2022
[27] Thesocialshepherd, Article: 22 Essential Twitter Statistics You Need to Know in 2022, https://thesocialshepherd.com/blog/twitter-statistics#twitter-has-3965-million-users-globally, last visit 30/11/2022

Chapter 2

2.1 The Sentiment Analysis Problem

Sentiment analysis is an active area of research, especially during the later years. Its main goal is to develop efficient automatic tools and computational methods that extract subjective information such as sentiment or opinion from textual data. Sentiment analysis is also referred to as opinion mining (Shaalan, et al., 2017), even though researchers regard that there is a subtle difference in the definition of sentiment and opinion. These terms are used interchangeably among researchers, whilst in business, the term sentiment analysis is used almost exclusively (Liu, 2015).

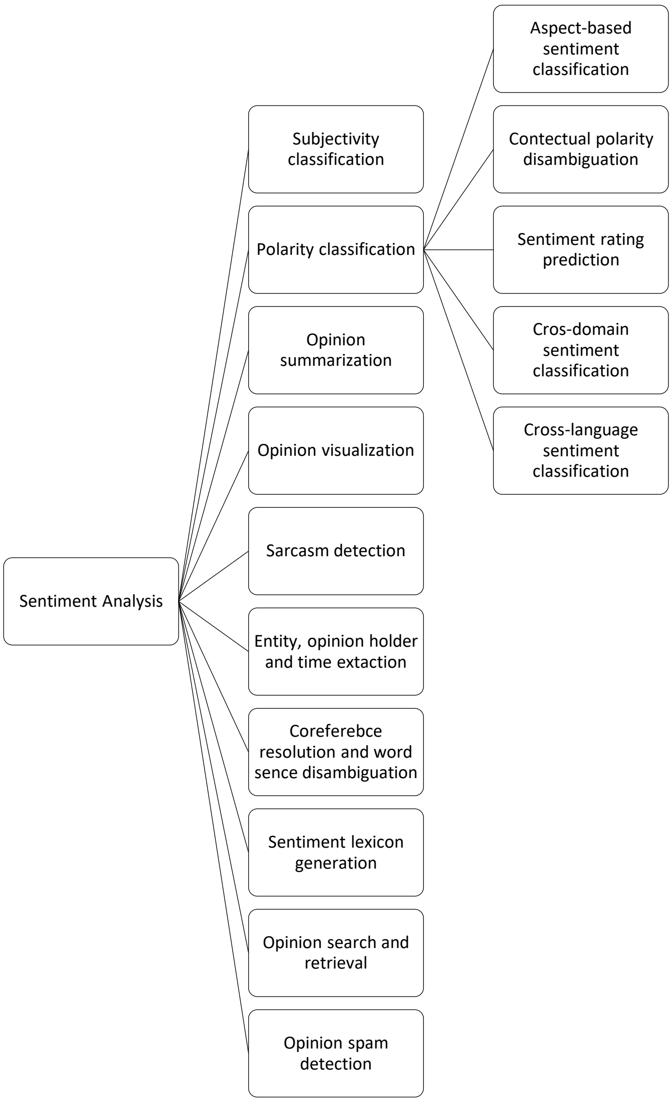

Figure 2. Sentiment Analysis Tasks (Pozzi, et al., 2017)

Semantically, to be precise, according to the Cambridge Dictionary[28], opinion is defined as the thought, belief, or judgment of someone about a particular matter, whereas sentiment albeit that it also means the above, expresses the feelings derive from the matter. Undoubtedly, each contains elements of the other. Depending on the field of application, sentiment analysis can be

[28] https://dictionary.cambridge.org/

also found as opinion mining, opinion extraction, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, and review mining (Pozzi, et al., 2017).

2.1.2 Sentiment Analysis Tasks

As mentioned above, sentiment analysis aims to identify the whole attitude or polarity of a person in relation to a product, brand, or service, extracted by his review or text posted (Tedmori & Awajan, 2019), and the main focus is on opinionated texts. Thus, researchers have implemented several methods, subtasks, and techniques to facilitate the overall procedures. (Pozzi, et al., 2017) as we can see in figure 2 provided a detailed taxonomy of the most frequent subtasks, which will be discussed in this chapter.

2.1.2.1 Subjectivity classification

Subjectivity classification is referred to as the capability to differentiate subjective inputs - in our case sentences - from objective (Cambria, et al., 2017). It is usually assumed to be the first step in sentiment analysis, and considered highly beneficial. Its goal lies to exclude objective (factual) data from subsequent processing (Kamal, 2013). It allows a detailed view of various opinions in data by locating subjective clues, hints, ideas, and sentiment words that carry emotions or notions like “amazing” or “tedious”, “cheap” or “expensive” and “hard” or “easy” (Ligthart, et al., 2021; Chenlo & Losada, 2014).

2.1.2.2 Polarity classification

The polarity classification is the subtask of sentiment analysis which identifies the polarity of a given text. These polarities could be classified as positive, negative, or neutral. Depending on whether is important, categories could also be expanded, according to the polarity score of the text, and include, very positive, positive, neutral, negative, and lastly, very negative. Occasionally, polarity classification is represented as a 5-star rating. A detailed taxonomy is presented by (Yadollahi, et al., 2017), which discriminates the task in opinion polarity classification and emotion polarity classification, in case there is some.

2.1.2.3 Opinion summarization

As reviews and online text data are constantly increasing nowadays, an automatic text summarization tool is needed (Pankaj, et al., 2016). In practice, absorbing this excessive amount of text data and summarizing them manually is a quite tricky task for a human being (Ambedkar, et al., 2019; Lerman, et al., 2009). Visual representations such as charts or histograms could be proved to be helpful, however, such representations occasionally lack disambiguation of information. On the other hand, the text retains complex information that cannot be easily interpreted from charts (Nishikawa, et al., 2010). There are two main approaches to sentiment summarization, as presented by (Premakumara, et al., 2019):

- Extractive summarization, which extracts the summary from the entire document.
- Abstractive summarization, which firstly understands the whole meaning of the input and then extracts the summary.

2.1.2.4 Opinion visualization

A lot of appealing patterns and trends could be identified from visualization. In real-world problems, the enormous amount of data complicates procedures and makes data difficult to digest. Hence, visualization methods are a substantial subtask in sentiment analysis, in which researchers can gain insights from visualizing eye-catching charts (Xu, et al., 2022). There is a plethora of charts to fill the arsenal of a researcher, such as word clouds which show the most frequent words in a dataset, scatter plots which can present for example how similar or not some words are and confusion matrices which present the overall performance of classifiers.

2.1.2.5 Sarcasm detection

According to (Whalen, et al., 2013; Whalen, et al., 2011) sarcasm is considered a sub-term of irony and both terms of figurative language (Skalicky & Crossley, 2018). Sometimes, irony is rather to have a positive meaning than sarcasm which is inclined to be more aggressive and offensive (Farias & Rosso, 2017). In most occasions, sarcasm is difficult to deal with. Sarcastic sentences' meaning is the opposite of the written text. Namely, when someone posts a positive review about a product, it is actually negative, and vice versa (Liu, 2015). Another factor that makes sarcasm detection rather difficult is the many factors that affect it, which could be tone, situation, or background. (Majumder, et al., 2019) study demonstrates how important is sarcasm detection and classification in sentiment analysis and how an efficient sarcasm detector can help in the whole performance.

2.1.2.6 Entity, opinion holder, and time extraction

As an entity, we define objects in a sentence that are discrete and recognizable (Phan, et al., 2023). So, when referring to an entity, we mean names of products, services, persons, events that may occur, and organizations (Liu, 2010). Opinion holders are any natural or legal persons that express a specific opinion in a sentence or text. Mostly, in online reviews or blogs, opinion holders are the authors of the posts, whereas in news articles opinion holders usually state an opinion of another person or organization (Liu, 2015). This is why opinion holders in news articles are regarded to be more important than the prior ones (Bethard, et al., 2004; Choi, et al., 2005; Kim & Hovy, 2004). Lastly, time extraction is also a substantial hint, especially when we elaborate on web context.

The above three aspects contribute to the opinion definition by (Liu, 2015) which will be presented in a later chapter. In addition, their extraction tasks are studied extensively during the past years, (Sarawagi, 2008) constituting a clear and comprehensive survey.

2.1.2.7 Word sense disambiguation

Word sense disambiguation is the task that differentiates the words in a document used with subjective sense from those used with objective sense. It is considered to be one of the difficult tasks in sentiment analysis, and occasionally it is omitted and other ways to express sentiment are preferred. For example, according to (Wilks & Stevenson, 1997), part-of-speech tagging could be a rough form of semantic disambiguation.

2.1.2.8 Sentiment lexicon generation

Sentiment lexicons are a key concept in sentiment analysis. Words that convey polarities either positive or negative are called sentiment words. (Qiu, et al., 2009). For example, words like pretty, kind, or good are positive sentiment words, and ugly, discourteous, and bad are negative sentiment words. Respectively to words, some phrases and idioms convey polarity and usually are treated the same as sentiment words. (Liu, 2015) distinguishes sentiment words into two categories, the base type, and the comparative type. The prior, corresponds to the above example words, while the latter is used to express opinions in comparative and superlative forms. Namely, words like better and best are the comparative types of the base adjective good.

2.1.2.9 Opinion search and retrieval

Opinion search combines two disciples, information retrieval and sentiment analysis. According to a specific query, (Zhang, et al., 2007), relevant documents must be retrieved that match the criteria, which in this case are negative or positive opinions. (Liu, 2012) indicates the two kinds of opinion searches. Queries about a specific entity or its aspect and queries about legal or natural persons (opinion holders) about an entity or its aspect.

2.1.2.10 Spam detection

Nowadays, e-commerce, social media, and generally review websites are a part of consumers' everyday life. Hence, reviews about a product or service are a primary factor that clearly affects the decision about purchasing a product or not (Crawford, et al., 2015). Those reviews opinions and reviews thus are considered to be highly valuable for a company or manufacturer. Since anyone has the freedom to post his reviews and opinions anonymously, it is not infrequent malicious users to create fake reviews (Birjali, et al., 2021) which can discredit the popularity of a product and the opposite. Users with that activity are called opinion spammers, and their actions, opinion spamming (Jindal & Liu, 2007). This phenomenon is regarded to be a major issue in social media. Apart from independent users who post fake reviews for their own reasons, there are also commercial companies whose part of their profession is to promote a brand of their client, even if it does not worth it. There are various types of spam, such as email spam which is one of the most common types, web spam in which the spammer tries to trick the ranking algorithm of a search engine for his profit, social network spam, and many more (Chen & Chen, 2015). Opinion spammers with flowery ways will try to conceal the review's genuineness, leading to well-structured and carefully-written reviews in most cases.

According to (Ravi & Ravi, 2015) effective spam detection depends on three characteristics in the context of a fake review. Firstly, the content of the review, second the meta-data of the review, and last the genuine knowledge about the product. Hence, it is very important to develop a detection system that identifies false reviews from true ones.

2.1.3 Levels of Analysis

Generally, sentiment analysis is distinguished into three major levels depending on the analytical depth and the specific objectives of the application (Wang, et al., 2014): Document level, Sentence level, and Aspect and Entity level, as shown in *figure 3*.

Sometimes, an additional level is added concerning granularity, which is sorted between sentence level and aspect level (Wankhade, et al., 2022). It is considered a hot topic of researchers during the later years and aims in extracting sentiment words at a phrase level.

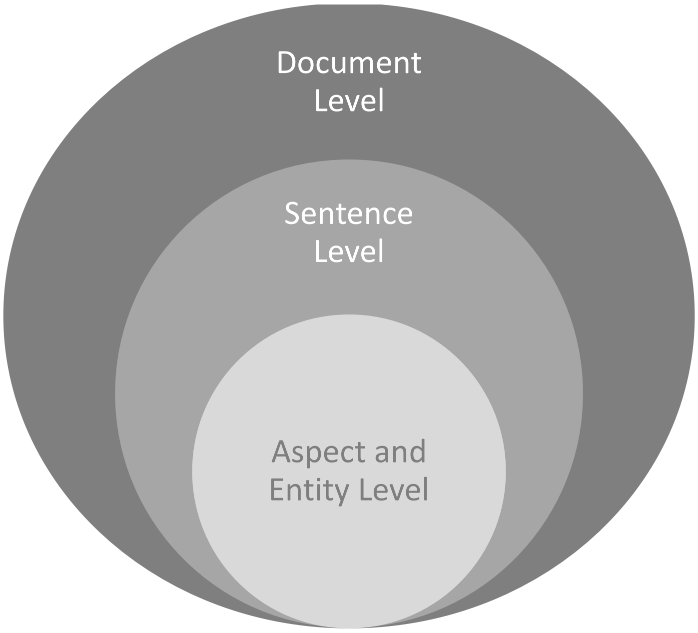

Figure 3. Levels of Analysis

2.1.3.1 Document Level

In this level of analysis, a document is considered to express a single opinion, positive or negative about a single entity (Liu, 2012). This task is also known as document-level sentiment classification. As an outcome, it determines the overall sentiment that derives from the whole document. This is why, document level classification is suitable when dealing with texts written by one person, or classifying chapters or pages of a book and article, and so forth (Birjali, et al., 2021; Wankhade, et al., 2022). Hence, even though it is perhaps the most extensively studied level, particularly in sentiment analysis early days (Pang, et al., 2002; Turney, 2002), it cannot be implemented in texts and opinions in which comparison and evaluation of multiple entities take place (Behdenna, et al., 2016). Its simplicity apropos the other levels justifies the fact that in real-world problems is not used a lot. Although, document sentiment classification can be implemented with both supervised and unsupervised learning approaches, as presented extensively by (Bhatia, et al., 2015).

2.1.3.2 Sentence Level

This level of analysis determines which sentences in the documents have positive, which negative, and which neutral sentiments are associated with them. Sentence level is also related to subjective classification (Rao, et al., 2018; Wiebe, et al., 1999), in which sentences that express factual information (namely, objective) are distinguished from those that have subjective information. Although, any confusion between subjectivity and sentiment would be a mistake, assuming that occasionally objective sentences insinuate sentiments or opinions, for instance, “I went hiking last Sunday and scratched my knee in some bushes”. Likewise, there are subjective sentences that lack opinions or sentiments, for example, “I think Argentina will win World Cup” (Liu, 2015). Another approach regarding the goal of subjectivity classification (Kumar & Sebastian, 2012) is to entirely exclude sentences that do not contain sentiment or opinion. This level is particularly helpful when a unique sentence contains multiple sentiments in it (Yang & Cardie, 2014).

2.1.3.3 Aspect and Entity Level

In most real-life applications, document and sentiment levels are considered insufficient. The reason lies in the difficulty of those levels of analysis to discern sentiment or opinion targets or allocate sentiments to the targets (Liu, 2015). Aspect level helps bridge this insufficiency gap created by the previous levels, by first identifying all the aspects in a sentence and then assigning a sentiment degree to each of those aspects (Behdenna, et al., 2016; Lu, et al., 2011; Schouten & Frasincar, 2016). By discovering how important those aspects are, we then have a clear view of the sentiment analysis problem. It is also known as aspect-level sentiment classification and in the early years was called feature level. In addition, it is considered to perform a fine-grained analysis, and it is applied to the most recent real-world applications.

For example, a sentence such as: "Although the burger was very tasty, delivery took more than expected". On the one hand, we can assume a positive tone is expressed, but we cannot say that it has the same tone for the entire sentence. Hence, someone who cares about the delivery time probably will choose a different restaurant. In the above example, "burger" and "delivery" are called opinion targets and are usually described by entities (e.g., "burger") and/or their aspects (e.g., delivery). After aspect-level sentiment classification, we can produce a summary of sentiments about all aspects and entities in a sentence (Liu, 2015).

2.1.4 Terminology

2.1.4.1 Opinion Definition

As discussed above, sentiment analysis is the computational study of opinions, sentiments, and emotions that are reflected in texts. Opinion is a term that contains various concepts such as sentiment, evaluation, appraisal, and attitude. Also, it covers exogenous information like the individual that expresses the opinion as the opinion target and the sentiment that derives from it. When dealing with multiple opinion holders an opinion summary is needed. Depending on the application and the desired level of analysis, different representations of definitions of opinions could be implemented. A definition that is generally acceptable for most sentiments analysis applications is defined as follows, according to (Liu, 2015):

$$(e,\ a,\ s,\ h,\ t)$$

where,

- e is the target entity,
- a is the target aspect of the entity e on which the opinion has been given,
- s is the sentiment of the opinion on aspect a of entity e,
- h is the opinion holder,
- t is the opinion posting time

2.1.4.2 Target entity and target aspect

It is the first component of the opinion definition. It is the object on which the opinion is expressed. An entity has its own set of attributes, features, or parts that constitute it, called aspects. An entity could be various things such as individuals, products, goods, services, companies, and so forth. For example, a car can be an entity, and its aspects can be the maximum speed, the dimensions, the consumption, etc.

2.1.4.3 Sentiment of opinion

Sentiment is the feeling that derives or attitude or appraisal or emotion related to an opinion. There are three types of sentiments, linguistic-based, psychology-based, and consumer research-based (Liu, 2015). The most frequent type is the latter, which also could be distinguished into two types (Chaudhuri, 2006): rational sentiment and emotional sentiment.

- Rational sentiments, also called rational opinions, do not express emotions. They state rationality and tangibility.
- Emotional sentiments express deep feelings and go deep into someone's psychological state of mind.

Sentiment could also be represented as:

$$(y, o, i)$$

where,

- y is the type of sentiment as mentioned above,
- o is the orientation, expressed as positive, negative, or neutral,
- i is the sentiment's intensity, which can be expressed by sentiment words or by using intensifiers and diminishers to change the degree of the sentiment

A practical example of rational and emotional types in real-world applications is the usage of a 5-star rating to express the intensity of sentiment. Usually, neutral sentiment is in the middle (three stars), rational sentiment is conveyed by a rating of 2 or 4 stars, and emotional sentiment by either zero stars or five stars.

2.1.4.4 Opinion holder

Usually, the opinion holder is the author of the sentence or review, hence the person who expresses the opinion. Opinion holders are also called opinion sources (Kim & Hovy, 2004). Occasionally, the opinion holder differs from the author. In those cases, the author conveys the sentiment of another person in the review.

2.1.4.5 Time of opinion

Timestamps of an opinion are widely used when dealing with social media applications. They are keeping track of the changeability of opinions over time, and they are useful for locating the specific time when opinion trends are proliferating.

2.2 Sentiment Analysis Approaches

There are 2 major sentiment analysis approaches, depending on the technique used, the Lexicon-based approach and Machine Learning-based approach. Although, most literature research now, adds an extra category, the Hybrid approach, which combines the previous two. Therefore, it is clear that those methods are being modified and updated frequently since this is a very active area of research and new techniques are arising constantly. The main goal of those approaches is to obtain better accuracies in conjunction with minimum available computational costs. Figure 4 shows an overview of the approaches and their techniques.

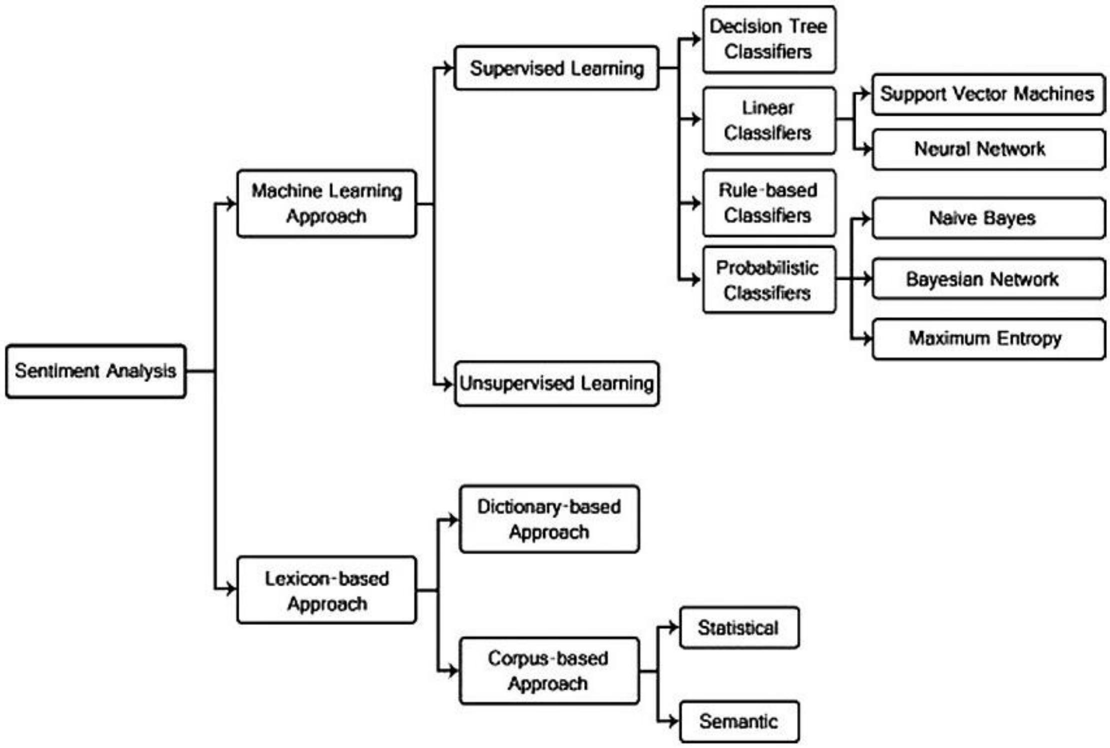

Figure 4. Sentiment analysis approaches (Medhat, et al., 2014)

2.2.1 Lexicon-based approach

Lexicon-based approach is an unsupervised methodology, and it is also known as knowledge-based approach (Birjali, et al., 2021). It belongs to the first methods implemented for sentiment analysis and is divided into two categories: The dictionary-based approach and the corpus-based approach. It draws on information from a lexicon, called opinion lexicon (Medhat, et al., 2014). Lexicons could be either built independently from different corpora, or they could be created from the same. Lexicons consist of tokens, namely words, which are associated with their semantic orientation. Apart from the above two main categories, a lexicon could be built manually but it is a very time-consuming procedure and rarely used in practice. These orientations, in most cases, are scores, which indicate if a word is neutral, positive, or negative (Kiritchenko, et al., 2014; Hu & Liu, 2004; Jurek, et al., 2015). Hence, semantic orientation is a measure of subjectivity and opinion in the document (Taboada, et al., 2011). Usually, a score could be either a simple polarity value, ranging from -1 to 1 for example, or a value corresponding to the word's sentiment intensity (Birjali, et al., 2021). Those tokens could be words, phrases, or even idioms. Generally, tokens with the highest scores are adjectives and adverbs. Lastly, to obtain the final sentiment of the document, we calculate or aggregate the semantic orientation of the tokens that compose the lexicon.

Lexicon-based approach is regarded to be particularly efficient at sentence and aspect levels of sentiment analysis. Its main advantage is that it does not need any training data to work. Although, building opinion lexicons is a process that needs thoroughness because of the complexity of the meanings of words and phrases. A word can have a positive meaning when used in a particular domain and exactly the opposite meaning when is used in a different one. For example, the word "small" could either has a negative meaning when used in a sentence like "The portions in that restaurant were too small", or a positive for a sentence like "There is only a small chance of making a mistake". As a rule, this obstacle can be overcome by either creating a domain-specific opinion lexicon or by modifying an existing vocabulary (Wankhade, et al., 2022).

2.2.1.1 Dictionary-based approach

The fundamental idea behind the dictionary-based approach is to build a vocabulary by utilizing a small set of sentiment words, also called seeds (Taboada, et al., 2011; Turney, 2002) through an iteration procedure. Initially, the first sentiment words, which their polarity scores are known, are collected manually. Then, the first iteration takes place by looking at an online dictionary like WordNet (Miller, et al., 1990) or another famous corpus, for synonyms and antonyms for each word in the set. The assumption here, polarities are the same for synonyms and the opposite for antonyms. When the first iteration is finished the newly found words are appended to the seed set, and then are used as input for the next iteration. The loop ends when the algorithm is not able to find new words to append to the set (Valitutti, et al., 2004; Hu & Liu, 2004). After the iteration is over, an inspection could be done, for cleaning up the set and removing possible errors.

An advantage of dictionary-based approach is that filling the requested vocabulary is a quick and painless process. Though, it comes with two major drawbacks. The first is that cleaning up the set from errors demands time and effort. The second is that building dictionaries with this approach eliminates the domain dependency needed for efficient semantic orientation of words, as mentioned above. However, corpus-based approaches assist in dealing with this disadvantage as we will see. Below, we can see in table 1 some examples of the most popular online dictionaries.

Table 1. Some well-known online dictionaries

Lexical Resource	Description
WordNet[29]	English lexical database which groups synonyms into sets called synsets. It consists of 117000 synsets, which are linked to other synsets via conceptual relations.
SentiWordNet 3.0[30]	Presented by (Baccianella, et al., 2010). A lexical resource directly for facilitating sentiment analysis applications. It annotates WordNet synsets according to their sentiment orientation.
SenticNet 2[31]	Presented by (Cambria, et al., 2012). SenticNet 2 provides one of the most comprehensive semantic resources. It associates 14000 concepts with their corresponding cognitive and affective information as well as their semantics.
NerthusV3[32]	Presented by (Arista, et al., 2016). Online database of Old English. It contains 31298 files which include various information for each word.
MPQA[33]	Presented by (Deng & Wiebe, 2015). Contains a list of subjective clues, namely express subjective opinion or judgment.
Opinion Lexicon by Liu Bing[34]	A list of positive and negative sentiment words, around 6800 words.

[29] https://wordnet.princeton.edu/
[30] https://aclanthology.org/L10-1531/
[31] https://sentic.net/
[32] https://varieng.helsinki.fi/CoRD/corpora/Nerthus/
[33] https://mpqa.cs.pitt.edu/corpora/mpqa_corpus/
[34] https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

2.2.1.2 Corpus-based approach

Contrary to the dictionary-based approaches, corpus-based methods aim to bootstrap dictionaries according to syntactic patterns and co-occurrences between words. This technique helps with the issue of context-dependent orientations we encounter with the aforementioned approach. Initially, a predefined list of words (called seeds) with their corresponding orientations is used to examine and find similar sentiment words conveying their orientations from a domain corpus. The main intuition of this approach is to measure the semantic distance between words in order to estimate their sentiment polarity (Darwich, et al., 2019). The use of linguistic constraints is particularly important regarding the addition of new sentiment words to the list. The initial paper was proposed by (Hatzivassiloglou & McKeown, 1997), in which a list of seed adjectives with their orientations in a corpus was exploited to find new adjectives to expand the list. The idea was that parts of speech, in this case adjectives, when they co-occur with a conjunction (e.g. AND), they have the same orientation. For example, regarding the sentence “The colors of this television are beautiful and vivid”, where “beautiful” and “vivid” usually have the same orientation. Likewise, orientations changes when words are connected with different conjunctions like OR, EITHER-OR, BUT etc. This notion is called sentiment consistency. Whilst corpus-based approach wins over dictionary approaches regarding context-dependent orientations, it is very difficult to create such a large corpus to cover all English words. Corpus-based method is distinguished into two categories, depending on its implementation, statistical approach and semantic approach.

2.2.1.3 Statistical approach

Statistical approach’s core idea is that the main factor determining a word’s sentiment orientation is that if a word appears more in positive texts then it is more likely to be positive and the opposite. Hence, if two words co-occur frequently in the same sentences, then it is more probable to share the same polarities. In the (Turney & Littman, 2003) seminal paper, the frequency of word co-occurrences was calculated to obtain their unknown polarities. The study relies on the phrase “a word is characterized by the company it keeps” quoted by the famous English linguist J.R. Firth (Firth, 1957). Specifically, they calculate the orientation of a given word based on its probability to co-occur with a word in a set of 7 positive words (excellent, superior, good, correct, positive, nice, and fortunate) minus the probability to co-occur with a word in a set of 7 negative words (unfortunate, bad, poor, inferior, nasty, negative and wrong). The methodology implements two different statistical measures of word association:

1. Pointwise Mutual Information (PMI) (Church & Hanks, 1991). The PMI measures the degree of statistical dependence between two words $word_1$ and $word_2$ as defined:

$$PMI(word\_1, word\_2) = log\_2\left(\frac{p(word_1 \wedge word_2)}{p(word_1)p(word_2)}\right)$$

   where, $p(word_1 \wedge word_2)$ is the probability of $word_1$ co-occurs with $word_2$. Provided that the words are statistically independent, their co-occurrence probability is given by the product $p(word_1)p(word_2)$. If the binary logarithm of the ratio is positive, then there is a tendency between these two words to co-occur with each other, otherwise if negative then it is more possible the presence of the first word repels the presence of the other.

2. Latent Semantic Analysis (LSA) (Landauer, et al., 1998) to calculate the strength of the semantic relationships between words. LSA deploys Singular Value Decomposition (SVD) (Golub & Van Loan, 1996) to analyze those relationships between words in the document. Before that, the algorithm begins by creating a matrix $\boldsymbol{X}$**,** with words as rows and documents as columns. Then Term Frequency – Inverse Document Frequency (Van Rijsbergen, 1979) is used to weight the matrix (we will discuss TF-IDF later in the thesis). SVD is the next step, where the decomposition of matrix $\boldsymbol{X}$ takes place:

$$X_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^{T}_{n\times n}$$

where, U and V are both orthogonal and Σ is a diagonal matrix of $singular\ values$. This technique generates the required patterns to measure the similarity between two words. The three matrices $U_{m\times m}\Sigma_{m\times k}V^{T}_{k\times k}$, is an approximation of the original matrix $\boldsymbol{X}_{m\times n}$ with the top k singular values.

Finally, the similarities between two words $LSA(word_1, word_2)$ are calculated using the cosine similarity distance of their corresponding row vectors (Deerwester, et al., 1990; Bartell, et al., 1992; Schütze, 1992; Landauer & Dumais, 1997):

$$\cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

where,

- A and B represent the word vectors
- A_i, B_i represent index i of that vector

Several studies and papers have adopted the above technique to perform sentiment analysis and build sentiment lexicons.

2.2.1.4 Semantic approach

Semantic approach, also known as ontology-based approach, differs from the aforementioned in the methodology it follows in the context of how to compute the similarities scores between words. The notion relies on words that are semantically close to each other share similar sentiment values. This technique, uses thesaurus and dictionaries such as WordNet to draw on synonyms, antonyms, or even relationships between tokens, namely words. Then, the iteration begins in which similar words are appended to the initial set. Lastly, a corresponding sentiment polarity or value of a word is given based on the relative number of positive and negative synonyms of the word in the dictionary (Kim & Hovy, 2004).

Statistical and semantic approaches are frequently combined in sentiment analysis. Such work is presented by (Zhang, et al., 2012) in which both techniques were used to find weaknesses of the products among Chinese reviews. They built an expert system based on aspect-based sentiment analysis and similarity calculated using HowNet (Zhendong, et al., 2010). The features were extracted and grouped and then PMI method was used to identify implicit features (aspects).

2.3 Machine Learning-based approach

In this approach sentiment polarities are classified automatically through a training procedure, with labeled source texts. In most cases as before, polarities are from three categories, positive, negative and neutral. (Zhang & Zheng, 2016) suggested that machine learning-based approaches could be divided into two substantial groups. Firstly, traditional methods, which contain the classic machine learning techniques (Dang, et al., 2020), excessively presented by (Pang, et al., 2002). Some of the most popular and effective algorithms, which we will discuss them later in this chapter, are naïve Bayes classifier (Malik & Kumar, 2018), maximum entropy classifier (Mehra, et al., 2002; Wu, et al., 2017), or Support Vector Machines (SVM) (Firmino Alves, et al., 2014). Though, on the later method, deep learning models usually provide more accurate results, by the use of deep neural networks (also known as artificial networks). Generally, deep neural networks are widely used in sentiment analysis. When the training process is completed, then evaluation of the model takes place by testing how it performs with new inputs, namely texts. Admittedly, machine learning approaches in relation to the others are more generally used now. This is something which we can also observe in figure 4 below. The (Kulkarni & Rodd, 2021) study examines the frequencies of the published papers, in the context of which sentiment analysis approach is used in Hindi.

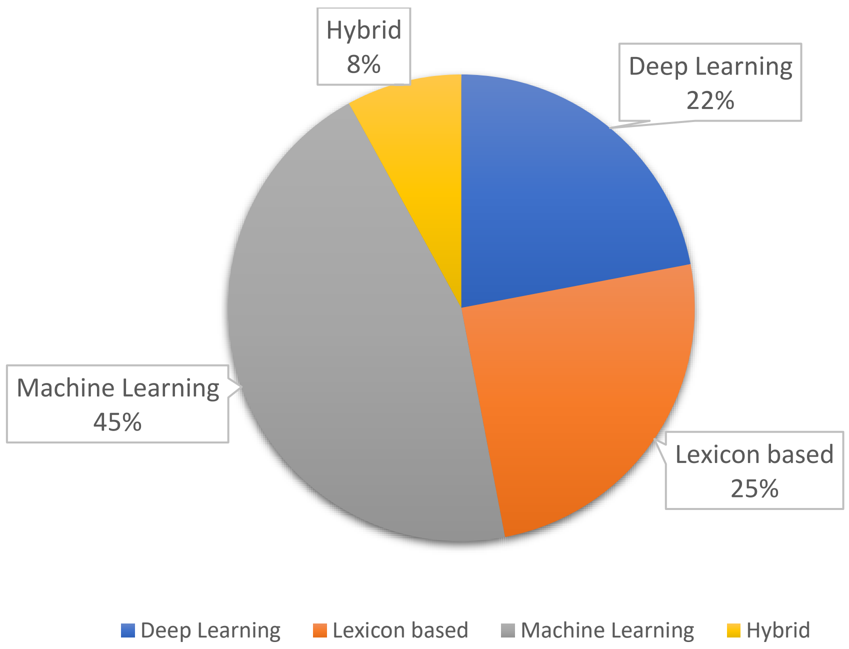

Figure 5. Frequencies of research work using Sentiment Analysis in Hindi.

Machine learning approach is distinguished also in four sub-categories (Yusof, et al., 2015):

1. Supervised learning
2. Unsupervised learning
3. Semi-supervised learning
4. Reinforcement learning

In the next section we will discuss some of the most common machine learning algorithms used in sentiment analysis.

2.3.1 Supervised learning

Supervised learning methods are the most frequently methods used in sentiment analysis from the above four categories of machine learning approach, and apparently is the most accurate (Wankhade, et al., 2022). The provision of labeled data for an effective classification is essential. This requirement, despite the fact that supervised learning outperforms the other methods, sometimes it makes it inefficient. Depending on the methodology supervised sentiment classification methods are also divided in two main classes: 1) Probabilistic classification, which includes classifiers such as naïve Bayes, Maximum entropy or Bayesian network, and Non-probabilistic classification which includes classifiers such as Support Vector Machines, Nearest Neighbor, Decision Trees etc. (Hemmatian & Sohrabi, 2019).

2.3.1.1 Logistic Regression

Logistic regression is one of the most commonly algorithms used, in social and natural sciences, and close relative with neural networks (Jurafsky & James H., 2009). It classifies an observation into one of two classes 0 or 1, so it is mainly used for binary classification tasks, like distinguishing positive or negative sentiments. In fact, logistic regression takes as an input a linear regression, and applies the sigmoid function to the output. It can be divided in two phases:

1. Training: The algorithm is trained on the weights and biases. It is optimized by using usually stochastic gradient descent, and because it is a classification task, its loss function is cross-entropy.
2. Testing: It computes $p(y|x)$ for an observation x and returns the higher probability class $y = 0$ or $y = 1$.

For linear regression:

$$z = b + w_0x_0 + w_2x_2 + \cdots + w_Nx_N = W^TX + b$$

where, w are the weights, namely the parameters, and b the bias term, also called intercept, and is a real number.

Logistic regression is of the form:

$$sigmoid(z) = \frac{1}{1 + e^{-z}}$$

Sigmoid function named after its visual representation, as we observe in the following graph, figure 6. We can see that it is nearly linear when it takes values near zero and one.

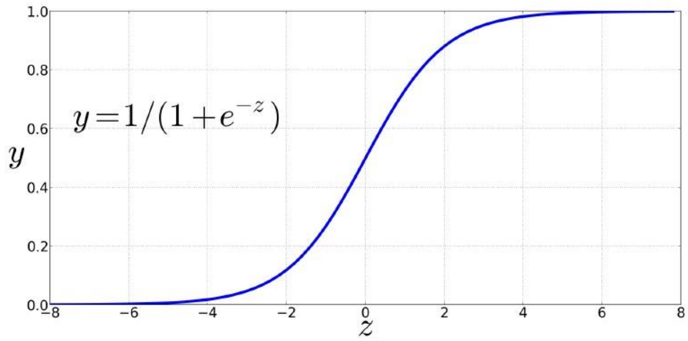

Figure 6. Logistic Regression Curve

The cutoff point of logistic regression is at 0.5, which means if the function takes a value below that point, it classifies the observation as 0 and above 0.5 as 1.

2.3.1.2 Naïve Bayes

Naïve Bayes (NB) belongs to the probabilistic classifiers, and it took its name after the Bayes' theorem, when first implemented by (Mosteller & Wallace, 1963) for text classification. It is considered as a simple classifier and is also very commonly used in those problems (Medhat, et al., 2014).

One suitable for machine learning task word representation (we will speak later in another section) is bag-of-words (BoW). According to BoW, words can be represented by their frequency in the document or corpus, which is simply the count of each word. We assume that the probability of a word to occur is independent to its context and position in the corpus (Mccallum & Nigam, 2001). Hence, for a document d, the probability of assigning it to the class $\hat{c}$ out of all possible classes, where $c \in C$, is by finding its maximum posterior probability:

$$\hat{c} = argmax_{c \in C} P(c|d)$$

Knowing the Bayes Theorem formula, we can adjust it according to our problem. So, we could rewrite it as:

$$\hat{c} = argmax_{c \in C} P(c|d) = argmax_{c \in C} \frac{P(d|c)P(c)}{P(d)}$$

In the above formula the denominator can be dropped, because it will be applied for each possible class. So, the most possible class given the documents, is resulting to be the highest product of two probabilities: the likelihood of the document $P(d|c)$ and the prior probability of $P(c)$. Furthermore, we can generalize without any loss $P(d|c)$ and represent d as $f\_1, f\_2, \dots, f\_n$. In addition, based on the independence assumption said above, the probabilities $P(f\_i|c)$ could be written as: $P(f\_1|c) \cdot P(f\_2|c) \cdot \ldots \cdot P(f\_n|c)$. Lastly, to avoid numerical underflow we introduce the log in the equation. Also, if we take into account the position of each word in the document the last equation is expressed as (Jurafsky & James H., 2009):

$$c_{NB} = argmax_{c \in C} log \overbrace{P(c)}^{prior} + \sum_{i \in positions} log \overbrace{P(w_i|c)}^{likelihood}$$

Whilst Naïve Bayes classifiers are difficult to be implemented in real-world situations due to the conditional independence assumption, it still obtains outstanding results in text categorization (Pang, et al., 2002). Finally, in addition to its simplicity, Naïve Bayes is substantially fast compared to other more complex methods. There are various implementations of Naïve Bayes, most common are:

- Multinomial Naïve Bayes (MNB):
    - Used especially for text classification tasks. Suitable for multinomially distributed data. MNB also works well with Term Frequency-Inverse Document Frequency (tf-idf) vectors apart from BoW.
- Gaussian Naïve Bayes:
    - The likelihood of the inputs is assumed to be Gaussian. Furthermore, mean (μ) and variance (σ^2) are estimated using maximum likelihood.
- Complement Naive Bayes (CNB):
    - CNB is an adjustment of MNB for tasks with imbalanced data. To compute the model's weights CNB deploys statistics from the complement of each class.
- Bernoulli Naïve Bayes:
    - Implemented when data is distributed based on the multivariate Bernoulli distributions. Each feature is assumed to be a binary-valued variable (Manning, et al., 2008).

2.3.1.3 Maximum entropy

Maximum entropy (ME) is another probabilistic algorithm for text classification tasks. It is considered to be an alternative method of NB, which is an acceptable solution and sometimes even outperforms NB. Contrary to NB, Maximum Entropy is exempted from independence assumptions, something which is a crucial factor for ME, to be opted in case those assumptions are not met (Pang, et al., 2002). Maximum entropy estimates $P(c|d)$ with the following exponential form:

$$P_{ME}(c|d) = \frac{1}{z(d)} exp\left(\sum_i \lambda_{i,c} f_{i,c}(d,c)\right)$$

where,
$z(d)$ is a normalization function,
$\lambda_{i,c}$ is a parameter for the weight to ensure the observed features match the expected,
$f\_\{i,c\}$ is a feature function for f_i and class c as defined:

$$f_{i,c}(d,\acute{c}) = \begin{cases} 1, n_i(d) > 0 \; and \; \acute{c} = c \\ 0, otherwise \end{cases}$$

A large value for $\lambda_{i,c}$ is interpreted as the i^{th} feature is considered an important indicator for the class c.

2.3.1.4 Support Vector Machines

Support Vector Machines (SVM) (Cortes & Vapnik, 1995) belong to the family of non-probabilistic classifiers and is suited for both continuous and discrete variables. This model is one of the most preferred algorithms for sentiment analysis, and is very similar to logistic regression. They are both driven by a linear function $x^T x + b$, but differ on the fact that SVM provides a class identity instead of probabilities (Goodfellow, et al., 2016). SVM are able distinguish data either linearly or non-linearly. The goal of an SVM classifier is to identify the hyperplane which perfectly distinguishes the data into distinct classes. The optimum distinction for a hyperplane occurs when it obtains the maximum margin to the closest point from either class.

A hyperplane in an n-dimensional space, is a flat affine subspace of $n-1$ dimensions. For example, for two-dimensions the hyperplane is a flat line. Generally, for n-dimensions a hyperplane is defined as:

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n = 0$$

When X does not satisfy the above equation, two situations happen. Firstly, observe inequality below 0, which means that X lies to one side of the hyperplane, or secondly, inequality above 0, which means the opposite. Below, on the right of the figure 7, we can observe a hyperplane dividing a space of two-dimensions, by (James, et al., 2014).

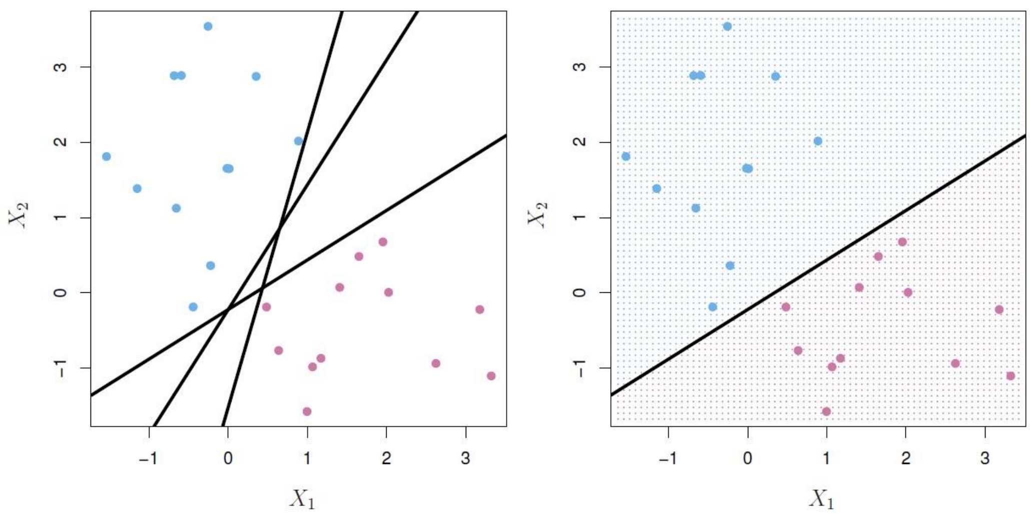

Figure 7. Left: 2 classes, 3 hyperplanes. Right: Same classes but one hyperplane.

Assuming an initial hyperplane that separates linearly the data is constructed, then we could easily infer that an infinite number of hyperplanes could be created from the first. This happens because a hyperplane could do a little spin move, or move up or down, but without any of the observations being touched by it. The left side of figure 7 shows three hyperplanes separating the data in the two classes.

Consequently, a hyperplane that is furthest from the observations can create, which is called maximal margin hyperplane (also called optimal separating hyperplane). Firstly, the distance of each element from a given separating hyperplane is computed. This distance is called the margin. The goal is to find the maximal margin hyperplane, or maximal margin classifier, which is the separating hyperplane that has the largest margin.

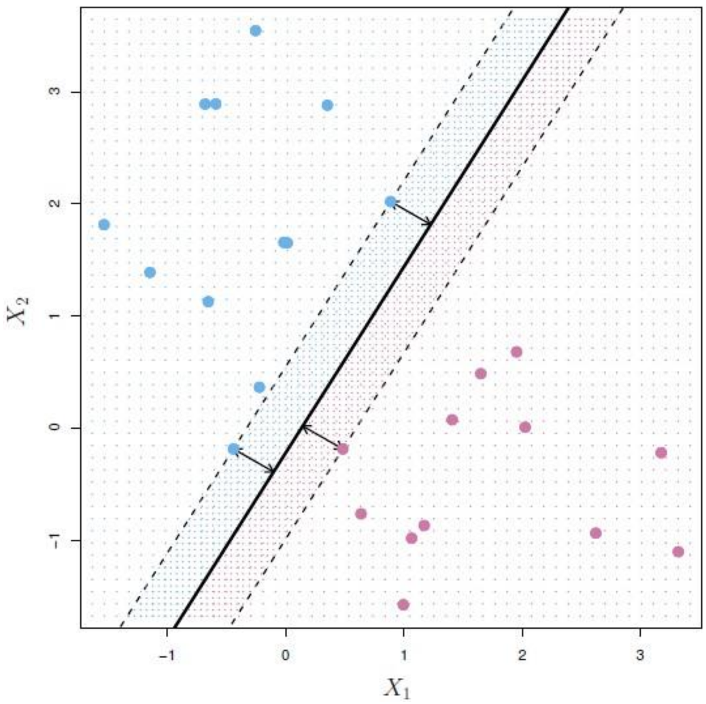

Figure 8. Support Vector from the previous figure.

In *figure 8* (James, et al., 2014), we can observe the maximal margin hyperplane as a solid line. The area inside the maximal margin hyperplane and the dashed lines of the margin. We can also observe that 2 observations from the blue class and 1 from the red class, determine the width of the margin and at the same time have the same distance from the maximal margin hyperplane. Those observations are called support vectors. The area also inside the dashed lines indicates the decision rule by a classifier on this separating hyperplane.

Support vector machines also have implementations for observations that are not linearly separable. In these cases, (Cortes & Vapnik, 1995) introduced the soft margin method, that is a variable which for each misclassified observation gives a penalty to increase the distance from the observation's support vector. Furthermore, to achieve non-linear classification, SVM provide a way of mapping data to a higher dimensional space, called kernel trick. This can be obtained by applying a kernel function to the data. Most common kernel functions are Radial basis function (RBF), Sigmoidal kernel and Polynomial kernel (Manning, et al., 2008).

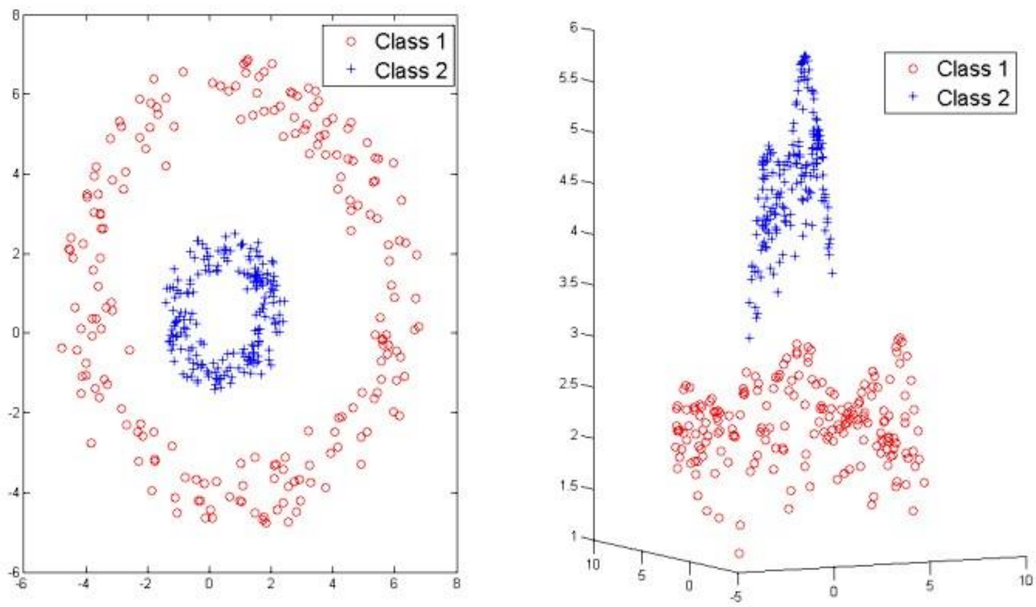

Figure 9. Data remapped using RBF kernel

Figure 9 (Fletcher, 2009) is shown how the data are remapped using RBF kernel. For better adjustments two parameters are crucial:

- Parameter γ, controls the radius of influence for each observation,
- Parameter C, which is the misclassification penalty.

2.3.2 Unsupervised learning

In contrast to supervised learning, in unsupervised learning the data are not associated with a supervision signal, namely the labels. Available data is not always well structured, especially when it comes to the context of natural language. Sometimes is preferable to deploy algorithms that can discover on their own, aspects of the features that interlock with each other to develop relations. Generally, unsupervised learning aims at answering questions like:

- What visualization is the most informative for the data?
- Are there any relations between the observations of the variables?

Theoretically, those answers are described from subjectivity. There is not an objective goal, such as prediction or output. Hence, according to (James, et al., 2014) the task is frequently part of exploratory data analysis.

Usual tasks of unsupervised learning include density estimation, data visualization, denoising data from a distribution, clustering observations into subgroups, etc. (Goodfellow, et al., 2016). Some of the most common unsupervised learning algorithms are:

- K-means:
    - Clustering algorithm. Separates the elements in n clusters of variances as equal as possible.
- DBSCAN:
    - DBSCAN is also a clustering algorithm. Separates areas of low density from those areas of high density.

- Empirical Covariance:
  - The classical maximum likelihood estimator (or empirical covariance approximates the covariance matrix of the data.

In the following part we will discuss an unsupervised learning algorithm, the Principal Component Analysis (PCA) which is widely used in sentiment analysis.

2.3.2.1 Principal Component Analysis

Principal component analysis (PCA) is a method that projects the observations, in reduced dimensions. Very frequently word representations consist of a large number of dimensions, for example a typical length of a word vector range between 100 and 300 dimensions. A key concept of principal component analysis, is that despite the dimensionality reduction, compressed data keeps the maximum information about the original word vectors counterparts. That is, most of the variation in the data is explained in fewer dimensions. PCA also complies with the criterion of independent representations, which means elements have not linearly correlated with each other.

Usually, the preferred representation is in two dimensions. This is because plotting the data in x and y axes, and visually examine them helps a lot, in identifying clusters and associations between the data. The usual process for PCA is described as follows:

1. Normalization of the data. A preferred normalization is subtracting their mean from all examples.
2. Compute the covariance matrix of the data.
3. Compute the eigenvectors and the eigenvalues of the covariance matrix.
4. Multiplication of the first desired k eigenvectors by the normalized data.

In the *figure 10,* we can observe 11 example words represented as word vectors reduced in two-dimensional space, in order to be examined visually. Those words are automatically clustered in four categories depending what sentiment is represented. We can deduce that words that express emotion are clustered at up-left corner, while totally different meaning words like gas, oil, and petroleum are clustered on the other side. Furthermore, there are two remaining clusters which are similar to each other. The first consist of words like city, town, and village, while the other words like country and continent. The visualization explains to us that while the two categories differ from each other, the words that form the two clusters are semantically very close.

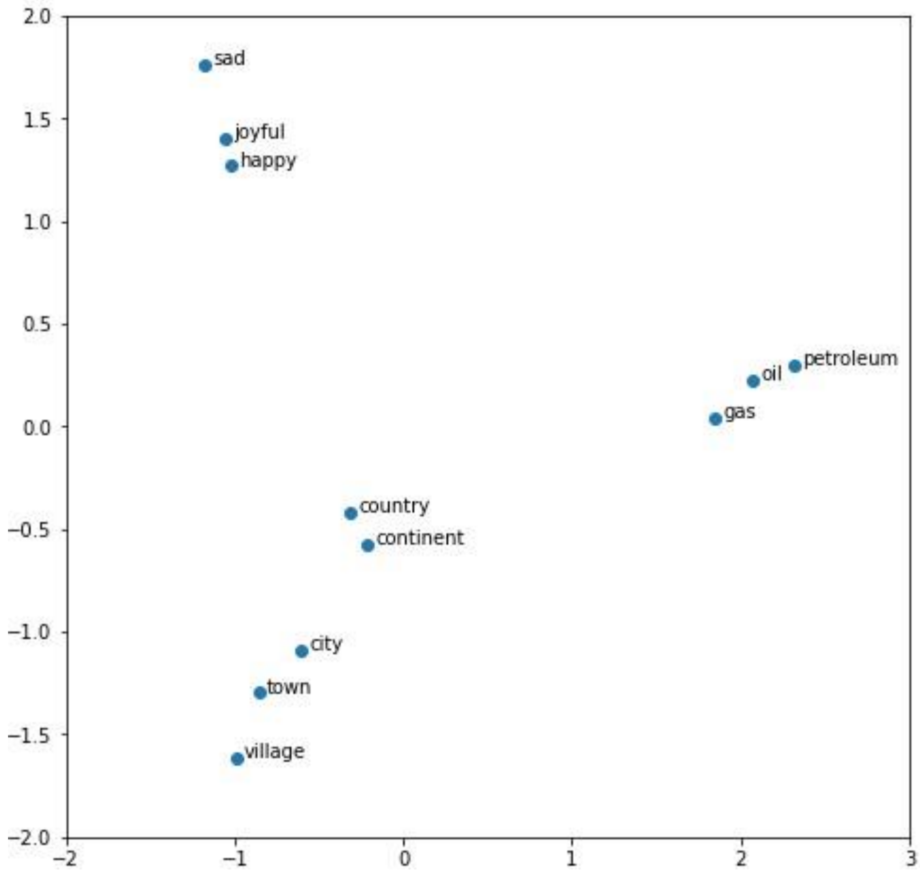

Figure 10. Word vectors clustered in four categories

2.3.3 Semi-supervised learning

Semi-supervised algorithms apply learning from both labeled and unlabeled data in a single model. While most popular classifiers use labeled data, there are some crucial drawbacks to take under consideration. Labeling for example, requires a lot of effort, and strong domain knowledge. Some other times, labels could be difficult to be found, or expensive or need a lot of time to be obtained. On the other hand, unlabeled data may be abundant, spread around the web or other sources, but there are not as many approaches as supervised learning (Hemmatian & Sohrabi, 2019). Hence, semi-supervised learning, gathers a lot of unlabeled data, along with some little labeled to build the classifiers. The general idea is that instead of manually label the whole dataset, a model can be trained on a small part of it and then the classifier could expand it, to label all the remaining examples. Some of the most known algorithms are:

- Self-training:

  Self-training is a widely used approach. Firstly, the model is trained with the few labeled data, and then the classifier will classify the rest unlabeled dataset. Then, the procedure will be repeated with the whole dataset. In the end, data with the highest confidence of sentiment label are picked and appended to the labeled data.

- Co-training:

  It is considered to be an improved version of self-training. Though, the training process differs from the first because two individual classifiers are trained based on two views of data, which are independent from each other. Hence, they provide mode different information about each example.

- Graph-based methods:

  Graph-based methods are quite new, since they gained attention from researchers in the last decade. Vertices represent instances in the graph and edges represent similarities among instances. Instances that share strong connections with each other are inclined to belong to the same class (Hajmohammadi, et al., 2015).

2.3.4 Reinforcement learning

In reinforcement learning the crucial notion is that the learning procedure gets through a trial and error mechanisms. That is, there is no prior knowledge which are the proper patterns an agent should follow and the algorithm learns them via reward-based reinforcement, depending on the agent's previous actions. In reinforcement learning, the surrounding environment is taking into account, and how an agent interacts with it. It is a dynamically developing procedure. Its notion is that the complexity of those systems makes them difficult to be modeled explicitly, but simple enough to reward the proper actions and encourage an agent (Aggarwal, 2018). In addition, the algorithm penalties an agent with negative values in case the action or move is not the desired. Hence, the algorithm does not need labels to learn the output but it learns from ever-changing and dynamic situations. According to (Birjali, et al., 2021), the main advantage of this approach lies in the way of the algorithms learn, which is very similar in how human beings learn, which is very desirable especially in sentiment analysis discipline.

Reinforcement learning applications range mostly in game development, robot control and self-driving cars. In sentiment analysis, a very interesting application is by (Liu, et al., 2018), in which physiological signals were considered to predict emotion states in real-time. The paper explains that although there are prominent methods that predict emotion states in real-time, there is a shortage of those which deliberate rewarding current operation in each iteration. In the end, this method saved a significant amount of time, in addition to noticeable performance.

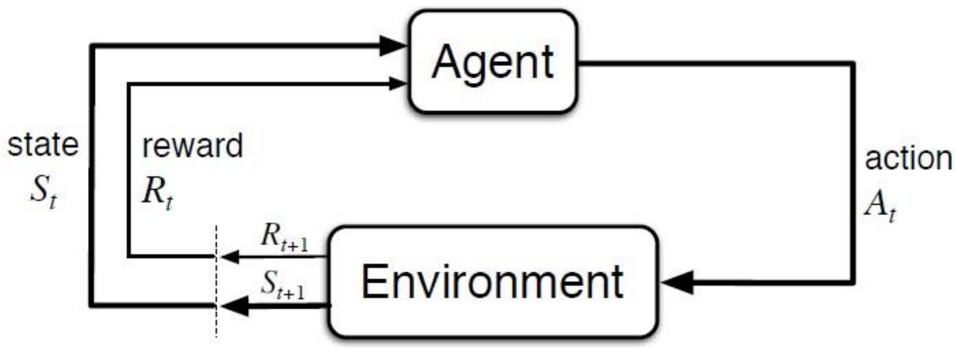

Figure 11. State diagram of agent/environment (Sutton & Barto, 2018)

In *figure 11*, we can observe the state diagram (Sutton & Barto, 2018) in which describes the interactions between agent and environment on each time step, $t = 0,1,2,\ldots,n$. For every time step, a state environment, in which $S_t \in \mathcal{S}$, sends its representation to the agent, and an action is chosen, where $A_t \in A$, where A is the set of possible actions. In the next step, gets a reward, $R_{t+1} \in \mathcal{R}$, and flows to the next state, S_{t+1}.

2.3.5 Model Evaluation

In order to evaluate the learning process, there is a wide range of different measures for sentiment analysis from information retrieval (Baeza-Yates & Ribeiro-Neto, 1999).:

- Precision:

  Precision is the ratio of the positive observations that predicted correctly to the total number of observations that predicted as positive (Meivel, et al., 2022). Precision indicates the strength of the prediction.

$$Precision = \frac{True_{positive}}{True_{positive} + False_{positive}}$$

- Recall:

  Recall (also known as sensitivity) is the ratio of the positive observations that predicted correctly to the total number of positive observations in the data (Meivel, et al., 2022). Recall is inversely proportional to precision.

$$Recall = \frac{True_{positive}}{True_{positive} + False_{negative}}$$

- F1-score:

  F1-score is the harmonic mean of the above two measures. It takes values from 0 to 1, the closer to 1 the F1-score is the better performance of the model

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

- Accuracy:

  Accuracy is the ratio of the observations that predicted correctly to the total predictions in the data. It is the most common classification metric.

$$Acc = \frac{True_{positive} + True_{negative}}{Truepositive + True_{negative} + False_{positive} + False_{negative}}$$

2.4 Natural Language Processing

The choice of which methodology to choose conserving a sentiment analysis task, it is a fundamental concept. Natural language processing is the field that understands natural unstructured language from the computer’s perspective, hence the preprocessing part should not be omitted. In this section, we will discuss the most common preprocessing techniques and methods used for sentiment analysis.

2.4.1 Preprocessing

Preprocessing techniques aim to prepare the data in a way, that non-only computers are able to interpret it but also humans through preprocessing and visualization techniques. Especially data that collected via internet may contain a lot of noise, apart from grammatical and spelling issues. Hence, for efficient text classification, documents should follow the same structure and patterns, in the whole data. Preprocessing techniques may include:

- Tokenization:

  A word, according to (Meyer, 2009), is the smallest meaningful unit (Zhiyuan, et al., 2020). By this, and in order to facilitate the procedure sentences are divided into smaller units, called tokens.

- Annotation:

  Annotation helps in the identification of named entities, part-of-speech or key phrases within a text.

- Tagging:

  Part-of-Speech (POS) tags are the syntactic category a work belongs to (White, et al., 2018). Although, POS tags may differ, since there are various different tag groups. One of the most deployed tag set is the English Penn Treebank[35].

- Lemmatize

  Lemma is the base form of a word as presented in a lexicographical resource (White, et al., 2018). It is very related to stem, which is the process of reducing words to their root form.

- Stop words

  Stop words generally constitute words that are the most commonly used in a language. They are considered to be insignificant hence, they do not help with the learning process.

2.4.2 Word representations

In order to have an effective solution in sentiment analysis tasks, an optimum representation of words is needed. Those representations usually contain words represented as vectors, and they are substantial in sentiment analysis. There are various word representations in terms of the task and the chosen procedure. In the next section we discuss some common representations in sentiment analysis.

2.4.2.1 One-Hot encoding

One-hot encoding is the simplest form of word representation. All words in a corpus are stored in a set, which is basically the vocabulary $V = \{w_1, w_2, \ldots, w_V\}$ of the corpus. Then each word represented by a vector $\vec{w}$ of V dimensions. $\vec{w}$ can take only two values. The value 1 at its index in the sorted vocabulary set, and for all other dimensions 0. Formally, a word $\vec{w}$ is represented as (Zhiyuan, et al., 2020):

$$w_i = \begin{cases} 1 & if\ w = w_i \\ 0 & everywhere\ else \end{cases}$$

[35] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

2.4.2.2 Term frequency – Inverse document frequency (Tf-idf)

In most cases sentiment analysis tasks perform better when the word representations are not just raw indexes. Tf-idf is the most common reweighting scheme in sentiment analysis and in NLP generally. Tf-idf tends to punish the words that occur many times in a document, because they are considered to be insignificant. On the other side words that exist fewer times tend to get higher values in the end. Tf-idf formula is as follows:

$$TF = P(w|d)$$

$$IDF = log\left(\frac{|D|}{|d \in D|w \in d|}\right)$$

$$TF - IDF = TF \cdot IDF$$

where, $|D|$ is the number of documents in a corpus
$|d \in D|w \in d|$ is the number of documents containing the word w.

2.4.2.3 Word embeddings

In contrast with the previous methods of word representations, word embeddings is the method in which a word is represented as a real-valued vector. Each word's vector is designed in such a way that other words' vectors with similar representations have also similar word meanings (Dang, et al., 2020). Word embeddings are widely used in sentiment analysis with deep learning approaches. In addition, they are representations of dense matrices, that is most values are above zero. Word2Vec (Mikolov, et al., 2013) is one of the most widely used word embeddings. It is categorized in two methods. The first, Continuous Bag-of-Words (CBOW) (Mikolov, et al., 2013) is a shallow neural network that predicts a word from its surrounding words. On the contrary, the second method called Skip-gram (SG), presented by (Mikolov, et al., 2013) is directly the opposite, that is, predicts the surrounding words given an input word. Another word embedding representation is Global Vectors (GloVe), presented by (Pennington, et al., 2014), in which the vectors are developed through unsupervised learning approach.

2.5 Deep learning-based approaches

Deep learning approaches (DL) are considered to be a branch of machine learning approach. It is a re-branded name of artificial neural networks (ANN), which is a simulation of the learning mechanism in human organism. Our nervous system contains neurons[36] (also called nerve cells) which are connected with each other and consist of three main parts. Apart from their main body, consist of axons[37] which are the output parts of neurons and they are responsible to pass electric messages to adjacent neurons. Thirdly, the dendrites[38] which are the receiving parts of neurons. Depending on the input received, a dendrite will decide whether fire an action or not.

Likewise, deep learning algorithms consist of units, which are the nerve cells counterpart, and they are organized in layers. The neurons of each layer are connected with each other and their

[36] The University of Queensland, Article: What is a neuron? https://qbi.uq.edu.au/brain/brain-anatomy/what-neuron, last visit 14/12/2022
[37] The University of Queensland, Article: Axons: the cable transmission of neurons, https://qbi.uq.edu.au/brain/brain-anatomy/axons-cable-transmission-neurons, last visit 14/12/2022
[38] Biology dictionary, Article: Dendrite, https://biologydictionary.net/dendrite/ last visit 14/12/2022

output may feed into the inputs of the following one or more neurons (Goldberg, 2015). This kind of neural networks is called feedforward networks, and they are the most common kind of neural networks. A fully connected neural network, is a network in which all neurons in every layer, receive as inputs the outputs of all the neurons of the previous layer, and generally this is a typical neural network structure.

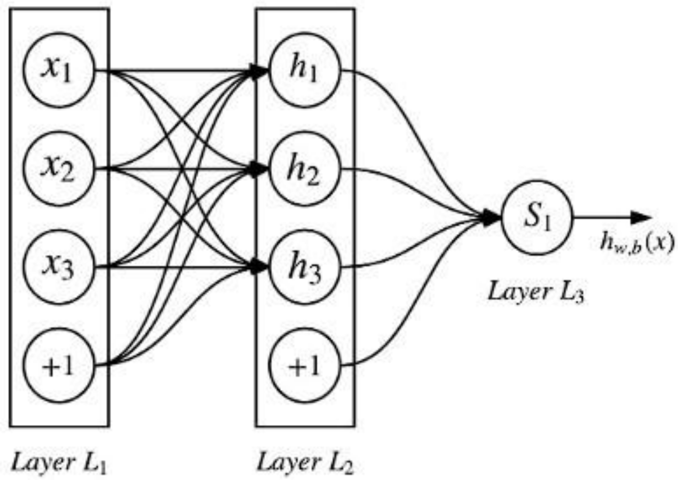

Figure 12. Simple feedforward neural network (Zhang, et al., 2018)

In the *figure 12* (Zhang, et al., 2018) we can see the formation of a simple neural network. L_1, L_2, and L_3 are the three layers that constitute the network. The first layer is called input layer and the third is the output. L_2, the second layer is also called the hidden layer. Neural networks that have a large number of hidden layers are called deep neural networks.

Each unit of the network, is assigned some computations for the learning process (Jurafsky & Martin, 2020). Firstly, it receives the input $x_1, x_2, \dots, x_m$ vector. For every input, the units have the corresponding weights $w_1, w_2, \dots, w_n$, which use them to first multiply them with the input vector, and then sum the results along with a bias term b. For a fully connected network, all the outputs of the previous layer, are summed by every individual neuron by the next. Hence, we can represent the input of a unit as:

$$z = \sum_i w_i x_i + b$$

Or in a more convenient way, as the dot product of the vectors:

$$z = w \cdot x + b$$

Afterwards, an additional function is used to generate the output, called activation function. Linear functions, such as z, are not helpful in the learning process of the neural network, so we need non-linearities (Goldberg & Hirst, 2017) to reach the final output for the next layer. Hence, the activation function is:

$$a = f(z)$$

The choice of activation function is a crucial part in deep learning network design. According to the desired output value, the choice should be opted thoroughly. For example, to find a real number, then a valuable choice is to pick the identity activation function, that is a least-squares regression (Aggarwal, 2018). In the following section we discuss the most common activation functions.

2.5.1 Activation Functions

As already said, activation functions are commonly used to brake the linearities. There is not a correct choice for a given task every time, and it is mostly an empirical question (Goldberg & Hirst, 2017).

2.5.1.1 Sigmoid

We have already seen sigmoid function in the section of logistic regression algorithm. It is used for binary classification, since its output is a probability ranging from 0 to 1:

$$sigmoid(z) = \frac{1}{1 + e^{-z}}$$

where, $z \in (-\infty, \infty),\ sigmoid(z) \in (0,1)$

2.5.1.2 Tanh

Hyperbolic tangent (tanh) is often used in the hidden layers of the neural network (RASCHKA & VAHID, 2019). It is also a sigmoidal function and it is considered as an alternative to sigmoid, since it usually performs faster in practice. The main difference from sigmoid is that tanh outputs values ranging from -1 to 1, as we see in figure 13.

$$tanh(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

where, $tanh(z) \in (-1,1)$

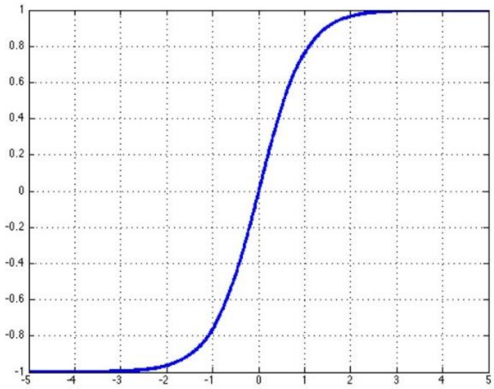

Figure 13. Hyperbolic Tangent

2.5.1.3 ReLU

The Rectified Linear Unit (ReLU) is a very common activation function in neural networks, because it addresses the issue of vanishing gradients. An issue in which the gradient terms during backpropagation a very close to 0. It is a simple function, namely when z is positive its output is z, and when z is negative its value is always 0 (Jurafsky & Martin, 2020), as we observe in *figure 14.*

$$rect(z) = max(0, z)$$

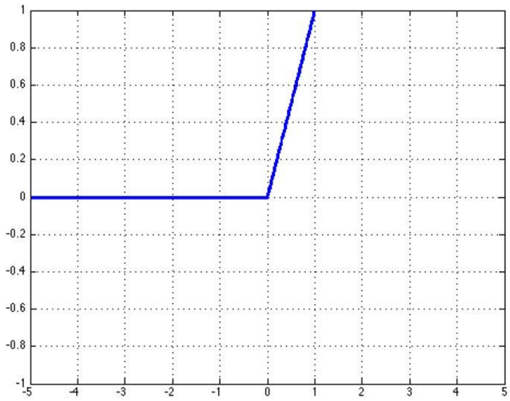

Figure 14. Rectified Linear Unit

2.5.1.4 Softmax

Traditionally, softmax function is used by the final layer of a deep neural network. It is a generalization of the sigmoid function, and it is used to represent the probability distribution over n different classes (Goodfellow, et al., 2016), with each value ranging from 0 to 1. Hence, for an input vector $z = [z_1, z_2, \dots, z_k]$ of dimensionality k:

$$softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} , 1 \leq i \leq k$$

2.5.2 Building an artificial neural network

In *figure 12* we saw a simple feedforward neural network, also referred as shallow neural network. In real-world problems deep learning networks consist of a lot more hidden layers. A fundamental concept of an ANN is backpropagation. In a classification problem, information must pass through the layers of a neural network as many times as it needs for efficient prediction of the observations $x = x_1, x_2, \dots, x_n$, based on the corresponding labels given $y = y_1, y_2, \dots, y_n$. Hence, the algorithm needs a function $f()$, which is called the loss function, that achieves this mapping of input observations with the desired labels, that is $\hat{y} = f(x)$ (Goldberg & Hirst, 2017). In a more formal way, given the expected output y and a predicted output y the loss function $L(\hat{y}, y)$ assigns a scalar, which is an indicator of the accuracy of the model for the specific iteration. Its goal, is to minimize the scalar in every iteration. Generally, when the loss function reaches a point where there is no more significant minimization then the iterations stop. Loss functions in supervised learning are distinguished into two categories depending on the task. Below are presented some of the most common loss functions. (White, et al., 2018) For regression tasks the loss function is the squared error loss defined by:

$$SE(y, \hat{y}) = (y - \hat{y})^2$$

For binary classification though the loss function is the cross-entropy loss, as:

$$CE(y, \hat{y}) = -\big((1 - y)log(1 - \hat{y}) + ylog(\hat{y})\big)$$

For circumstances where the classification classes are more than two the cross-entropy loss function is adjusted to categorical cross-entropy, as follows:

$$CE(y, \hat{y}) = -\sum_{i=1}^{n} y_i \cdot log(\hat{y}_i)$$

In every iteration, the forward computation ends when the loss function calculates the scalar (loss-node), which represents how well the model performs. Afterwards, begins the backward pass, in which the algorithm will compute all the gradients of the parameters with respect to the value of the loss-node. The backward pass is called back-propagation algorithm, and its usefulness is that it spreads the information of the cost to flow through the network (Goodfellow, et al., 2016). Back-propagation exploits the well-known chain-rule of calculus (Rumelhart, et al., 1986). By calculating the partial derivatives of the network's weights $\frac{\partial Loss}{\partial W_{i,j}}$ for any $W_{i,j}$, or biases, enables the algorithm to save the gradients in order that update the parameters in every next iteration. That is, in every loop, the parameters are continually updating their values in the direction, in which the gradient minimizes the loss (White, et al., 2018). In this point, we reach a new concept of neural networks, where specific algorithms use the stored gradients to optimize all the parameters of the network.

2.5.3 Optimization algorithms

As already mentioned, in order to obtain the optimum results for our training process, the neural network needs to minimize a loss function suited for the problem. Hence, backpropagation algorithm finds and saves the derivatives of the previous iteration. Since, the derivative of a differential function finds the slope of $f(x)$ for x (Goodfellow, et al., 2016), the neural network will know whether the loss function is decreasing or increasing, so in the next iteration move towards the correct direction. When $f'(x) = 0$, we know that there is now more improvement in our loss function, as shown in the *figure 15.* Though, this is not always the solution to the problem, that is we have not found the global minimum which is our goal. Those cases are called either local minimum or local maximum, and their values are known as called critical points. An answer to this puzzle, is to use the second derivatives of those critical points. By finding the slope of the derivative we will know the improvement of the gradient based on the previous steps, hence we can determine if we reached the global minimum. The matrix of the second derivatives is called Hessian matrix, and defined as follows (Goodfellow, et al., 2016):

$$\mathrm{H}(f)(x)_{i,j} = \frac{\partial^2}{\partial x_i \partial x_j} f(x)$$

Every next iteration, the neural network updates the parameters of the previous one. This method is known as gradient descent. This is achieved with the use of a constant, known as the learning rate. The formula, or the update rule is:

$$\vec{W}_{t+1} = \vec{W}_t - \eta \frac{\partial L}{\partial \vec{W}_t}$$

where, $\frac{\partial L}{\partial \vec{W}}$ is the gradient of the previous parameter (weight),
η is the learning rate.

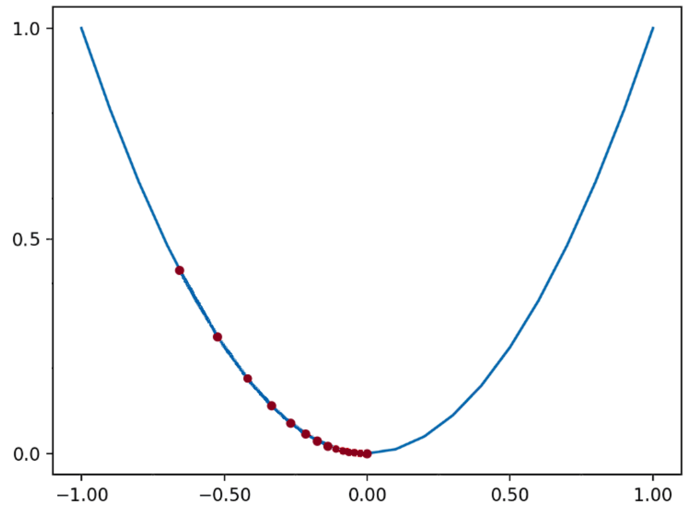

Figure 15. Plot of Gradient descent progress[39]

Most of the times simple gradient descent method cannot overcome situations such the one described above with local minimum or maximum. In addition, it is considered to be a slow algorithm, because it has to update in every iteration all parameters in the neural network.

Stochastic gradient descent (SGD) usually solves this problem. SGD works with the concept of batches. Batching is the procedure when the data are not fed directly in an iteration, but split in n batches (sets). So, in every iteration the update of the weights is done n times instead of one like gradient descent. Though, on the grounds of working with batches, SGD loses a little bit of accuracy, something which is inconsiderable in relation to the benefits we gain by using it.

Nowadays, optimization techniques and strategies are a lot and vary depending on the problem. We will discuss some of the most popular in the following section.

2.5.3.1 Momentum

Even though stochastic descent methods are considered to be effective solutions, constant learning rates that they use may present approximation issues. That is, low learning rates take an important amount of time to reach desired accuracies, while high learning rates may deviate from optimal solutions. In order to overcome this setback, the algorithm takes into account in the update rule a new parameter called momentum which is basically a kind of velocity (Goodfellow, et al., 2016). Momentum considers the value of the previous update step and calibrates the next update accordingly. Momentum update formula is (Aggarwal, 2018):

$$\vec{V}_{t+1} = \alpha \vec{V}_t - \eta \frac{\partial L}{\partial \overrightarrow{W}_t}$$

where, $\alpha \in (0,1)$ and it is the momentum value,
$\vec{V}$ is the velocity vector.

Then, the velocity vector contributes to the update rule, which is reformed as:

$$\overrightarrow{W}_{t+1} = \overrightarrow{W}_t + \vec{V}_{t+1}$$

---

[39] Machine Learning Mastery, Article: How to Implement Gradient Descent Optimization from Scratch, https://machinelearningmastery.com/gradient-descent-optimization-from-scratch/, last visit 17/12/2022

2.5.3.2 Adaptive gradient algorithm (AdaGrad)

AdaGrad (Duchi, et al., 2011) algorithm dynamically calibrates a learning rate for each parameter of neural network (Kochenderfer & Wheeler, 2019). Roughly, learning rates for weights with large derivative values decrease rapidly, whilst a relative smaller reduction provided for those with small derivative values (Goodfellow, et al., 2016). Formula is as follows:

$$S_{i+1} = S_i + \left(\nabla \mathrm{f}(\mathrm{w_i})\right)^2$$

And the update rule is:

$$w_{i+1} = w_i - \frac{\eta}{\sqrt{S_{i+1} + \epsilon}} \nabla(w_i)$$

where, ϵ is a small smoothing constant to avoid division with zero.

2.5.3.3 RMSProp

Root mean square propagation algorithm (RMSProp) is considered to be an extension of AdaGrad, though, it slightly differs regarding implementation. Its formula is:

$$S_{i+1} = \rho s_i + (1-\rho)\left(\nabla f(w_i)\right)^2$$

where, $\rho \in (0,1)$, the update rule is exactly the same as AdaGrad:

$$w_{i+1} = w_i - \frac{\eta}{\sqrt{S_{i+1} + \epsilon}} \nabla(w_i)$$

2.5.3.4 Adam

The adaptive moment estimation method (Adam) (Kingma & Ba, 2014) constitutes a combination of AdaGrad and RMSProp. Adam calibrates the learning rate in every iteration, and consists of 2 base parameters β_1 and β_2, where both $\in (0,1)$. β_1 generally, handles expectations of gradients, and the latter is the sum of squares of the gradients (Yadav, 2021). The algorithm contains four steps, from which the first two provide information for the gradient in the previous step v and similarly to RMSProp, sums the squares of the gradients, s. The last two steps refer to the correction of the bias, since Adam operates under a strong assumption, that is all gradients come from the same distribution. Adam goes as follows:

1st moment estimates without bias correction:

$$v_{i+1} = \beta_1 v_i + (1-\beta_1)\left(\nabla f(w_i)\right)$$

2nd moment estimates without bias correction:

$$s_{i+1} = \beta_2 s_i + (1-\beta_2)\left(\nabla f(w_i)\right)^2$$

Correction of bias for1st moment:

$$v_{i+1} = \frac{v_{i+1}}{1-\beta_1^i}$$

Correction of bias for 2nd moment:

$$s_{i+1} = \frac{s_{i+1}}{1-\beta_2^i}$$

Finally, the update rule is as follows:

$$w_{i+1} = w_i + \frac{\eta v_{i+1}}{\sqrt{s_{i+1}} + \delta}$$

where, δ is a constant to avoid division with zero.

2.5.4 Neural Network architectures

Neural networks are distinguished in specific types that specialize in different tasks. For example, a task that contains face recognition is applied with convolutional neural networks (CNN) or a task for machine translation would use recurrent neural networks (RNN). In the following sections we will have a brief overview of the most common neural network architectures and their uses.

2.5.4.1 Shallow Neural Networks

Generally, a shallow neural network (SNN) consists of one hidden layer. Some of its main uses, include machine learning models which we saw in previous sections, such as logistic regression, support vector machines. A typical architecture of an SNN we saw in *figure 12*.

2.5.4.2 Deep Neural Networks

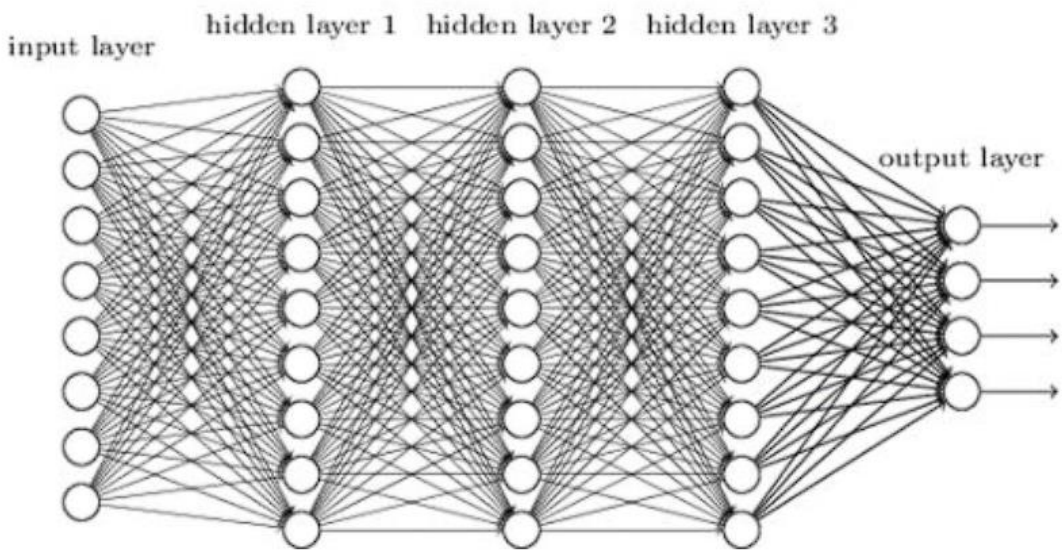

Figure 16. Deep Neural Network structure[40]

As Deep neural network (DNN) (also called multilayer perceptron) is considered a network with more than one hidden layer. Their use is for more sophisticated tasks since multiple layers add more complexity to the network. DNNs are divided in two categories, feedforward and backward depending on their flow process (Birjali, et al., 2021). In *figure 16*, we can see a typical DNN structure. Apart from the input layer where the features are fed into the network, it also consists of three hidden layers, fully connected with the nodes of the next and previous layer. As an output layer of the network, it uses a layer which has four nodes, namely four outputs.

[40]The Windows Club, Article: What is Deep Learning and Neural Network, https://www.thewindowsclub.com/deep-learning-and-neural-network, last visit 20/12/2022

2.5.4.3 Convolutional Neural Networks

Convolutional neural networks (CNN) were first invented for tasks containing computer vision (Lecun, et al., 1998), and constitute the main neural network choice regarding this field. Nowadays though, CNN's utilization is expanded also in natural language processing, with many novel papers now obtain promising results. In the paper by (Kim, 2014), is explained the way in which pre-trained word vectors are used for training a CNN for sentence-level classification. In its basic form, a CNN consists of multiple convolutional layers in its hidden layers, which are responsible for various tasks like, feature extraction, resolution reduction for feature independency from noise and insignificant changes (Birjali, et al., 2021).

2.5.4.4 Recurrent Neural Networks

Recurrent neural networks (RNN) are very popular option in sentiment analysis and natural language processing generally. The main leverage of this kind of neural networks is that they make use of information of the previous time step in order to use it as memory to estimate the current time step (Wankhade, et al., 2022). This feature is described as "memory" which gives RNN the ability to read inputs one after another, and remember the information passed from the previous layers. Hence, RNNs are categorized as a highly effective solution in sequence tasks, which are mainly documents are sentences in sentiment analysis. An example of RNN is shown in *figure 17*.

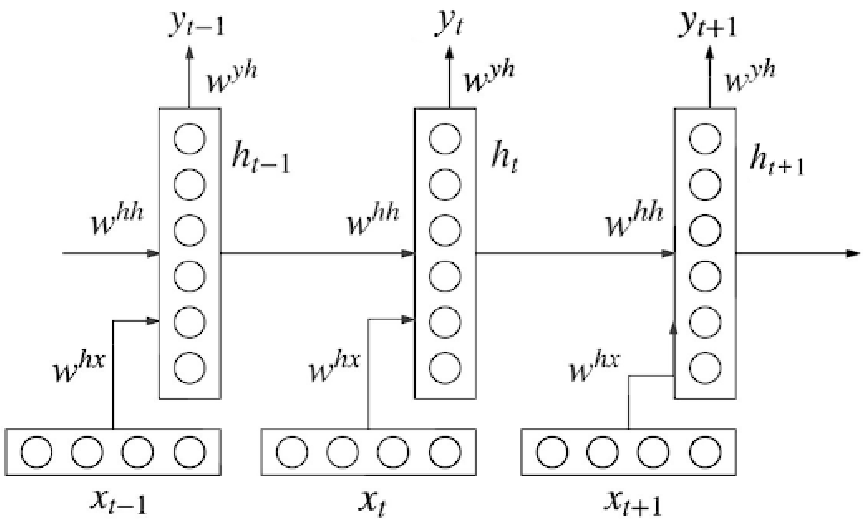

Figure 17. Typical RNN example (Zhang, et al., 2018)

In the above figure we observe three layers of a typical RNN example. Because RNNs work on sequences, the number of layers (time steps) are tied with the number of inputs (words in a sentence). The vector X_t is the input at time t, and likewise layer h_t is the hidden layer of the same time step. The inputs of h_t are based on the calculations of the previous layer and the input of the current X_t. The predictions y in a typical RNN are taking place every time step. According to (Zhang, et al., 2018) the equation of h_t for the current time step is as follows:

$$h_t = act(w^{hx}x_t + w^{hh}h_{t-1})$$

where, act is the activation function, which is in most cases tanh or ReLU.

Because recurrent neural networks are very deep, the network could face vanishing gradient or exploding gradients problems. This means that the derivatives either decrease or increase exponentially. A solution to this issue is Long Short Term Memory networks (LSTM). Those networks use gates in each node that determine which term will be outputted and which will not:

- Forget gate, which signs a value based on Bernoulli distribution and determines if the information will be kept or forgotten.
- Input gate that indicates which of the values will be updated.
- Output gate, which states the prediction of the current time step.

Chapter 3

3.1 Dataset Description

The dataset, called sentiment140, first presented by (Go, et al., 2009) consists of 1.600.000 tweets extracted from Twitter API. The target labels are 2, tweets with positive sentiment classification with the value of 4 and tweets with negative sentiment classification with value of 0. It contains 6 columns:

1. target: The sentiment of the tweet as described above.
2. ids: The id of the tweet.
3. date: Timestamp indicating when tweet is composed.
4. flag: The query.
5. user: The user composed the tweet.
6. text: The text of the tweet

3.1.1 Preprocessing Steps

In order to prepare the data accordingly, we proceeded to some preprocessing steps, in order that words obtain a formal structure. Initially, the first 4 columns of the dataset (ids, date, flag and user) are not needed so we omitted those and kept only target and text. Then, we used regular expressions to find and replace or remove patterns of words that are insignificant to the model. Such patterns may include, URL removal, remove punctuation, hashtag removal and more. We also tried to remove contractions and misspellings as much as possible and restore their formal way. In natural language, users tend to write a lot using contractions and abbreviations. Some examples below showing how we handled this situation. In the first aspect, a word presented in its initial form how we encountered it, while in the second the modification of it:

- wasn’t ⟶ was not
- weren’t ⟶ were not
- hasn’t / hasn’t ⟶ has not
- I’m / I’m / im ⟶ I am
- here’s ⟶ here is
- some1 ⟶ someone
- wanna ⟶ want to
- gonna ⟶ going to
- favourite ⟶ favorite
- ty ⟶ thank you
- you’re / youre ⟶ you are

The above process also repeated for slang words replacement, as some examples are shown:

- u ⟶ you
- plz / pls ⟶ please
- thx / thnx ⟶ thanks
- nah ⟶ no
- 2morrow ⟶ tomorrow
- 4got ⟶ forgot
- cud ⟶ could
- bday ⟶ birthday

Each tweet is tokenized during preprocessing and found the corresponding POS tags for each word. We used the POS tag in order that determine if the word “4” was meaning either the number or the preposition. If the meaning was the later, then I would change the number “4” to “for”. Same procedure was repeated also for word “2”. The preprocessing steps are described also in pseudocode in *table 2.*

Table 2. Pseudocode for Preprocessing function

ALGORITHM 1 DATA PREPROCESSING

procedure preprocess_tweet
    Input: $\mathrm{X} = x_1, x_2, \dots, x_n$ representing the tweets
    Output: Tweets preprocessed
        $[reg]$ ⟵ Define the regular expressions pattern list
        $[slang]$ ⟵ Define the slang word list
        $[punct]$ ⟵ Define the punctuation list
        $[stop\_words]$ ⟵ Define stop words list
        for $i = 1, \dots, n$ **do**
            if any pattern of $[reg] \in \boldsymbol{x_i}$
                x_i ⟵ actions to replace x_i
            if any word from $[slang] \in \boldsymbol{x_i}$
                $\boldsymbol{x_i}$ ⟵ replace with the proper word
            if any punctuation form $[punct] \in x_i$
                $\boldsymbol{x_i}$ ⟵ remove punctuation
            if any stop word $\in x_i$
                $\boldsymbol{x_i}$ ⟵ remove stop word
            $[\boldsymbol{tokens}]$ ⟵ Tokenize x_i
            Find the POS tag of every token of $[tokens]$
            Correct numeric tokens 4 and 2 if necessary based on the POS tags of $[tokens]$
            $[tweet\_preprocessed]$ ⟵ Stem words in $[tokens]$
        end for
        return $[tweet\_preprocessed]$

In *table 3*, is described the function for the correction of numerical representations of words “4” and “2”.

Table 3. Correction function of preposition words represented as numbers.

ALGORITHM 2 DATA PREPROCESSING

procedure correct_numeric_word
  Input: (POS_tags) Tuple with (tokens, corresponding POS tags)
  Output: $[Tweet_{c}orrected]$
    for $\boldsymbol{i = 1, \ldots, len(tokens)}$ AND $token$ IN $[tokens]$ **do**
      if $token = 4$
        Define if token is a PREPOSITION indeed
        $token \longleftarrow$ “for”
      if $\boldsymbol{token = 2}$
        Define if token is a PREPOSITION or ADVERB indeed
        $token \longleftarrow$ to
      $[\boldsymbol{tweet\_corrected}] \longleftarrow$ $\boldsymbol{append\ the\ corrected\ words}$
    end for
    return $[\boldsymbol{tweet\_corrected}]$

The platforms that we used for developing the code are Anaconda[41] and as an interface Jupyter Lab[42], which both are open-source distributions and suitable for data analysis. The programming language we used is Python 3, which provides a vast variety of tools and libraries regarding sentiment analysis and natural language processing more generally. More specifically, below we present the most important libraries used for this thesis:

- NLTK (Natural Language Toolkit)[43]:
  - NLTK is a substantial library for developing tasks such as data analysis, tagging, tokenization and most of the required preprocessing steps needed in general.
- Scikit Learn[44]
  - Scikit-learn is the commonly-held machine learning library. Most of the machine learning algorithms are supported by scikit-learn for supervised or unsupervised learning. In addition, it provides a variety of tools for model selection or analysis metrics.
- Numpy[45]
  - One of the most widely used libraries for numerical computations and mathematical functions. Numpy is very efficient, regarding the time it needed for operations in matrices and broadcasting.
- Pandas[46]
  - Pandas is the most popular library for manipulation of tabular data. It is regarded to be very fast and efficient and provides automatic tools for analysis in order to reduce the amount of surplus code.
- Tensorflow[47]
  - Tensorflow is an artificial intelligence library and provides developers with the required tools for developing and creating neural networks.

[41] https://www.anaconda.com/
[42] https://jupyter.org/
[43] https://www.nltk.org/
[44] https://scikit-learn.org/stable/
[45] https://numpy.org/
[46] https://pandas.pydata.org/
[47] https://www.tensorflow.org/

3.1.2 Exploratory Data Analysis

In this section, there was an effort to acknowledge the data and determine if the was any issue during the previous steps. After preprocessing steps, the words are divided into tokens.

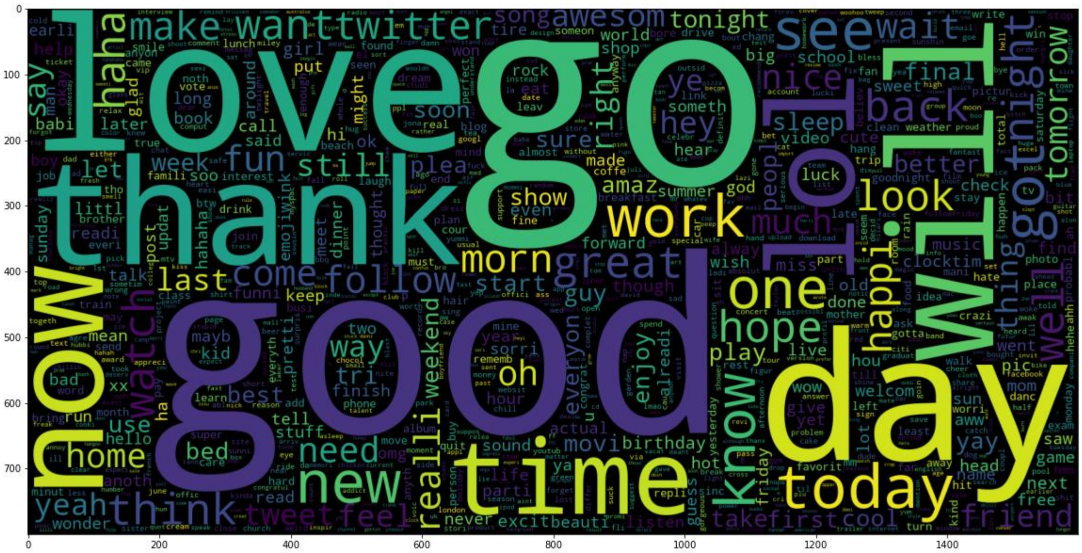

Figure 18. Words that appear mostly in positive label sentences.

In *figure 18,* we used Word Cloud[48] library in order to plot the words which appear mostly in sentences with positive sentiments. While in *figure 19,* we can see words that appear mostly in negative sentences. As we can observe there are words that appear in both tables, and they are the most common words.

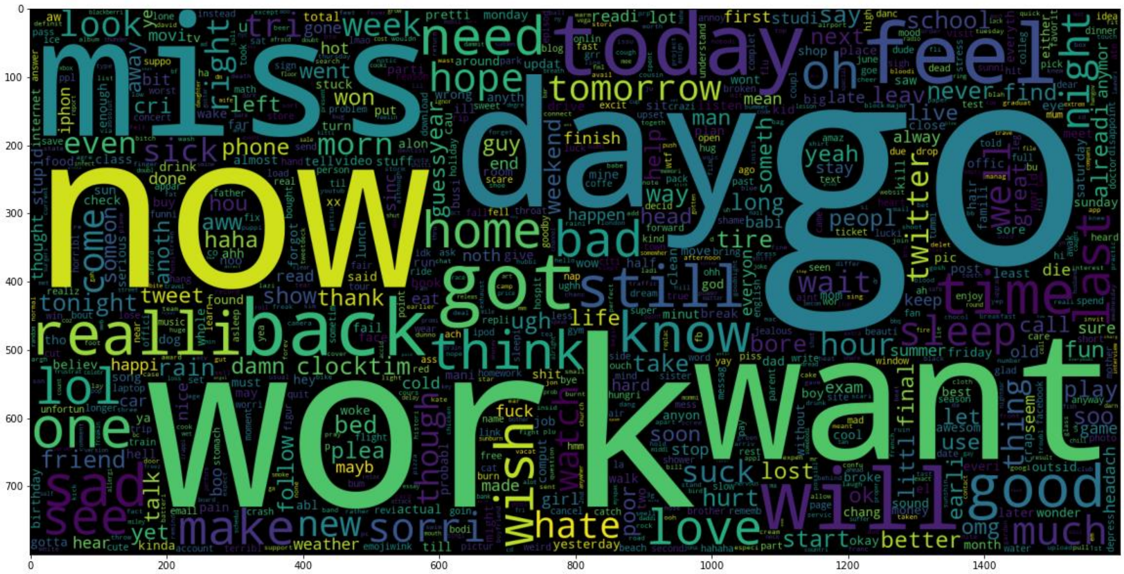

Figure 19. Words that appear mostly in positive label sentences.

[48] http://amueller.github.io/word_cloud/

Chapter 4

4.1 Feature extraction

Depending on the sentiment analysis approach, we used two kinds of word representations:

1. In machine learning-based approach words are represented with two features (columns). The first is the number each word occurs in a sentence with positive sentiment. The second is the number each word appears in a negative sentiment sentence. These word representations are used for probabilistic approaches.
2. In deep learning-based approach, we used word embedding from both word2vec technique and GloVe.

4.2 Experimental Results

In this section we discuss the experimental results from the experiments divided in two categories, according to the approach. Firstly, we used machine learning-based approach and implemented the following algorithms:

- Logistic Regression
- Multinomial Naïve Bayes

4.2.1 Machine learning results

4.2.1.1 Logistic regression implementation

Our first experiment is deployed using logistic regression. Below, we can see the classification report from the results. As we can observe, logistic regression implementation performed better regarding tweets labeled as positive than the negative ones. The accuracy of this implementation scored 71%.

LOGISTIC REGRESSION

	Precision	Recall	F1-score	Accuracy
0	0.74	0.66	0.70	
1	0.69	0.77	0.73	
				0.71

4.2.1.2 Multinomial naïve Bayes implementation

In this experiment we deployed Multinomial Naïve Bayes algorithm. It is a variation of classic naïve Bayes algorithm and is suitable for multinomially distributed data. Multinomial naïve Bayes requires also a smoothing parameter α, which prevents computations when probabilities are zero.

The formula is as follows:

$$\hat{\theta}_{yi} = \frac{N_{yi} + \alpha}{N_y + \alpha n}$$

where, N_{yi} is how many times i appears in class y
N_y is total count of all features in y
α smoothing parameter.

When parameter α is set to 1 the smoothing technique is called Laplace smoothing. Below we present the classification reports for multinomial naïve Bayes. In contrast with logistic regression, multinomial naïve Bayes algorithm performs slightly better in negative labeled tweets. Though, the overall accuracy of this model is 77%, which indicates that it performs sufficiently better than the logistic regression.

MULTINOMIAL NAÏVE BAYES				
	Precision	Recall	F1-score	Accuracy
0	0.76	0.79	0.78	
1	0.78	0.76	0.77	
				0.77

4.2.2 Deep learning results

For deep learning-based approach we followed the baseline preprocessing steps as we did for machine learning approach. Though, we trained each model with different word representations. Overall, we developed 8 different models to evaluate the training procedures and the total accuracies. For the implementation of the first two models, we used word2vec of 100 dimensions. For the rest 6 models, we used GloVe with various dimensions though, in order to estimate the behavior of each model.

Furthermore, regarding the architectures of the models, apart from LSTM layers, which described above, we used also:

- Bidirectional Layer
  - A bidirectional layer connects two hidden layers, with the input sequence of the first layer is in the opposite direction of the second. This is an efficient way for a neural network to gain more information. That is, we need the whole sequence in order to use a bidirectional layer, which means it is not suitable for speech recognition.
- Dropout layer
  - Dropout layer is used for regularization of the neural network. In order to prevent overfitting a dropout layer sets a percentage of the input units to 0 randomly. A variation of a classic dropout layer is Spatial dropout layer which is suitable for convolutional networks.
- GlobalMaxPooling1D layer
  - GlobalMaxPooling1D layer outputs the maximum value as vector.

Below, there is a comprehensive presentation of the methodologies used for the experiments:

Experiment 1:

Proportion of the data used for the experiment:	100%
Word embedding	Word2vec
Word embedding dimensions	100
Neural network architecture	• Embedding layer • Dropout 50% of units • LSTM with 100 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	86

Experiment 2:

Proportion of the data used for the experiment:	50%
Word embedding	Word2vec
Word embedding dimensions	100
Neural network architecture	• Embedding layer • Dropout 50% of units • LSTM with 100 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	42

Experiment 3:

Proportion of the data used for the experiment:	70%
Word embedding	GloVe
Word embedding dimensions	50
Neural network architecture	• Embedding layer • Spatial Dropout 1D • Conv1D • Bidirectional LSTM with 64 units • Dense layer with 512 units • Dropout 50% of units • Dense layer with 512 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	34

Experiment 4:

Proportion of the data used for the experiment:	70%
Word embedding	GloVe
Word embedding dimensions	100
Neural network architecture	• Embedding layer • Spatial Dropout 1D • Conv1D • Bidirectional LSTM with 64 units • Dense layer with 512 units • Dropout 50% of units • Dense layer with 512 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	36

Experiment 5:

Proportion of the data used for the experiment:	70%
Word embedding	GloVe
Word embedding dimensions	200
Neural network architecture	• Embedding layer • Spatial Dropout 1D • Conv1D • Bidirectional LSTM with 64 units • Dense layer with 512 units • Dropout 50% of units • Dense layer with 512 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	37

Experiment 6:

Proportion of the data used for the experiment:	70%
Word embedding	GloVe
Word embedding dimensions	300
Neural network architecture	• Embedding layer • Spatial Dropout 1D • Conv1D • Bidirectional LSTM with 64 units • Dense layer with 512 units • Dropout 50% of units • Dense layer with 512 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	43

Experiment 7:

Proportion of the data used for the experiment:	100%
Word embedding	GloVe
Word embedding dimensions	200
Neural network architecture	• Embedding layer • Spatial Dropout 1D • Conv1D • Bidirectional LSTM with 64 units • Dense layer with 512 units • Dropout 50% of units • Dense layer with 512 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	55

Experiment 8:

Proportion of the data used for the experiment:	100%
Word embedding	GloVe
Word embedding dimensions	200
Neural network architecture	• Embedding layer • Bidirectional LSTM with 64 units • Conv1D • GlobalMaxPooling1D • Dense layer with 16 units • Dense layer with 1 output unit
Epochs	10
Time elapsed for training in minutes	111

The epochs are chosen to be 10 for all the experiments to maintain equality for the experiments. Also, we selected various proportions of the data, to check how is affects training, time and overall accuracies. The last rows in the experiment tables show the time needed for training each model.

In the following table we show the overall training losses for each experiment:

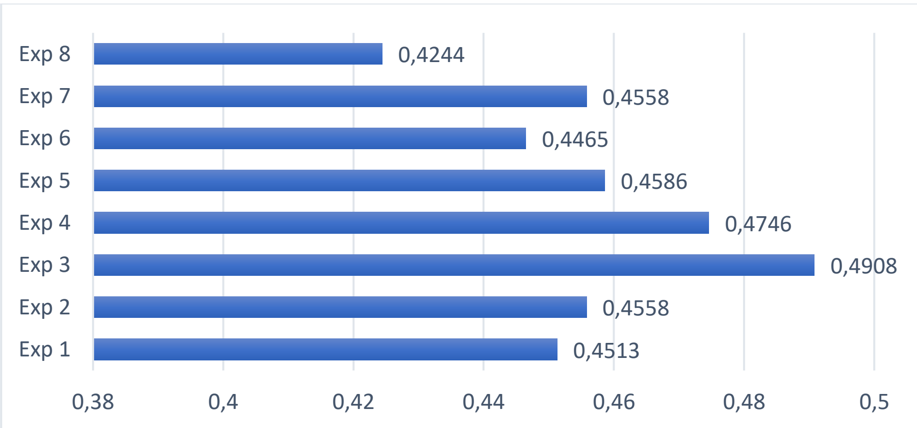

Figure 20. Training losses for all the experiments

Likewise, below we can see the improvement of training and validation accuracies of the experiments every epoch.

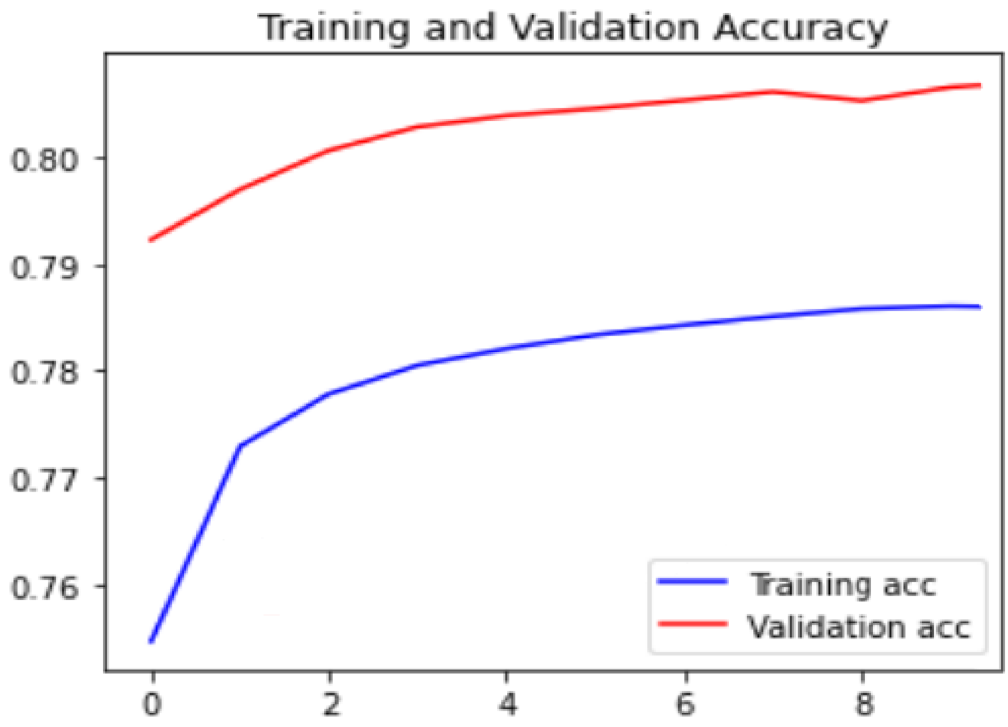

Figure 21.Training and validation accuracy for experiment 1

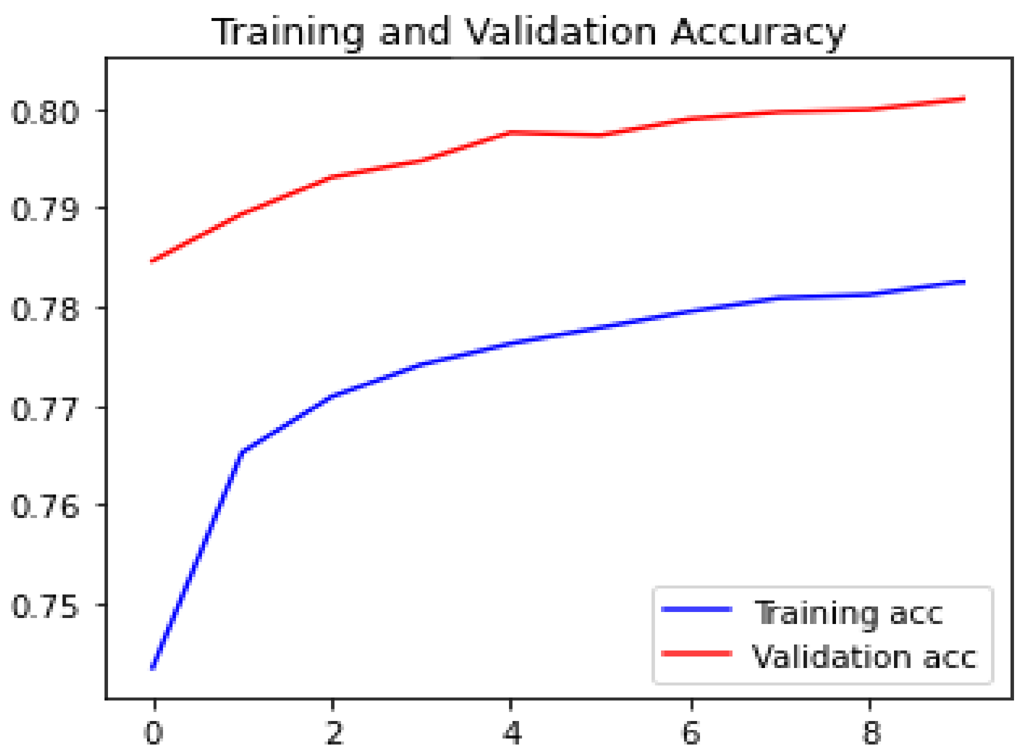

Figure 22. Training and validation accuracy for experiment 2

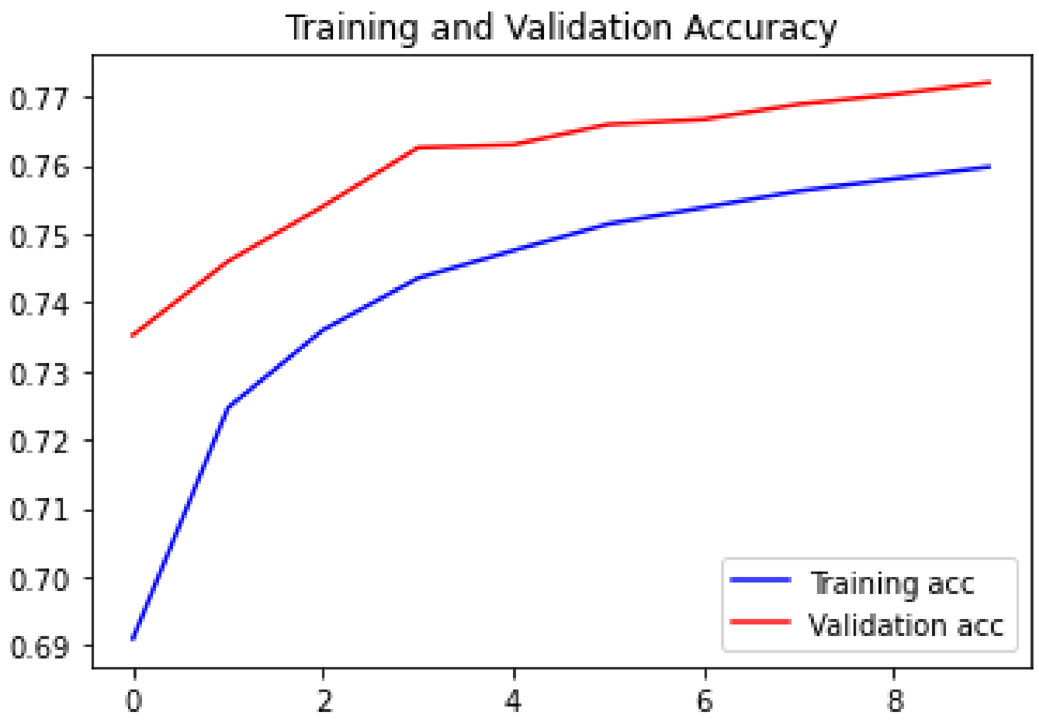

Figure 23. Training and validation accuracy for experiment 3

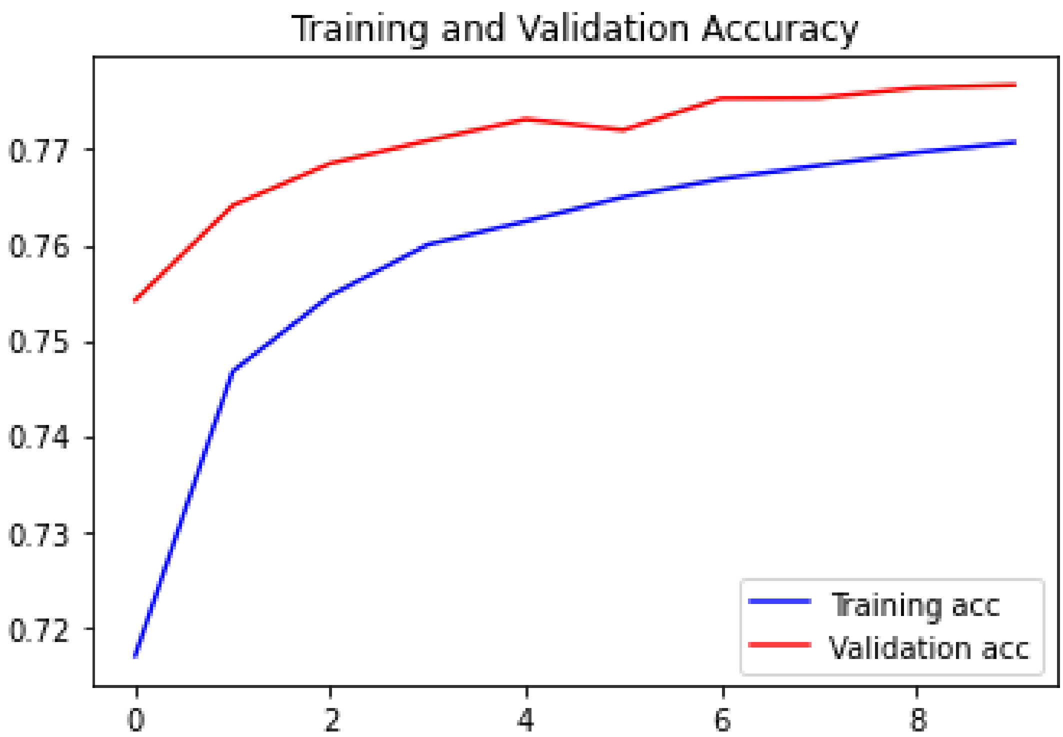

Figure 24. Training and validation accuracy for experiment 4

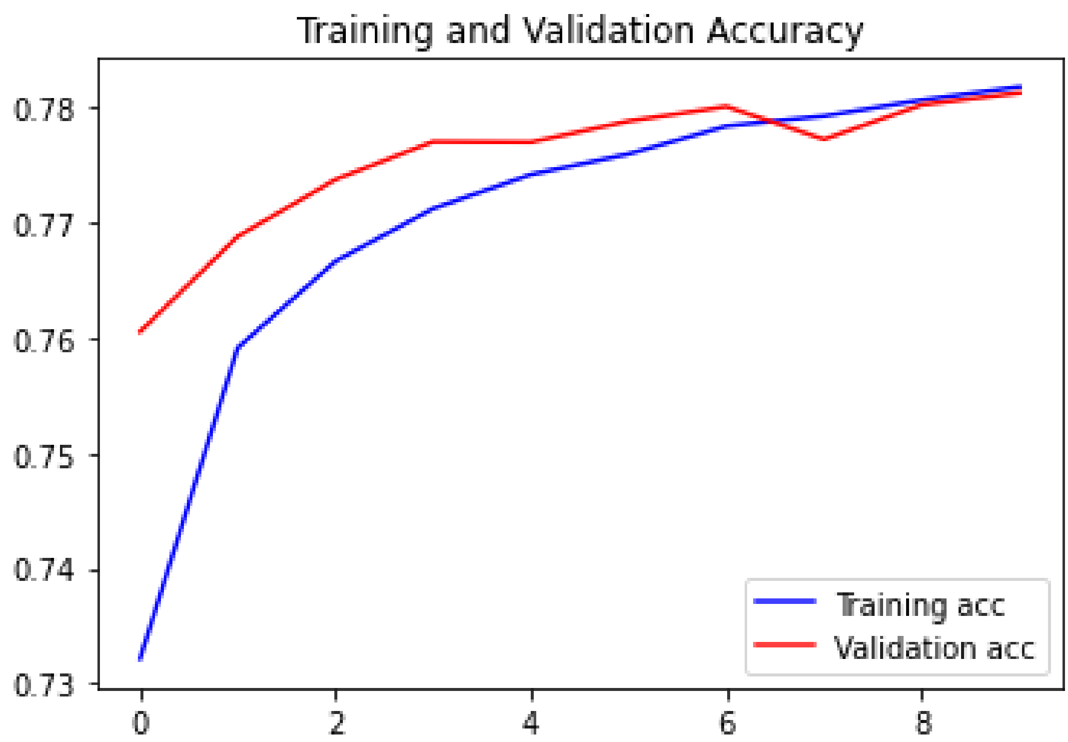

Figure 25. Training and validation accuracy for experiment 5

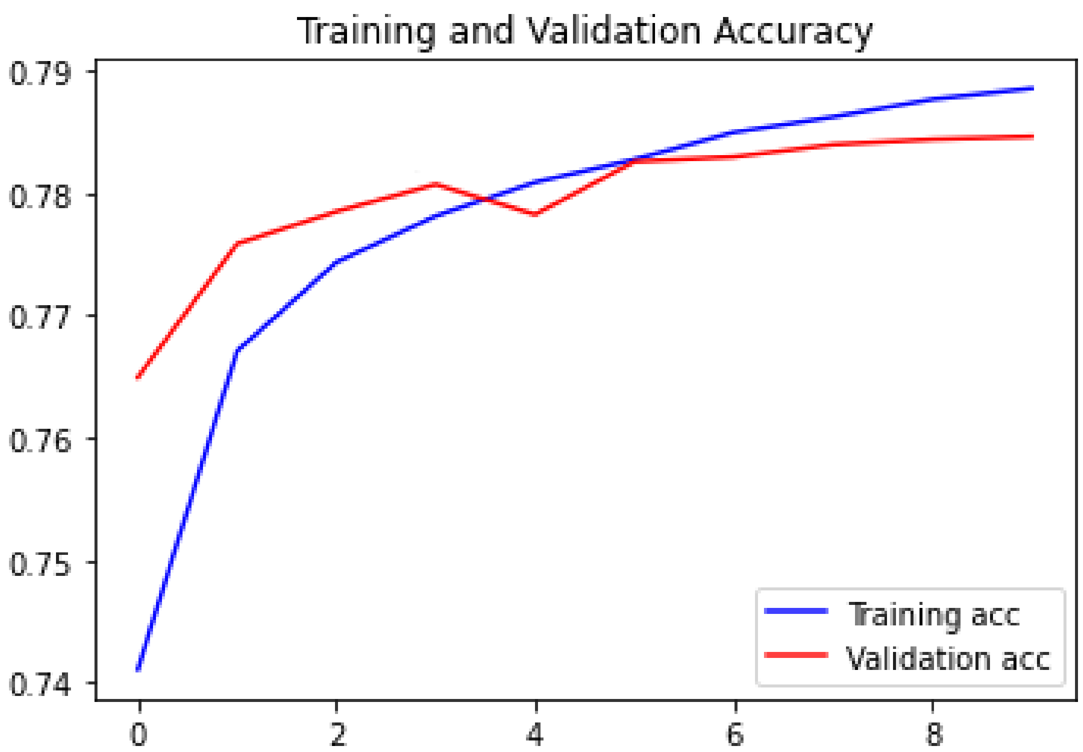

Figure 26. Training and validation accuracy for experiment 6

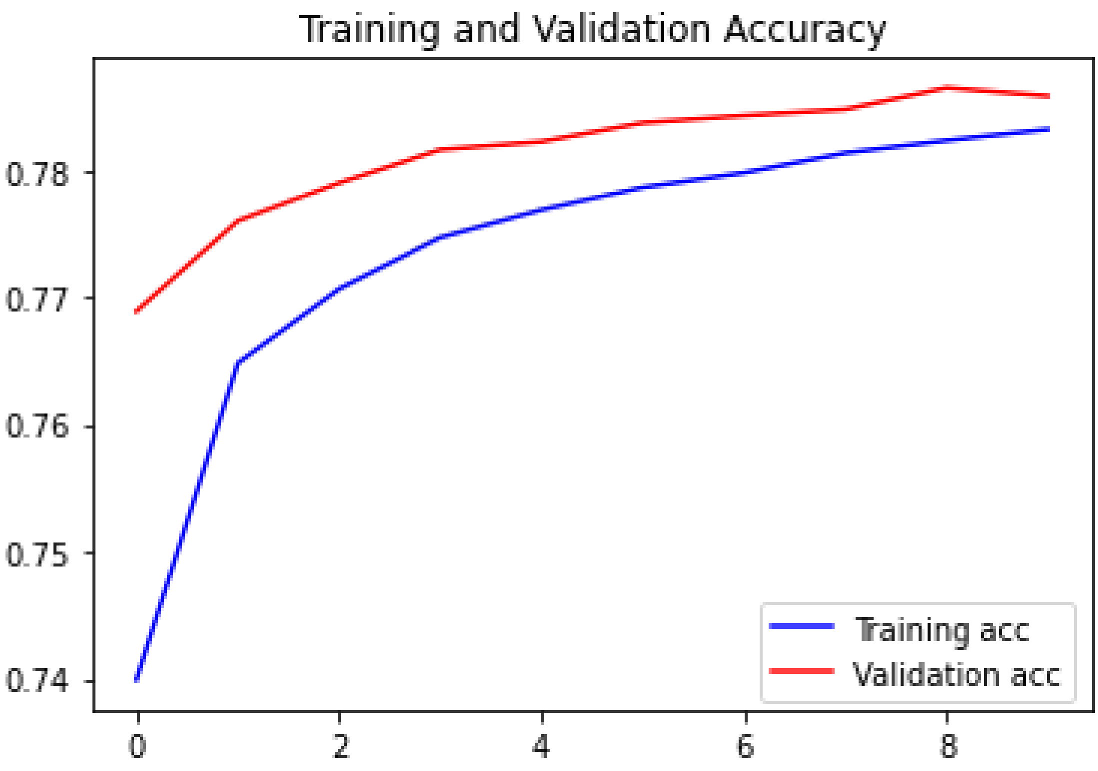

Figure 27. Training and validation accuracy for experiment 7

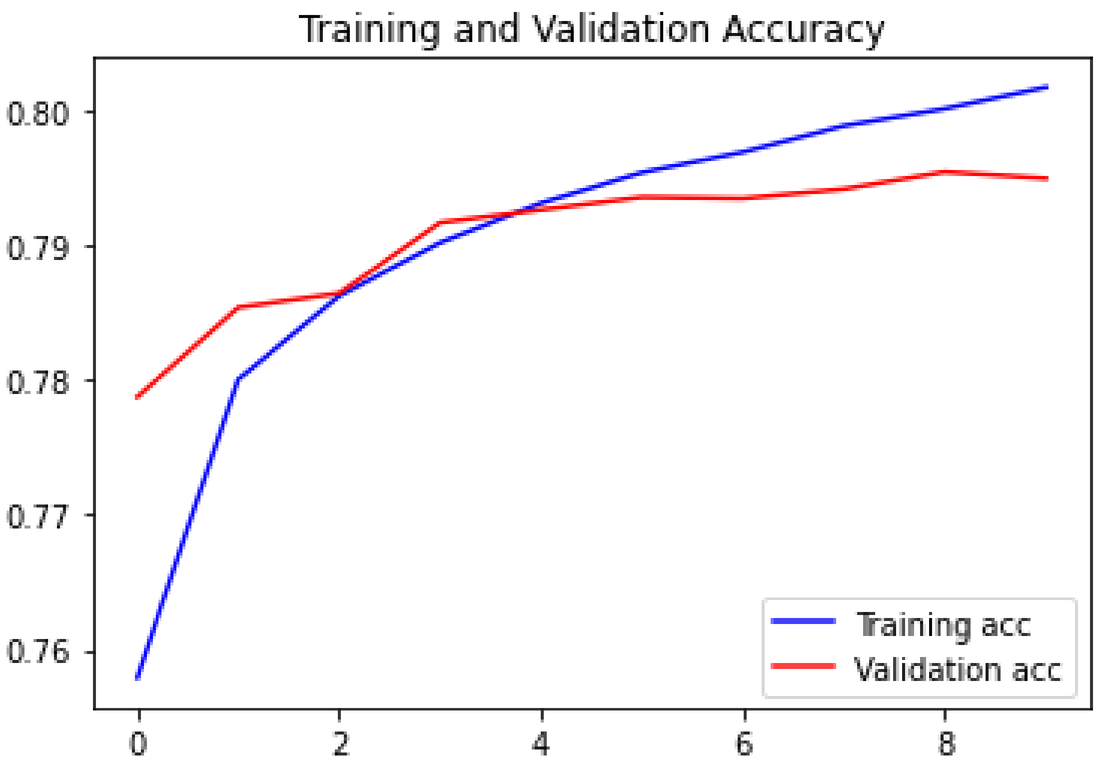

Figure 28. Training and validation accuracy for experiment 8

In the following chart, we can observe the models' accuracies evaluated in test data:

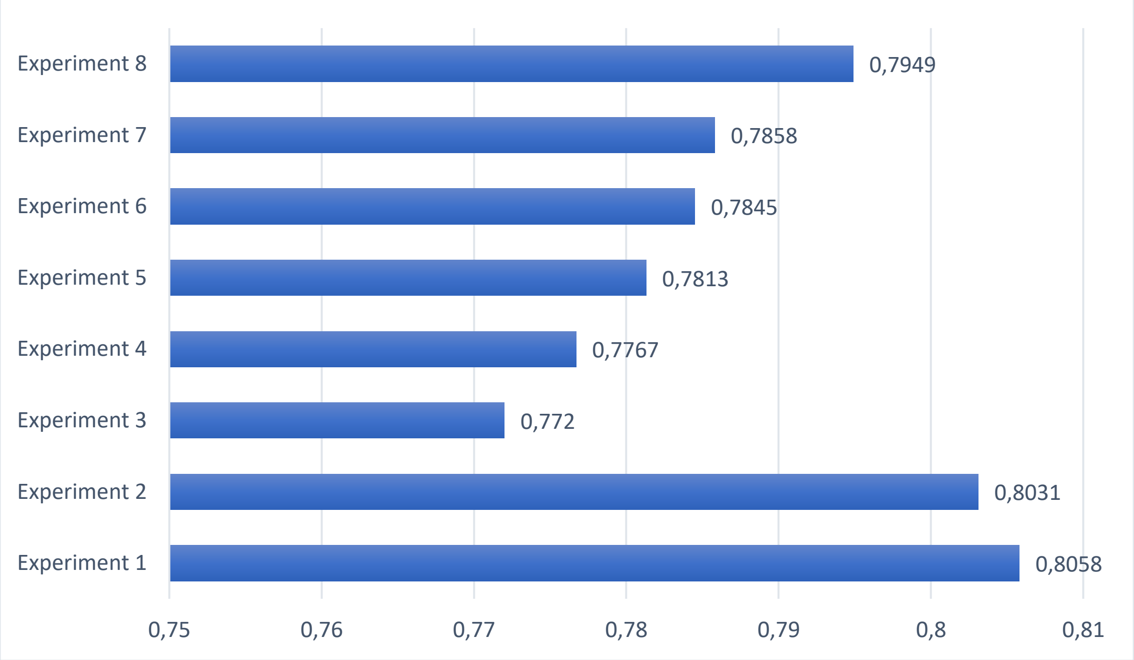

Figure 29. Predicted Accuracies

As we can observe, the first two models which were developed with word2vec embeddings, perform better for this particular training data, scoring 0.8058 and 0.8031 respectively (*figure 29*). The accuracies of those 2 experiments are remarkably close, despite the fact that the 2nd experiment significantly reduced the elapsed time to the half from the 1st. In addition, the training and validation accuracy curve for the 2nd experiment indicates that it could absorb more training, while in the 1st experiment the curve seems to be flattened after the 6th epoch.

Models developed with GloVe embeddings (3rd, 4th, 5th, and 6th experiment) achieve gradually increments in their predicting accuracies, depending on the dimensions chosen for each word embeddings. In experiment 3, in which we chose a 50-dimension GloVe, we observe an accuracy score of 0.772, which is significantly less than the other models. In the contrary, a GloVe embeddings with 300-dimensions, in experiment 6 predicted with 0.7845 accuracy.

The last two experiments 7 and 8, developed with the 200-dimensions Glove embeddings. Experiment 7 has the same network architecture and hyperparameters as experiment 5, in which we used the same GloVe embedding size. They differentiate though, in the size of the data used. Experiment 7 scored 0.7858 accuracy while experiment 5 scored 0.7767, however experiment's 7 elapsed time for training was 1.49 times longer than experiment's 5. A different network architecture was selected for experiment 8. In particular, we omitted Dropout layers and reduced the number of Dense layers to prevent overfitting. The accuracy it scored is 0.7949, but acquired the lowest training loss, 0.4244 (*figure 20*) compared with all the experiments.

Chapter 5

5.1 Conclusion and Future Work

In the first chapters of this thesis, we discussed the problem of sentiment analysis, specifically on Twitter. We elaborated on the most significant current applications and the fundamental concepts regarding the theoretical background for developing such applications. We presented several types of sentiment analysis approaches along with the levels of analysis. Also, some of the most popular preprocessing tools and techniques for efficient word representations are presented, in chapter 2. Furthermore, we discussed in excess the machine learning approaches and the most well-known algorithms used for sentiment analysis, some of which we deployed for our experiments. Then, we introduced deep learning approaches, and presented some of the key factors to build an efficient neural network. Lastly, we evaluated the performance of various experiments on sentiment analysis using machine and deep learning approaches.

During the last years we have witnessed a massive interest in natural language processing and especially sentiment analysis. The papers and also the applications are constantly updating and bring the task of sentiment analysis to a new level every day. In this thesis I have tried to present an overall approach to sentiment analysis ranging from a general perspective to a more sophisticated presentation. We discussed the main approaches to sentiment analysis, accompanied with a brief introduction to the domains in which sentiment analysis find implementations. Definitely, SA is a field that is studied deeply and profoundly and undoubtedly it could not cover all of its aspects. Although, the presentation’s purpose is the clarification of its fundamental notions and unpacking some approaches in order to have a better understanding of the implementations.

Definitely, there are more to come in the field of sentiment analysis despite the fact that it is in the eye of research a lot through the years. The future of sentiment analysis is bright, while innovative ideas and applications in business industry are yet to be implemented. For example, employee burnout is a very common obstacle in the progress of business teams nowadays. Sentiment analysis automates procedures and help to overcome such issues. In addition, a field that is in growth, is the artificially developed emotions. Mental health treatment could be promising knowing the fact that people are tend to keep their feeling inside them and express them mainly via online chats.

Without any doubt, there is a variety of implementations in sentiment analysis, and in conjunction to the constant need of reducing the time needed for concluding projects, automated systems will also be in front line. And this is why sentiment analysis will be there to fill the gap in any field that may arise.

Bibliography

Aggarwal, C. C., 2018. *Neural Networks and Deep Learning - A Textbook.* s.l.:Springer.

Ahmad, A., 2010. Is Twitter a useful tool for journalists?. *Journal of Media Practice*, 08, Volume 11, pp. 145-155.

Aich, S., Choi, K.-W. & Kim, H.-C., 2017. An Approach to Investigate the Impact of Political Change on the Economy of South Korea Using Twitter Sentiment Analysis. *Advanced Science Letters*, 10, Volume 23.

Aisopos, F., Papadakis, G. & Varvarigou, T., 2011. Sentiment Analysis of Social Media Content Using N-Gram Graphs. In: *Proceedings of the 3rd ACM SIGMM International Workshop on Social Media.* Scottsdale, Arizona, USA: Association for Computing Machinery, p. 9–14.

Ambedkar, K., Sukomal, P. & Rajendra, P., 2019. Text summarization from legal documents: a survey. *Artificial Intelligence Review*, 03, Volume 51.

Arista, M. et al., 2016. NerthusV3. Online Lexical Database of Old English. In: *Nerthus Project.* s.l.:Universidad de La Rioja.

Baccianella, S., Esuli, A. & Sebastiani, F., 2010. {S}enti{W}ord{N}et 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation ({LREC}'10).* Valletta, Malta: European Language Resources Association (ELRA).

Baeza-Yates, R. A. & Ribeiro-Neto, B. A., 1999. *Modern Information Retrieval.* s.l.:ACM Press / Addison-Wesley.

Bartell, B. T., Cottrell, G. W. & Belew, R. K., 1992. Latent Semantic Indexing is an Optimal Special Case of Multidimensional Scaling. In: *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* s.l.:Association for Computing Machinery, p. 161–167.

Behdenna, S., Barigou, F. & Belalem, G., 2016. Sentiment Analysis at Document Level. In: *Smart Trends in Information Technology and Computer Communications.* Singapore: Springer Singapore, pp. 159--168.

Bethard, S. et al., 2004. *Automatic Extraction of Opinion Propositions and their Holders,* s.l.: s.n.

Bhatia, P., Ji, Y. & Eisenstein, J., 2015. Better Document-level Sentiment Analysis from {RST} Discourse Parsing. In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.* s.l.:Association for Computational Linguistics, pp. 2212--2218.

Birjali, M., Kasri, M. & Beni-Hssane, A., 2021. A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, Volume 226, p. 107134.

Blumenthal, S., 1982. *The Permanent Campaign.* New York: Simon and Schuster.

Bollen, J., Mao, H. & Zeng, X., 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, Volume 2, pp. 1-8.

Brown, N., 2002. The Permanent Campaign and Its Future. *Rhetoric & Public Affairs*, 01, Volume 5, pp. 549-551.

Burbach, L., Halbach, P., Ziefle, M. & Calero Valdez, A., 2020. Opinion Formation on the Internet: The Influence of Personality, Network Structure, and Content on Sharing Messages Online. *Frontiers in Artificial Intelligence*, Volume 3.

Cambria, E., Das, D., Bandyopadhyay, S. & Feraco, A., 2017. Affective Computing and Sentiment Analysis. In: *A Practical Guide to Sentiment Analysis.* s.l.:Springer International Publishing, pp. 1--10.

Cambria, E., Havasi, C. & Hussain, A., 2012. SenticNet 2: A semantic and affective resource for opinion mining and sentiment analysis. *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25*, 01, pp. 202-207.

Chaudhuri, A., 2006. Emotion and Reason in Consumer Behavior.

Chenlo, J. M. & Losada, D. E., 2014. An empirical study of sentence features for subjectivity and polarity classification. *Information Sciences*, Volume 280, pp. 275-288.

Chen, Y.-R. & Chen, H.-H., 2015. Opinion Spam Detection in Web Forum: A Real Case Study. In: *Proceedings of the 24th International Conference on World Wide Web.* s.l.:International World Wide Web Conferences Steering Committee, p. 173–183.

Choi, Y., Cardie, C., Riloff, E. & Patwardhan, S., 2005. Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing.* Vancouver(British Columbia): Association for Computational Linguistics, pp. 355-362.

Church, K. & Hanks, P., 1991. Word association norms, mutual information and lexicography. *Computational Linguistics*, 01, Volume 16, pp. 22-29.

Cortes, C. & Vapnik, V., 1995. Support-Vector Networks. *Machine Learning*, 09, 20(3), p. 273–297.

Crawford, M. et al., 2015. Survey of review spam detection using machine learning techniques. *Journal of Big Data*, 10, Volume 2, p. 23.

Dang, N. C., Moreno-García, M. N. & De la Prieta, F., 2020. Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, 9(3).

Dang, N. C., Moreno-García, M. N. & De la Prieta, F., 2020. Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*, 9(3).

Darwich, M., Mohd Noah, . S. A., Omar, N. & Osman, N., 2019. Corpus-Based Techniques for Sentiment Lexicon Generation: A Review. *Journal of Digital Information Management*, 10, Volume 17, p. 296.

Das, S. & Chen, M., 2007. Yahoo! For Amazon: Sentiment Extraction from Small Talk on the. In: *Management Science.* s.l.:s.n.

Deerwester, S. et al., 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), pp. 391-407.

Deng, L. & Wiebe, J., 2015. MPQA 3.0: An Entity/Event-Level Sentiment Corpus. *Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies*.

Duchi, J., Hazan, E. & Singer, Y., 2011. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.*, jul, Volume 12, p. 2121–2159.

Enas, K., Enas, E. H. & Hoda, M., 2021. Sentiment Analysis Tasks and Approaches. *International Journal of Computer Science and Information Security*, 11, Volume 19.

Farias, H. D. & Rosso, P., 2017. Irony, Sarcasm, and Sentiment Analysis. In: *Sentiment Analysis in Social Networks.* 113-128: Morgan Kaufmann, pp. 113-128.

Firmino Alves, A. L. et al., 2014. A Comparison of SVM Versus Naive-Bayes Techniques for Sentiment Analysis in Tweets: A Case Study with the 2013 FIFA Confederations Cup. In: *Proceedings of the 20th Brazilian Symposium on Multimedia and the Web.* s.l.:Association for Computing Machinery, p. 123–130.

Firth, J. R., 1957. A synopsis of linguistic theory 1930-55. In: *Studies in Linguistic Analysis (Special Volume of the Philological Society).* Oxford: The Philological Society, pp. 1-32.

Fletcher, T., 2009. *Support Vector Machines Explained,* s.l.: s.n.

Go, A., Bhayani, R. & Huang, L., 2009. *Twitter sentiment classification using distant supervision.* s.l.:CS224N Project Report, Stanford.

Goldberg, Y., 2015. *A Primer on Neural Network Models for Natural Language Processing,* s.l.: arXiv.

Goldberg, Y. & Hirst, G., 2017. *Neural Network Methods for Natural Language Processing.* s.l.:Morgan & Claypool Publishers.

Golub, G. H. & Van Loan, C. F., 1996. *Matrix Computations.* Third ed. s.l.:The Johns Hopkins University Press.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. *Deep Learning.* s.l.:MIT Press.

Guess, A. M. & Lyons, B. A., 2020. Misinformation, Disinformation, and Online Propaganda. In: *Social Media and Democracy: The State of the Field, Prospects for Reform.* Cambridge: Cambridge University Press, p. 10–33.

Hajmohammadi, M. S., Ibrahim, R. & Selamat, A., 2015. Graph-Based Semi-supervised Learning for Cross-Lingual Sentiment Classification. In: *Intelligent Information and Database Systems.* Cham: Springer International Publishing, pp. 97--106.

Hatzivassiloglou, V. & McKeown, K. R., 1997. Predicting the Semantic Orientation of Adjectives. In: *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the {E}uropean Chapter of the Association for Computational Linguistics.* s.l.:Association for Computational Linguistics, pp. 174--181.

Hemmatian, F. & Sohrabi, M., 2019. A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 10, Volume 52.

Hu, M. & Liu, B., 2004. Mining and Summarizing Customer Reviews. In: *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Seattle, WA, USA: Association for Computing Machinery, p. 168–177.

Ilyas, S. H. W. et al., 2020. Analyzing Brexit's Impact Using Sentiment Analysis and Topic Modeling on Twitter Discussion. In: *The 21st Annual International Conference on Digital Government Research.* Seoul, Republic of Korea: Association for Computing Machinery, p. 1–6.

James, G., Witten, D., Hastie, T. & Tibshirani, R., 2014. *An Introduction to Statistical Learning: with Applications in R.* s.l.:Springer New York.

Jindal, N. & Liu, B., 2007. Review Spam Detection. In: *Proceedings of the 16th International Conference on World Wide Web.* s.l.:Association for Computing Machinery, p. 1189–1190.

Jurafsky, D. & James H., M., 2009. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition.* 2 ed. s.l.:Prentice Hall, Pearson Education International.

Jurafsky, D. & Martin, . J., 2020. *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition.* 3rd Edition draft ed. s.l.:s.n.

Jurek, A., Mulvenna, M. & Bi, Y., 2015. Improved lexicon-based sentiment analysis for social media analytics. *Secur Inform*, 4(9).

Kalyani, J., Bharathi, N. & Jyothi, R., 2016. Stock Trend Prediction Using News Sentiment Analysis. *International Journal of Computer Science and Information Technology*, 06, Volume 8, pp. 67-76.

Kamal, A., 2013. *Subjectivity Classification using Machine Learning Techniques for Mining Feature-Opinion Pairs from Web Opinion Sources,* s.l.: arXiv.

Khatua, A., Khatua, A. & Cambria, E., 2020. Predicting political sentiments of voters from Twitter in multi-party contexts. *Applied Soft Computing*, Volume 97, p. 106743.

Kim, S.-M. & Hovy, E., 2004. Determining the Sentiment of Opinions. In: *Proceedings of the 20th International Conference on Computational Linguistics.* Geneva, Switzerland: Association for Computational Linguistics, p. 1367–es.

Kim, S.-M. & Hovy, E., 2004. Determining the Sentiment of Opinions. In: *Proceedings of the 20th International Conference on Computational Linguistics.* s.l.:Association for Computational Linguistics, p. 1367–es.

Kim, Y., 2014. Convolutional Neural Networks for Sentence Classification. In: Doha, Qatar: Association for Computational Linguistics, pp. 1746--1751.

Kingma, D. & Ba, J., 2014. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, 12.

Kiritchenko, S., Zhu, X. & Mohammad, S., 2014. Sentiment Analysis of Short Informal Text. *The Journal of Artificial Intelligence Research (JAIR)*, 08, Volume 50.

Kochenderfer, . M. J. & Wheeler, T. A., 2019. *Algorithms for Optimization.* s.l.:The MIT Press.

Koukaras, P., Nousi, C. & Tjortjis, C., 2022. Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning. *Telecom*, 3(2), pp. 358--378.

Kulkarni, D. S. & Rodd, S. S., 2021. Sentiment Analysis in Hindi—A Survey on the State-of-the-Art Techniques. *CM Trans. Asian Low-Resour. Lang. Inf. Process.*, 11, Volume 22.

Kumar, A. & Sebastian, T., 2012. Sentiment Analysis: A Perspective on its Past, Present and Future. *International Journal of Intelligent Systems and Applications*, 09, Volume 4.

Kusen, E., Stremberck, M. & Conti, M., 2019. Emotional Valence Shifts and User Behavior on Twitter, Facebook, and YouTube. In: *Influence and Behavior Analysis in Social Networks and Social Media.* s.l.:Springer.

Landauer, T., Foltz, P. & Laham, D., 1998. An Introduction to Latent Semantic Analysis. *Discourse Processes*, 01, Volume 25, pp. 259-284.

Landauer, . T. K. & Dumais, S. T., 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition induction, and representation of knowledge. *Psychological Review*, 104(2), p. 211–240.

Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P., 1998. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 12, Volume 86, pp. 2278 - 2324.

Lerman, K., Blair-Goldensohn, S. & McDonald, R., 2009. Sentiment Summarization: Evaluating and Learning User Preferences. In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics.* USA: Association for Computational Linguistics, p. 514–522.

Ligthart, A., Catal, C. & Tekinerdogan, B., 2021. Systematic reviews in sentiment analysis: a tertiary study. *Artificial Intelligence Review*, 54(7), pp. 4997-5053.

Lilleker, D., Pack, M. & Jackson, N., 2010. Political Parties and Web 2.0: The Liberal Democrat Perspective. *Politics*, 06, Volume 30, pp. 105 - 112.

Liu, B., 2010. Sentiment Analysis and Subjectivity. In: *Handbook of Natural Language Processing.* s.l.:s.n.

Liu, B., 2012. Opinion Search and Retrieval. In: *Sentiment Analysis and Opinion Mining.* Cham: Springer International Publishing, pp. 107-111.

Liu, B., 2012. Sentiment Analysis: A Fascinating Problem. In: *Sentiment Analysis and Opinion Mining.* s.l.:Springer International Publishing, pp. 1--8.

Liu, B., 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.* s.l.:s.n.

Liu, B., 2020. Introduction. In: *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions.* s.l.:Cambridge University Press, pp. 1-17.

Liu, W., Zhang, L., Tao, D. & Cheng, J., 2018. Reinforcement online learning for emotion prediction by using physiological signals. *Pattern Recognition Letters*, Volume 107, pp. 123-130.

Lu, B., Ott, M., Cardie, C. & Tsou, B. K., 2011. Multi-aspect Sentiment Analysis with Topic Models. In: *2011 IEEE 11th International Conference on Data Mining Workshops.* s.l.:s.n., pp. 81-88.

Majumder, N. et al., 2019. *Sentiment and Sarcasm Classification with Multitask Learning,* s.l.: arXiv.

Malik, V. & Kumar, A., 2018. Communication. Sentiment Analysis of Twitter Data Using Naive Bayes Algorithm.. *Int. J. Recent Innov. Trends Comput. Commun.*, 6(4), pp. 120-125.

Manning, C. D., Raghavan, P. & Schütze, H., 2008. *Introduction to Information Retrieval.* s.l.:Cambridge.

Mäntylä, Mika V.; Graziotin, Daniel; Kuutila, Miikka, 2018. *The evolution of sentiment analysis—A review of research topics, venues, and top cited papers,* s.l.: Computer Science Review.

Mccallum, A. & Nigam, K., 2001. A Comparison of Event Models for Naive Bayes Text Classification. *Work Learn Text Categ*, 05, Volume 752.

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

Mehra, N., Khandelwal, S. & Patel, P., 2002. *Sentiment Identification Using Maximum Entropy Analysis of Movie Reviews,* s.l.: s.n.

Meivel, s. et al., 2022. Mask Detection and Social Distance Identification Using Internet of Things and Faster R-CNN Algorithm. *Computational Intelligence and Neuroscience*, 02, Volume 2022.

Meyer, C. F., 2009. *Introducing English Linguistics.* Cambridge: Cambridge University Press.

Mikolov, T., Chen, K., Corrado, G. & Dean, J., 2013. Efficient Estimation of Word Representations in Vector Space. *Proceedings of Workshop at ICLR*, 01, Volume 2013.

Mikolov, T. et al., 2013. Distributed Representations of Words and Phrases and Their Compositionality. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2.* Lake Tahoe, Nevada: Curran Associates Inc., p. 3111–3119.

Miller, G. A. et al., 1990. *Wordnet: An on-Line Lexical Database.* Oxford University Press ed. s.l.:s.n.

Mosteller, F. & Wallace, D. L., 1963. Inference in an Authorship Problem. *Journal of the American Statistical Association*, 58(302), pp. 275-309.

Nasukawa, T. & Yi, J., 2003. Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In: *Proceedings of the 2nd International Conference on Knowledge Capture.* Sanibel Island, FL, USA: Association for Computing Machinery, p. 70–77.

Nishikawa, H., Hasegawa, T., Matsuo, Y. & Kikui, G., 2010. Optimizing Informativeness and Readability for Sentiment Summarization. In: *Proceedings of the {ACL} 2010 Conference Short Papers.* Uppsala: Association for Computational Linguistics, pp. 325-330.

Norris, P., 2000. *A Virtuous Circle: Political Communications in Postindustrial Societies.* Massachusetts: Cambridge University Press.

Pang, B. & Lee, L., 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*, Volume 2, pp. 1-135.

Pang, B., Lee, L. & Vaithyanathan, S., 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In: *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing.* s.l.:Association for Computational Linguistics, p. 79–86.

Pankaj, G., Ritu, T. & Nirmal, R., 2016. Sentiment analysis and text summarization of online reviews: A survey. In: *2016 International Conference on Communication and Signal Processing (ICCSP).* s.l.:s.n., pp. 0241-0245.

Pennington, J., Socher, R. & Manning, C., 2014. Glove: Global Vectors for Word Representation. *EMNLP*, 01, Volume 14, pp. 1532-1543.

Phan, H. T., Nguyen, N. T. & Hwang, D., 2023. Aspect-level sentiment analysis: A survey of graph convolutional network methods. *Information Fusion*, Volume 91, pp. 149-172.

Pozzi, F. A., Fersini, E., Messina, E. & Liu, B., 2017. Challenges of Sentiment Analysis in Social Networks: An Overview. In: *Sentiment Analysis in Social Networks.* Boston: Morgan Kaufmann, pp. 1-11.

Premakumara, N. et al., 2019. Application of Summarization and Sentiment Analysis in the Tourism domain. In: *2019 IEEE 5th International Conference for Convergence in Technology (I2CT).* s.l.:s.n., pp. 1-5.

Qiu, G., Liu, B., Bu, J. & Chen, C., 2009. Expanding Domain Sentiment Lexicon through Double Propagation. *IJCAI International Joint Conference on Artificial Intelligence*, 01, pp. 1199-1204.

Rao, G., Huang, W., Feng, Z. & Cong, Q., 2018. LSTM with sentence representations for document-level sentiment classification. *Neurocomputing*, Volume 308, pp. 49-57.

RASCHKA, S. & VAHID, M., 2019. *Python Machine Learning;Machine Learning and Deep Learning with Python Scikit-Learn and Tensorflow 2.* 3rd ed. s.l.:PACKT Publishing.

Ravi, K. & Ravi, V., 2015. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems*, Volume 89, pp. 14-46.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J., 1986. Learning representations by back-propagating errors. *Nature*, Volume 323, pp. 533-536.

Sarawagi, S., 2008. Information Extraction. *Found. Trends Databases*, Volume 1, p. 261–377.

Schouten, K. & Frasincar, F., 2016. IEEE Transactions on Knowledge and Data Engineering. *Survey on Aspect-Level Sentiment Analysis*, 28(3), pp. 813-830.

Schütze, H., 1992. Word Space. In: *Advances in Neural Information Processing Systems.* s.l.:Morgan-Kaufmann.

Shaalan, K., Hassanien, A. E. & Tolba, F., 2017. *Intelligent Natural Language Processing: Trends and Applications.* 1st ed. s.l.:Springer Publishing Company, Incorporated.

Skalicky, S. & Crossley, S., 2018. Linguistic Features of Sarcasm and Metaphor Production Quality. In: *Proceedings of the Workshop on Figurative Language Processing.* New Orleans, Louisiana: Association for Computational Linguistics, pp. 7-16.

Sutton, R. S. & Barto, A. G., 2018. *Reinforcement Learning: An Introduction.* Second ed. s.l.:The MIT Press.

Taboada, M. et al., 2011. Lexicon-Based Methods for Sentiment Analysis. *Computational Linguistics*, 37(2), pp. 267-307.

Tedmori, S. & Awajan, A., 2019. Sentiment Analysis Main Tasks and Applications: A Survey. *Journal of Information Processing Systems*, 15(3), pp. 500-519.

Turney, P. D., 2002. Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics.* Philadelphia, Pennsylvania: Association for Computational Linguistics, p. 417–424.

Turney, P. D. & Littman, M. L., 2003. Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Trans. Inf. Syst.*, 10, 21(4), p. 315–346.

Valitutti, A., Strapparava, C. & Stock, O., 2004. Developing affective lexical resources. *PsychNology Journal*, 2(1), p. 61 – 83.

Van Rijsbergen, C., 1979. *Information Retrieval.* 2 ed. London: Butterworths: s.n.

Vergeer, M., Hermans, L. & Sams, S., 2013. Online social networks and micro-blogging in political campaigning: The exploration of a new campaign tool and a new campaign style. *Party Politics*, Volume 19, pp. 477-501.

Volkova, S., Wilson, T. & Yarowsky, D., 2013. Exploring Sentiment in Social Media: Bootstrapping Subjectivity Clues from Multilingual Twitter Streams. *ACL 2013 - 51st Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, 08, Volume 2, pp. 505-510.

Wang, G. et al., 2014. Sentiment classification: The contribution of ensemble learning. *Decision Support Systems*, Volume 57, pp. 77-93.

Wankhade, M., Rao, A. & Kulkarni, C., 2022. A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 02, 55(7), pp. 1-50.

Whalen, J. M., Pexman, P. M., Gill, . A. J. & Nowson, S., 2013. Verbal irony use in personal blogs. *Behaviour and Information Technology*, 32(6), p. 560 – 569.

Whalen, J., Pexman, P., Gill, A. & Nowson, S., 2011. Verbal irony use in personal blogs. *Behaviour & Information Technology*, 10, Volume 2012.

White, L., Togneri, R., Liu, W. & Bennamoun, M., 2018. *Neural Representations of Natural Language.* 1st ed. s.l.:Springer Publishing Company, Incorporated.

Wiebe, J., Rebecca, F. B. & Thomas, P. O., 1999. Development and use of a gold-standar. *roceedings of the Association for Computational Linguistics*.

Wilks, Y. & Stevenson, M., 1997. The Grammar of Sense: Using Part-of-Speech Tags as a First Step in Semantic Disambiguation. *Natural Language Engineering*, 12, Volume 4.

Woolley, S. C., 2020. Bots and Computational Propaganda: Automation for Communication and Control. In: *Social Media and Democracy.* s.l.:s.n.

Wu, H., Li, J. & Xie, J. (., 2017. *Maximum entropy-based sentiment analysis of online product reviews in Chinese,* s.l.: s.n.

Xia, E., Yue, H. & Liu, H., 2021. *Tweet Sentiment Analysis of the 2020 U.S. Presidential Election,* Ljubljana, Slovenia: Association for Computing Machinery.

Xing, F. Z., Cambria, E. & Welsch, R. E., 2018. Natural Language Based Financial Forecasting: A Survey. *Artificial Intelligence Review*, 06, 50(1), p. 49–73.

Xu, Q. A., Chang, V. & Jayne, C., 2022. A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, Volume 3, p. 100073.

Yadav, K., 2021. *A Comprehensive Study on Optimization Strategies for Gradient Descent In Deep Learning,* s.l.: arXiv.

Yadollahi, A., Shahraki, A. G. & Zaiane, O. R., 2017. Current State of Text Sentiment Analysis from Opinion to Emotion Mining. *ACM Computing Surveys*, 03, 50(2).

Yang, B. & Cardie, C., 2014. Context-aware Learning for Sentence-level Sentiment Analysis with Posterior Regularization. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* Baltimore, Maryland: Association for Computational Linguistics, pp. 325--335.

Yaqub, U., Chun, S. A., Atluri, V. & Vaidya, J., 2017. Sentiment Based Analysis of Tweets during the US Presidential Elections. In: *Proceedings of the 18th Annual International Conference on Digital Government Research.* Staten Island(NY): Association for Computing Machinery, p. 1–10.

Yusof, N. N., Mohamed, A. H. & Rahman, S. A., 2015. Reviewing Classification Approaches in Sentiment Analysis. In: *International Conference on Soft Computing in Data Science.* s.l.:s.n.

Zhang, L., Wang, S. & Liu, B., 2018. Deep Learning for Sentiment Analysis: A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Volume 8.

Zhang, W., Xu, H. & Wan, W., 2012. Weakness Finder: Find Product Weakness from Chinese Reviews by Using Aspects Based Sentiment Analysis. *Expert Syst. Appl.*, September, 39(11), p. 10283–10291.

Zhang, W., Yu, C. & Meng, W., 2007. Opinion Retrieval from Blogs. In: *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management.* s.l.:Association for Computing Machinery, p. 831–840.

Zhang, X., Fuehres, H. & Gloor, P., 2011. Predicting Stock Market Indicators Through Twitter – “I Hope it is Not as Bad as I Fear. *Procedia - Social and Behavioral Sciences*, 12, Volume 26, p. 55–62.

Zhang, X. & Zheng, X., 2016. Comparison of Text Sentiment Analysis Based on Machine Learning. In: *2016 15th International Symposium on Parallel and Distributed Computing (ISPDC).* s.l.:s.n., pp. 230-233.

Zhendong, D., Qiang, D. & Changling, H., 2010. HowNet and Its Computation of Meaning. In: *Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations.* s.l.:Association for Computational Linguistics, p. 53–56.

Zhiyuan, L., Yankai, L. & Maosong, S., 2020. *Representation Learning for Natural Language Processing.* s.l.:Springer Singapore.