NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF SOFTWARE ENGINEERING

A COMPARATIVE STUDY OF DEEP LEARNING MODELS USING N-GRAMS, AND WORD EMBEDDINGS.

M.Sc. THESIS

Anvi Alex EPONON

Supervisor
Assoc. Prof. Dr. KAMIL DIMILILER

Nicosia

November, 2022

ANVI ALEX EPONON A COMPARATIVE STUDY OF DEEP LEARNING MODELS USING N-GRAMS, AND WORD EMBEDDINGS. MASTER THESIS 2022

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF SOFTWARE ENGINEERING

A COMPARATIVE STUDY OF DEEP LEARNING MODELS USING N-GRAMS, AND WORD EMBEDDINGS.

M.Sc. THESIS

Anvi Alex EPONON

Supervisor
Assoc. Prof. Dr. KAMIL DIMILILER

Nicosia

November, 2022

Approval

We certify that we have read the thesis submitted by Anvi Alex Eponon titled **"A Comparative Study of Deep Learning models using n-grams, and Word embeddings."** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Software Engineering.

Examining Committee	Name-Surname	Signature
Head of committee:	Assoc. Prof. Dr. Boran Şekeroğlu	....
Committee member:	Assist. Prof. Dr. Elbrus Imanov	....
Supervisor:	Assoc. Prof. Dr. Kamil Dimililer	....

Approved by the Head of the Department

14/12/2022

Assoc. Prof. Dr. Boran Şekeroğlu

Head of Department

Approved by the Institute of Graduate Studies

NEAR EAST UNIVERSITY INSTITUTE OF GRADUATE STUDIES 1988

...../...../2022

Prof. Dr. Kemal Hüsnü Can Başer

Head of the Institute

Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of Institute of Graduate Studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

EPONON ANVI ALEX

-----/----/2022

Day/Month/Year

Acknowledgements

I would like to show my utmost gratitude to my supervisor, Assoc. Prof. Dr. KAMIL DIMILILER, Near East University, Northern Cyprus. He gave me expert guidance, insightful suggestions, and offered timely and helpful feedback throughout the whole process. I also want to thank the Head of the Software Engineering Department, Assoc. Prof. Dr. BORAN ŞEKEROĞLU, and all my lecturers in the department. Finally, I would like to thank my family and friends for their love, sacrifices, prayers, and moral support throughout my graduate studies and thesis writing.

EPONON ANVI ALEX

ABSTRACT

A COMPARATIVE STUDY OF DEEP LEARNING MODELS USING N-GRAMS, AND WORD EMBEDDINGS.

Anvi Alex Eponon

Master, Department of Software engineering

November, 2022, 73 pages

Sentiment analysis is an area of natural language processing. It involves figuring out the polarity of a corpus, usually a corpus of writing. The most popular text categorization technology is sentiment analysis. Sentiment analysis might help discover more about how people generally feel about a product. This is frequently impossible to do manually because of the enormous amount of data. Thanks to specialist SaaS tools, for businesses, getting deeper insights into their text data has gotten easier. This might incorporate everything from customer reviews to employee surveys to social media updates. Sentiment information from various sources can be used to inform crucial business decisions. Sentiment analysis is widely used today. Platforms like Youtube, Netflix, Facebook, and many others use sentiment analysis to build recommender systems to help their users make decisions. In the business industry, sentiment analysis gives reliable feedback to companies to improve the quality of their services and increase customer satisfaction. Recently, many studies have been conducted on sentiment analysis for several purposes. However, we noticed a lack concerning the presence of comparative studies of feature extraction techniques such as n-grams and word embeddings.

In our study, we update the literature with a comparative study of three deep learning models (CNN, RNN (LSTM), and Bert) along with two feature extraction techniques (n-gram and word embeddings). First, we preprocessed the data by removing any artifacts (words, symbols, numbers, etc.) that could interfere with model training. Then, we extracted features via bigram and word embeddings, the latter using Glove. Finally, we trained our models according to their configurations. We found that CNN models performed the best during our experiment, and word embedding seems to be the best choice for sentiment analysis. The maximum accuracy we reached was 90.30%. This study helps extend our knowledge of the overview performance of sentiment analysis with deep learning.

Keywords: NLP, deep learning, Cnn, Lstm, sentiment analysis.

Table of Contents

CHAPTER IV: MATERIALS AND METHODOLOGY

List of tables

List of figures

List of abbreviations

CNN	:	Convolutional Neural Network
LSTM	:	Long Short-Term Memory
BERT	:	Biderectional Encoder Representation Transformers
NLP	:	Natural Language Processing
POS	:	Part-Of-Speech
ASR	:	Automated Speech Recognition
SVM	:	Support Vector Machine
IMDB	:	Internet Movie Database
TF-IDF	:	Term Frequency- Inverse Document Frequency
IWV	:	Improved Word Vectors
KNN	**:**	K-Nearest neighbors
EMD	:	Empirical Modal Decomposition
ELSTM	:	Emotion-enhanced LSTM
ABSA	:	Aspect-Based Sentiment Analysis
HMM	:	Hidden Markov Model
VoC	:	Voice of Customer
NPS	:	Net Promoter Score

CHAPTER I

Introduction

1.1 Introduction

The goal of natural language processing (NLP), a field at the nexus of artificial intelligence, linguistics, and computer science, is to make it easier for machines and people to communicate with one another through spoken and written human languages. Machines can absorb and comprehend human language with the help of natural language processing, enabling them to carry out jobs more efficiently. The 1950s saw the introduction of the field of Natural Language Processing with Alan Turing's article "Computing Machinery and Intelligence." The suggested article is a test that entails the creation of natural language and automated interpretation. The following is a historical breakdown of NLP:

Symbolic NLP was the foundation of NLP in the 1950s and 1990s. By using a set of rules on the data, symbolic NLP simulates natural language understanding. Then came Statistical NLP from the 1990s to the 2010s. Statistical NLP is the use of machine learning techniques to address machine translation problems. It is focused on unsupervised and semi-supervised learning.

Neural language processing now makes up most NLP. Due to neural networks' excellent state-of-the-art performance on a variety of language problem (Goldberg, Yoav, 2016) (Goodfellow, et.al., 2016).
We can use NLP in a variety of ways in our daily lives. We focus on sentiment analysis in this work, which is one of these applications. A method for determining the polarity of the data is sentiment analysis, which is a natural language methodology. Determining the data's positivity, negativity, or neutrality is what it signifies, in other words. In many different ways and in many industries, sentiment analysis is now widely used. The common applications of sentiment analysis include text analysis, computational linguistics, natural language processing, and biometrics.

1.2 Statement of the problem

Natural language processing is a broad topic. It involves several subjects and fields, such as: Machine Translation, summarization, Part-Of-Speech tagging (POS), sentiment analysis, answering queries, Automated Speech Recognition (ASR), etc. And each topic by itself has several concepts, methods, and models to address specific problems. With such tremendous information, it becomes hard for a non-experienced and young researcher to start contributing to the NLP community by tackling real-world problem scenarios.

Young researchers may encounter another obstacle in addition to this one that prevents them from beginning to address real-world issues. It is a dearth of material resources and intelligible source code. The study of NLP has advanced significantly over the past ten years, and several topics are now regarded as "solved cases." The materials and resources required for this enormous accomplishment are quite sophisticated, particularly the study team's programming and the stack they used. This particular situation creates a serious challenge in front of new researchers who may not have considerable experience in understanding the current state-of-art codes in order to provide valuable support. In fact, few materials are aimed at young researchers by providing comprehensive source code and documentation related to their research that is fairly and easily approachable by beginners.

Also, several studies have been conducted in the field of sentiment analysis, but few have compared the performance of natural language techniques such as n-grams and word embedding in conjunction with deep learning models such as Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Bert.

We can find several studies made on hybrid models to solve particular NLP tasks such as translation and text generation. We can also find several reviews on deep learning and machine learning models in sentiment analysis, but most of the work comes with the unavailability of the code source as mentioned earlier.

All of these situations, combined, create frustration for the researcher whose aim is about to start his journey in the field of Natural Language Processing.

1.3 Purpose of the thesis

When starting with a new subject, new researchers ask several questions. Among them, we have the question of the reliability and precision of the information contained in the study but also its level of comprehensiveness and complexity.

This study provides new researchers in the field of NLP, precisely in sentiment analysis, an easy approach to the topic; a specific problem to solve with the core and necessary information needed to understand the context; but also a correct implementation of the state-of-the-art techniques used to solve problems in sentiment analysis. The study's goal is to provide new researchers with everything they need to quickly work on real-world sentiment analysis problems. By so doing, the current study's source code has been made available for further analysis, complete with comments.

In addition to that, we want to close the gap in the literature by making a comparative study of several models of CNN, LSTM, and Bert along with two feature extraction techniques, n-grams and word embeddings. This comparative study will help better extend our knowledge of the performance of models in combination with feature extraction techniques.

1.4 Importance of the Thesis

Several techniques and models exist in NLP, precisely in sentiment analysis. Depending of the targeted problem, a specific technique or model may be the most suitable. Our study aims at giving more insights into the performance of specific models such as CNN, and RNN when applied with techniques like n-grams and word embeddings. These insights will help researchers make decisions on either using a specific combination of models and techniques over others. It will facilitate the research process of future studies and contribute to the creation of innovative solution. Also, with the effort done on making the source code comprehensive and available, the study will give an opportunity to new researchers to easy enter the field of NLP.

1.5 Limitation of the thesis

The current study is a Natural Language Processing comparative study on CNN, LSTM, and Bert models with n-grams and word embeddings. During this study we selected specific techniques and models based on previous works in the literature. This means that we didn't studied all the feature extraction techniques. We limited ourselves to study n-grams, specifically, bigrams and word embeddings which have been intensively used in the literature.

We also, put limits on the number of models used and studied. We selected three models, CNN, LSTM and Bert for our study. From our research, they are the most used models in Sentiment Analysis. About the preprocessing phase, we didn't make use of all the techniques. For example, we didn't make use of lemmatization as we think it would not be necessary to process the data with it.

Finally, there is a limit in the analysis of the results obtained by our models with the help of bigrams and word embedding techniques. It doesn't contain results from other studies in the core analysis. Although we compare certain results obtained by previous studies with ours.

1.6 Overview of the Thesis

The following is how the work in this thesis has been set up.

The literature review is included in the second chapter. Three major topics are covered in the literature's three sections. First, we'll talk about the three layers of extraction methodologies used in sentiment analysis. Attribute levels include word, sentence, and aspect. We will highlight three different feature extraction methods used in sentiment analysis in the second section. N-grams, word embedding, and TF-IDF are all present.

Finally, in the final section, we discuss three models that are currently employed for sentiment analysis.

In the third chapter, we will discuss about the different types of sentiment analysis but also the different applications of sentiment analysis nowadays.

We will be showing our methods and materials in the fourth chapter. We will show the setup of our computer lab. We will explain the method used for each model accompanied with figures.

In the fifth chapter, we are showing our results from the experiment we conducted but also we are discussing them in comparison to previous studies.

Finally in the sixth chapter, we draw conclusions on the work done and discuss improvements that can be made as future work on the model to increase its effectiveness.

CHAPTER II

Literature Review

Artificial intelligence is a branch of computer science that allows computers to mimic or even outperform humans in certain tasks. Recently, several fields of study and industries have been revolutionized by artificial intelligence. Tremendous studies in the field of medical science (Dimililer K., et.al., 2016) (Yoney K.E., et.al., 2016), also in the field of education (Boran S., et.al., 2019) (Yoney K.E., Dimililer K., 2017) have seen breakthroughs with the help of artificial intelligence. Machine learning which is a subfield of artificial intelligence, employs statistics and computational algorithms assist computers in making decisions (Kayali D., et.al., 2022) (Sekeroglu B., et.al., 2022) (Dimililer K., et.al., 2022).

The goal of the machine learning discipline known as "Natural Language Processing" (NLP) is to make human-machine interaction simpler. To solve different problems related to human language, NLP is divided into multiple sub-tasks such as translation, summarization, and sentiment analysis. (Farzindar, A., Inkpen, D., 2015).

NLP's task of sentiment analysis involves taking views or sentiments from a corpus. (text, audio, video, etc.) (Kaur, H. et al., 2017). The outputs are called polarities, and they can be positive, negative, or neutral (Nhan C.D. et al., 2020). The goal behind extracting polarities from corpora is to be able to understand how to behave with a particular document. An example can be analyzing a product review to improve its quality and the client’s satisfaction (Thematic., 2020) . The process of extraction can be done at different levels. We identify three common levels, which are word-level, sentence-level, and aspect-level (Nhan, C.D. et al., 2020).

At word-level, the polarity of the document depends on the average polarity of each word (Nikos E. et al., 2011).

In the "impact of features extraction on sentiment analysis," the author extracted the polarity of different tweets at the word level, which gives a performance 3.4% higher compared to other techniques studied in the paper. (Ahuja et al., 2019).

As for Nikos Engonopoulos and his team, they introduced a new method. a technique for entity-level sentiment analysis at the word level.They made it possible by training a conditional random field model (CRF) to detect the sentiment of words in a specific corpus. The outcomes are then used to ascertain the entity's attitude. With this technique, they reached 70.2% accuracy overall. (Nikos E. et al., 2011).

By using Reinforcement Learning, a group of researchers presented a new framework called World-level Sentiment LSTM. The results show it is possible to predict the sentiment of each word. The method reached 95.5% accuracy for the WS-LSTM model. (Ruiqi C., et al., 2019).

To facilitate the use of word-level sentiment annotations to enhance Telugu sentiment analysis tasks, the Sreekavitha team created an annotated corpus in 2018. This paper's main goal is to validate and look into the use of word-level sentiment annotations and machine learning methods for autonomous sentiment identification tasks.They achieved 78.97% accuracy (Sreekavitha P., Vijjini A. R., Radhika M., 2018).

A recent study developed a new neural network architecture called Sentiment Explainable Neural Network (SINN) to achieve greater interpretability and utility when analyzing data. Due to its ability to extract raw word-level sentiment along with its local and global word-level context, SINN is interpretable (Tomoki I., Kota T., Hiroki S., Tatsuo Y., Kiyoshi I., 2020).

Additionally possible are word-level fusion, reinforcement learning, and multimodal sentiment analysis. A GME-LSTM(A) model with two modules that incorporates temporal attention. When there are noisy modalities, we can use Gated Multimodal Embedding to help with fusion. Word level fusion between the input modalities may be performed by the Long Short-Term Memory (LSTM) with TA (Temporal Attention) while paying attention to the most crucial time steps. As a result, the model can more precisely depict several structure modes of speech across time and perform better when understanding sentiment.(Minghai Chen, Sen W., Paul P. L., Tadas B., Amir Z., Louis-Philippe M., 2017).

When it comes to sentence level, the focus is put on each sentence in the corpora in order to extract the overall polarity. (K. Veselovská, 2011) Sentiment analysis at the phrase level was performed on Amazon product reviews by Xuechun Li and his team using a bidirectional LSTM model. The model reaches an accuracy of up to 96%. (Li, X. et al., 2021).

The goal of a 2019 study proposal is to classify financial hazards at the sentence level using data from financial reports. The results demonstrate that the classification performance is enhanced by the combination of senti-phrases with embedding models. They obtained a maximum accuracy of 88.79% (Chi-Han D., et al., 2019).

There has been yet another sentiment analysis. This time, employing distributed text representations and multi-instance learning on financial news. Compared to similar research, the method presents superior performance with an accuracy of up to 69.90% (Bernhard L., et al., 2018).

A research focused on Sentence-level sentiment analysis in Persian to compensate for the lack of attention dedicated to languages other than English. The research tackles the issue of scarcity by proposing two new resources: SPerSent, a sentence-level dataset for sentiment analysis in Persian, and CNRC, a new Persian lexicon. SPerSent has 150000 sentences, each with two labels: a binary label denoting the sentence's polarity and a five-star rating. A lexicon-based technique is used to generate these labels automatically. To categorize each sentence, three lexicons are employed separately. The results are then aggregated using the majority voting and average procedures for polarity and five-star labeling, respectively. Finally, the SPerSent is evaluated using a well-known machine learning approach, Nave Bayes. A 92% accuracy rate has been achieved (Mohammad E. B., Arman K.,2017).

Oscar Täckström and Ryan McDonald developed two distinct kinds of semi-supervised models for sentiment analysis. Both models leverage a small number of human-produced sentence labels and a lot of natural supervision in the form of review ratings to build sentence-level sentiment classifiers. The recommended model is a cross between a structured conditional model that is completely supervised and one that is partially supervised. As a result, extremely effective estimation and inference techniques with thorough feature descriptions are made possible. When compared to all baselines, they experimentally show that both modifications significantly improve results for sentence-level sentiment analysis. They analyze the two variations as well as their component models (McDonald R., Täckström O., 2016).

Despite extensive research efforts and broad interest in lexical resources and learning approaches, almost all current systems are designed to only work with English text. The bulk of current lexical resources in each language are changed as part of recognized solutions, sometimes without proper validations or basic baseline comparisons. A study was done to evaluate current initiatives for language-specific sentiment analysis. In order to do this, they compared two language-specific algorithms with the twenty-one methodologies for sentence-level sentiment analysis presented for English. They offer a comprehensive quantitative examination of the most advanced multi-language techniques based on nine language-specific datasets. The results showed that simply translating the input information from a particular language to English and using one of the English methodologies already in use would be more effective than the language-specific efforts now under consideration (Matheus A., Julio R., Adriano P., Fabricio B., 2016).

We can also extract polarities at the aspect level. This technique tries to extract not only one polarity from a sentence but different polarities identified in each part of the sentence. (Pang, G., et al., 2021Using Amazon customer review data as an example, we can concentrate on extracting aspect phrases from each review while also identifying the speech portion in order to obtain the polarity. (S. Vanaja, and M. Belwal,2018)

In the paper proposed by Yequan Wang, an aspect-level sentiment capsule model is studied. The approach use the attention mechanism to pay attention to words that are both aspect-related and sentimental to each aspect. With a maximum accuracy of 87.2%, the study demonstrates cutting-edge performance. (Yequan W., et al., 2019).

A study conducted by a group of researchers introduced a new model for aspect-level sentiment analysis. They combined SentiWordNet with n-gram feature extraction and two alternative linguistic feature selections. The result gives an accuracy of 78.7% (V.K. Singh et al., 2013).

A recent research proposed a method for incorporating word dependencies into aspect-level sentiment analysis, in which type information is represented using key-value memory networks and multiple dependencies are appliedThe usefulness of the method is demonstrated by experimental findings on five benchmark datasets, where it consistently outperforms baseline models and performs at the cutting edge on three of them (Yuanhe T., Guimin C., Yan S., 2021).

A Research established a two-stage paradigm that may be completed in two phases to handle aspect-level extraction: (1) Position attention is modeled explicitly in the Stage 1 model to represent the relationship between the aspect and its context words; and (2) The Stage 2 model looks at how to use position attention to represent various characteristics in a single opinionated sentence. Comparing the proposed method to other state-of-the-art attention-based algorithms, the SemEval 2014 datasets used for empirical evaluation show that it delivers a significant performance boost (Xiao M.,Jiangfeng Z., Limei P., Giancarlo F., Yin Z., 2019).

A study on a novel type of aspect-level sentiment analysis on movie reviews using domain-specific feature-based heuristics was published. It has created an approach that is aspect-focused and analyzes text reviews of movies to assign an emotion to each aspect. An overall net sentiment profile of the movie is created by averaging the ratings from many reviews for each aspect. Adjectives, adverbs, and verbs were among the two different language feature options available in their SentiWordNet-based model, which they also used to extract n-gram features. The results show that the approach produces a sentiment profile that is more precise and focused than the sentiment analysis performed on individual documents (V. K. Singh; R. Piryani; A. Uddin; P. Waila., 2013).

As we have demonstrated, opinions may be extracted from any corpus at various levels, however in order to increase the models' accuracy, several techniques like TF-IDF or n-grams are frequently used (Imperial, J., et al., 2019). An N-gram is the occurrence of a sequence of words (n-words) in a document. The concept comes from the bag of words that measures the presence of known words. N-grams give a certain level of understanding of words, which helps in building predictions. S. Srinidhi 2021).

In "Search for optimal Feature in Political Sentiment Analysis", the team of Aman Ullah compares the performances of four features such as unigrams, bigrams, trigrams, and opinion words. They found using unigrams outperformed other features with an accuracy of 81% in their study (Aman U., et al., 2020).

Another study used machine learning techniques to conduct a comparative sentiment analysis of sentence embedding. The classifiers they utilized included Multinomial Naive Bayes, SVM, and Logistic Regression. With an accuracy of 86.23%, the results favor models that incorporated n-gram and bigram (A. Poornima et al., 2020).

Using Bigram collocation with several classifiers such as Maximum Entropy, Nave Bayes, and Support Vector Machine, a group of researchers predicted tweet sentiments with an accuracy of 88.42% (Sumaya I. M., et. al., 2021).

A 2012 study demonstrated that bigram features provide consistent gains on sentiment analysis tasks when used in conjunction with Support Vector Machine or Naive Bayes. An accuracy of 91.22% has been reached on the IMDB dataset(Sida W., Christopher D. M., 2012).

Another study by Pavitra and Kalaivaani coupled topic detection and bigram sentiment analysis with document-level sentiment categorization. The Latent Dirichlet Allocation model, which is based on the weakly supervised Joint Sentiment-Topic model, is enhanced by the sentiment layer. They thought of Bigrams to increase sentiment analysis's accuracy. To assess the sentiment polarity of the bigrams, they also developed a sentiment thesaurus comprising positive and negative lexicons. This concept can be used in different fields. This is experimentally shown to outperform existing semi-supervised methods in four different domains (R Pavitra; P C D. Kalaivaani, 2015).

A huge dataset of one million messages posted on the microblogging service StockTwits was used by a group of academics to test the efficacy of various preprocessing methods and machine learning algorithms for sentiment analysis in finance. The efficacy of sentiment classification is greatly increased by the use of bigrams and emojis, they found (Thomas R., 2019).

TF-IDF is a technique related to the bag of word concepts (BoW) that helps to extract the importance of words inside a corpus (Huilgol, P. 2020). The main idea behind TF-IDF is to give more importance to words or tokens that appear less frequently than others. (MonkeyLearn Blog, 2019).

In "A Comparative Study on the TF-IDF Feature Weighting Method and its Analysis Using Unstructured Datasets," Mamata Das' team looked into feature weighting techniques like TF-IDF and N-grams for text classification on unstructured data, and found that when using a combination of cutting-edge machine learning classifiers, TF-IDF outperforms with a score of 93.81% (D. Mamata et al. 2021).

Provider By.U is a telecommunication service that provides several products and services to its users, such as internet access and music streaming, in Indonesia. With the use of TF-IDF and Support Vector Machine, a trio of academics did a sentiment analysis on customer evaluations for the By-U provider on the Google Play Store. They reached an average accuracy of 84.7% (Susanti F., et al., 2020).

Srividya and A.Mary Sowjanya conducted an aspect-based sentiment analysis using Part of Speech Tagging and TF-IDF. In their study, they developed two models. the first without using TF-IDF and the second using TF-IDF. The second model, which had a maximum accuracy of 94.52%, fared the best, according to the data. (S. Kotagiri, and S. A. Mary 2019).

A study looked at usage patterns like tweet frequency and length, as well as terms used and their length, to see what they said about users' attitudes toward BrexitTo ascertain whether there was a relationship between text data features and Twitter user characteristics, regression analysis and sentiment analysis, specifically TF-IDF, were applied. As a result, it discovered precise usage patterns shared by both sides of the debate, as well as systematic but minor differences in how the two sides use language (Alexander M., et.al, 2021).

Internet users have benefited greatly from the tourism sector's ability to offer reasonable and comparable hotel rates. Ram Krishn Mishra's team used cosine similarity and term frequency-inverse document frequency (TF-IDF) in their paper to present more hotels based on reader reviews (Ram K. M., et.al., 2019).

Gang Li introduces a novel technique for sentiment analysis known as clustering-based sentiment analysis. With the use of a voting system, term scores can be imported, and a TF-IDF weighting approach can produce an acceptable and stable clustering outcome. It performs better than the two types of existing approaches, supervised learning techniques and symbolic techniques. Their method for solving sentiment analysis issues is well-executed, effective, and non-human participation (Gang L.; Fei L, 2010).

In natural language processing, the word embedding approach is used to represent words, where each word is mapped to one real-valued vector that represents the meaning of the word. This definition means that words that are close in the vector space will have the same representation but also have the same meaning (Bin W., et al., 2019).

Eman Saeed and Norah Saleh conducted a study in which they implemented an unsupervised learning approach for aspect-level sentiment classification based on semantic language. This method shows how to leverage the full capacity of word embedding generated by models like Glove. The results gave an accuracy of 83.04% (Eman S., Norah S., 2021).

In the study titled "More than bags of words: Sentiment analysis with word embeddings", To gauge the degree of negativity in parliamentary remarks, a team of academics used word embeddings in a supervised machine method.. They reached 58% accuracy (Elena R., et al., 2018).

In his model, Aytug Onan coupled a CNN and LSTM architecture with a TF-IDF weighted by Glove which is word embedding approach. The study evaluates the effectiveness of a number of word embedding approaches, including word2vec, fatsText, Glove, and DOC2vec. The maximum accuracy reached is 93.85% with the glove.

A team of academics focused on sentiment analysis using improved pre-trained word embeddings. IWV which means Improved Word Vectors, a novel technique they created, raises the precision of pre-trained word embeddings in sentiment analysis. They base their methods on a number of strategies, including Word2Vec-GloVe techniques, lexicon-based procedures, and word placement algorithms. They tested the accuracy of their approach using a range of deep learning models and benchmark sentiment datasets. The findings demonstrate that Improved Word Vectors (IWV) can do sentiment analysis (Seyed M. R., Rouhollah R., Ali G., Hadi V., 2019).

Erion Ano and Maurizio Morisio looked at the relationship between the training method, training corpus size, and thematic significance of texts to determine the efficiency of word embedding features on sentiment analysis of tweets, music lyrics, movie reviews, and product reviews. They also looked into a number of training or post-processing techniques that could be applied to improve word embeddings' performance in particular tasks or domains. The best models for addressing syntactic and semantic word comparison problems, according to their empirical findings, are those trained on massive, vocabulary-rich multithematic texts. (Erion Ç., Maurizio M., 2019).

The researchers developed a different approach based on distributed word embeddings. As part of a supervised machine learning process that determines the degree of negativity in parliamentary statements, they used word embeddings. Crowdcoded training words and a study of negative speech patterns in Austrian legislative speeches are used, respectively, to assess the accuracy of the procedure and external validity. The results demonstrate the potential of sentiment analysis in social sciences using the word embeddings approach (Elena R., et.al., 2018).

In addition to feature extraction techniques, we add models or a combination of models for prediction purposes. Long-Short Term Memory (LSTM), Transformers, specifically BERT, and Convolutional Neural Networks (CNN) are the three main topics of our study.

Convolutional Neural Networks (CNNs) are deep learning algorithms that accept an input, some text in our case, and assigns a certain weight and bias in order to be able to recognize the difference between other text inputs (Shahid D. 2021). It is possible to use CNN in sentiment analysis due to its ability to perform semantic segmentation. (Alam, M., et al., 2021).

Shiyang Liao's team presents an approach to predict user satisfaction with a product, happiness with some particular environment, or destroying a situation after a disaster. The experiment reaches 75.39% accuracy in a CNN model (Shiyang L., et al., 2017).

On the IMDB dataset, a study from 2019 used a MLP also called a multilayer perceptron, a CNN, a LSTM recurrent neural network, and finally a hybrid CNN with LSTM model. The data has been preprocessed using Word2vec. The results showed that the CNN model got the second highest accuracy with 87.7%, after the CNN-LSTM model (Nehal M. A., Marwa M. A. E. H., and Aliaa Y., 2019).

Hu Xu's team presented in 2018 a novel CNN model comprised of a general-purpose embedding and a domain-specific embedding. With just this configuration, the model produced cutting-edge performance, with a maximum F1 score of 81.59% (Hu X., Bing L., Lei S., Philip S. Y. 2018).

A recent work covered aspect-based sentiment analysis, a key element of sentiment analysis. A robust model for aspect-based sentiment analysis using convolutional neural networks was suggested in this work by fusing the capacities of language resources and gating mechanisms. To start, it is possible to ascertain the polarity of an aspect's attitude using the regularizers given by real-world linguistic resources. Second, when the given aspect is employed as direction, the gating technique may better regulate the sentiment features. The core building blocks of the model are convolutional neural networks, which can efficiently carry out concurrent operations during training. Experimental results on SemEval 2014 Restaurant Datasets show that the strategy can deliver excellent results for aspect-based sentiment analysis (Zeng D., et.al., 2019).

In China, researchers presented a unique multi-channel convolutional neural network with multi-head attention mechanism-based sentiment analysis model. This model independently combines several features, and dependency syntax characteristics to create three new combination features that help it learn the sentiment information in the text more thoroughly. The multi-channel convolutional neural (MCNN) is then fed with these three new merged features, together with the multi-head attention mechanism. Last but not least, the experiments use two Chinese text data sets. The experimental findings show that, in comparison to other baseline models with an accuracy of 86%, the studied model has a higher classification accuracy and a more affordable training time cost. (Yue F., Yan C., 2021).

Another study that examined sentiment made use of the Convolution Neural Network (CNN) technique, which was developed by Alex K. According to Alex K's analysis, the error rate is reduced by 15%, whereas in 2017 it was only reduced by 5%. The best training data will be chosen using a CNN approach that has been optimized and uses a threshold (CNN-T). With this approach, more than one aspect can be produced from a single data test. In comparison to CNN and the three traditional machine learning methods, SVM, Naive Bayes, and KNN, the average outcome of this experiment employing CNN-T obtained improved F-Measure. With an overall F1 score of 0.71, CNN-T outperforms other comparable approaches (Budi M. M.; Widyantoro D. H., 2018).

The Long Short-Term Memory also called LSTM is the most used recurrent neural network in the artificial intelligence industry for several purposes like handwriting recognition, speech recognition, machine translation, video games, etc. (Scientific Integrity, 2021). What makes an LSTM model a good algorithm for natural language processing is its ability to keep hold of context by remembering previous inputs without suffering high processing times like a normal recurrent neural network. (Shekhar, S. 2021).

Shervin Minaee and Elham Azimi proposed an ensemble model comprised of a CNN and an LSTM model to understand public opinions or trends from reviews or tweets. The proposed model reached 90% accuracy on the IMDB dataset.(M. Shervin, A. Elham, 2019).

The LSTM recurrent neural network has also been intensively used in the financial industry. A group of three researchers proposed a deep learning model that considers an investor's emotions. The use of empirical modal decomposition (EMD) has been used by the group to decompose the complex sequence of stock prices. The results show that the emotional tendencies of investors effectively improve the predicted results (Zhigang J., et. al., 2020).

Another study focused on the prediction of the volatile price movement of cryptocurrency. The work is done by examining social media sentiment and establishing a link between it and price movements. The experiment has been done on Sina-Weibo, a Chinese social media platform. Modern auto-regressive based models were demonstrated to perform worse than the suggested LSTM model by 15.4% and 18.5% respectively in recall and precision score (Xin H., et. al., 2021).

For text sentiment analysis, a brand-new model called AEC-LSTM has been released. By fusing Emotional Intelligence (EI) and attentional mechanisms, it seeks to improve the LSTM network. To be more precise, an emotion-enhanced LSTM known as ELSTM is created by using emotional intelligence to augment the feature learning capacity of LSTM networks. The proposed emotion modulator and emotion estimator enable ELSTM to achieve its emotion modulation of the learning system. To more successfully capture various structure patterns in text sequences, ELSTM is also used with other techniques including convolution, pooling, and concatenation. The weight of the text concealed representation is then recommended to be adaptively adjusted using a topic-level attention method. Research on actual data sets demonstrates that our method can achieve 96% accuracy while greatly outperforming state-of-the-art deep learning (Huang F., et.al., 2021).

A high-quality training set is believed to be provided in the majority of existing approaches. However, creating a high-quality training set with extremely accurate labels is difficult in practical applications. The reason for this challenge is that text samples typically include intricate sentiment representations, and their annotation is arbitrary. In other studies, this problem is addressed by building a sentiment classifier using a two-level LSTM network and a new labeling technique. Experiments show that the suggested method surpasses cutting-edge algorithms with a maximum accuracy of 84% on benchmark English data as well as our gathered Chinese data (Wu O., Yang T., Li M., Li M., 2020).

The study by You Zhang's group introduces a recurrent attention-based LSTM neural network to iteratively discover an attention zone encompassing the key sentiment terms. By gradually decreasing the attention range and the amount of tokens, the model can use the weight of the significant sentiment words for final classification. The document-level corpora from IMDB, Yelp, and Amazon are used in the comparative trials. With 74.3% accuracy, the results demonstrate that the suggested model surpasses a number of cutting-edge techniques in document-level sentiment categorization (Zhang Y., Wang J., Zhang X., 2021).

Finally, Transformers are semi-supervised deep learning models. They are commonly used in NLP, especially in text summarization and translation. Transformers can process all the input sequences at once, and they can run multiple sequences in parallel with a self-attention mechanism that keeps the context for a better prediction. (Vaswani, et.al., 2017)

With 95.6% accuracy, Usman Naseem and his team present a method based on transformers for sentiment analysis with a bi-directional LSTM in " Transformer-based Deep Intelligent Contextual Embedding for Twitter Sentiment Analysis," published in 2020 (Naseem U., et al., 2020).

The Bert model was used in a study by Mrityunjay's team to do a sentiment analysis on how the coronavirus affects social life. They performed the study on two datasets. The first is made up of tweets from all around the world, while the second is made up of Indian tweets. Their final accuracy is approximately 94% (Mrityunjay S., Amit K. J., Shivam P., 2021).

A deep learning study has been conducted in finance again. To give pertinent information for decision-making, the authors performed sentiment analysis on news items using a bidirectional encoder representation from Transformers Bert. After tuning the parameters of the studied model, they achieved 72.5% of the F-score (Matheus G. S., et. al., 2021).

The task of tweet sentiment analysis is taken into consideration in this work. The pre-processing of tweets is carefully looked at in order to take use of information concealed in emoticons, emojis, and hashtags while avoiding noise created by various web constructs like URLs and mentions as well as by other text fragments. In order to assess the statistical significance of the impact that various tweet pre-processing techniques have on the results of sentiment analysis, several research utilizing a cutting-edge classification model (BERT) are planned. The findings allow for the selection of the tweet pre-processing technique that is most useful, enhancing sentiment analysis technology in both languages. The achieved accuracy is 75% (Marco P., Mirko V., Hamido F., Massimo E., 2021).

The more difficult process of aspect-based sentiment analysis (ABSA) entails determining both sentiments and aspects. According to some studies, using word representations in context from the pre-trained language model BERT in conjunction with a precise tuning method using additional generated text, it is possible to solve out-of-domain ABSA and outperform previous state-of-the-art results on some SemEval tasks in 2015 and 2016 (Hoang M., Bihorac O. A., Rouces J., 2019).

The performance of the E2E-ABSA task's modeling was investigated in a 2019 study employing contextualized embeddings from previously trained language models, such as BERT. In order to deal with E2E-ABSA, it created a number of straightforward yet informative neural baselines. The experimental results show that the BERT-based architecture can outperform cutting-edge efforts even with a simple linear classification layer (Li X., et.al., 2019).

In our current study, we are going to make a comparative study of three different models, CNN, LSTM and Bert along with two feature extraction techniques such as Bigrams and word embeddings (Glove).

CHAPTER III

Sentiment Analysis

The Chapter 3 focuses on the definition and importance of sentiment analyisis. It helps better understand how sentiment analysis is being applied in the business industry but also to know the different approaches of performing sentiment analysis.

3.1 Sentiment analysis approaches

Sentiment analysis is a method of obtaining information from a corpus. The ability to assess the polarity of a specific text is the major goal. Positive, negative, and neutral polarities are the various potential states.

We must select one approach out of three in order to be able to extract those polarities mentioned above. The three are the word level, the sentence level, and the aspect or feature level. However, there are three ways to tackle a problem with sentiment analysis. strategies based on lexicons, machine learning, and hybrid approaches (Bhavitha B., et al., 2017).

Lexicon-based techniques were the first used in sentiment analysis (Nhan, C.D., et al., 2020). It has two approaches. Both a corpus-based method and a dictionary-based approach rely on statistical content analysis. The corpus-based approach is represented by methods such as K-nearest neighbor (KNN) (Huq, M.R., et al., 2021), Conditional Random Field (CRF) (Pinto, D., et al., 2003), and Hidden Markov Model (HMM) (Soni, S.; Sharaff, A., 2015).

The lexicon-based approach is easier to implement than the machine learning-based approach, but it requires the involvement of humans in the text analysis process (Azeema S., 2018).

The machine learning-based approach gives better accuracy than the lexicon-based approach. However, the results are highly dependent on the quality and volume of the training dataset (Azeema S., 2018). Two groups can be identified in this approach. Traditional models, which refer to classical machine learning techniques and deep learning models (Nhan, C.D., et al., 2020),

The hybrid approach combines both the machine learning techniques and the lexicon approach. Here, sentiment lexicons play a major role in this strategy (Nhan, C.D., et al., 2020).

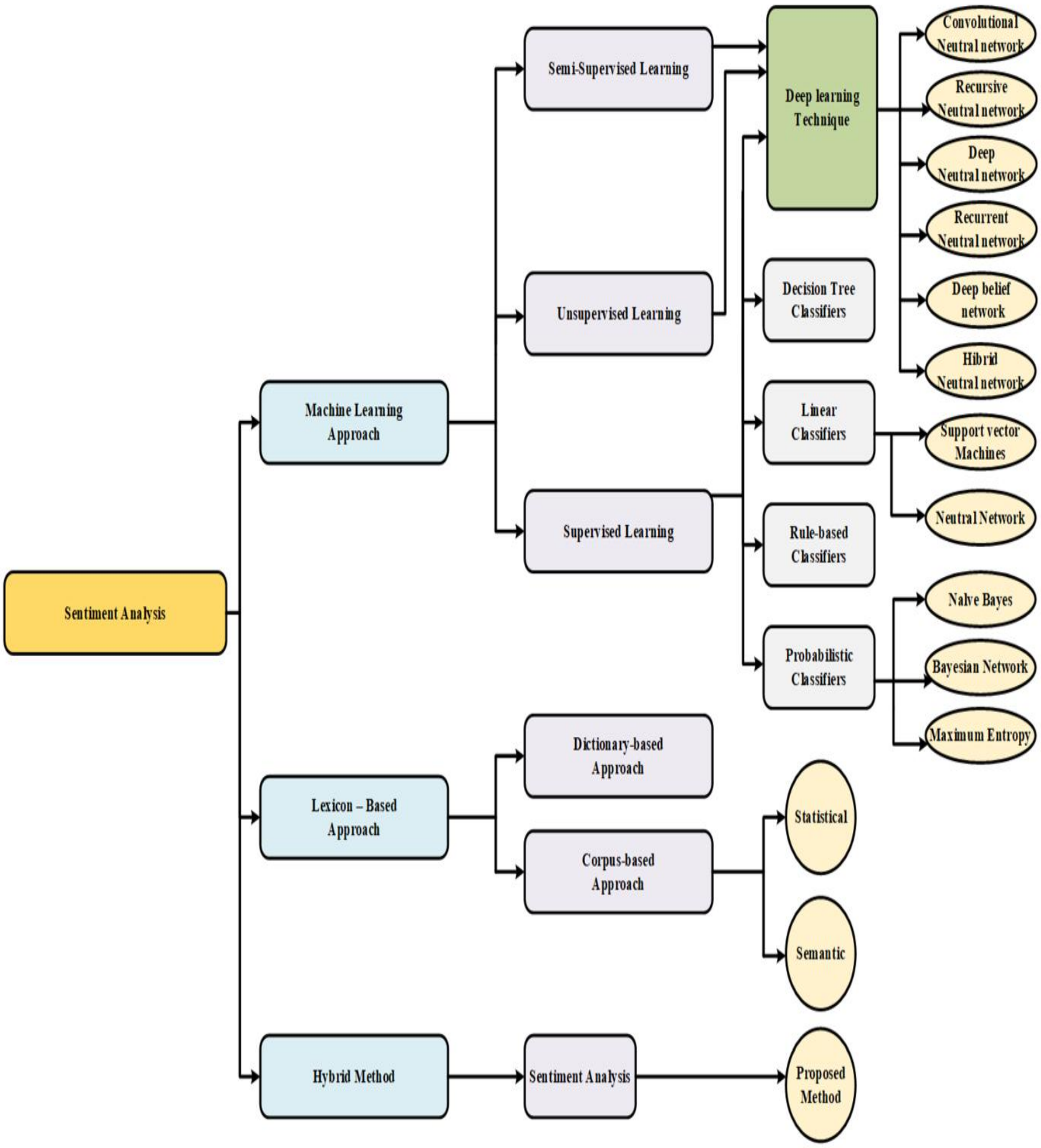

Figure 1: Diagram of the different approaches to Sentiment Analysis. Source (Nhan C.D., et.al., 2020)

3.2 Applications of sentiment analysis

These days, sentiment analysis is widely used across a variety of sectors, including banking, business, and healthcare. Through the association between social media and prices, sentiment research is being employed in the financial sector to forecast the volatile price of cryptocurrencies. 2021 H. Xin et al. By analyzing news stories, another sentiment analysis project assists investors in making wise selections (Matheus G., et. al., 2021).

In the business realm, sentiment analysis is incredibly helpful for organizations to get honest feedback about the quality and other parameters of their products and services. Companies can use these insights to better their goods and services, and it also gives them a chance to expand their market. A framework for sentiment analysis using machine learning techniques was described in a suggested study by Jain and Dandannavar (Jain, A.P., Dandannavar, P., 2016). Sentiment analysis has also been useful in the field of medicine. Sentiment analysis was the topic of a study by Sabarmathi and Chinnaiyan to gauge patient satisfaction. The work's objective was to provide a prescription medication evaluation recommendation that will assist any patient in choosing the best drug based on the specified criteria (Sabarmathi G., Chinnaiyan R., 2021).

3.3 Business application of Sentiment Analysis

Businesses can better understand the qualitative data they routinely gather from a number of sources by using sentiment analysis. We'll look at some of the most common commercial applications now.

3.3.1 Voice of Customer (VoC) Programs

It's urgent for companies to understand how the target market feels about their products and services. Because such information can enhance the client experience or resolve issues with the products or services. The department that aims to collect this information is called Voice of Customer (VoC) (Qualtricks, 2022).

A common way to access how customers feel about the services and products is to use the Net Promoter Score (NPS). Customers are often faced to answer a similar question like "How likely are you to recommend us to a friend?". The answer is most of the time represented by a number on the scale of 1 to 10 where customers who answered with a high score like 10 is considered to be a "promoter". They are the one who will be more enclined to recommend the product to a friend or a family member. The objective is to reduce the number of paid customer and give more spaces to word-of-mouth advertising. Generally, high NPS means the retention of customers is also high (Customer alliance, 2022).

NPS surveys have the downside of not providing you with much information about the true reasons why your consumers feel the way they do. Along with the NPS rating questions, there are open-ended inquiries. These factors reflect whether customers are inclined or unlikely to recommend particular goods and services. This text's drivers of NPS are revealed through sentiment analysis (Customer alliance, 2022).

NPS is merely one sort of VoC survey. Any measure that you would be interested in—such as customer effort score, customer satisfaction, etc.—applies in the same way. What metric is employed really doesn't matter all that much. It's more crucial to understand what causes the metric's ups and downs. (Cleave P., 2022)

A fantastic VOC program takes into account customer feedback from all channels. You can see how even for a mid-size B2B company, it may quickly balloon to hundreds of thousands of pieces of feedback. To make sense of this data, sentiment analysis is essential.

Finally, businesses can swiftly identify clients who are expressing very unfavorable experiences and address pressing problems. We may spot nascent concerns and address them before they grow into greater difficulties by monitoring the customers' mood over time (Customer alliance, 2022).

3.3.2 Customer Service Experience

A corporation can succeed or fail based on its ability to provide excellent customer service. Customers want to know that their inquiry will be handled quickly, effectively, and professionally. Companies' customer service operations can be streamlined and improved with the aid of sentiment research (Vishnoi L., 2022).

Customer support discussions can benefit from sentiment analysis as well as text analysis. Conversations can be ranked automatically by urgency and topic using machine learning algorithms. Consider a community where people report technical problems. The posts with the most negative sentiment can be found using a sentiment

analysis method. These inquiries can be given higher priority for an inside expert. Other community members can respond to common queries (MonkeyLearn, 2022).
As we can see, by sending the proper kinds of questions to the right people, sentiment analysis can shorten processing times and boost productivity. In the end, we can lower churn rates and provide consumers with a better support experience (MonkeyLearn, 2022).

3.3.3 Product Experience

Sentiment analysis can identify how your customers feel about the features and benefits of your products. This can help uncover areas for improvement that you may not have been aware of.

Sentiment analysis can reveal how consumers feel about a product's advantages and characteristics. This can show us places where we might need to make improvements but weren't aware of. (MonkeyLearn, 2022)
For instance, we discovered that mobile check deposit was the most crucial function when we examined the tone of US banking app reviews. It's interesting that most apps struggled with this feature. Such information could be used in the marketing messages of companies that have the fewest complaints about this feature (Team H.C., 2022). Product managers can make changes to the functionality over time. After that, they can employ sentiment analysis to see whether customers are noticing an improvement in the use and dependability of the check deposit. (Merryweather E., 2021)

3.3.4 Brand Sentiment Analysis

Sales, attrition rates, and the likelihood that consumers will promote a certain brand to others can all be influenced by how consumers feel about it. The "Super Size" documentary, which followed director Morgan Spurlock for 30 days while he solely consumed McDonald's cuisine, was released in 2004. The company's profits in the UK dropped to their lowest levels in 30 years as a result of the accompanying media frenzy and other unfavorable press. In response, the business started a PR campaign to enhance its reputation (The Business Professor, 2022).

Brands may keep track of how their consumers feel about them by using sentiment analysis. They can monitor the reputation of their brand by analyzing

online forums, groups, and social media sites. They could also ask their clients in surveys what concerns are important to them (Anušić I., 2022).
To develop a long-term awareness of brand image, businesses also monitor mentions of their names, those of their products, and those of their rivals. This enables businesses to evaluate the effects of PR initiatives or the introduction of new products on consumer perceptions of their brands (Das R., 2022).

3.3.5 Social media sentiment analysis

Social networking is a potent tool for connecting with current consumers and attracting new ones. Positive customer feedback and social media posts inspire other customers to make purchases from the business. But the opposite is also accurate. Negative social media comments or reviews can cost your company a lot of money. According to Convergys Corp. research, a bad review on Facebook, Twitter, or YouTube can cost a business roughly 30 clients. Large financial losses might also result from defamatory social media posts against a business. One enduring instance is Elon Musk's tweet from 2020, which complained that the stock price of Tesla was excessive (Dynamic Business, 2021).

In just a few hours, Tesla's valuation dropped by $14 billion as a result of the trending tweet. These kinds of problems can be quickly identified using sentiment analysis to prevent escalation. Businesses can then take prompt action to limit financial costs and offset any harm to their brand's reputation (Trending F.P., 2020).

3.3.6 Market research

Companies can use sentiment analysis to investigate new markets, study rivals, and spot developing trends. Businesses could desire to investigate customer opinions of rivals' goods or services. This data can be subjected to sentiment analysis to determine what customers like and dislike about the goods of their rivals. These revelations may show you how to acquire a competitive advantage. For instance, sentiment research may show that customers of rivals are dissatisfied with their laptops' short battery life. The business might then emphasize in its marketing campaign how long its batteries last (Qualtricks, 2022).

To identify fresh prospects, sentiment analysis could also be used to analyze market reports and business magazines. For instance, research into real estate market statistics may show that a specific location is getting more favorable press. This information can imply that business insider’s view this region as a promising investment prospect. Then, by making investments before the rest of the market, these insights might be exploited to secure a head start (Qualtricks, 2022).

CHAPTER IV

Materials And Methodology

In this chapter, which is the chapter 4, we are going to describe and present our materials used for conducting the experiment but also display the methodology. It will be divided in two major parts, one for the materials and the last one for the methodology.

4.1 MATERIALS

4.1.1 Computer Specifications

To conduct our experiment, we needed to have access to the university lab with specific setup. Due to the high computation required for machine learning and deep learning tasks, such decision is well understood. We used a machine of brand Imac of the year 2019. The screen of the machine is made of a Retina 5K with 27 inches. The processor of the Imac is 3.6 GHz with 8 cores Intel. As a Memory, we used a 16 GB 2667 MHz DDR4. The graphics technology is made with a Radeon Pro 580X 8GB.

4.1.2 Descriptions of core Software

4.1.2.1 Tensorflow

Tensorflow's initial version was released in February 2017. It is a machine learning software library that was created by Google and is open-source. In other words, Tensorflow makes it simple to create and refine machine learning models. It is accessible on a variety of operating systems, including Windows, macOS, and Linux.

4.1.2.2 Keras

Tensorflow's Python API is called Keras. It is used to easily create deep neural networks because it comprises the majority of the standard neural network pieces such as activation functions, loss and cost functions, optimizers, layers, etc.

4.1.2.3 Nltk

Working with human language is made easier by the Natural Language Toolkit (Nltk), a framework made up of numerous libraries. Because of its libraries, Nltk is frequently used in Natural Language Processing (NLP) to handle a variety of tasks as tokenization, parsing, stemming, semantic reasoning, etc.

4.1.2.4 Numpy

Travis Oliphant founded the open-source Numpy project in 2005. It is mostly employed to manipulate intricate arrays and matrices. In order to make using various tools, such the Fourier Transform, easier, it also offers a large library of mathematical functions.

4.1.2.5 Pandas

Pandas is a Python software library released in January 2008. Its main objectives are data analysis and data manipulation. We can work with several formats, such as JSON, comma-separated values, SQL database tables or queries, Microsoft Excel, etc…

4.1.2.6 Huggingface-hub

The Hugging Face Hub is a platform that provides tools for building, training, and deploying machine learning models. The platform has more than 6,000 models and datasets. The objective of the platform is to share, explore, and experiment.

4.2 METHODOLOGY

4.2.1 DATASETS

In order to perform our study, we selected three benchmark datasets heavily used in the literature and previous works in the area. The selected datasets are:

- IMDB movie reviews
- AMAZON polarity Reviews
- Tweet U.S Airline Sentiment

4.2.1.1 IMDB movie reviews

The IMDB movie reviews dataset is a sizable collection of film reviews that have been categorized by polarities: "1" for positive and "0" for negative. The dataset is split into two equal sets, each with 25.000 samples. The training set and testing set are these sets, respectively (Andrew L. Maas et al., 2011).

4.2.1.2 AMAZON polarity reviews

The Amazon review dataset is a collection of reviews from over 18 years of Amazon. It contains approximately 35 million reviews with different polarities ranging from 1-5 where 1 and 2 scores are considered to be negative reviews. 4 and 5 scores are positive reviews and samples with 3 scores are neutral (J. McAuley and J. Leskovec, 2013).

4.2.1.3 Tweet U.S Airlines reviews

Around 55.000 reviews were gathered in February 2015 for the Airline Twitter Sentiment dataset, which is a decent collection. To determine and categorize each tweet's polarity as either good, negative, or neutral, the goal was to identify it (Crowdflower, 2016).

4.2.2 PROPOSED MODEL

To conduct our study, the below figure shows our proposed model. Firstly, the dataset is preprocessed during phase 1. This phase consists of cleaning the data, mostly texts, and modifying certain values of specific datasets to make the model creation easy. Secondly, we extract the features during phase 2. In this phase, we generate two kinds of features. The first one is word embedding, and the second one is a bag-of-bigrams. We generate these features for the model training. Lastly, after extracting the features, we train our models on phase 3. We used three classifier models, which are: Convolutional Neural Network (CNN), Long Shot-Term Memory (LSTM – RNN) and the pretrained model Bert. The classifiers at the end provide us with predictions which are either "positive" or "negative" to know the sentiment of the text.

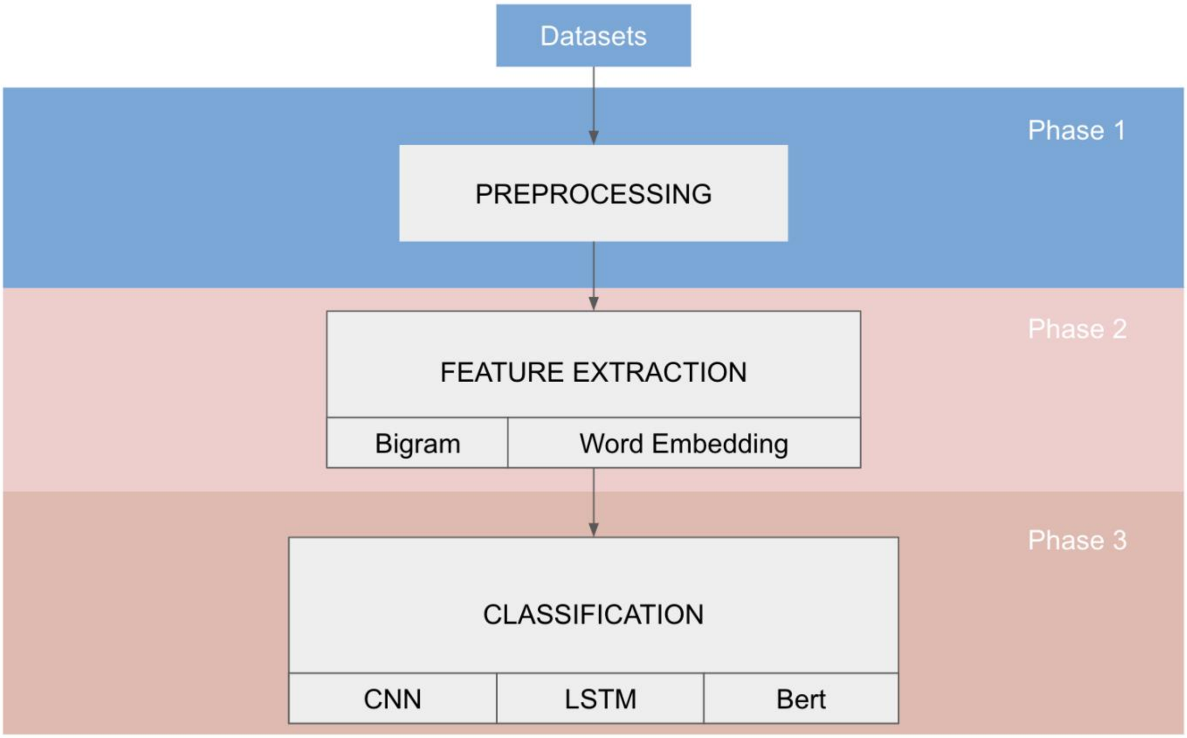

Figure 2: Diagram of the proposed model

4.2.3 PREPROCESSING

The preprocessing phase is the first phase in our model study. Phase 1's goal is to prepare the data for later processing and model creation. In real world data analysis, particularly in natural language processing, the data doesn’t come in the form that is ready for any processing, precisely in sentiment analysis. For example, a tweet may contain special characters and symbols that we need to get rid of in order to access only the important words that affect the sentiment of the tweet.

In our study, we go through several steps to preprocess our selected data.

After importing the raw data and extracting samples, we first clean the data, tokenize it, and finally use the stemming technique to apply stemming to the words contained in each text.

Another preprocessing task we did was to replace certain string values in our dataset with integers to facilitate the later processing.

We replaced in the Tweet U.S. Airline dataset, "negative" to 0, "positive" to 1, and neutral to "0". We also replaced 1 to 0 and 2 to 1 in the Amazon Polarity review dataset.

4.2.3.1 Text cleaning

One step in the preprocessing stage is text cleaning, which involves removing words or other elements that don't contain pertinent information and lower the efficacy and standard of the sentiment analysis study. The texts observed contain extra white spaces, symbols or special characters, punctuations but also stop words. In order to clean each text, we follow these steps listed below:

- Remove numbers from text;
- Remove English stop words and punctuations;
- Remove extra white spaces;
- Remove symbols;
- Convert the text to lower case;

4.2.3.2 Tokenization

In natural language processing, the act of tokenizing is used to separate a text into individual words for later processing. We can use them as features to train our classifier, as we will see later. We used the *word_tokenize* function from the Natural Language Toolkit to tokenize our texts from each dataset.

4.2.3.3 Stemming

In NLP, the process of stemming is utilized to get to the word's root. For example, if we have these three words: "ask," "asked," and "asking". The result after applying the stemming method is "ask." We used the *PorterStemmer* function from the Natural Language Toolkit to apply stemming to each word in our dataset.

4.2.4 FEATURE EXTRACTION

Feature extraction is the step just after the preprocessing phase that generates numerical features from raw data that can be processed, for example, to train a classifier without losing too much information. This technique has better results than processing the raw data directly from previous work in the area. In our study, we used two feature extraction techniques, such as Bigram from the bag of words concept and Word Embedding.

4.2.4.1 Bigrams

As a special case of N-grams, bigrams are a combination of two words (2-grams) in the corpus. The bag of words notion, which counts the number of well-known words in a document, is where the idea originated. The bigrams from the line "Let's get to school" are: "Let's get," "get to," and "to school."

In this study, we developed a global function that extracts bigrams from each text in the dataset using the bigrams function from the Natural Language Toolkit.

Due to its popularity in earlier studies in the field, we decided to use this methodology in our research. According to the text in Neural Network Methods in Natural Language Processing, page 75, "A bag-of-bigrams representation is much more powerful than a bag-of-words, and in many cases proves very hard to beat."

To include more features for training our models, we use Bigrams in our models. This occurs once the data has undergone preparation and is thoroughly cleansed and prepared for processing.

4.2.4.2 Word Embedding (Glove)

Word embedding is a method for representing words in natural language processing where each word is mapped to a single real-valued vector that symbolizes the meaning of the word. This definition means that words that are close in the vector space will have the same representation but not have the same meaning. In our study, we used the Glove algorithm to generate word embeddings. It stands for Global vectors for representation. Glove is an unsupervised learning algorithm created by Stanford University for generating word embeddings.

After the preprocessing phase, we added for each model a word embedding layer from Glove to take into account more features for the training phase.

4.2.5 CLASSIFIERS

After preprocessing the data, which means cleaning and adding some modifications to our datasets, we extracted more features from an embedding word or from a bag-of-bigrams as stated previously. The next step in our study was to make different configurations which pertain to each model. We built exactly nineteen models. Nine models are Convolutional Neural Network models (CNN), other nine models are Long Short-Term Memory (LSTM) models with different variations and the last one is our pretrained Bert model. The CNN models are as follow:

- Three models directly use the cleaned data as features.
- Three models in addition to the cleaned data use word embeddings generated by Glove algorithm.
- Three models use bag of bigrams as features.

The LSTM (RNN) models follow the same architecture.

4.2.5.1 Convolutional Neural Network (CNN)

Convolutional neural networks are artificial intelligence precisely deep learning algorithms that are able to take an image as input and give importance to the numerous objects in the image while being able to recognize them from each other. When assigning importance, it creates learnable weights and biases that it attaches to the objects. When compared to other classification methods, CNNs require fewer preprocessing tasks and are able to learn the features by themselves.

Deep neural networks of the type convolutional neural networks are frequently employed in the field of computer vision. They are made up of fully linked layers with many nodes, each of which has an output that is altered in accordance with the input value. The nodes are arranged so that the input of one node is also the output of another (O'Shea et al., 2015). The Convolutional Neural Network algorithm is composed of several layers with a minimum of 4 layers, all interconnected. We have the input layer, the convolution layer, the pooling layer and the output layer.

4.2.5.1.1 Input Layer

The input layer represent the raw data, for example, the raw image we pass to the convolutional neural network.

4.2.5.1.2 Convolution Layer

Convolutional Layers are always the initial layer in a convolutional neural network. Convolutional layers perform repetitive tasks with the help of the kernel that is called convolution operation. This operation is done on the input and then the outcome is sent to the following layer in the networkA convolution turns each pixel in its receptive area into a single value. When a convolution is applied to an image, for example, the image size is reduced and all field data is compressed into a single pixel. The final convolutional layer is a vector. Depending on the type of problem we have to solve and the traits we want to learn, we can use a variety of convolutional structures.

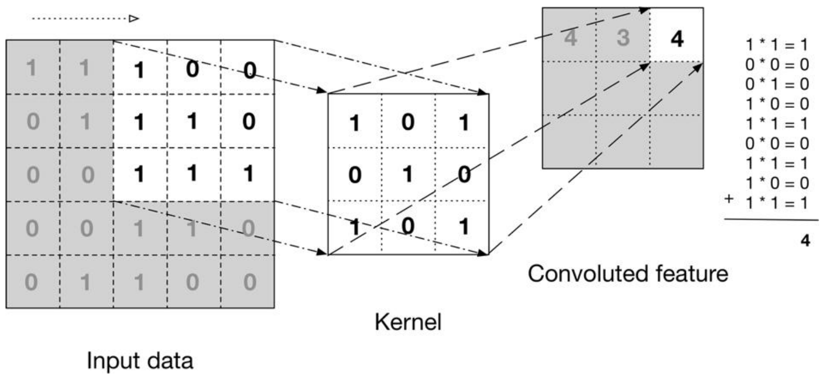

Figure 3 : Convolution layer illustration. Source (Pratiwi, N. C., et.al., 2020)

4.2.5.1.3 Pooling layer

Like the convolution layer, the pooling layer shrinks the size of the feature map output. The main goal is to reduce the computational power required to process the data. Additionally, it supports the model's efficient training process by assisting in the extraction of prominent features that are rotational and positional invariant.
We distinguish three different type of pooling. We have Minimum pooling, Average pooling and Maximal pooling. When using the maximum pooling, the space covered by the kernel returns the largest value while it returns the smallest value when using minimum pooling.

Concerning the average pooling, it is the average of all the values from the area covered by the kernel which is returned.
Max Pooling performs as a Noise Suppressant as well. The noisy activations are completely discarded after being de-noised, dedimensionalized, and reduced in size. Average Pooling, on the other hand, simply uses dimensionality reduction as a form

of noise suppression. Therefore, we may say that Max Pooling performs better than Average Pooling.

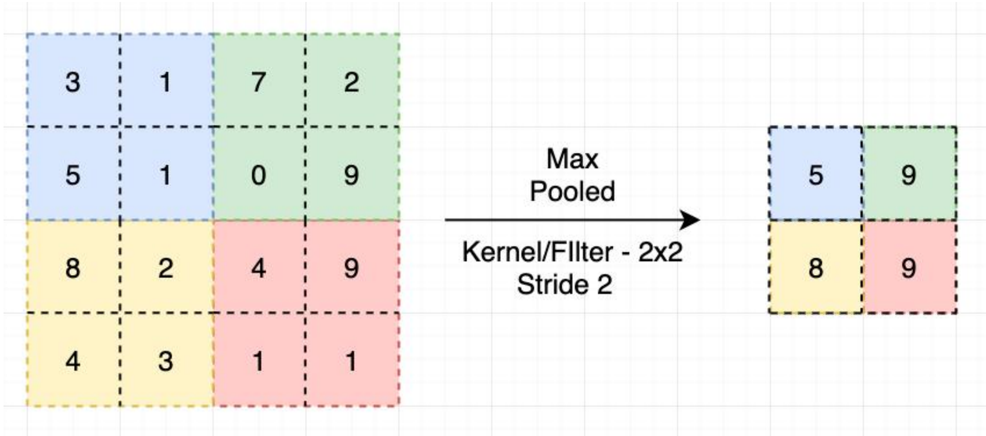

Figure 4: Max pooling illustration. Source (Podareanu, D., et.al., 2019)

4.2.5.1.4 Output Layer

The neural network's output layer, which is the last layer, is where the intended predictions are made. The final prediction that is wanted is produced by one output layer in a neural network. Prior to deriving the final result, it has its own set of weights and biases that are applied.

4.2.5.1.5 Proposed CNN Architecture

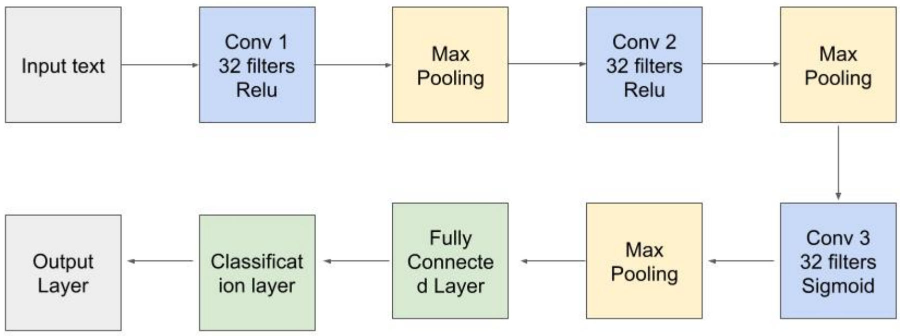

Figure 5: Proposed CNN model

4.2.5.1.6 Standard CNN

The models that directly employ the cleaned data as features are the typical CNN models. The three datasets receive roughly the same setups. The configurations we created for the conventional CNN for each dataset are described in the lines that follow. With a batch size of 15, each CNN model—whether conventional, CNN + Word Embedding, or CNN + Bigrams—was trained on five iterations.

4.2.5.1.7 IMDB movie reviews dataset

The weight of the embedding layer is initially set to 100. This function will learn the embedding of every words in the training dataset. The vocabulary size for the IMDB movie reviews dataset is 37759 words. The output dimension is set to 100, and the text can be as long as 1482 characters. This results in a matrix of 1482 x 100. Three CNN models are arranged in a stack, which we specify. A filter with a kernel size of 1, 2, or 3 is present in each model. For each model, we define 64 filters as well. A 1482x32 neuron matrix is the output of the first model's neural network layer. We should take note of the fact that this outcome will be input into the subsequent CNN model, which will have all of the same parameters defined with the exception of the kernel size, which will be increased by 1, and the maximum length, which will be 1481.

A 1481X32 matrix is the output of the second layer. The third layer's output matrix, which uses the same logic, is 1480 x 32.

A dropout layer and a max pooling layer were added. Their task is to stop the data from being overfit. We have picked a size of 4 and a dropout of 0.5 for the maximum pooling layer. For each model, we added a 370X32 max pooling matrix. The first model for the dropout layer is created as a matrix of 1482 x 32, followed by the second model's matrix of 1481 x 32, and the last model's matrix of 1480 x 32.

The vector of 10 is finally reduced to 1 for the final layer using a sigmoid activation function, yielding a result that can be either "positive" or "negative."

Layer (type)	*Output Shape*	*Number of Parameters*
Input_1 (inputlayer)	(None, 1482)	0
Input_2 (inputlayer)	(None, 1482)	0
Input_3 (inputlayer)	(None, 1482)	0
Embedding (embedding)	(None, 1482, 100)	3775900
Embedding_1 (embedding)	(None, 1482, 100)	3775900
Embedding_2 (embedding)	(None, 1482, 100)	3775900
Conv1d (conv1d)	(None, 1482, 32)	3232
Conv1d_1 (conv1d)	(None, 1481, 32)	6432
Conv1_2 (conv1d)	(None, 1480, 32)	9632
Dropout (dropout)	(None, 1482, 32)	0
Dropout_1 (dropout)	(None, 1481, 32)	0
Dropout_2 (dropout)	(None, 1480, 32)	0
Max_pooling1d (maxpooling1d)	(None, 370, 32)	0
Max_pooling1d_1 (maxpooling1d)	(None, 370, 32)	0
Max_pooling1d_2 (maxpooling1d)	(None, 370, 32)	0
Flatten (flatten)	(None, 11840)	0
Flatten_1 (flatten)	(None, 11840)	0
Flatten_2 (flatten)	(None, 11840)	0
Concatenate (concatenate)	(None, 35520)	0
Dense (dense)	(None, 10)	355210
Dense_1 (dense)	(None, 1)	11

Table 1: Summary of the CNN model's structure for the IMDB dataset.

4.2.5.2 Long Short-Term Memory (LSTM - RNN)

LSTM stands for Long Short-Term Memory. It is an artificial intelligence and deep learning recurrent neural network algorithm. In addition to analyzing single data points like photographs, this algorithm can also evaluate entire data sequences like speech or video. We use LSTM for several purposes like speech recognition, robot control, machine translation, video games etc...

Input, output, forget, and cell are the components of general LSTM unit. The three gates regulate information flow into and out of the cell, enabling the cell to store values over a wide range of time periods.

An LSTM network is comprised of what we call cells. Each cell transfers two types of information: the cell state and the hidden state. The memory blocks are what allow the algorithm to remember things, and there are three main gates that allow you to manipulate this memory.

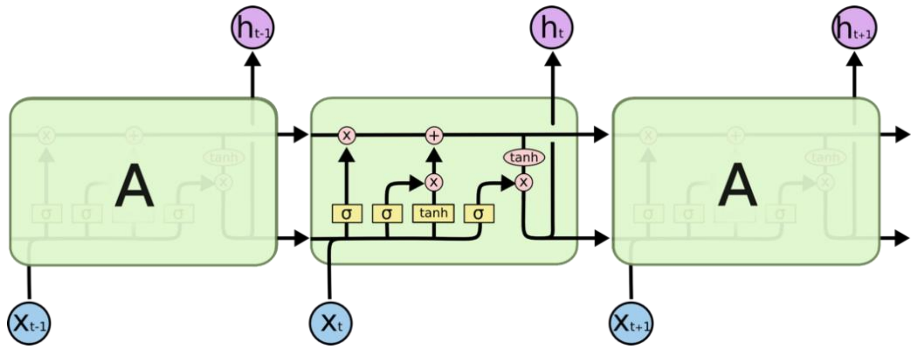

Figure 6: LSTM structure illustration. Source (Fu, J., et.al., 2019)

4.2.5.2.1 Forget Gate

The function of a forget gate is to remove data which exist in the state of the cell. By multiplying a filter, information that is no longer required for the LSTM to understand anything or information of lesser relevance is removed. This is an important step in order to improve the LSTM network's performance.

4.2.5.2.2 Input Gate

The input gate is responsible for adding new data into the cell's state.
Every potential value that could be added to the cell state is included in the vector that the algorithm creates. The sigmoid function will then be used to this vector to determine which values should be added to the cell state. In order to achieve this, the output of the regulatory filter (the sigmoid gate), which has values between -1 and +1, is multiplied by the output of the tanh function, which creates values between -1 and +1, and added to the cell state.

4.2.5.2.3 Output Gate

Again, the operation of an output gate can be divided into three steps:

- After scaling the values to the range of -1 to +1 using the tanh function, a vector is produced.
- Using the values of h t-1 and x t to create a filter that can regulate the values that need to be generated from the vector created previously. Once more, this filter makes use of the sigmoid function.

- multiplying the vector created in step 1 by the value of this regulatory filter, adding the result, and transferring the result to the subsequent cell's output and hidden state.

4.2.5.2.4 Proposed LSTM model

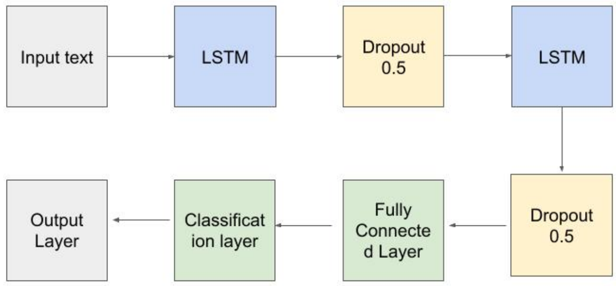

Figure 7: Proposed LSTM model

4.2.5.2.5 Standard LSTM

We initialized the embedding layer with an output dimension of 128 units. Then we define an LSTM model of 128 units to match the embedding layer. We added a dropout layer to avoid over fitting, and finally, we added a sigmoid activation function to have only one prediction. We trained all LSTM models on three epochs with a batch size of 64.

4.2.5.2.6 IMBD movie reviews dataset

Layer (Type)	*Output Shape*	*Number of parameters*
Embedding (Embedding)	(None, None, 128)	4833152
Lstm (LSTM)	(None, 128)	131584
Dense (Dense)	(None, 1)	129

Table 2: Standard LSTM model summary on IMDB movie reviews dataset.

4.2.5.3 Biderectional Encoder Representations Transformers (BERT)

Bidirectional Encoder Representations from Transformers is known as BERT. It is a method of understanding natural language processing using artificial intelligence. The system was created in 2018 by the Google AI Language Researchers Team and may be used for over eleven common language applications, including sentiment analysis and named entity recognition. BERT's ongoing performance has been aided by a sizable dataset of 3.3 billion words.

BERT was specifically trained using the 800M words and 2.5B words combined from Google's BooksCorpus and Wikipedia. These substantial informative datasets helped BERT develop a thorough understanding of not only the English language but also of our planet.

We used the pretrained model Bert-base-multilingual-uncased-sentiment in our study. The model can be found on the hugging face platform. The pre-trained model helps us to easily construct a model that will predict the sentiments in our datasets.

CHAPTER V
Results

In this chapter, we are going to show the results of our experiments and discuss each set of results in order to better understand the implications of the results.In our study, we considered two metrics, which are accuracy and the duration time of training the models. We discussed them in detail and compared them with previous work in the area.

5.1 CNN

5.1.1 Standard CNN

With the standard CNN models, we had at the end of each training six results for three datasets. In other words, this means three results for the training datasets and three results for the test datasets. The obtained accuracy results are shown in the tables below:

Train sets:

Tweet U.S Airlines	***IMDB movie review***	***Amazon Polarity R.***
99%	99.45%	99.90%

Table 3: Training accuracy of each standard CNN model for each dataset.

Test sets:

Tweet U.S Airlines	***IMDB movie review***	***Amazon Polarity R.***
89%	85.96%	76.95%

Table 4: Test accuracy of each standard CNN model for each dataset.

5.1.2 CNN + Word Embedding

The CNN + Word Embedding models are the models that use, in addition to the cleaned data, generated word embeddings as features. This adds more details to the features for model training. The below tables show the accuracy results :

Train sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
97.30%	99.27%	99.50%

Table 5: Training accuracy of each CNN + Word Embedding model for each dataset.

Test sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
90.30%	85.58%	74.75%

Table 6: Test accuracy of each CNN + Word Embedding model for each dataset.

5.1.3 CNN + Bigrams

The CNN + bigrams models add more features to the training of the models by using the extracted bigrams from the datasets. We show the accuracy results below.

Train sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
99.74%	99%	99%

Table 7: Training accuracy of each CNN + Bigrams model for each dataset.

Test sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
86.01%	50%	51.74%

Table 8: This table shows the test accuracy of each CNN + bigrams model for each dataset.

The next tables show a resume of the different accuracies for the CNN models.

Train sets:

	CNN		
	Standard CNN	***CNN + Word embedding***	***CNN + Bigram***
Airline tweets	99%	97.30%	99.74%
IMDB movies review	99.45%	99.27%	99%
Amazon review	99.90%	99.50%	99.00%

Table 9: Resume of train accuracies for each CNN model

Test sets:

	CNN		
	Standard CNN	***CNN + Word embedding***	***CNN + Bigram***
Airline tweets	89%	90.30%	86.01%
IMDB movies review	85.96%	85.58%	50%
Amazon review	76.95%	74.75%	51.74%

Table 10: Resume of test accuracies for each CNN model

CNN			
	Standard CNN	***CNN + Word embedding***	***CNN + Bigram***
Airline tweets	0:00:09	0:00:05	0:00:16
IMDB movies review	0:12:49	0:13:10	1:54:45
Amazon review	0:00:25	0:00:14	0:01:15

Table 11: The processing time for training each CNN model.

5.2 RNN-LSTM

5.2.1 Standard RNN-LSTM

The models are using only the cleaned data as features, without any additional work. Below, the results we had after training the models.

Train sets:

Tweet U.S Airlines	***IMDB movie review***	***Amazon Polarity R.***
48.85%	50.49%	48.75%

Table 12: Training accuracy of each standard RNN-LSTM model for each dataset.

Test sets:

Tweet U.S Airlines	***IMDB movie review***	***Amazon Polarity R.***
39%	49.50%	51.89%

Table 13: Test accuracy of each standard RNN-LSTM model for each dataset.

5.2.2 RNN-LSTM + Word Embedding

As we did for the CNN models, we generated word embeddings and trained stacked LSTM models with them. Below are the results.

Train sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
87.29%	50.49%	53.75%

Table 14: Training accuracy of each RNN LSTM + Word embedding model for each dataset.

Test sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
88.45%	49.50%	49.59%

Table 15: Test accuracy of each RNN LSTM + Word embedding model for each dataset.

5.2.3 RNN-LSTM + Bigrams

We trained RNN LSTM models on bigrams directly extracted from each dataset. These are the results:

Train sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
98.88%	50%	90.54%

Table 16: Training accuracy of each RNN LSTM + bigrams model for each dataset.

Test sets:

Tweet U.S Airlines	*IMDB movie review*	*Amazon Polarity R.*
13.98%	50%	48.25%

Table 17: Test accuracy of each RNN LSTM + bigrams model for each dataset.

RNN-LSTM			
	Standard LSTM	***LSTM + Word embedding***	***LSTM + Bigram***
Airline tweets	49%	87.29%	98.88%
IMDB movies review	50.49%	50.49%	50%
Amazon review	48.75%	53.75%	90.54%

Table 18: Resume of train accuracies for each RNN-LSTM model

RNN-LSTM			
	Standard LSTM	***LSTM + Word embedding***	***LSTM + Bigram***
Airline tweets	39%	88.45%	13.98%
IMDB movies review	49.50%	49.50%	50%
Amazon review	51.89%	49.59%	48.25%

Table 19: Resume of test accuracies for each RNN-LSTM model

RNN-LSTM			
	Standard LSTM	***LSTM + Word embedding***	***LSTM + Bigram***
Airline tweets	0:00:08	0:00:08	0:00:09
IMDB movies review	0:30:07	0:22:11	0:29:00
Amazon review	0:00:29	0:00:18	0:00:27

Table 20: Processing time for training each RNN-LSTM model.

5.3 BERT

We used the pretrained model Bert-base-multilingual-uncased-sentiment in our experiment. For the sake of simplicity, we directly tested the model. These are the results.

Test sets:

Tweet U.S Airlines	IMDB movie review	Amazon Polarity R.
75.69%	87%	90.25%

Table 21: Test accuracy of the pretrained Bert model for each dataset.

CHAPTER VI

Discussion

The primary goal of our research is to provide the community with the most recent comparative study of the most commonly used deep learning models in sentiment analysis. This study arose as a result of the fact that most recent studies, particularly comparative studies, did not include a model known as Bert. Nhan C.D. et al. compiled a review of 32 studies in the sentiment analysis field. In their research, the majority of works did not include a transformer model. In addition, based on the different performances observed, this study provides a new perspective on what to use when confronted with a sentiment analysis problem. The study also excels at confirming the various performances of each model on the various datasets.

Again, we did things differently than previous researchers in the field by comparing word embeddings and n-grams. The majority of the studies examined focused solely on n-grams due to their claimed performance. We used n-grams and word embeddings techniques to fully comprehend their performance and effects.

6.1 Preprocessing

In our implementation, preprocessing was consistent across all datasets. We cleaned up the dataset by removing stopwords, as well as white spaces, symbols, and numbers. We also used a technique known as stemming to improve the model's performance. We had a choice between lemmatization and stemming at the start of the process. After running some tests on samples from each dataset, we ultimately decided to use stemming. By incorporating this layer of transformation into the dataset's structure, regularities are created that aid the model in recognizing patterns in the datasets (Gini, G., et al. 2019).

Another observation we made at this stage, particularly on the U.S. Tweet Airline Sentiment dataset, is the impact of our cleaning strategy on model performance. The dataset U.S. Tweet Airline Sentiment has three labels: positive, negative, and neutral. When we train the model on these labels, we see a significant overfitting of the model, with the training accuracy being nearly 100% and the test accuracy being 20%. Our solution was to convert the neutral labels into negative labels in the dataset.

This transformation significantly improved the accuracy of the models. We still saw overfitting in the models, but we realized that this was due to the nature of the dataset itself. This demonstrates that the nature of the dataset influences model performance (Choi, Y., and Lee H.J., 2017).

6.2 Feature Extraction

We use two feature extraction techniques, as stated above: n-grams and word embeddings. The implementation of the word embedding technique proved to be more difficult than the implementation of the n-gram technique. However, models that extracted features from word embeddings performed better. This means that all of the models can detect patterns in the data. We also discovered that training models on bigram-processed datasets took longer than training models on word embedding datasets. This implies that using a word embedding for sentiment analysis is by far the best option.

6.3 Classification

The tables presented in the results section overall tell us that the Word Embedding technique produced higher performances compared to standard models or bigram techniques. The bigrams technique performs the worst among the different techniques used in our study, with a longer duration time for training the models. It took almost two hours with the bigrams technique to train the CNN model on the IMDB movie reviews dataset.

On the other hand, we recorded a shorter duration time with the embedding technique on the Tweet U.S. Airline Sentiment dataset.

The Bert model performed relatively well, with high accuracy on each dataset. Also, we don't ignore that the type of dataset influences considerably the results of a sentiment analysis. (Choi, Y., and Lee H.J., 2017).

The CNN model was easy to implement and fast to train. However, the duration time may drastically increase when the density of the dataset also increases. A good example is the processing time contrast between the IMDB movie reviews dataset and the Tweet U.S. Airline Sentiment dataset. The average accuracy of CNN models is between 75% and 90% for the test dataset and 97% to 99% for the training set.

In contrast to the study of Nhan C. (Nhan C., et.al., 2020), RNN has the lowest reliability when either word embedding, or bigram is applied. In addition, it has the highest processing time. This last result supports the results obtained by the same study.

We also observed overfitting of the data when applying the bigrams technique on the Tweet U.S. Airline Sentiments dataset and the Amazon Polarity Reviews dataset. The average accuracy recorded is between 14% and 88% on the test set.

In several comparative studies, we noticed that only accuracy has been taken as the mode of evaluation technique. However, the work done by Nhan C. added time as a crucial mode of evaluation technique.

We consider that with the complexity and the density of today's data we encounter, taking time into consideration is one of the most important criteria for selecting the best model.

In our study, as opposed to Nhan C's work, three techniques (Word Embedding, bigrams and standard) are examined on two deep learning algorithms for the purpose of extending our knowledge of overview performances of sentiment analysis with deep learning.

Finally, we can summarize our findings as follow:

- CNN models give the highest scores on the three datasets using all three techniques. Its processing time is also noticeable as it is the lowest. Compared to the work done by *Nhan C. (Nhan C., et.al, Sentiment Analysis Based on Deep Learning: A comparative study, 2020*), we observe the same similarity of performances between CNN and RNN.
- We observed that the word embedding is the best technique to use compared to bigrams.
- We noticed that the Tweet U.S Airlines Sentiment dataset gave on average the highest score compared to other datasets used in the study.

CHAPTER VII

Conclusion

Sentiment analysis is now ubiquitous. Industries are using it to improve their products and services, increase customer satisfaction, and find solutions to more complex problems in the finance and medicine domains.

In our study, we focused on the analysis of deep learning models such as CNN, LSTM, and Bert, but also feature extraction techniques like Word embeddings and n-grams. We first of all preprocessed the data by cleaning it with various techniques. The first one was to delete all the symbols and special characters contained in the corpus. White spaces and numbers were also removed. After the text had been cleaned, we extracted the essential features using Word embeddings and Bigrams. We used the Glove model to inlay words. The various models were created and trained as the final phase. We developed three different model types: CNN, LSTM, and Bert. Each model was developed in a number of different ways.

We found that CNN models perform the best in terms of accuracy and processing time, but also that word embeddings seem to be the best choice when performing sentiment analysis. The maximum accuracy obtained is 90.30%. As social beings, we humans rely mostly on our emotions to make decisions. With the help of sentiment analysis, we can tremendously improve the way we live every day.

REFERENCES

Abid F., Alam M., Yasir M., Li C.J. (2019). Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. Future Gener. Comput. Syst. 95, 292-308.

Ahuja, R., Chug, A., Kohli, S., Gupta, S., Ahuja, P. (2019). The Impact of Features Extraction on the Sentiment Analysis. Procedia Computer Science, 152, 341–348.

Alam, M., Wang, J. F., Guangpei, C., Yunrong, L., & Chen, Y. (2021). Convolutional Neural Network for the Semantic Segmentation of Remote Sensing Images. Mobile Networks and Applications, 26(1), 200–215.

Alexander M., Elmina H., Francisco C., Ofer E. (2021).Sentiment analysis using TF–IDF weighting of UK MPs' tweets on Brexit.

Alharbi A.S.M., de Doncker E. (2019). Twitter sentiment analysis with deep neural network: An enhanced approach using behavioral information. Cogn. Syst. Res. 54, 50-61.

Aman Ullah, M., Arif Hasnayeen, M., Shan-A-Alahi, A., Rahman, F., & Akhter, S. (2020). A Search for Optimal Feature in Political Sentiment Analysis. *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. (2011). Learning Word Vectors for Sentiment Analysis. The 49th Annual Meeting of the Association for Computational Linguistics (ACL 2011).

Anušić, I. (2022, May 13). Brand Reputation 101: Monitoring, Analysis, and Management Tools. Blog.

Aytuğ Onan. (2020). Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks.

Azeema S., Fariha K., Fatima B. (2018). An Overview of Lexicon-Based Approach For Sentiment Analysis.

Bhavitha B., Rodrigue, A.P., Chiplunkar, N.N. (2017). Comparative study of machine learning techniques in sentimental analysis. In Proceedings of the 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India; pp. 216–221.

Bin W., Angela W., Fenxiao C., Yuncheng W., Jay K. C.-C. (2019). Evaluating word embedding models: methods and experimental results.

Brand Loyalty - Explained. (n.d.). The Business Professor, LLC. Retrieved October 26, 2022.

Budi M. M.; Widyantoro D. H. (2018). Aspect-Based Sentiment Analysis Approach with CNN.

Chi-Han D., Ming-Feng T.; Chuan-Ju W. (2019). Beyond Word-level to Sentence-level Sentiment Analysis for Financial Reports.

Choi, Y., Lee, H.J. (2017). Data properties and the performance of sentiment classification for electronic commerce applications. Inf. Syst. Front. 19,993-1012.

Cleave, P. (2022, September 23). What Is Net Promoter Score? SmartSurvey.

Customer A. (2022, July 6). Customer Satisfaction Metrics Explained:Customer Effort Score (CES).

Das, R. (2022, June 9). Brand Monitoring: How To Improve Brand Awareness Statusbrew. Statusbrew Blog.

Dimililer, K., Ever, Y. K., & Ugur, B. (2016). ILTDS: Intelligent Lung Tumor DetectionSystem on CT Images. Advances in Intelligent Systems and Computing, 225–235. https://doi.org/10.1007/978-3-319-47952-1_17

Dimililer K., Yoney, K. E., & Haithm, R. (2016, August 30). Intelligent eye tumour detection system. Science Direct Assets.

Dimililer K., Teimourian H., Al-Turjman F. (June 2022). Radio Galaxies Classification System, Using Machine Learning, Journal of Experimental & Theoretical Artificial Intelligence, Taylor Francis (SCI-E), DOI: https://10.1080/0952813X.2022.2080277.

Elena R., Martin H., Matthias W., Marcelo J., Štefan E., Michael S. (2018). More than Bags of Words: Sentiment Analysis with Word Embeddings.

Eman S.A., Norah S.A. (2021). Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings.

Erion Ç., Maurizio M. (2019). Word Embeddings for Sentiment Analysis: A Comprehensive Empirical Survey.

Farzindar, A., & Inkpen, D. (2015). Natural Language Processing for Social Media. *Synthesis Lectures on Human Language Technologies*, *8*(2), 1–166.

Fu, J., Chu, J., Guo, P., & Chen, Z. (2019). Condition monitoring of wind turbine gearbox bearing based on deep learning model. Ieee Access, 7, 57078-57087.

Gang L., Fei L. (2010). A clustering-based approach on sentiment analysis.

Gini, G., Zanoli, F., Gamba, A., Raitano, G., & Benfenati, E. (2019). Could deep learning in neural networks improve the QSAR models?. SAR and QSAR in Environmental Research, 30(9), 617-642.

Goldberg, Yoav (2016). "A Primer on Neural Network Models for Natural Language Processing". *Journal of Artificial Intelligence Research*. **57**: 345–420.

Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron (2016). *Deep Learning*. MIT Press.

Hu X., Bing L., Lei S., Philip S.Y. (2018). Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction.

Huang F., Li X., Yuan C., Zhang S., Zhang J., Qiao S. (2021). Attention-Emotion-Enhanced Convolutional LSTM for Sentiment Analysis.

Huilgol, P. (2020, December 23). *Quick Introduction to Bag-of-Words (BoW) and TFIDF for Creating Features from Text*. Analytics Vidhya.

Huq, M.R.; Ali, A.; Rahman, A. (2017). Sentiment analysis on Twitter data using KNN and SVM. IJACSA Int. J. Adv. Comput. Sci. Appl., 8, 19–25.

Imperial, J. M., Roxas, R. E., Campos, E. M., Oandasan, J., Caraballo, R., Sabdani, F. W., & Almaroi, A. R. (2019). Developing a machine learning-based grade level classifier for Filipino children's literature. *2019 International Conference on Asian Language Processing (IALP)*.

Jain, A.P., Dandannavar, P. (21–23 July 2016). Application of machine learning techniques to sentiment analysis. In Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Karnataka, India; pp. 628–632.

Kaur, H., Mangat, V., & Nidhi. (2017). A survey of sentiment analysis techniques.

Kayali, D., Olawale O. P., Ever Y. K., Dimililer K. (September 2022). "The effect of Compressor-Decompressor Networks with different image sizes on Mask Detection using Convolutional Neural Networks - VGG-16 "Akıllı Sistemlerde Yenilikler ve Uygulamaları Konferansı, ASYU2022.

Krizhevsky, I. Sutskever and G. E. Hinton. (2012). "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, pp. 1097-1105.

Li L., Goh T.-T, Jin D. (2018). How textual quality of online reviews affect classification performance: A case of deep learning sentiment analysis. Neural Comput. Appl., 1-29.

Li X., Bing L., Zhang W., Lam W. (2019). Exploiting BERT for End-to-End Aspect-based Sentiment Analysis.

Li, X., Sun, X., Xu, Z. (2021). Explainable Sentence-Level Sentiment Analysis for Amazon Product Reviews. 2021 5th International Conference on Imaging, Signal Processing and Communications (ICISPC).

Lutz B., Pröllochs N., Neumann D. (2018). Sentence-Level Sentiment Analysis of Financial News Using Distributed Text Representations and Multi-Instance Learning,

M. Hoang, O. A. Bihorac, and J. Rouces. (2019). Aspect-Based Sentiment Analysis using BERT. In Proceedings of the 22nd Nordic Conference on Computational Linguistics, pages 187–196, Turku, Finland. Linköping University Electronic Press.

Mamata Das, Selvakumar Kamalanathan and P.J.A. Alphonse. (2021). A Comparative Study on TF-IDF Feature Weighting Method and its Analysis using Unstructured Dataset.

Marco P., Mirko V., Hamido F., Massimo E. (2021). Multilingual evaluation of preprocessing for BERT-based sentiment analysis of tweets.

Mary S., Srividya K. (2019). Aspect Based Sentiment Analysis using POS Tagging and TFIDF.

Matheus A., Julio R., Adriano P., Fabricio B. (2016). An evaluation of machine translation for multilingual sentence-level sentiment analysis.

Matheus G.S., Kenzo S., Lucas de Souza R., Pedro H.M., Eraldo R.F., Edson T.M. (2019). BERT for Stock Market Sentiment Analysis.

McDonald R., Täckström O. (2016). Semi-supervised latent variable models for sentence level sentiment analysis.

Merryweather, E. (2021, October 18). 5 Reasons Why It’s Important to Understand Data as PM. Product School.

Minghai Chen, Sen W., Paul P. L., Tadas B., Amir Z., Louis-Philippe M. (2017). Multimodal sentiment analysis with word-level fusion and reinforcement learning.

Mohammad E. B., Arman K. (2017). Sentence-level sentiment analysis in Persian.

MonkeyLearn Blog. (2019). Understanding TF-ID: A Simple Introduction.

Mrityunjay S., Amit Kumar J., Shivam P. (2021). Sentiment analysis on the impact of coronavirus in social life using the BERT model.

Naseem, U., Razzak, I., Musial, K., Imran, M. (2020). Transformer based Deep Intelligent Contextual Embedding for Twitter sentiment analysis. Future Generation Computer Systems, 113, 58–69.

Nehal M. A., Marwa M.A.E.H., Aliaa Y. (2019). Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models.

Nhan C.D., María N. M-G., Fernando De la P. (2020). Sentiment Analysis Based on Deep Learning: A comparative study.

Nikos E., Angeliki L., Georgios P. (2011). ELS: A Word-Level Method for Entity-Level Sentiment Analysis.

O'Shea, K., & Nash, R., An Introduction to Convolutional Neural Networks., 2015.

Oscar B.D., William A.A., Felix A.L., A. Jeffery A. (2018). Sentiment Analysis with Word Embedding.

Pang, G., Lu, K., Zhu, X., He, J., Mo, Z., Peng, Z., & Pu, B. (2021). Aspect-Level Sentiment Analysis Approach via BERT and Aspect Feature Location Model. *Wireless Communications and Mobile Computing*, *2021*, 1–13.

Pavitra R., Kalaivaani P C D. (2015). Weakly supervised sentiment analysis using joint sentiment topic detection with bigrams.

Pinto, D., McCallum, A., Wei, X., Croft W.B. (28 July–1 August 2003).Table extraction using conditional random fields. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval, Toronto, ON, Canada; pp. 235–242.

Podareanu, D., Codreanu, V., Sandra Aigner, T. U. M., van Leeuwen, G. C., & Weinberg, V. (2019). Best Practice Guide-Deep Learning. Partnership for Advanced Computing in Europe (PRACE), Tech. Rep, 2.

Poornima; K. Sathiya P. (2020). A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques.

Pratiwi, N. C., Ibrahim, N., Fu'adah, Y. N., & Masykuroh, K. (2020, December). Computer-Aided Detection (CAD) for COVID-19 based on Chest X-ray I mages using Convolutional Neural Network. In IOP Conference Series: Materials Science and Engineering (Vol. 982, No. 1, p. 012004). IOP Publishing.

Ram K. M., Siddhaling U., Angel A. J. J. (2019). A Sentiment analysis-based hotel recommendation using TF-IDF Approach.

Ruiqi C., Yanquan Z., Liujie Z., Xiuyu D. (2019). Word-level sentiment analysis with reinforcement learning.

Sabarmathi . G, Dr.R.Chinnaiyan. (2021). Sentiment Analysis for Evaluating the Patient Medicine Satisfaction.

Satuluri V.; Meena B. (2018). Aspect-Level Sentiment Analysis on E-Commerce Data.

Scientific Integrity, Turing Lecture. (2021,December 31). Turing Award for Deep Learning.Supsi.

Sekeroglu B., Ever Y. K., Dimililer K., Al-Turjman F.(July 2022). "Comparative Evaluation and Comprehensive Analysis of Machine Learning Models for Regression Problems ", Data Intelligence, MIT Press, (E-SCI),Vol:4, Issue:3, pp. 620-652, DOI:https://doi.org/10.1162/dint_a_00155.

Sekeroglu B., Dimililer K., Tuncal K., (2019), Artificial Intelligence in Education: application in student performance evaluation.

Seyed M. R., Rouhollah R., Ali G., Hadi V. (2019). Sentiment analysis based on improved pre-trained word embeddings.

Shahid D. (2021). Convolutional Neural Network - Towards Data Science. Medium, December 8.

Shekhar, S. (2021, June 30). LSTM for Text Classification | Beginners Guide to Text Classification.Analytics Vidhya.

Shervin M., Elham A., AmirAli A. (2019). Deep Sentiment: Sentiment Analysis Using Ensemble of CNN and Bi-LSTM Models.

Shiyang L., Wang J., Yu R., Sato K. and Cheng Z. (2017). "CNN for situation understanding based on sentiment analysis of twitter data", 8th International Conference of Advances in Information Technology IAIT2016, vol.111, pp. 376-381,

Sida W. and Christopher D. (2012). Manning, Baselines and Bigrams: Simple, Good Sentiment and Topic Classification.

Singh V. K., Piryani R.; Uddin A. ;Waila P. (2013). Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification.

Sohangir S., Wang D., Pomeranets A., Khoshgoftaar T.M. (2018). Big Data: Deep Learning for financial sentiment analysis. J. Big Data 2018; 5, 3.

Soni S., Sharaff A. (2015).Sentiment analysis of customer reviews based on hidden markov model. In Proceedings of the 2015 International Conference on AdvancedResearch in Computer Science Engineering & Technology (ICARCSET 2015), Unnao, India, 6 March 2015; pp. 1–5.

Sumaya I.M., Md. Sadiqur R. M., Zannatun N., Md. Al Mamun. (2021). Sentiment Analysis Of English Tweets Using Bigram Collocation.

Susanti F., Rianto R., Acep I. G. (2020). Sentiment Analysis Provider By.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method.

Sreekavitha P., Vijjini A. R., Radhika M. (2018). BCSAT : A Benchmark Corpus for Sentiment Analysis in Telugu Using Word-level Annotations.

Srinidhi, S. (2021, December 12). Understanding Word N-grams and N-gram Probability in Natural Language Processing. Medium.

Team, H. C. (2022, October 26). 11 Reasons Why Business Communication is Critical to Your Company's Success. Haiilo.

Thematic. (2019,March 20). Sentiment Analysis, Comprehensive Beginners Guide, Thematic. .

Thomas R. (2019). Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages.

Tomoki I., Kota T., Hiroki S., Tatsuo Y., Kiyoshi I. (2020).Word-Level Contextual Sentiment Analysis with Interpretability.

Toshitaka H., Hamido F. (2019). Word Embeddings-based Sentence-Level Sentiment Analysis considering Word Importance.

Vanaja, S., & Belwal, M. (2018). Aspect-Level Sentiment Analysis on E-Commerce Data. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA).

Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L.,Gomez A. N., Kaiser L., Polosukhin, Illia (2017-06-12). "Attention Is All You Need".

Veselovská K. (2011). Sentence-Level Polarity Detection in a Computer Corpus. WDS'11 Proceedings of Contributed Papers, Part I, 167–170, 2011.

Wu O., Yang T., Li M., Li M. (2020). Two-Level LSTM for Sentiment Analysis With Lexicon Embedding and Polar Flipping.

Xiao M.,Jiangfeng Z., Limei P., Giancarlo F., Yin Z. (2019). Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis.

Xin H., Wenbin Z., Xuejiao T., Mingli Z., Jayachander S., Vasileios I., Zhen L., Ji Z. (2021). LSTM Based Sentiment Analysis for Cryptocurrency Prediction.

Yequan W., Aixin S., Minlie H., Xiaoyan Z. (2019). Aspect-level Sentiment Analysis using AS-Capsules.

Yoney K. E., Dimililer K., (2017), The effectiveness of a new classification system in higher education as a new e-learning tool.

Yuanhe T., Guimin C., Yan S. (2021). Enhancing Aspect-level Sentiment Analysis with Word Dependencies.

Yue F., Yan C. (2021). Short Text Sentiment Analysis Based on Multi-Channel CNN With Multi-Head Attention Mechanism.

Zeng D., Dai Y., Li F., Wang J., Sangaiah A. K. (2019). Aspect based sentiment analysis by a linguistically regularized CNN with gated mechanism.

Zhang Y., Wang J., Zhang X. (2021). Conciseness is better: Recurrent attention LSTM model for document-level sentiment analysis.

Zhigang J., Yang Y., Yuhong L. (2019). Stock closing price prediction based on sentiment analysis and LSTM.

Alex Eponnon MSc After Presentation

INBOX | NOW VIEWING: NEW PAPERS ▾

Submit File

Online Grading Report | Edit assignment settings | Email non-submitters

AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
Alex Eponnon	ABS 121222_v2	0%	--	--		1978977044	12-Dec-2022
Alex Eponnon	ALLTH 121222	8%	--	--		1978882170	12-Dec-2022
Alex Eponnon	CH1 12122022	2%	--	--		1978869480	12-Dec-2022
Alex Eponnon	CH2 121222	9%	--	--		1978869510	12-Dec-2022
Alex Eponnon	CH3 121222	13%	--	--		1978869537	12-Dec-2022
Alex Eponnon	CH4 121222	8%	--	--		1978869569	12-Dec-2022
Alex Eponnon	CH5 121222	0%	--	--		1978869554	12-Dec-2022
Alex Eponnon	CH6 121222	0%	--	--		1978869586	12-Dec-2022
Alex Eponnon	CNC 121222	0%	--	--		1978869616	12-Dec-2022

Assoc. Prof. Dr.
Kâmil Dimililer