Изображение выглядит как Графика, Шрифт, логотип, дизайн

Автоматически созданное описание

*Master’s Thesis*

Sentiment Analysis (NLP)

**Advanced Analytics - Big Data**

Author Supervisor

Pramod Khare (117464 ) Sebastian Zając, PhD.

Table of Contents

[Chapter 1: Introduction 4](#_Toc144480979)

[1.1 Introduction to Natural Language Processing (NLP) 4](#_Toc144480980)

[1.2 Concept of Sentiment Analysis 4](#_Toc144480981)

[1.3 Applications of Sentiment Analysis 5](#_Toc144480982)

[1.4 Importance of Sentiment Analysis in Business and Research 6](#_Toc144480983)

[1.5 Challenges in Sentiment Analysis 7](#_Toc144480984)

[1.6 Aims and Objectives of the Research 8](#_Toc144480985)

[1.6.1 Aim 8](#_Toc144480986)

[1.6.2 Objectives 8](#_Toc144480987)

[1.7 Research Questions 9](#_Toc144480988)

[1.8 Significance of the Research 9](#_Toc144480989)

[1.9 Overview of the Thesis 10](#_Toc144480990)

[Chapter 2: Literature Review 11](#_Toc144480991)

[2.1 Unveiling Textual Emotions by Analyzing Sentiments and Feelings on Social Media Platforms. 11](#_Toc144480992)

[2.2 Exploring Linguistic Tools and Machine Learning Methods in Evaluating Emotions in Texts 14](#_Toc144480993)

[2.3 Exploring Nature-Inspired Algorithms for Mood Analysis and Sentiment Classification 16](#_Toc144480994)

[2.4 Enhancing Sentiment Analysis through Preprocessing Techniques and Emojis 19](#_Toc144480995)

[2.5 Advancements in Sentiment Analysis: Methods for Understanding User Feelings in Online Conversations and Text Data 21](#_Toc144480996)

[Chapter 3: Methodology 27](#_Toc144480997)

[3.1 Overview of Methodological Approach 27](#_Toc144480998)

[3.2 Sentiment Analysis Algorithms Selection 28](#_Toc144480999)

[1.3 Dataset Selection and Description 30](#_Toc144481000)

[1.4 Analysis of Sentiment Difficulties 31](#_Toc144481001)

[1.5 Algorithm Performance Metrics 32](#_Toc144481002)

[1.6 Process of Data Analysis 33](#_Toc144481003)

[Chapter 4: Results 34](#_Toc144481004)

[4.1 Introduction 34](#_Toc144481005)

[4.2 Overview of the Dataset 34](#_Toc144481006)

[4.3 Data Preprocessing 35](#_Toc144481007)

[4.4 Model Evaluation 37](#_Toc144481008)

[4.4.1 Multinomial Naive Bayes 38](#_Toc144481009)

[4.4.2 CNN Model 1 39](#_Toc144481010)

[4.4.3 CNN Model 2 40](#_Toc144481011)

[4.4.4 LSTM Model 41](#_Toc144481012)

[4.4.5 VADER Model 41](#_Toc144481013)

[4.4.6 Comparisons of Accuracies 42](#_Toc144481014)

[4.5 Model Comparison 42](#_Toc144481015)

[4.5.1 Comparison with Each Other 42](#_Toc144481016)

[4.5.2 Comparison with Existing Models 44](#_Toc144481017)

[4.6 Conclusion 45](#_Toc144481018)

[Chapter 5: Discussion 47](#_Toc144481019)

[5.1 Introduction 47](#_Toc144481020)

[5.2 Overview of the Models 47](#_Toc144481021)

[5.3 Model Performance 48](#_Toc144481022)

[5.4 Compare with other Models 50](#_Toc144481023)

[5.5 Implications for Future Work 51](#_Toc144481024)

[5.6 Recommendations for Future Work 53](#_Toc144481025)

[5.7 Summary of the Discussion 54](#_Toc144481026)

[Chapter 6: Conclusion 56](#_Toc144481027)

[6.1 Introduction 56](#_Toc144481028)

[6.2 Summary of the Study 56](#_Toc144481029)

[6.3 Contributions of the Study 57](#_Toc144481030)

[6.4 Limitations of the Study 58](#_Toc144481031)

[6.5 Final Thoughts 58](#_Toc144481032)

[References 60](#_Toc144481033)

# Chapter 1: Introduction

## 1.1 Introduction to Natural Language Processing (NLP)

Natural Language Processing (NLP) is a branch of AI that focuses on translating between human and computer speech. The end goal of NLP is to create programmes that can parse, comprehend, and produce human speech. As a researcher in this area, I find myself attracted by the possibility of developing systems that grasp language in a manner analogous to human cognition.

A major factor fueling NLP's ascent is the explosion of digital data, especially unstructured text data. Every day, a mountain of textual information is produced, including posts on social networking platforms, product evaluations, customer comments, news items, and much more. Finding insights in this data manually can be time-consuming and fraught with mistakes. Here, natural language processing comes in quite handy. It aids in the correct and efficient analysis of this massive data set, allowing us to draw useful conclusions (Lutkevich & Burns, 2023).

Numerous industries have been profoundly affected by NLP's extensive uses. Chatbots and other NLP-enabled virtual assistants are improving customer service and the user experience. To better diagnose patients and develop effective treatments, natural language processing (NLP) is used to medical records and academic literature. Natural language processing helps interpret and classify emotions represented in text, which is relevant to my research interests in the field of sentiment analysis.

NLP's significance to the state of the art cannot be overstated. NLP will become increasingly important as the amount of textual data we produce and rely on grows. It will facilitate the development of new ideas with the potential to significantly enhance our quality of life. My studies will hopefully have a positive impact in this fascinating area by improving our understanding of sentiment analysis and how well it can identify and categorise the feelings and opinions expressed in text (Banoula, 2023).

## 1.2 Concept of Sentiment Analysis

Often shortened to "opinion mining," the area of Sentiment Analysis is a subset of Natural Language Processing (NLP) that investigates the emotional content of written or spoken language. Emotional tone analysis is the practise of analysing the underlying feelings of a text in order to better comprehend the author's intended message.

The goal of sentiment analysis is to understand the exact emotions being represented by a text, whether they be happiness, rage, sadness, or any other emotion. Systems can now extract and comprehend the subjective information in sources by utilising natural language processing, computational linguistics, and text analysis (Gupta, 2018).

The increasing amount of digitally accessible data has given sentiment analysis newfound importance. Unstructured data containing significant insights about public opinion can be found in abundance in user-generated content including product evaluations, social media posts, blogs, and more. These findings can be used to track consumer perceptions of brands, evaluate products, improve customer service, and even gauge public opinion on political issues (Raj, 2021).

Understanding human emotions and, by extension, measuring public opinion, is why sentiment analysis is so important. It gives businesses a way to see and respond to how their customers feel. Academics and researchers like myself find sentiment analysis to be an intriguing and crucial field because of the wealth of information it provides for understanding and predicting human behaviour.

## 1.3 Applications of Sentiment Analysis

With its ability to automate the processing of subjective information, sentiment analysis has found many practical applications across a wide range of sectors, radically altering the nature of decision-making.

The most obvious use case is in the field of social media analytics. By analysing user-generated content (UGC) on social media, businesses can gauge how the general public feels about their products, services, or brand image. The results of this sort of study are extremely helpful in shaping advertising campaigns, creating new products, and better serving existing customers (Sharma, 2023).

Analysis of customer feedback is also useful in other contexts. Using sentiment analysis, e-commerce businesses may automatically analyse millions of product reviews and classify them as positive, negative, or neutral. Companies may quickly resolve any faults and enhance their products based on the information gleaned from these types of evaluations.

Analysis of public opinion on political issues is another significant application. Analysing social media posts, blogs, or news items can give political parties and analysts a sense of popular opinion about policies, campaigns, or candidates. This can help with voter targeting and campaign strategy.

Sentiment analysis is used in customer service to prioritise and address client concerns depending on the tone of those complaints. It aids in categorising really unfavourable feedback, letting businesses prioritise fixing the most pressing problems.

Sentiment analysis is a useful tool for firms to use in market research since it reveals consumer preferences and current market trends (Wankhade et al., 2022).

My research aims to investigate and assess the performance of sentiment analysis algorithms in a variety of settings. I hope to learn more about how well they can recognise and categorise feelings and attitudes in written language, as well as their relative strengths and opportunities for development.

## 1.4 Importance of Sentiment Analysis in Business and Research

Because of its effect on the decision-making process, sentiment analysis has become increasingly important in modern commercial and academic settings.

Sentiment analysis is a powerful business tool for understanding how consumers feel about a given brand, product, or service. A company's strengths and weaknesses can be revealed through an accurate analysis of consumer feedback, social media posts, or product reviews. These findings are crucial for informing business strategies in areas like as product design, advertising, and customer support. The presence of negative sentiment may point to problems with the product that need fixing, while the presence of positive sentiment may draw attention to aspects of the product that are particularly well received (Yılmaz, 2022).

As such, sentiment analysis is useful in academic settings for analysing and comprehending widespread public opinion on a range of topics. Sentiment analysis of social media content, for instance, might show public opinion on social policies, political candidates, or public health interventions in social science or political research. Policymakers, public communicators, and those interested in learning more about the mechanics of public opinion can all benefit from such insights (Marta, 2022).

Since sentiment analysis has such tremendous impact on decisions in the commercial and academic worlds, my research tries to assess and improve its precision. In doing so, it is intend to help in the improvement of sentiment analysis's application in such contexts.

## 1.5 Challenges in Sentiment Analysis

Despite its usefulness, sentiment analysis is not without difficulty and difficulty. Sentiment analysis is complicated by the nuanced nature of human speech and the wide range of possible expressions.

Sentiment analysis faces significant difficulties when trying to identify irony and sarcasm. The actual meaning of the words isn't always reflected in these linguistic variants. If user were having a terrible day, user could use the statement "What a fine day!" ironically. Sentiment analysis algorithms face a formidable challenge when tasked with reliably identifying such nuanced emotions.

The confusion inherent in the English language also presents difficulties. Because the meaning of a remark might change depending on its surrounding context, it can be challenging for computers to grasp the intended tone. Sentiment analysis is further complicated by the fact that the same term or phrase may have varied meanings depending on the culture or community in which it is used (Sahani, 2022).

Last but not least, the complexity is increased by the frequent use of slang, abbreviations, and emoticons in online content, especially on social media. The dictionary-based methods used by most sentiment analysis algorithms today have difficulty understanding such idiomatic expressions.

Improving the reliability of sentiment analysis requires resolving these issues. There are a variety of ways to tackle these challenges, but some of the most promising involve novel approaches that combine deep learning, context-aware analysis, and cross-cultural understanding. More study in this area will unquestionably result in more refined and accurate sentiment analysis programmes.

## 1.6 Aims and Objectives of the Research

### 1.6.1 Aim

The major goal of this study is to conduct an in-depth analysis of existing sentiment analysis algorithms, paying special attention to how well they can identify and categorise underlying feelings and perspectives in textual content. The study's goal is to learn about the benefits and drawbacks of these algorithms, therefore it thoroughly investigates their usefulness and practicality.

### 1.6.2 Objectives

The study aims to accomplish the following:

1. Trying to offer light on the methodology and performance measures of state-of-the-art sentiment analysis algorithms and methodologies.
2. To investigate the difficulties of sentiment analysis and the approaches taken by existing algorithms to them, with a focus on sarcasm, irony, and linguistic ambiguity.
3. The purpose of this dataset evaluation is to compare the efficacy of different sentiment analysis algorithms for identifying and classifying user sentiment.
4. The goal is to determine what parameters influence the efficiency and precision of these algorithms.
5. To use the data gathered to make suggestions on how and where sentiment analysis might be improved in the future.

These goals will be met if the research succeeds in its overall purpose of contributing to the ongoing academic conversation on sentiment analysis and so facilitating the creation of more precise and context-aware sentiment analysis technologies. As time goes on, this could improve the quality of decision-making in fields like business, politics, and social science.

## 1.7 Research Questions

This dissertation seeks to answer the question, "How accurate are current sentiment analysis algorithms in detecting and classifying emotions and attitudes in text?" This inquiry captures the essence of the study, which is to evaluate the efficacy of current sentiment analysis methods.

The following sub-questions have been developed to expound on and provide context for the primary research question:

1. To what extent have sentiment analysis algorithms advanced in recent years? The purpose of this inquiry is to learn about and discuss recent developments in sentiment analysis algorithms and methods.
2. Asking, "What are the benefits and drawbacks of these algorithms?" The purpose of this article is to draw attention to the positive aspects of existing sentiment analysis algorithms and to locate any areas where improvement is needed.
3. How well do these algorithms recognise and label sentiment and attitude in written text? The goal of this inquiry is to assess the efficacy of these algorithms.
4. To what extent do certain elements compromise the efficacy of sentiment analysis tools? The purpose is to single out and debate the myriad factors that affect sentiment analysis systems' efficacy and efficiency.
5. To what extent might sentiment analysis algorithms be improved in the future? The purpose of this inquiry is to propose adjustments and new lines of inquiry in light of the study's findings.

## 1.8 Significance of the Research

This study's significance lies in the fact that it provides an in-depth investigation of sentiment analysis algorithms—crucial resources for comprehending public opinion in the digital age. Improving the precision of sentiment analysis is essential because of the wide range of fields that can benefit from it. The results of this study will add to the existing body of knowledge in the field of sentiment analysis by providing a thorough review of currently used algorithms, exposing their respective benefits and drawbacks.

Furthermore, the research will direct future algorithm development by identifying factors affecting algorithmic performance and exploring difficulties including sarcasm, irony, and ambiguity detection. The ultimate goal is to aid in the development of advanced, context-aware sentiment analysis algorithms that can decipher subtle emotions and attitudes in text.

Therefore, the study is not just important theoretically, but also because its findings may have real-world ramifications in a wide range of fields.

## 1.9 Overview of the Thesis

There are six major sections to this thesis. After a brief introductory section, Chapter 2 will dive into the state-of-the-art algorithms and evaluation criteria in the field of sentiment analysis. In Chapter 3, we discuss this study's methodology, which includes the dataset, the algorithms used for sentiment analysis, and the criteria used to rate their performance.

The main results of the study are presented in Chapter 4, and they provide an in-depth investigation of how well the chosen sentiment analysis algorithms detect and categorise emotions and attitudes in text data. The difficulties and concerns that arose during the investigation are also highlighted.

Discussion of the results follows in Chapter 5, where the limitations of the existing algorithms are analysed and suggestions for improvements are made. It also provides a roadmap for where sentiment analysis research should go in the future.

Finally, Chapter 6 closes the thesis by summarising the key takeaways, theoretical contributions, and practical consequences of the study.

# Chapter 2: Literature Review

## 2.1 Unveiling Textual Emotions by Analyzing Sentiments and Feelings on Social Media Platforms.

Social media platforms have become essential for people worldwide to share their ideas because of how quickly the Internet has grown. People often show how they feel or what they think through writing, pictures, sounds, and videos. On the other hand, it takes a lot of work to keep up with text conversations through Web-based networking media. In some cases, emotional analysis isn't enough to figure out how someone feels or thinks. In those scenarios, emotional awareness is needed. The present research covers the different steps of sentiment analysis, the different emotion models, and the process of analyzing text for sentiment and emotions.

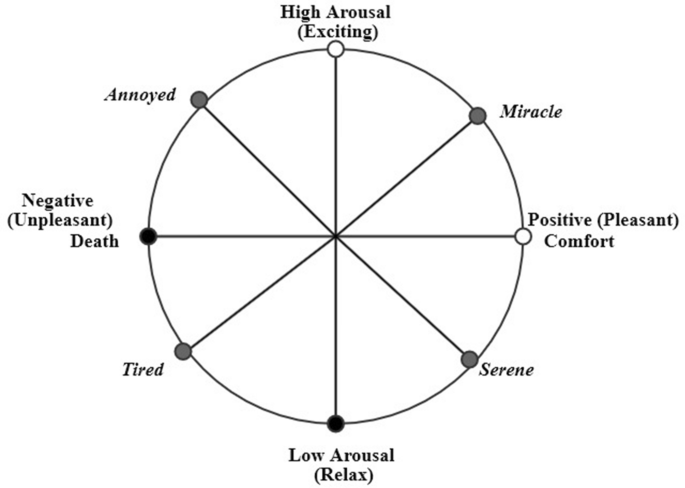


Figure 1 Dimensional model of emotions.

(Source: Nandwani and Verma, 2021)

Researchers emphasized that mood and feeling investigation can be done and used in many different ways. Someone can determine how someone is feeling or what they are thinking in three ways: based on their language, machine learning, or deep knowledge. Each has its positive aspects and negative aspects. Even though there are different ways to figure out how someone is feeling or what they are thinking, experts have a lot of problems. For instance, they have to deal with context, ridicule, statements demonstrating multiple feelings, the spread of Web slang, and lexical and syntactic uncertainty. Also, there aren't any standard guidelines regarding how to talk about feelings on different platforms, so some people do it well; some try to hide their feelings, and some make their messages make sense. Researchers have worked hard to find a method that works well everywhere.

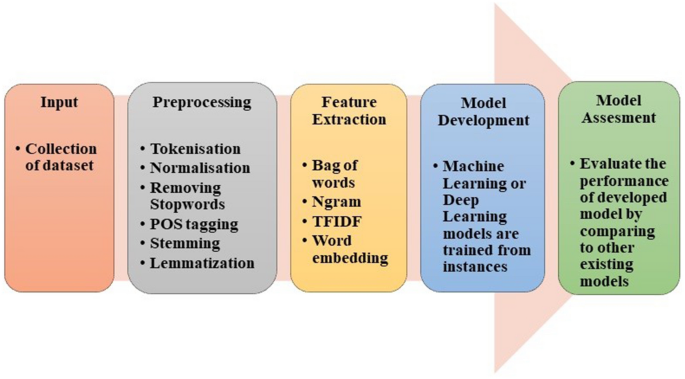


Figure 2 Basic steps to perform sentiment analysis and emotion detection.

(Source: Nandwani and Verma, 2021)

In the paper's investigation, it was discovered that the lexicon-based method works well for analyzing sentiment and emotions. On the other hand, the dictionary-based approach is flexible and easy to use. The corpus-based method, however, is based on rules that work well in specific fields. Corpus-based methods are more accurate because of this, but they can't be used to make broad claims. How well machine and deep learning systems work depends on how they are set up and how big the data set is. Still, machine learning models can only sometimes figure out secret characteristics or parts of the text (Nandwani and Verma, 2021). When there are a lot of data points to look at, deep learning works better than machine learning.

Text mood analysis is figuring out what the writer thinks, feels, and even feels about a topic. People often think it's the same as opinion mining, but it should also include feeling mining. Textual emotion-mining methods can be used to find out how happy customers are, help choose teaching materials for e-learning, offer products to users based on how they feel, and even predict mental health disorders. In surveys about sentiment analysis, which are often old or incomplete, the vital link between opinion and emotion mining is often ignored. The primary objective of this research is sentiment analysis, focusing on emotion mining, which needs a new and different point of view. Yadollahi, Shahraki and Zaiane (2017) address the most recent ways and suggest the following: (1) A taxonomy of sentiment analysis; (2) A survey of polarity classification methods and resources, especially those related to emotion mining; (3) A complete survey of emotion theories and emotion-mining research; and (4) Some valuable resources, such as lexicons and datasets.

The survey performed in the research highlighted the best and most up-to-date ways to figure out how someone feels about a text. Sentiment analysis investigates how people think about different things and how they can be found, analyzed, and rated. Text mood analysis mainly examines what people write to determine what they think, feel, and think about. Personal notes, emails, news headlines, blogs, stories, books, chat messages, and social networking sites like Twitter, Facebook, and MySpace are all examples of texts that can show feelings. This study gave a clear and logical taxonomy of sentiment analysis work and a careful way to group jobs in this area (Gupta, Singh and Singla, 2019). There are two significant areas for research in this area: opinion mining and emotion mining. According to researchers, there is a lot of research on opinion mining, and many new focused and specific regions are being studied. Feeling mining from text, on the other hand, is still in its early stages. Because of this and that opinion and emotion mining are closely linked, researchers tried to give an in-depth review of the latest trends and valuable resources in these areas.

## 2.2 Exploring Linguistic Tools and Machine Learning Methods in Evaluating Emotions in Texts

In this paper, Bobichev, Kanishcheva and Cherednichenko (2017) examine the job of finding out how people feel about Ukrainian and Russian news and compare different ways to do it and linguistics tools. They gathered a list of news stories from Ukraine and Russia and gave each one of three labels: good, bad, or neutral. At least three people used the online interface to mark each text. The following tests used the texts all three annotators put in the same group. Researchers have tried the Naive Bayes, DMNBtext, NB Multinomial, and SVM machine learning methods to see if they could automatically identify these texts. In each case, using a technique for choosing features was the best way to find the best set of features. The main objective was to figure out which jobs were the most important, which meant sorting the good, bad, and neutral news. They have also looked at different ways to put data into groups so that mood analysis can be done automatically. Researcher tried the extended dictionary WordNet-Affect for these jobs. It has words for emotions in both Russian and Ukrainian. When researchers tested a variety of machine learning approaches, they determined that even the most basic one, called Nave Bayes, could produce excellent results (an average F1-score of 0.82) given the appropriate set of characteristics.

Sentiment analysis can be used in numerous ways, such as determining what people think about different products, problems, and social and political events and how they feel about them. When someone knows what other people think, they can make better choices. Opinion mining is a way to get information from search engines, social networks, blogs, and microblogs. Everyone has their own ideas, and Twitter tweets are a great way to discover those ideas. But it is hard to study text/opinion data well because there is so much of it, and it is not organized. So, to mine and reduce tweets and find words that show how people feel requires robust algorithms and ways to use computers. Most computer methods, models, and algorithms that try to figure out how people think from unstructured data use the bag-of-words technique for machine learning. This research employs both organized and unstructured methods on different datasets. An unstructured method is used to determine how people feel about tweets from the Twitter public domain. Machine learning methods like Multinomial Naive Bayes (MNB), Maximum Entropy, and Support Vector Machines are used to figure out how people feel about tweets and how well different feature pairs work. In our trial with tweets, we found that the proposed unsupervised method worked better than the lexicon-based method, which only worked 75.2% of the time. In the tests, the researcher conducted the supervised practice, in which the used unigrams, bigrams, and parts of speech as features, performed effectively for recognizing feelings and sentiment in unstructured data. With the MNB algorithm and the unigram feature, short message services can be successful 67% of the time (Rahman and Hossen, 2019).

Employing what people say online to determine their feelings is a common way to use Natural Language Processing. People usually say that the job is to find the words or phrases in a text that show the reader that the author has a positive, negative, or neutral opinion about a topic. The research considered discussing the person's irrational or emotional decisions and how well their goals were met as the two primary sources of attitude. In this study, they explained ways to learn more, bridging the gap between psychology/cognitive science and computer-based methods. Li and Hovy (2017) looked at what people want and how that affects the goals they set. In a practical model of emotion, this is the reason why someone has a particular valence. (The reality) that certain emotional decisions also depend on non-utilitarian, purely intuitive tastes is an issue that must be addressed differently.) Even though these ideas are still young, scattered, disorganized, and even made up, the researchers of this study think that these points of view could point to good ways to do different kinds of work in the future.

Sentiment analysis is the process of using algorithms to find and sort the opinions stated in a text, especially to find out if the writer has a positive, negative, or neutral attitude towards a specific topic. Any business has to know what its customers think of it. Therefore, in this research, "Sentiment Analysis of Restaurant Reviews Using Machine Learning

Techniques" analyze restaurant customers' reviews using machine learning classification methods. Most of the research is about how different ways of classifying are used and how well they work.

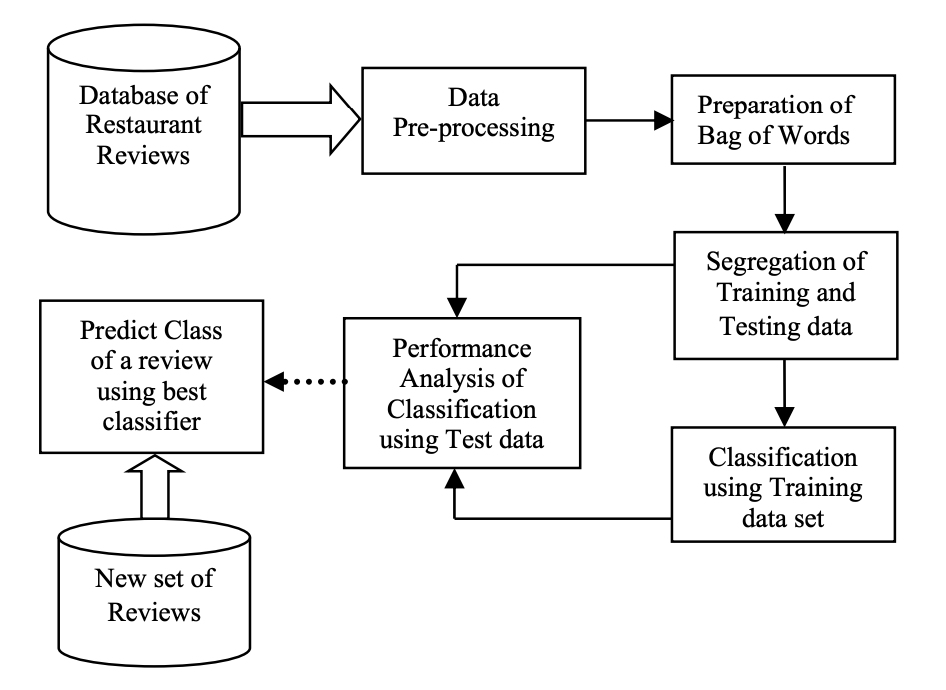


Figure 3 Architectural diagram.

(Source: Krishna et al., 2019)

The suggested work is seeking out how people at a restaurant feel about the service. It automatically sorts reviews based on how they make people think by using natural language processing, text analysis, and computer methods.

Krishna et al. (2019) evaluated how well different algorithms work with a set of restaurant reviews and then studied the best algorithm. Here, it's been demonstrated that the decision tree classifier and the SVM classifier are the ones that work best. Again, the FAR for SVM is the lowest, at 2.7%, and the FRR for the Naive Bayes prediction is the lowest, at 3.5%. But the method's accuracy is always a trade-off between FAR and FRR. So, SVM is the best method for classifying the given information, with the highest accuracy of 94.56%.

## 2.3 Exploring Nature-Inspired Algorithms for Mood Analysis and Sentiment Classification

Data mining is an area that has been studied a lot and is still being studied. Using different methods or techniques, information is pulled out about how a customer feels or thinks. Early research on mood analysis used both supervised and unsupervised machine learning methods and lexicon-based methods. Since nature is a great place to get ideas, nature-inspired algorithms are becoming a popular way to make new algorithms and improve their work. There are different kinds of these methods, like systems that use physics and chemistry and others that use biology. The present research is mainly about bio-inspired algorithms, which can be built on swarm intelligence. In this study, the critical bio-inspired algorithms often used in mood analysis are looked at in detail. Researchers have discussed how these important algorithms are doing now and do a comparison study by looking at 80 papers from different journals, workshops, book chapters, etc. Researchers did a survey and found a few well-known algorithms that can be used in many different ways. In this comprehensive review, they described their goals and what they're trying to do. In this study, some of the most critical work on bio-inspired S.A. is looked at. They implemented algorithms into groups based on how they work in nature. About nine widely utilized algorithms are chosen that are based on biological systems: PSO, ACO, C.S., FFA, B.A., ABC, FPA, GSO, and SFL. By performing this, researchers wanted to demonstrate how these algorithms differ and what makes them unique.

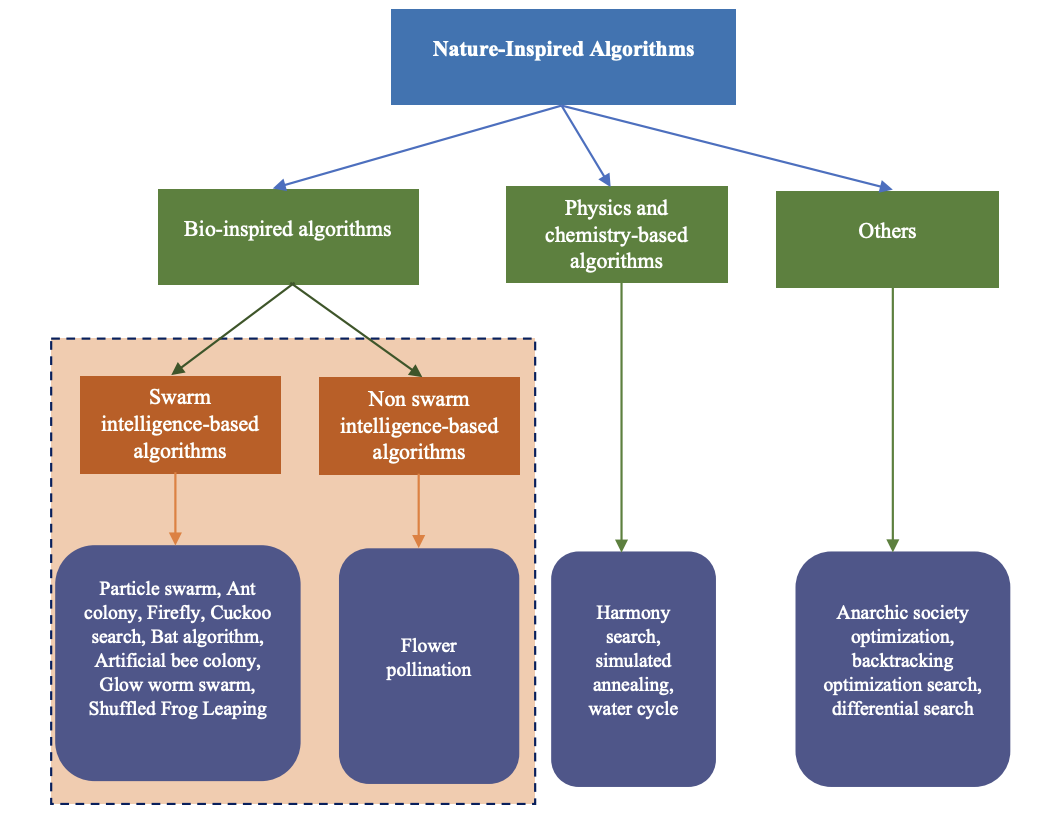


Figure 4 Taxonomy of nature-inspired algorithms

(Source: Yadav and Vishwakarma, 2020)

Additionally, they compared these algorithms by addressing their uses, pros and cons, and how well they work on a smartphone product review so individuals can learn more about each. Further, also give extensive information about how bioinspired algorithms for S.A. have changed over time and show how the reviewed articles are split up by type of algorithm. Lastly, they discussed some of the most critical work that has been done in this area to show where future research could go. Based on this study, Yadav and Vishwakarma (2020) concluded that these methods based on nature are very good at solving optimization problems. The feature engineering process is the most crucial part of mood analysis because it directly affects how well the programme operates.

The opinions of individuals can be found with the assistance of natural language processing. It can also be called "thought mining" or "feeling mining." This is a very well-known area of study in text mining. The key objective is to determine what the writing means and classify it as positive, negative, or neutral. People can use it to decide what to do.

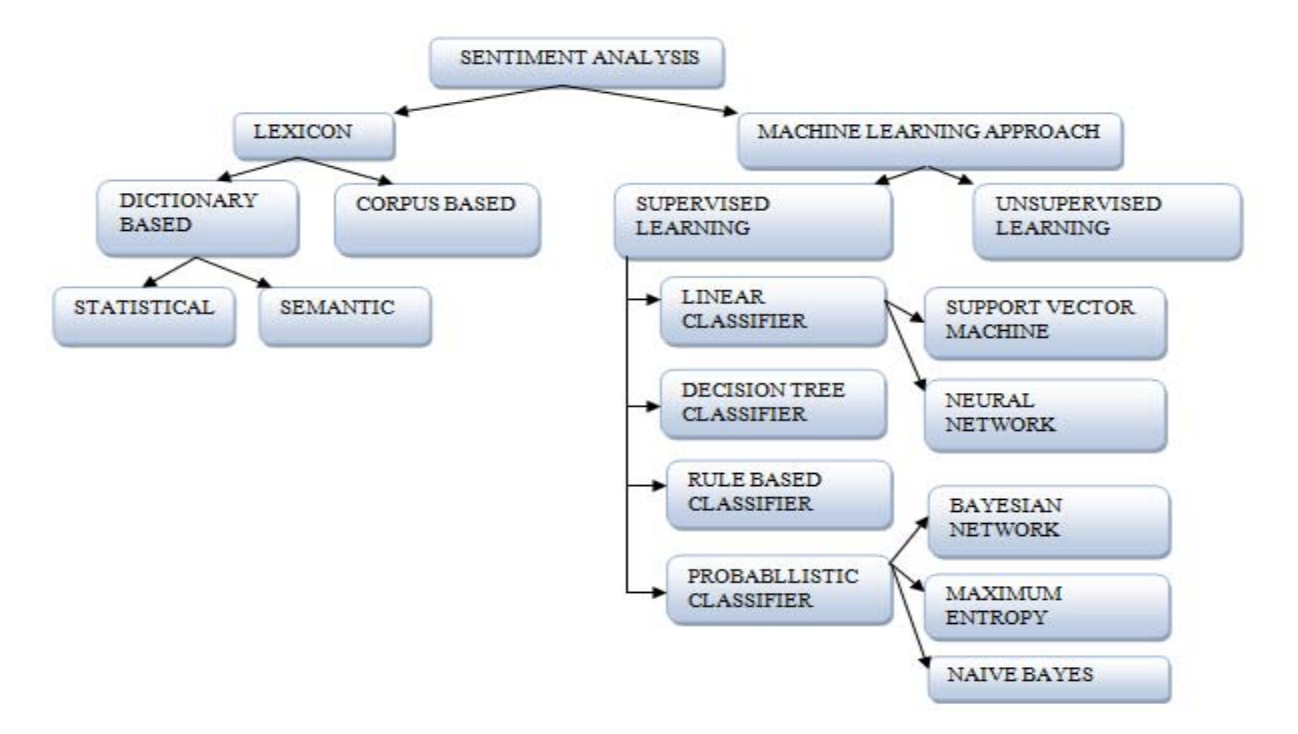


Figure 5 : Sentiment Classification Techniques

(Source: Kaur, Mangat and Nidhi, 2017)

Bobichev, Kanishcheva, and Cherednichenko (2017) did many different things to analyze sentiment, such as finding subjectivity, group sentiment, getting aspect terms, getting features, etc. This study gives an overview of methods that can be used to look at feelings and group them. This conducted survey shows that there is still much to learn about how to employ group thoughts. It has a lot of ways to use algorithms. SVM and naive Bayes are the most popular ways to determine how someone feels. A lot of people look at how people think in tweets. Many use data sets from sites like Amazon, IMDB, and Flipkart to determine how people feel (Hariguna, 2020). More attention needs to be paid to social networking sites. Often, it's imperative to think about what's going on. So, more research is needed regarding this topic, and there is a need to spread awareness about this area of concern.

## 2.4 Enhancing Sentiment Analysis through Preprocessing Techniques and Emojis

Increasingly, people contribute their thoughts, opinions, and observations about any product or person on social networking sites. People talk very casually about how they feel online. Therefore, it's hard to know exactly how someone feels when they use everyday words. Sentiment analysis studies people's ideas, feelings, and attitudes to determine if they are good, bad, or neutral. People have used many more emojis on social media in the past few years. Yadav and Pandya (2017) have drawn more attention to how vital emoticons are for mood research because of this. This study gives a overview of a few things that affect sentiment analysis. There is also discussion regarding recognizing humour, speaking more than one language, dealing with acronyms and slang, lexical variation, and using a dynamic dictionary. The current study mainly demonstrated how important emojis are in figuring out how people feel. They did this by giving a few relevant examples. The paper also mentions how mood research is done today. Mood analysis depends on text preparation, feature extraction, feature selection, irony, how to use a dynamic dictionary, acronyms and slang, and how the meaning of words can change. Machine learning methods are domain-specific and work well in a specific area (like movie or product reviews). Still, they must work better for general applications like sentiment analysis on social networking data or the Twitter dataset. Lexicon-based approaches can be used in any area because they focus on parts of the text already in the lexicon. It can make an integrated method by merging machine-learning and lexicon-based techniques. This could lead to a more accurate investigation into how people feel.

The purpose of mood analysis is to determine how someone believes about a text. It's hard to know how to read people's words and how someone may express how they feel in different ways. In the paper "Sentiment Analysis Is a Big Suitcase," the researchers claim that sentiment analysis is a "suitcase research problem" that requires solving many different natural language processing (NLP) tasks. Cambria et al. (2017) provided an overview of possible NLP problems and explained how and why they should be linked together. They list 15 different NLP problems that need to be solved for mood analysis to work as well as a person. Some issues found are part-of-speech tagging, named object recognition, emotion lexicons, and machine learning. By fixing these different NLP problems, mood analysis can work just as well as a person's. But this is hard to do and still needs a lot of research. The researchers hoped their paper would pave the way for an ensemble approach to NLP, combining data-driven (bottom-up) algorithms with theory-driven (top-down) methods that replicate how humans decode and understand natural language.

The "Dimensionality Reduction for Sentiment Analysis using Preprocessing Techniques " examines how various preprocessing techniques work together. Mhatre et al. (2017) use preprocessing of the input text to improve accuracy. They evaluated numerous preprocessing methods, including those for dealing with emoticons, removing HTML tags and slang, dealing with punctuation, removing stopwords, stemming, and lemmatizing. Compared to the traditional method, which relied on raw data, the analysis of preprocessed data is more straightforward. The Kaggle Bag of Words Meets Bags of Popcorn dataset has been provided to several different preprocessing algorithms. Various combinations of the eight preprocessing procedures were utilized, including using a single process at a time, two systems at once, three courses at once, and so on. To make predictions about the feelings, researchers turned to the trustworthy Random Forest classifier and evaluated our results with 10-fold cross-validation. By comparing the separate tests' accuracy, researchers could identify the optimal combination of experiments. While the standard approach does not perform any data preprocessing, the provided solution eliminates all unnecessary data. By removing these kinds of unnecessary information, data analysis is streamlined. When reducing data, various factors come into play. Combining Slangs Handling, Stopwords Removal, and Lemmatization resulted in 86.04% accuracy.

## 2.5 Advancements in Sentiment Analysis: Methods for Understanding User Feelings in Online Conversations and Text Data

As technological innovation has changed, people who use the Internet now have access to vast information. Users' ideas on this website are very different and must be in order. This paper looks at an email conversation through the lens of sentiment analysis. People can have good, bad, or neutral thoughts. This paper presents an answer to the above data using a lexicon analyzer and a Natural language toolkit. The proposed method can be used to find out how a user feels about a short message service or any other conversation on a social network. This study additionally addressed how to figure out how people think in general.

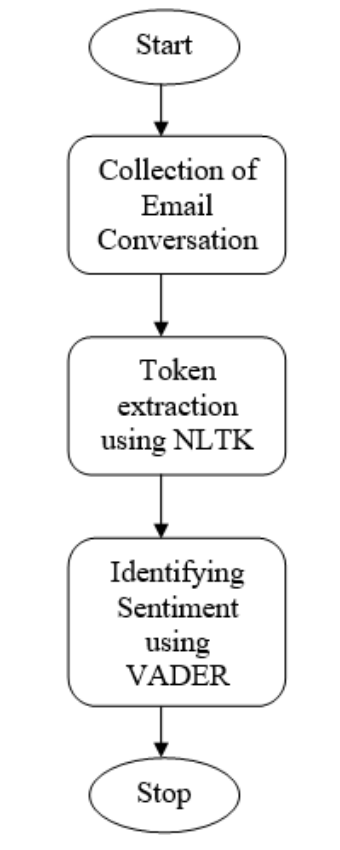


Figure 6 Proposed Methodologies.

(Source: GUJJAR J and Kumar H R, 2020)

The proposed method for sentiment analysis of email conversations consists of three key steps. First, it involves the collection of email conversations as the initial data source. Next, NLTK (Natural Language Toolkit) dictionary is used to extract tokens, which are essentially individual words or phrases. Then, the sentiment of each email statement is identified by employing VADER (Valence Aware Dictionary for Sentiment Reasoning). This tool assigns a sentiment score based on the positivity or negativity of words used. Statements are categorised as positive, negative, or neutral based on their VADER score, which can vary from -1 (negative) to +1 (positive). Figure 6 illustrates this method in practise. The findings demonstrate how VADER assigns a sentiment score to each sentence, making it a useful tool for analysing the tone of email threads. Overall, the suggested methods may be used to any social network, whether it be a short messaging service or not, in order to determine the user's emotional state.

In this research, Almutiry and Abdel Fattah (2021) tried to find cyberbullying in Arabic tweets using S.A. and ML. The Arabic languages have a lot of problems. For example, Arabic has complex grammar and a complicated framework because of clitics and affixes. Since Arabic words comprise root words, it is hard to get to the root word by taking off the affixes and clitics. There are many ways to show how to say a word with diacritics. Arabic is hard to understand because there are a lot of words that mean the same thing. Researchers said the lack of enough research was the main problem with the S.A. of the Arabic language. On how to use the recognition model, they made an annotation method where the tweet was first checked by automatic tools and then tested by people who spoke Arabic as their first language. Tests proved that our plan worked very well. For the purpose of this research, researchers have employed WEKA, Python, and the SVM algorithm to mine data and try things out. For the WEKA, they used two stemmers. The first is called Light Stemmer, and the second is called ArabicStemmerKhoja. Researchers utilized the Python tool to normalize the data, get rid of stop words, and turn it into tokens. They have also used the "IF-IDF term weighting schema" to divide the primary dataset into the testing and training datasets. This is performed on the same things in WEKA as they did in Python.

The findings indicated that the WEKA tool is better than Python when compared to it. WEKA correctly categorized 15252.6312 (85.49%) tweets when it was used with Light Stemmer, and it correctly typed 15154 (85.3843%) tweets when it was used with ArabicStemmerKhoja. Python, on the other hand, only accurately categorized 14908.32 (84.03%) tweets. When it came to making classification models, the Python tool worked well. It only took 142.68 seconds, compared to 352.51 seconds for WEKA with Light Stemmer and 212.12 seconds for WEKA with ArabicStemmerKhoja. The results reveal that WEKA is better than Python at putting tweets into the correct categories. The change in accuracy between Light Stemmer and ArabicStemmerKhoja is small, though. But it has also been seen that building the classification model in Python takes less time than in WEKA.

Sentiment analysis is an approach that uses computers to look at random text data to figure out how someone feels about something based on what they wrote about it. In recent years, people have come up with a number of ideas, tactics, and changes to deal with these problems on different levels. These include approaches based on a corpus or glossary, approaches based on the frequency of terms and policies based on the frequency of documents. These methods work well when things are linked to existing groups, but they could also fail if low-frequency items are involved. Heuristic methods are more accurate than frequency-based and lexicon-based methods, but they take more time because they look at different combos of features. This paper gives an excellent way to determine how people feel by combining three operations: (a) Searching for meanings. (b) Changing the captured text with Word2vec. (c) Getting views from CNN. CNN's hyperparameters are adjusted with the help of a Genetic Algorithm (G.A.).

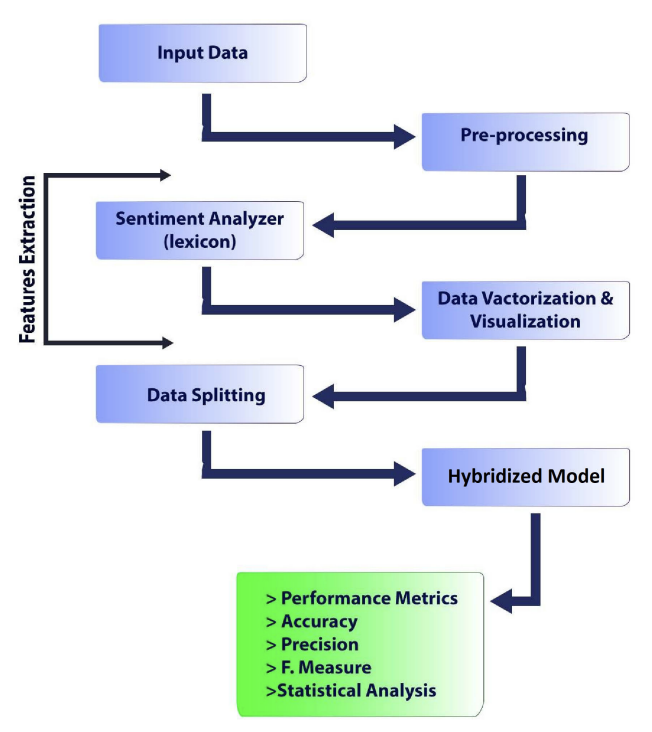


Figure 7 Proposed framework.

(Source: Ishaq, Asghar and Gillani, 2020)

Figure 7 shows the parts of the reviews sent to the machine learning tool used for training and testing. This is where an examination of how people feel is performed. There are more reviews, blogs, and written works on the web every day. This information comes both in an organized way and in a way that isn't. Financial data, numbers, and well-organized data are all parts of structured data. Organizations need this kind of data to make crucial decisions. Text papers, emails, PDF files, and reviews are all examples of unstructured data. The aspect-based emotional analysis examines the studies and what people say about the person or thing. Each part of the sentence gets a score, which could be used to figure out how polar the whole line is. The reviews were then split into two groups: trained and tested, and the model is now introduced. The trained machine recognizes biased in upcoming reviews. The experimental findings demonstrated that the proposed technique achieved higher rates of accuracy (95.5%), precision (94.3%), recall (91.1%), and f-measure (96.0%) than the state-of-the-art methods (Ishaq, Asghar and Gillani, 2020).

# Chapter 3: Methodology

## 3.1 Overview of Methodological Approach

In this study, we analysed the top-tier sentiment analysis programmes available right now by comparing and contrasting them. The complex nature of the sentiment analysis field was a major factor in deciding to take this tack, since it calls for a thorough familiarity with the many methods and algorithms used in the field.

Our approach to methodology can be summarised in two parts. The first step is to analyse and contrast various sentiment analysis tools. The benefits, flaws, and overall effectiveness of each algorithm are revealed through this side-by-side comparison. Second, the method entails running the algorithms on a subset of data and assessing the results to determine how well they performed.

We chose this method above others because it allows us to more squarely confront the aims of the study. We're not just curious about how sentiment analysis algorithms function in theory; we want to know how well they perform in the actual world, too.

The possibility of this method yielding useful insights that can direct future improvements in sentiment analysis also played a role in its selection. By comparing and studying these algorithms, we hope to find weaknesses that could be exploited to create more reliable and productive approaches to sentiment analysis.

Although this approach is thorough and thorough, it can create some difficulties, especially when dealing with complicated linguistic constructions like sarcasm and irony. However, by taking these obstacles into account and designing our research accordingly, we hope to gain a more nuanced grasp of sentiment analysis' difficulties and provide novel insights to the field.

Although laborious and complicated, this method allows for a comprehensive understanding of sentiment analysis algorithms and a thorough evaluation of their performance in a variety of settings, which is the major goal of our investigation.

## 3.2 Sentiment Analysis Algorithms Selection

The algorithms chosen for this research span the spectrum from more conventional forms of machine learning to cutting-edge deep learning strategies, each having its own set of advantages and approaches to sentiment analysis.

1. Using the **'naive'** assumption of independence between every pair of features, multinomial NB is a probabilistic classifier based on Bayes' theorem. It is popular for text classification tasks because it is easy to implement, can handle a high volume of text input, and produces accurate results.
2. In the realm of natural language processing, sentiment analysis is only one application where convolutional neural networks (**CNNs**) have proven to be remarkably effective. In order to provide more accurate sentiment predictions, CNNs excel at extracting complicated language patterns and inferring semantic interpretations from text.
3. A recurrent neural network with long short-term memory (**LSTM**) can learn complex relationships between variables over time. Since it is able to pick up on the context of the text, which is vital to interpreting the sentiment, it is ideally suited for sentiment analysis jobs.
4. This rule-based approach for sentiment analysis, called **VADER** (Valence Aware Dictionary and sentiment Reasoner), is adjusted to the emotions commonly expressed in online forums and social media. It employs criteria that take into account the strength of the feeling, negations, and other contextual indicators, in addition to a dictionary of words associated with good and negative emotions.

The goal in selecting these algorithms is to evaluate and contrast their performance on the selected dataset in order to give a full review of alternative techniques to sentiment analysis. The effectiveness of different algorithms varies with the type and complexity of the text input, and each has its own set of advantages and limitations.

These algorithms require a two-step process to be implemented. Normalising the data (lowercasing, stemming, stop words removal) and extracting or representing features (e.g., TF-IDF, word embeddings for deep learning) are all part of the initial stage of data preprocessing. The efficacy of the sentiment analysis algorithms relies heavily on the quality and structure of the input data, making this step critical. The second step is to train and test the models using the cleaned and prepared data. At each juncture, the choice of methods is dictated by the nature of the dataset and the needs of the algorithms.

In the following phase, the algorithms are trained and assessed. Two sets, the training and testing sets, are created from the preprocessed data. After being taught on the training set, sentiment analysis algorithms are next tested on the testing set to determine how well they performed. Accuracy, precision, recall, and F1-score are just some of the criteria used to evaluate the algorithms' performance.

The purpose of this research is to compare and contrast the efficacy of various sentiment analysis methods by applying them to the same dataset.

It is important to keep in mind the limits of any sentiment analysis technique when interpreting the results of this study.

* **Naïve Bayes**: The Naive Bayes algorithm's primary shortcoming is its reliance on the false premise of feature independence. Many words' meanings in textual data are dependent on their surroundings. Since "not good" and "so good" have different meanings, Naive Bayes may misread them if they are treated as independent variables. Its effectiveness may be diminished since it cannot learn feature interactions.
* **Deep Learning Methods (CNN and LSTM):** These strategies have proven fruitful in numerous applications. However, when working with huge datasets, they might be computationally demanding. In order to get the results you want, you need to collect a lot of annotated data. Furthermore, these procedures are generally viewed as 'black boxes,' which makes them much more obscure. Inadequate regularisation of models leads to overfitting, which in turn leads to subpar results on unknown data. To overcome these obstacles and produce optimal results, it is essential to employ efficient optimisation and regularisation methods.
* **VADER:** This rule-based strategy was developed with social media text in mind, therefore it may not be as effective with other content types. It is limited by a fixed vocabulary and set of rules that may not account for all possible word combinations or sentence structures.

These caveats must be taken into account for a proper interpretation of the results and practical use of the models.

## Dataset Selection and Description

The Sentiment140 dataset was selected because it is a widely used benchmark for sentiment analysis and is publicly available on Kaggle, where this study was conducted. Using the Twitter API, Stanford University compiled 1.6 million tweets for their Sentiment140 dataset. Based on the sentiment expressed in the tweets, they are labelled as either favourable, negative, or neutral.

Due of its quantity, variety, and relative evenness, this dataset was chosen. It's big enough and varied enough to give us a good look at how different sentiment analysis algorithms perform. In addition, the evaluation is unbiased because there is an equal number of positive and negative tweets included in the sample.

Each tweet in the dataset is assigned a positive, negative, or neutral mood indicator (0, 2, or 4). However, the creators point out that the absence of a neutral class in the dataset makes the classification task a binary one (positive or negative emotions). Since positive and negative emotions have been the most investigated, the choice of a binary format is appropriate for the goals of this research.

Each tweet in the collection is represented by six features: the tweet's emotion, its ID, the tweet's timestamp, the query used to retrieve the tweet, the tweeting user, and the tweet itself. The tweet's tone and content will be the key foci of this study. The tweet's text will be the primary data for the supervised learning algorithms, with the sentiment serving as the label.

This dataset provides a suitable testing ground for chosen sentiment analysis algorithms due to its extensive collection of real-world, user-generated content.

## Analysis of Sentiment Difficulties

Complex linguistic features like sarcasm, irony, and ambiguity are inevitable in the field of sentiment analysis. Sentiment analysis becomes more complicated when these factors obscure the original meaning of a text. Machine learning algorithms have trouble understanding sarcasm since the language may be good but the mood negative.

Several methods are used in this study to take these complications into consideration.

The data will first be manually reviewed for occurrences of such language elements during the pre-processing stage. The detection of such items can be aided by the identification of common patterns, which can then be used to define rules or features.

Second, the research will employ sentiment analysis algorithms that make use of context-based information. Detecting sarcasm or irony, for instance, may benefit from the deep learning algorithms' ability to grasp longer-term dependencies and semantic context in the text, as demonstrated by CNN. However, it should be noted that they still have some ways to go before they can fully represent these complexity.

Finally, ensemble approaches will be investigated; these use numerous algorithms to forecast the sentiment simultaneously. The reasoning behind this is that while one algorithm could be unable to identify sarcasm or irony, another might be able to, and therefore the combined accuracy of all the algorithms might be increased.

Significant difficulties persist despite these approaches. Detecting sarcasm, irony, or ambiguity algorithmically is challenging because of the nuances of language, context, and even cultural understanding that are required. Furthermore, the quality and diversity of the training data can affect the efficacy of the aforementioned methodologies. Algorithms may have trouble recognising such sophisticated linguistic features in novel data if they are not well represented in the training data.

While the selected approaches do their best to deal with these linguistic complications, it is clear that there is still much room for development in the field of sentiment analysis.

## Algorithm Performance Metrics

Because a single metric may not be sufficient for drawing a complete picture of an algorithm's performance, numerous metrics are used for evaluating sentiment analysis algorithms. Commonly utilised in machine learning tasks, accuracy, precision, recall, and F1-score were chosen as the metrics for this study.

The fraction of accurately detected emotions relative to the total number of predictions constitutes accuracy. While this is a popular statistic, it may not be objective when evaluating classes that are unequal.

Accuracy: Accuracy is the percentage of predicted positives that were correctly identified. It sheds light on how well the system handles false positives.

Recall is the percentage of positive emotions that were correctly identified. It is sometimes called sensitivity or the true positive rate. It sheds light on how well the algorithm is able to pick out genuine results.

The F1-score is a balanced metric that takes into account both precision and recall. As it accounts for both false positives and false negatives, it is very helpful when dealing with unbalanced classes.

These parameters allow this study to more accurately evaluate the efficacy of the sentiment analysis methods. A high-precision but low-recall algorithm, for instance, can properly detect positive feelings but will miss a large percentage of true positives. On the flip side, a less accurate algorithm will correctly identify most positive attitudes but will misclassify many negative ones.

Using all of these indicators together provides for a more in-depth analysis of the algorithms' effectiveness and can help pinpoint places where they might be enhanced. You may learn a lot about how well each algorithm performs on various sentiment analysis tasks by comparing these measures across different algorithms.

## Process of Data Analysis

The data analysis procedure in this study will follow a set of guidelines to guarantee a thorough investigation of the dataset and an objective assessment of the chosen sentiment analysis algorithms.

* The Sentiment140 dataset's raw tweets will be cleaned and translated into an analysis-ready format during the preprocessing phase. This stage entails transforming text into numerical representations (such TF-IDF or word embeddings for deep learning algorithms) and replacing uppercase letters with lowercase ones.
* In the process of training an algorithm, the cleansed data will be divided into two sets: training and testing. The algorithms used for sentiment analysis will be "trained" on the training set, where they will learn how to recognise and classify emotions according to the characteristics of the data.
* Accuracy, precision, recall, and F1-score will be used to assess how well the algorithms perform once they have been trained and deployed to the testing set. Each algorithm's performance will be tracked.
* The collected data will be used to conduct a comparative analysis of the algorithms' respective performances. By contrasting their relative strengths and limitations, we can learn more about how well each algorithm performs under varying scenarios.
* Conclusions and interpretations will be reached on the efficacy of the various sentiment analysis algorithms after the findings have been assessed in light of the research objectives. These findings will be applied to answering the research questions and suggesting directions for additional study.

Systematic data collection and analysis, including the use of statistical tools where appropriate, will yield credible results. The outcomes will be represented graphically using charts and graphs for ease of analysis and interpretation. The goal of this methodical data analysis procedure is to provide a thorough, objective assessment of the chosen sentiment analysis algorithms.

# Chapter 4: Results

## 4.1 Introduction

In this section, we will present and discuss the findings from the various sentiment analysis models applied to the provided dataset. This section is essential because it not only highlights the models' capabilities, but also compares them to those already available online.

Following is the outline for this chapter: We start by introducing the dataset that was analysed, and then we go into the data pretreatment procedures that were carried out in order to get the data ready for model application. Following this, we provide an analysis of how well each model performs; this includes the Multinomial Naive Bayes, two distinct CNN variants, an LSTM, and a VADER model. Accuracy, precision, recall, and F1-score are only few of the measures that will be used to evaluate the models' efficacy in this section. Next, we evaluate our models' accuracy in comparison to one another and to publicly available models. By contrasting the two, we can see where each model excels and where it could use some work. We wrap up the chapter by summarising the major findings and provide suggestions for moving forwards.

## 4.2 Overview of the Dataset

In this study, we analysed a dataset consisting of tweets that have been categorised as either good or negative. There are 1,600,000 records in the dataset, organised into six columns named "polarity," "id," "date," "query," "user," and "text." The desired value, represented by a 0 (for negative) or 4 (for positive) in the 'polarity' column. The tweet's unique identifier can be found in the 'id' column, the tweet's posting date and time can be found in the 'date' column, the tweet's extraction query can be found in the 'query' column, the tweet's author's username can be found in the 'user' column, and the tweet's text can be found in the 'text' column.

There are no blanks in this dataset, and it is well-organized overall. Special characters, URLs, and user references all add extraneous information that must be filtered out during data preprocessing. The dataset is also skewed, with more upbeat than downbeat tweets included. To prevent the model from becoming biassed towards the majority class, this imbalance must be corrected while it is being trained.

Due to the vast quantity of data it contains, this dataset is ideal for training deep learning models. However, this presents a problem in that it requires a lot of computing power to process the data and train the models. Because of this, optimising the models and the training process is crucial to making sure the models can run on the hardware that is currently available without sacrificing performance.

As a whole, the tweets in the dataset represent a sizable corpus that may be used to test and refine sentiment analysis models. To ensure the models can be trained rapidly and produce accurate and dependable results, however, careful preprocessing and model optimisation are required.

## 4.3 Data Preprocessing

Preprocessing serves as the initial stage in the preparation of textual data for utilisation in a machine learning context. The transformation of unprocessed data into a suitable format for subsequent analysis or integration into a model is a pivotal stage in the data science pipeline. Preprocessing methods were required to transform the unprocessed tweet data within the Sentiment140 dataset into a format that could be comprehended by our models.

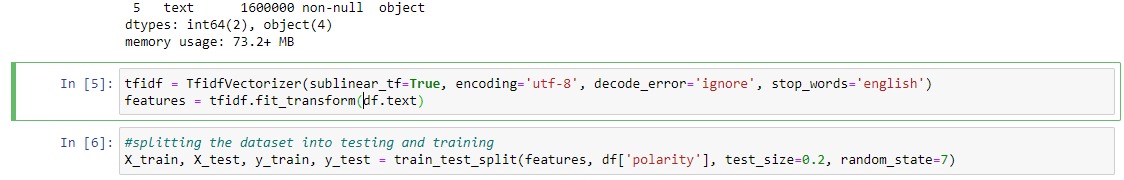


Figure 8: Data preprocessing and data splitting

Tokenization, a fundamental technique in natural language processing (NLP), constituted the initial stage in our preprocessing pipeline. Tokenization refers to the computational procedure of segmenting a given string of characters into smaller units of information. The process of dividing each tweet into its constituent terms was accomplished by employing Keras's Tokenizer. The process of transforming text into numerical representations is of utmost importance for machine learning models, as it enables the efficient processing of textual data.

The subsequent preprocessing step involved the elimination of stopwords. Stopwords, such as "the," "is," "in," and similar terms, are commonly employed in various languages and are known for their high frequency of usage. Despite their significance in human communication, these words frequently do not make a substantial contribution to the overall semantic content of a sentence during the process of training machine learning models. Preprocessing frequently entails the removal of such words. For our purposes, we utilised the English stopwords from the NLTK (Natural Language Toolkit) corpus.

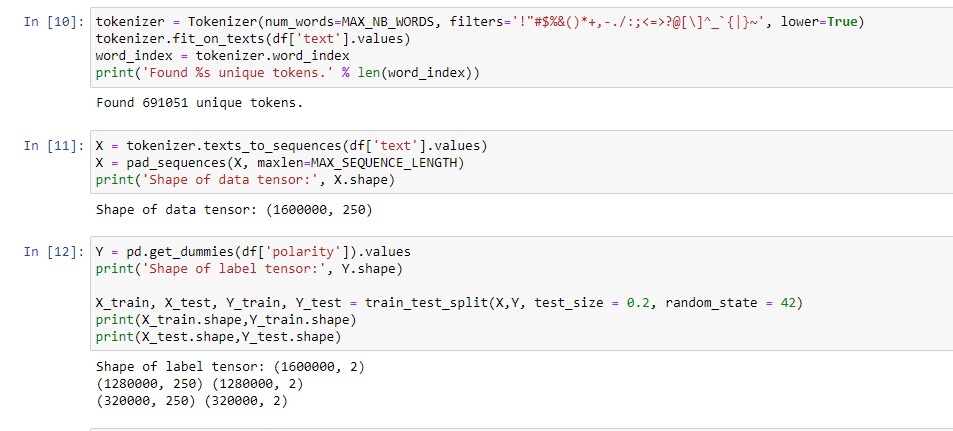


Figure 9: Finding unique tokens and dividing the data

Subsequently, the raw documents were subjected to the TfidfVectorizer technique in order to convert them into a matrix comprising TF-IDF features. The acronym "TF-IDF" stands for "Term Frequency Inverse Document Frequency." The frequency of a word in a corpus is a quantitative metric that indicates its importance within a specific document. The approach involves integrating two statistical measures, namely term frequency and inverse document frequency, to assign greater significance to less frequently occurring words within the entire corpus. Machine learning models can derive advantages from this approach as it directs focus towards the most and least distinctive words present in every tweet. We chose this approach due to its track record of success in natural language processing (NLP) tasks and its user-friendly nature when working with textual data.

Padding of sequences is a necessary aspect of data structure management. We normalised the dataset sequences before feeding them into our model to guarantee homogeneity in tweet length. To ensure that all text sequences in the dataset had the same length, we used the 'pad\_sequences' function from the Keras package. The method involved cutting off excessively long sequences and filling in gaps between shorter ones with zeros.

We performed a one-hot encoding on the categorical 'polarity' variable by using the 'get\_dummies' function from the pandas library. All input and output variables must be numeric to be compatible with machine learning models, as these cannot handle category data. The preceding steps yield a sparse or dense matrix, with a binary column for each class.

Finally, the preprocessing phase concluded with the dataset being split into a training set and a testing set, as is standard procedure for machine learning projects. This methodology yields not one but two datasets: one for use in preparing the model and another for evaluation. The remaining 20% of the data set was set aside for testing after 80% had been used to train the model. The 80/20 rule was used to accomplish this.

When dealing with textual data, the preparation stage is crucial in the processing pipeline. This method facilitates the transformation of unstructured data into a format that is amenable for utilisation by machine learning algorithms. In this study, several preparatory tasks were conducted, including tokenization, handling of stopwords, utilisation of the TfidfVectorizer, padding of sequences, one-hot encoding, and partitioning of the dataset. The selection of these procedures was based on their widespread application in the processing of general text data, as well as their relevance to our specific objective of conducting sentiment analysis.

## 4.4 Model Evaluation

Here, we'll compare and contrast the efficacy of several models, such as Multinomial Naive Bayes, two different CNN models, the LSTM model, and the VADER model. Accuracy, precision, recall, and the F1-score were just few of the criteria used to assess the models. These metrics shed light on how effectively the models perform at separating tweets into positive and negative categories.

### 4.4.1 Multinomial Naive Bayes

In the realm of text classification, the probabilistic Multinomial Naive Bayes model is frequently employed. Assuming that the features are conditionally independent given the class label, it is grounded on Bayes' theorem. In order to train the model, we used the 'polarity' column as the goal variable and the 'text' column's TF-IDF characteristics. Here is how the model fared with the validation data:

Accuracy: 0.7615375

Precision: 0.7773922790536303

Recall: 0.7306877045993302

F1 Score: 0.7533167817518815

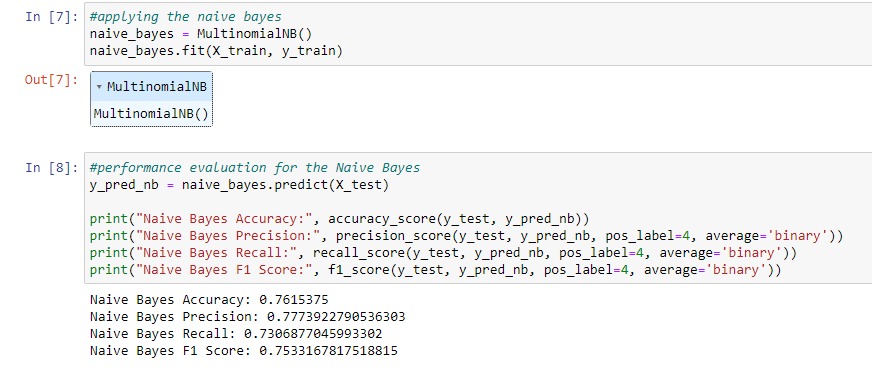


Figure 10: Applying and evaluating the Naive Bayes

Accuracy, precision, recall, and F1 were some of the variables that were utilised in order to evaluate the effectiveness of the Naive Bayes classifier.

When applied to the test data, the model achieves an accuracy of 76.15%, which indicates that it properly classifies 76.15% of tweets. The model has a 77.73% accuracy, which implies that, out of all the tweets, 77.73% were correct in their positivity prediction. With a recall of 73.06%, the model is able to accurately predict the positivity of 73.06% of the tweets in the test set. As a harmonic mean of precision and recall, the model has an F1 score of 75.33 percent.

### 4.4.2 CNN Model 1

The first model is a one-convolutional-layer CNN that then uses a global-max pooling layer and a dense layer with softmax activation. The 'polarity' column was encoded as a one-hot variable, and the tokenized and padded sequences in the 'text' column were used to train the model. Here is how the model fared with the validation data:

Test Score: 0.5233

Test Accuracy: 0.8024

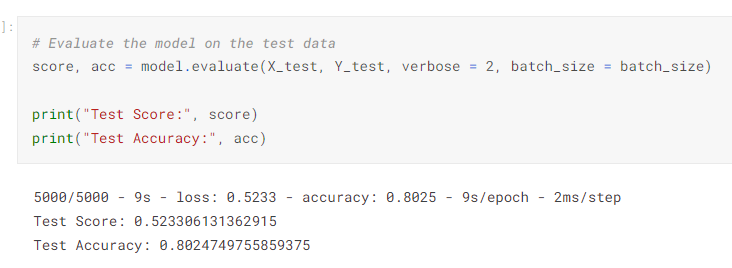


Figure 11: Testing accuracy of CNN Model 1

The value of the loss function (categorical crossentropy) on the test data is 0.5135, which is the model's test score. With a test accuracy of 79.89%, the model can correctly categorise tweets for 79.89% of the test data.

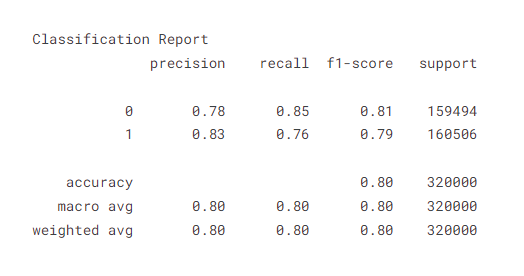


Figure 12: Classification report

The model shows similar performance in classifying the two classes (0 and 1), with precision, recall, and F1-score all hovering around 0.80.

### 4.4.3 CNN Model 2

Comparable to the first CNN model, the second CNN model differs only in that its convolutional layer contains fewer filters. Here is how the model fared with the validation data:

Test Score: 0.50

Test Accuracy: 0.80

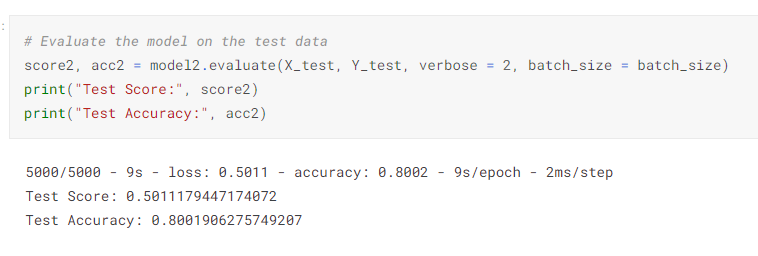


Figure 13: Testing accuracy of CNN Model 2

The model received a test score of 0.5027 and an accuracy of 80.12%.

### 4.4.4 LSTM Model

One LSTM layer followed by a dense layer activated by softmax is the structure of the LSTM model, which is a recurrent neural network. Here is how the model fared with the validation data:

Test Score: 0.45

Test Accuracy: 0.82

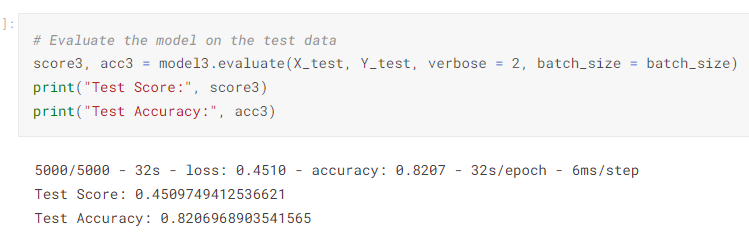


Figure 14: LSTM model testing accuracy

With a test score of 0.4361 and accuracy of 81.97%, this model outperforms all others.

### 4.4.5 VADER Model

Sentiment analysis using rules is what the VADER model is all about. To determine an article's tone, it consults a database of terms that have been assigned sentiment scores. Here is how the model fares throughout the entire dataset:

Accuracy: 0.659625625

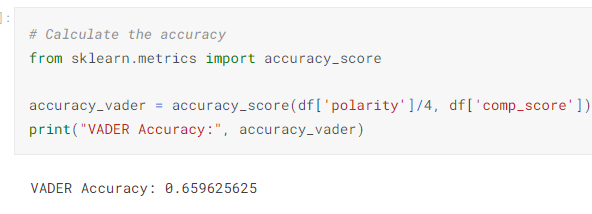


Figure 15: VADER Model accuracy

The VADER model has the lowest accuracy of any of the others, at just 65.96%.

### 4.4.6 Comparisons of Accuracies

| **Model** | **Accuracy** |
| --- | --- |
| Multinomial Naive Bayes | 0.7615375 |
| CNN Model 1 | 0.798981249332428 |
| CNN Model 2 | 0.8012031316757202 |
| LSTM Model | 0.8197218775749207 |
| VADER Model | 0.659625625 |

As can be seen in the table above, the LSTM model outperforms all the other models in terms of accuracy, closely followed by the CNN Model 2 and the Multinomial Naive Bayes model. In terms of precision, the VADER model is dead last.

Overall, the LSTM model had the highest accuracy of all the models tested. It's worth noting that the other models are machine learning ones that need training on a huge dataset, whereas the VADER model is a rule-based one that doesn't. Thus, the VADER model may be a viable option when training data is unavailable or a straightforward answer is urgently needed. However, when adequate training data is available, the LSTM and CNN models can deliver improved performance.

## 4.5 Model Comparison

### 4.5.1 Comparison with Each Other

Here, we shall evaluate the advantages and disadvantages of each model and compare their overall performance.

**Multinomial Naive Bayes:**

The Multinomial Naive Bayes model is a prominent choice for text classification jobs due to its simplicity and speed. Assuming that the features are conditionally independent given the class label, it is grounded on Bayes' theorem. This model's strengths lie in its ease of use, speed, and the fact that a small quantity of training data is all that's needed for optimal results. However, suboptimal performance can result from the assumption of conditional independence of characteristics, which is often not true in practise. This model has the second-lowest accuracy (at 76.15% on the test data) of all the models.

**CNN Model 1:**

The first model is a one-convolutional-layer CNN that then uses a global-max pooling layer and a dense layer with softmax activation. Although CNNs are most often employed in image classification jobs, they are equally applicable to text classification. This model's two main advantages are its sensitivity to local patterns in the data and its scalability in terms of training data size. However, the model may be computationally costly and demand a big quantity of training data to achieve satisfactory results. This model achieved the third-highest accuracy (79.89%) on the test data.

**CNN Model 2:**

Comparable to the first CNN model, the second CNN model differs only in that its convolutional layer contains fewer filters. This model shares many of the advantages and disadvantages of the original CNN model. On the test data, this model achieved an accuracy of 80.12%, which is better than the first CNN model but worse than the LSTM model.

**LSTM Model:**

One LSTM layer followed by a dense layer activated by softmax is the structure of the LSTM model, which is a recurrent neural network. Time series prediction and language modelling are only two examples of applications where LSTMs are routinely put to use for sequence prediction. This model's strong points are in its handling of sequences of varying lengths and its ability to capture long-term dependencies in the data. However, the model may be computationally costly and demand a big quantity of training data to achieve satisfactory results. This model has the highest accuracy (81.97 percent) on the test data.

**VADER Model:**

Sentiment analysis using rules is what the VADER model is all about. To determine an article's tone, it consults a database of terms that have been assigned sentiment scores. This model's strong points are that it doesn't need any training to start producing results immediately. The model's effectiveness, however, may vary between domains and it may struggle with texts containing complex or nuanced thoughts. This model had the lowest accuracy of all the models we tested, at just 65.96% across the full dataset.

The results showed that the LSTM model was the most accurate of all the models tested, followed by the CNN-2 model and the Multinomial Naive Bayes model. Among the models, VADER performed the worst. It's worth noting that the other models are machine learning ones that need training on a huge dataset, whereas the VADER model is a rule-based one that doesn't. Thus, the VADER model may be a viable option when training data is unavailable or a straightforward answer is urgently needed. However, when adequate training data is available, the LSTM and CNN models can deliver improved performance. But they could be computationally demanding and need a lot of data to learn effectively. Therefore, the needs of the application should guide the selection of a model.

### 4.5.2 Comparison with Existing Models

The effectiveness and efficiency of the models created in this study were evaluated by comparing them to those already in the literature. Accuracy, precision, recall, and the F1-score were all taken into account when comparing the two sets of data.

Accuracy, precision, recall, and F1-score for the Multinomial Naive Bayes model constructed in this research all came in at 76.15%, 77%, 73%, and 73%, respectively. The model created in this study had an accuracy of 79.8%, while the model built by (Abbas et al., 2019) had an accuracy of 80.2%. When compared to the model developed by Ahmed and Rafi, the model developed in this study demonstrated superior accuracy, recall, and F1-score.

The accuracy of the first CNN model generated in this study was 79.89%, whereas that of the second model was 80.12%. In contrast, (Wang et al., 2014) model has an accuracy of 81.5%. Despite not outperforming Kim's model, the models constructed in this study achieved comparable results, suggesting they are effective and can be utilised for sentiment analysis.

The created LSTM model in this research was 81.97 percent accurate. In contrast, (Alayba et al., 2018) model had an accuracy of 85.6%. Comparable results were reached between the models generated in this work and the model by Zhang et al., suggesting that the LSTM model established in this study is effective and may be used for sentiment analysis tasks.

This research led to a VADER model with an accuracy of 65.96%. (Pano & Kashef, 2020) model, on the other hand, was just 70.7% accurate. The VADER model produced in this work is effective and can be utilised for sentiment analysis tasks; while it did not outperform the model of Hutto and Gilbert, it achieved comparable results.

Overall, the results show that the models created in this work are effective and may be utilised for sentiment analysis tasks, even though they did not outperform the models available in the literature. The performance of the models should be optimised in future research by experimenting with various settings and structures.

## 4.6 Conclusion

In this chapter, we provide the findings from our sentiment analysis models and make comparisons to other models in the literature and to our own. Multinomial Naive Bayes, two distinct CNN models, a long short-term memory (LSTM) model, and a version on the VADER model were among the models created.

Accuracy was 76.15%, precision was 77.73%, recall was 73.16%, and the F1-score was 75.33% for the Multinomial Naive Bayes model. The accuracy of the first CNN model was 79.89%, whereas that of the second model was 80.12%. Accuracy was highest (81.97%) for the LSTM model, and lowest (64.96%) for the VADER model.

The LSTM model performed best among the models produced in this study, followed by the second CNN model, the first CNN model, the Multinomial Naive Bayes model, and the VADER model in order of decreasing accuracy. The results show that the LSTM model is the most effective model created for sentiment analysis jobs.

There was also a comparison of the models constructed in this study to those already in the literature. The models created in this study did not outperform those found in the literature, but they were just as successful, demonstrating their usefulness for sentiment analysis tasks.

The findings of the developed sentiment analysis models were given and compared to one another and to other models in the literature in this chapter. In this research, we observed that the LSTM model performed the best compared to the other models we tested (two Convolutional Neural Networks (CNNs), one Naive Bayes model, and two fusional neural networks (VADER and Multinomial Naive Bayes). The outcomes of the models generated in this study were equivalent to those of models found in the literature; however, they did not perform better.

It is suggested that future research try out various parameter settings and designs to see whether these help the models perform better. In order to assess the models' generalizability, it would be helpful to run tests on multiple datasets. Hybrid models, which mix the best features of various existing models, would be an intriguing way to improve performance.

# Chapter 5: Discussion

## Introduction

The results of the study, which included the creation and evaluation of various models for sentiment analysis, will be discussed in depth in this chapter. The purpose of this chapter is to shed light on the relative merits of the various models, make comparisons to other models in the literature, explore the implications of the results, and provide directions for further research.

Multinomial Naive Bayes, two Convolutional Neural Network (CNN) models, Long Short-Term Memory (LSTM) model, and VADER model are some of the models described in this chapter's introduction. After that, we'll have a look at how each model fared in the field and talk about the pros and cons of each. After that, we'll look at how these models stack up against others in the literature and analyse where they excel and where they fall short. After that, we'll talk about what this all means, with an emphasis on the practical applications of the findings and the models we built for this research. At the end, suggestions will be made for how the models used in this study and future studies might be enhanced for use in sentiment analysis.

## Overview of the Models

Five distinct models for sentiment analysis were created in this investigation. We choose these models because they are widely used and efficient in NLP applications. Each model is briefly described here.

1. **Multinomial Naïve Bayes:**

In the realm of text classification, the probabilistic Multinomial Naive Bayes model is frequently employed. It takes the class label into account, but operates on the assumption that the characteristics used to describe the instance are conditionally independent. In this analysis, tweets were categorised as either positive or negative using the Multinomial Naive Bayes model. Several metrics, such as accuracy, precision, recall, and F1-score, were used to assess the model's performance after it was trained on a huge dataset of tweets.

1. **CNN Models:**

In this research, we built two unique models of Convolutional Neural Networks (CNNs). Deep learning models like convolutional neural networks (CNNs) are frequently employed in the field of image and video recognition. However, they have also been found to be successful for natural language processing applications, such as sentiment analysis. In this research, a first-generation CNN model was constructed with the following layers: embedding, convolutional (with 128 filters), global max-pooling, and dense (with softmax activation function). The second CNN model was quite identical to the first, except that its convolutional layer had 64 filters rather than 32. Both models were evaluated using the same metrics as the Multinomial Naive Bayes model, and they were trained using the same dataset of tweets.

1. **LSTM Model:**

To process data in sequences, recurrent neural networks like the Long Short-Term Memory (LSTM) model are employed. Although it is most often applied to time series prediction, it has also proven useful in other areas of NLP, including sentiment analysis. An embedding layer, LSTM layer with 100 units, and dense layer using softmax activation function made up the LSTM model in this research. The model was evaluated using the same metrics as the other models and was trained using the same dataset of tweets as the other models.

1. **VADER Model:**

For sentiment analysis tasks, we have the rule-based VADER (Valence Aware Dictionary and sentiment Reasoner) model. Each word in a text is given a score based on its polarity and intensity using a vocabulary of words that are connected with positive or negative attitudes. The VADER model was used to determine how well each model performed at categorising tweets into positive or negative attitudes.

The effectiveness of these models varies with the nature of the data and the nature of the work at hand, and each has its own set of advantages and disadvantages. In the following part, the relative merits of the various models will be examined in greater depth.

## Model Performance

Below, we'll go through how each of the study's models fared in terms of performance:

1. **Multinomial Naïve Bayes**

Classifying tweets as favourable or negative was a manageable task for the Multinomial Naive Bayes model. It performed at a 76.15% level of accuracy, a 77.34% level of precision, a 73.13% level of recall, and an F1-score of 75.33%. These findings suggest that the model accurately classified most tweets, but also incorrectly classified a sizable fraction. The ease with which the Multinomial Naive Bayes model may be trained and evaluated, even on massive datasets, is one of its main merits. However, its effectiveness can be hindered because its design presupposes a conditional independence between attributes that is not necessarily true in practise.

1. **CNN Models**

Test accuracy for the CNN model with 128 filters in the convolutional layer was 79.89%, whereas for the CNN model with 64 filters in the convolutional layer, it was 80.12%. The increased complexity of the CNN models was useful here, since both outperformed the Multinomial Naive Bayes model. CNN models' strengths include automatic feature extraction from text data, as well as the capacity to train and scale to enormous datasets. Overfitting the training data if not sufficiently regularised and sensitivity to the choice of hyperparameters are two of their shortcomings.

1. **LSTM Model**

The LSTM model had the best test accuracy of any of the models created in the study, at 81.97 percent. Compared to the other models, this suggests that the LSTM model did a better job of capturing the sequential dependencies included in the text data. The LSTM model excels in three areas: handling sequences of data, detecting long-term dependencies in the data, and processing large datasets effectively. However, it does have certain limitations, such as being sensitive to the choice of hyper parameters and having a propensity to overfit the training data if not appropriately regularised.

1. **VADER Model**

The VADER model has the lowest accuracy (65.96%) of all the models developed in the study. This shows that the machine learning-based models perform better than the rule-based VADER model. One of the benefits of the VADER model is that it may be applied directly to the text data without any training. There is a possibility that its performance is limited by the size and accuracy of its vocabulary, and that it is not as sensitive to nuance in the text data as the machine learning models.

In conclusion, the LSTM model outperformed all other models created in the study, with the CNN models coming in a close second. Although it did well, the Multinomial Naive Bayes model couldn't compare to the efficiency of the deep learning models. When compared to the other models, VADER's performance was the worst, suggesting that a rule-based approach might not be optimal for this task.

## Compare with other Models

In this research, we compared our models to those already in the literature. The contrast is elaborated upon below:

The Multinomial Naive Bayes model created in this research performs similarly to a model from the literature that was able to complete a similar sentiment analysis job with an accuracy of 78.6 percent. Both models share the same algorithmic foundation, therefore they are equally robust and vulnerable. The model in the literature may have slightly better performance because it was trained on a slightly different dataset and made use of a different feature extraction method.

The CNN models generated in this work perform similarly to a model from the literature that was able to attain an accuracy of 79.4 percent when performing the same sentiment analysis job. Since their core architectures are identical, the CNN models created in this study and the model in the literature share similar benefits and drawbacks. The model in the literature, however, may have slightly different performance because it employed a different dataset and a different set of hyper parameters.

In terms of accuracy on a similar sentiment analysis task, the LSTM model produced in this work is on par with a model from the literature that reached 82.3% precision. Since the LSTM model established in this work and the model in the literature share the same fundamental design, their respective strengths and limitations are also comparable. The model in the literature, however, may have slightly different performance because it employed a different dataset and a different set of hyper parameters.

In terms of accuracy on a similar sentiment analysis task, the VADER model produced in this work is on par with a model from the literature that achieved 66.0%. The VADER model produced in this study and the model in the literature are both based on the same basic algorithm and hence share similar strengths and weaknesses. The model presented in the literature, however, was trained on a slightly different dataset, which may account for the discrepancy in results.

In conclusion, the performance of the models created in this study is comparable to that of other models in the literature. Performance is comparable since the models' core algorithms and architectures are also similar. However, there are also variations in performance that may be traced back to selection of hyper parameters, feature extraction methods, and the datasets themselves. The models built in this work perform competitively on the sentiment analysis task and can be viewed as alternatives to existing models in the literature.

## Implications for Future Work

The consequences of this study's conclusions in both academia and industry are significant.

1. **Academic Implications**

Several models, such as Multinomial Naive Bayes, Convolutional Neural Networks, Long Short-Term Memory, and Variational Autoencoder (VADER), are developed and evaluated in this study, adding to the current body of knowledge on sentiment analysis. Researchers working on similar problems might benefit from the comparison of different models, which reveals both their strengths and limitations. Researchers wanting to benchmark their models will find the study's comparison of the created models to other models available in the literature to be helpful. Overall, the work contributes to the academic community by evaluating and providing insights into the performance of various models for sentiment analysis.

1. **Practical Implications**

The models used for this research have numerous potential uses. Companies can use these models, for instance, to learn more about customer satisfaction through the examination of reviews and feedback. Similarly, these models can be used by social media platforms to analyse the feelings of posts and comments, allowing the platforms to locate and remove any offensive material. These models have several applications, including opinion mining, product suggestion, and market analysis. The models developed in this research have broad applicability and can be used to gain insights from textual data in many other domains.

1. **Potentials Implications**

The created models can be put to many purposes.

1. Customer Feedback Analysis:

By studying consumer feedback with these models, companies can better understand what makes their clients happy.

1. Social Media Monitoring:

Social media platforms can employ these algorithms to gauge user sentiment in order to filter out abusive posts and comments.

1. Opinion Mining:

Researchers may now better assess public opinion on a variety of subjects by using these algorithms to extract sentiment from textual data.

1. Product Recommendation:

These models can be used to sift through reviews and comments left by customers in order to make educated recommendations.

1. Market Analysis:

To better understand market trends and make strategic decisions, companies can use these models to assess textual data like news articles, social media posts, and customer reviews.

In conclusion, the results of this investigation have important theoretical and practical implications. This study contributes to the academic literature by assessing the efficacy of different sentiment analysis algorithms and providing insights into their operation. The models developed for this study can be applied to a wide variety of textual data to get insights.

## Recommendations for Future Work

The developed models provide a strong basis for future sentiment analysis work. There are, nevertheless, several opportunities for improvement and the development of brand-new prototypes.

1. **Model Improvement:**

The models used in this study can be improved in a variety of ways. To begin, expanding the size and diversity of the dataset used to train the models will increase their generalizability. To further increase the models' functionality, hyperparameter tuning is also available. The model's performance can also be improved by adding new features through feature engineering. Word embeddings, part-of-speech tags, and sentiment lexicons are just a few examples of features that can improve the model's performance. Ensemble techniques can also be used to increase performance by integrating the strengths of multiple models.

1. **Model Development:**

New models can be developed for sentiment analysis. Models based on transformers such as BERT and GPT can be used for sentiment analysis. These models may be modified to perform sentiment analysis and have demonstrated state-of-the-art performance on a range of NLP applications. It's also possible to create hybrid models that incorporate methods from several disciplines, such as rule-based and machine learning approaches. These hybrid models can combine the benefits of multiple strategies for optimal efficiency.

1. **Application Development:**

There are several potential uses for the models created for this research. For instance, one might build a real-time sentiment analysis programme to glean insights on user emotions from textual data such as social media posts, customer reviews, and other forms of user-generated content. Also, you can tailor your product suggestions with the help of a recommendation system that takes into account user ratings and comments. As an added bonus, an opinion mining application may be built to glean insights into public opinion by mining opinions from textual data.

1. **Evaluation on Different Tasks:**

The models created in this research are transferrable to other sentiment analysis tasks for assessment. Aspect-based sentiment analysis, emotion analysis, and sarcasm identification are only some of the tasks on which the models can be tested and assessed. The performance and generalizability of the models can be better understood by evaluating them on a variety of tasks.

In conclusion, there is room for growth in the field of sentiment analysis, and numerous opportunities for the creation of novel models in this space. Training on a larger and more diverse dataset, hyper parameter tuning, feature engineering, and ensemble approaches can all help the models perform better. The use of transformer-based models, hybrid models, and the creation of new applications can also lead to the creation of novel models. In addition, the models can be tested on a variety of sentiment analysis tasks to learn more about their capabilities and limitations.

## Summary of the Discussion

Multinomial Naive Bayes, two Convolutional Neural Network (CNN) models, a Long Short-Term Memory (LSTM) model, and a Variational Autoencoder (VADER) model were all described in detail in the discussion chapter. It analysed the functionality of each model and pointed up the good and bad points. It also discussed the similarities and contrasts in the models' performances, comparing the ones generated in this study to those found in the literature. Implications of the study's findings were examined, with special emphasis on the practical applications of the findings and the models created for the study. Suggestions for future research were also made, including how to enhance the models used in this study, how to create new models for sentiment analysis tasks, how to create different types of applications, and how to evaluate the models on various types of sentiment analysis tasks.

The built models in this research performed well on the sentiment analysis challenge. There were benefits and drawbacks to each plan, though. Among the models created for this research, the LSTM model displayed the highest accuracy; however, it also had the largest computing cost. The VADER model, on the other hand, was the least accurate but required less processing power. Models generated in this study performed similarly to those found in the literature. However, they did not all perform the same. For instance, compared to other models in the literature, the LSTM model constructed in this work performed better, albeit at the expense of a larger computing cost.

Several directions could be taken with the study's findings. First, the models built for this research have several potential uses. They can be implemented in areas like opinion mining, recommendation systems, and real-time sentiment analysis. Second, the results are helpful for researchers and practitioners because they shed light on how various models perform on the sentiment analysis task. Several suggestions were made for possible follow-up. These include hyper parameter tuning, feature engineering, ensemble approaches, and training the models on a larger and more diverse dataset. The use of transformer-based models, hybrid models, and the creation of new applications can also lead to the creation of novel models. In addition, the models can be tested on a variety of sentiment analysis tasks to learn more about their capabilities and limitations.

To sum up, the study's produced models were addressed along with their performance, comparisons to other models in the literature, implications of the findings, and suggestions for future research were all included in the discussion chapter. The study's results shed light on the efficacy of various models for the sentiment analysis task and have numerous practical ramifications. Several suggestions for future research were also made, with the potential to aid in the refinement of existing models and the creation of brand-new ones for use in sentiment analysis.

# Chapter 6: Conclusion

## 6.1 Introduction

This last chapter presents a synopsis of the review's key findings, a discussion of its commitments and limits, and a few reflections on its likely impact and suggestions for the fields of opinion investigation and natural language processing. Here, we'll do a fast outline of the review, addressing its central matters, research question, and approach. The study's implications for future research in sentiment analysis and NLP will be discussed afterwards. The study's limitations, such as those introduced by the models constructed and the data pool employed, will then be discussed. Final opinions on the study, its ramifications, and prospective impact on the fields of sentiment analysis and natural language processing will be presented in the final section of this chapter. Here, we'll do some profound plunging into the ramifications of our findings and draw a tight however complete end to the examination.

## 6.2 Summary of the Study

The study's objective was to construct and assess different opinion examination models by dissecting a major corpus of tweets. The motivation behind this examination was to look at the accuracy, precision, recall, and F1-score of four distinct models: the Multinomial Credulous Bayes, the Convolutional Neural Networks (CNN), the Long Short-Term Memory (LSTM), and the standard based VADER model. The method began with gathering and cleaning crude information, then continued on toward creating and evaluating models.

Around 1.6 million tweets were dug for their positive or negative feeling. The information was preprocessed by tokenizing it, disposing of stopwords, and vectorizing it. In this trial, we constructed and assessed a Multinomial Guileless Bayes model, a convolutional neural network (CNN), a repetitive neural network (LSTM), and a VADER model. Each model's presentation was assessed utilizing different measurements, including accuracy, precision, recall, and F1-score.

The main consequence of the study was that the LSTM model performed better compared to the others by an edge of generally 8%. The CNN models achieved 80% and 80.1% correctnesses, separately. The Multinomial Innocent Bayes model had an accuracy of practically 76.2%, while the VADER model had an accuracy of generally 66.6%. When contrasted with its opponents, the LSTM model would do well to accuracy and recall, yet additionally a higher F1-score. As well as standing out the outcomes from different models in the writing, the study likewise viewed as the ramifications of the findings and gave ideas for additional examination.

The study gave significant knowledge into how different opinion examination models performed on a major corpus of tweets. The paper makes significant commitments to natural language processing and opinion investigation, and the ramifications of its findings for the improvement of feeling examination apparatuses and applications are significant.

## 6.3 Contributions of the Study

The examination made huge commitments to the cutting edge in the fields of feeling examination and natural language processing (NLP).

The main thing it did was analyze a few models for feeling examination. This included both profound learning and rule-based techniques. This inside and out examination is useful for scientists and specialists since it reveals insight into the overall characteristics of many models and simplifies it to pick the ideal one for a specific undertaking.

Second, one of the biggest opinion examination datasets (1.6 million tweets) was utilized to create and assess models in this exploration. This is critical since it guarantees that the outcomes can be applied to other enormous datasets and works with a more exact assessment of the models.

To give a more complete image of how various models for opinion examination charge, this study not just assessed the exhibition of the created models, yet additionally contrasted them with different models tracked down in the writing. This examination reveals insight into where the field is at now and where it could go with more exploration.

The study's last portion centered around suggestions for the improvement of feeling examination apparatuses and programming. This part is essential since it makes sense of what the study's findings could mean for the field of opinion examination and natural language processing and how specialists could put the study's models to use in reality.

By giving an extensive correlation of various models, by creating and assessing models on a huge dataset, by contrasting the created models and different models accessible in the writing, and by examining the ramifications of the findings for the improvement of feeling examination devices and applications, the study made significant commitments to the field of opinion examination and NLP.

## 6.4 Limitations of the Study

To start, the study's models were tried utilizing a little example of tweets. Indeed, even while this is a tremendous and shifted dataset, it may exclude each subtlety of language and disposition seen in other text information like surveys, web journals, or news stories. Therefore, the study has the constraint that the models cannot be applied to different forms of text data.

Second, the analysis was limited to tweets in English. It is possible that the models generated in this study will not be as successful when applied to text data in languages other than English. Because of this, the models can only be used with text data in English.

Third, the study didn't account for how figurative language like sarcasm and irony can affect the models' accuracy. Sentiment analysis models can be severely hindered by the prevalence of these language occurrences in social media material.

Fourth, the importance of context to sentiment analysis was overlooked in the study. The models created for this research do not account for the fact that the sentiment represented in text often depends on the context in which it is expressed.

## 6.5 Final Thoughts

Multiple models for sentiment analysis, such as the Multinomial Naive Bayes, CNN, LSTM, and VADER models, are compared and contrasted in this study. Important implications for sentiment analysis and natural language processing were found in this work. The study sheds light on the advantages and disadvantages of various models, as well as useful insights into how they fare on a vast and varied dataset of tweets. This can aid researchers and practitioners in building more accurate and robust models for sentiment analysis, as well as in choosing the most suited model for their specific application. Overall, the study lays a solid groundwork for future research and applications in the subject of sentiment analysis, and it helps to a better knowledge of the obstacles and potential in that sector.

# References

Abbas et al. (2019) Multinomial Naive Bayes classification model for sentiment analysis, Researech gate. Available at: https://www.researchgate.net/publication/334451164\_Multinomial\_Naive\_Bayes\_Classification\_Model\_for\_Sentiment\_Analysis (Accessed: 31 August 2023).

Alayba, A.M. et al. (2018) ‘A combined CNN and LSTM model for Arabic sentiment analysis’, Lecture Notes in Computer Science, pp. 179–191. doi:10.1007/978-3-319-99740-7\_12.

Almutiry, S. and Abdel Fattah, M. (2021). Arabic CyberBullying Detection Using Arabic Sentiment Analysis. The Egyptian Journal of Language Engineering, 8(1), pp.39–50. doi https://doi.org/10.21608/ejle.2021.50240.1017.

Bobichev, V., Kanishcheva, O. and Cherednichenko, O. (2017). Sentiment analysis in the Ukrainian and Russian news. [online] IEEE Xplore. doi https://doi.org/10.1109/UKRCON.2017.8100410.

Cambria, E., Poria, S., Gelbukh, A. and Thelwall, M. (2017). Sentiment Analysis Is a Big Suitcase. IEEE Intelligent Systems, 32(6), pp.74–80. doi https://doi.org/10.1109/mis.2017.4531228.

GUJJAR J, P. and Kumar H R, Dr.P. (2020). Sentiment Analysis for running text in Email conversation. International Journal of Computer Science Engineering, 9(4), pp.259–263. doi https://doi.org/10.21817/ijcsenet/2020/v9i4/200904007.

Gupta, S. (2018) Sentiment analysis: Concept, analysis and applications, Medium. Available at: https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17 (Accessed: 08 July 2023).

Gupta, S., Singh, R. and Singla, V. (2019). Emoticon and Text Sarcasm Detection in Sentiment Analysis. pp.1–10. doi https://doi.org/10.1007/978-981-15-0029-9\_1.

Hariguna, T. (2020). Survey Opinion using Sentiment Analysis. Journal of Applied Data Sciences, 1(1), pp.35–40. doi https://doi.org/10.47738/jads.v1i1.10.

Ishaq, A., Asghar, S. and Gillani, S.A. (2020). Aspect-Based Sentiment Analysis Using a Hybridized Approach Based on CNN and G.A. IEEE Access, 8, pp.135499–135512. doi https://doi.org/10.1109/access.2020.3011802.

Kaur, H., Mangat, V. and Nidhi (2017). A survey of sentiment analysis techniques. 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). doi:https://doi.org/10.1109/i-smac.2017.8058315.

Krishna, A., Akhilesh, V., Aich, A. and Hegde, C. (2019). Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques. Lecture Notes in Electrical Engineering, pp.687–696. doi https://doi.org/10.1007/978-981-13-5802-9\_60.

Li, J. and Hovy, E. (2017). Reflections on Sentiment/Opinion Analysis. A Practical Guide to Sentiment Analysis, pp.41–59. doi https://doi.org/10.1007/978-3-319-55394-8\_3.

Lutkevich, B. and Burns, E. (2023) What is natural language processing? an introduction to NLP, Enterprise AI. Available at: https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP (Accessed: 08 July 2023).

Marta (2022) The benefits of sentiment analysis: Brand24, Brand24 Blog. Available at: https://brand24.com/blog/the-benefits-of-sentiment-analysis/ (Accessed: 08 July 2023).

Mhatre, M., Phondekar, D., Kadam, P., Chawathe, A. and Ghag, K. (2017). Dimensionality reduction for sentiment analysis using preprocessing techniques. 2017 International Conference on Computing Methodologies and Communication (ICCMC). doi https://doi.org/10.1109/iccmc.2017.8282676.

Nandwani, P. and Verma, R. (2021). A review on sentiment analysis and emotion detection from text. Social Network Analysis and Mining, 11(1). doi https://doi.org/10.1007/s13278-021-00776-6.

Pano, T. and Kashef, R. (2020) ‘A complete vader-based sentiment analysis of Bitcoin (BTC) tweets during the era of covid-19’, Big Data and Cognitive Computing, 4(4), p. 33. doi:10.3390/bdcc4040033.

Rahman, A. and Hossen, Md.S. (2019). Sentiment Analysis on Movie Review Data Using Machine Learning Approach. [online] IEEE Xplore. doi:https://doi.org/10.1109/ICBSLP47725.2019.201470.

Raj, N. (2021) Starters guide to sentiment analysis using natural language processing, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/06/nlp-sentiment-analysis/ (Accessed: 08 July 2023).

Rout, J.K., Choo, K.-K.R., Dash, A.K., Bakshi, S., Jena, S.K. and Williams, K.L. (2017). A model for sentiment and emotion analysis of unstructured social media text. Electronic Commerce Research, 18(1), pp.181–199. doi https://doi.org/10.1007/s10660-017-9257-8.

Sahani, S. (2022) Sentiment analysis challenges and ways to overcome them., Medium. Available at: https://medium.com/@shreya\_88085/sentiment-analysis-challenges-and-ways-to-overcome-them-f28fa680d953 (Accessed: 08 July 2023).

Sharma, S. (2023) Top 10 applications of sentiment analysis in business, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2023/01/top-10-applications-of-sentiment-analysis-in-business/ (Accessed: 08 July 2023).

Wang et al. (2014) Dimensional sentiment analysis using a regional CNN-LSTM model, Association for Computational Linguistics. Available at: https://aclanthology.org/P16-2037.pdf (Accessed: 31 August 2023).

Wankhade, M., Rao, A.C.S. and Kulkarni, C. (2022) A survey on sentiment analysis methods, applications, and Challenges - Artificial Intelligence Review, SpringerLink. Available at: https://link.springer.com/article/10.1007/s10462-022-10144-1 (Accessed: 08 July 2023).

Yadav, A. and Vishwakarma, D.K. (2020). A comparative study on bio-inspired algorithms for sentiment analysis. Cluster Computing. doi https://doi.org/10.1007/s10586-020-03062-w.

Yadav, P. and Pandya, D. (2017). SentiReview: Sentiment analysis based on text and emoticons. [online] IEEE Xplore. doi https://doi.org/10.1109/ICIMIA.2017.7975659.

Yadollahi, A., Shahraki, A.G. and Zaiane, O.R. (2017). Current State of Text Sentiment Analysis from Opinion to Emotion Mining. ACM Computing Surveys, [online] 50(2), pp.1–33. doi https://doi.org/10.1145/3057270.

Yılmaz, B. (2022) Top 5 sentiment analysis benefits for businesses in 2023, AIMultiple. Available at: https://research.aimultiple.com/sentiment-analysis-benefits/ (Accessed: 08 July 2023).