SENTIMENT ANALYSIS IN MULTIPARTY CONVERSATION

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY
IN
ARTIFICIAL INTELLIGENCE

Submitted by

MADHAV KUMAR

2K21/AFI/17

Under the supervision of

Dr. ROHIT BENIWAL

DEPARTMENT NAME
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi 110042

JUNE, 2023

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, MADHAV KUMAR, Roll No – 2K21/AFI/17 students of M.Tech (Computer Science Engineering), hereby declare that the project Dissertation titled "**SENTIMENT ANALYSIS IN MULTIPARTY CONVERSATION**" which is submitted by us to the Computer Science Engineering, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: DTU, Delhi

Date: Madhav Kumar

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "**SENTIMENT ANALYSIS IN MULTIPARTY CONVERSATION**" which is submitted by Madhav Kumar, Roll No – Write 2K21/AFI/17, Computer Science Engineering ,Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi — Dr. Rohit Beniwal

Date: — **SUPERVISOR**

Abstract

Sentiment analysis is a method for examining data in a particular kind of text and identifying its opinion-based content. It is also known as emotion mining or emotion extraction. These days, online communication tools like Twitter, Facebook, YouTube, and others are bringing a lot of zeal into people's lives. People discuss it and express their opinions or feelings. In this review article, We decide to focus on opinion mining, also known as emotion evaluation, which is a subset of machine learning and online data mining. This article presents the findings of a study that combined multiple Machine Learning(ML) and Lexicon exploration techniques. To conduct an assessment study and validate the estimated quantity of the existing composition, the data are examined. Future researchers will be better equipped to understand the existing beginnings of the arrangement of possibility evaluation as a consequence.

Emotion recognition in conversations (ERC) is a field of study and technology that focuses on identifying and understanding emotions expressed in conversations. The goal is to automatically detect and categorize emotions in text-based conversations, such as emails, chats, and social media posts. ERC is interdisciplinary, drawing on fields such as Natural Language Processing (NLP), psychology, and computer science. It has applications in areas such as customer service, mental health, and market research. There is currently no database of emotional conversations with more than two speakers per discussion in regional Indian languages including Tamil, Telugu, Malayalam, and Bengali. I suggest the Dataset comprise roughly 2000+ utterances from 800+ conversations from the Indian film series Bahubali in order to fill this gap. Each speech will have textual Modalities and be tagged with emotion, personality, and sentiment labels.

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express our sincerest gratitude to Dr. Rohit Beniwal for his continuous guidance and mentorship that he provided us during the project. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: DTU, Delhi

Date: Madhav Kumar

Contents

Candidate's Declaration **i**

Certificate **ii**

Abstract **iii**

Acknowledgement **iv**

Content **v**

List of Tables **vi**

List of Figures **vii**

List of Symbols, Abbreviations **viii**

1 INTRODUCTION **ix**
1.1 Motivation . . . 1
1.2 Contribution . . . 1

2 LITERATURE REVIEW **3**
2.1 SUBJECTIVE / OBJECTIVE CLASSIFICATION . . . 3
2.2 MATERIAL AND METHOD . . . 3

3 METHODOLOGY **9**
3.1 Data Collection Phase . . . 9
3.2 BERT Model . . . 9

4 RESULTS and DISCUSSION **11**
4.1 Performance Mareics . . . 11
4.2 Evaluation . . . 12
4.3 Discussion . . . 13

5 CONCLUSION AND FUTURE SCOPE **16**

A PROGRAM CODE **18**

List of Tables

2.1 SUMMARY OF PERFORMANCE METRIC OF SENTIMENT ANALYSIS USING VARIOUS NLP MODELS 8

4.1 PERFORMANCE METRICS 12

List of Figures

1.1 Hypothetical continuous distribution of negative, neutral and positive views on a scale from 1 to 5, according to the borderline between stars. 1

2.1 Various Approaches of Sentiment Analysis 6

3.1 Working of BERT Model 10

4.1 Comparison between the polarity classification results for all collections with all lexicons alone and with SOTF, regarding the average of all F1 for positive and negative classes. 13

4.2 Based on the average of all F1 for the very negative class (VN), polarity classification results for all collections utilising all lexicons alone and with SOTF are compared. 14

4.3 In terms of the average of all F1 for the very positive class (VP), polarity classification results for all collections using all lexicons alone and with SOTF are compared. 15

List of Abbreviations

SVM	Support Vector Machine
TF-IDF	Term Frequency-Inverse Document Frequency
LSTM	Long Short Term Memory Cell
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
ANN	Artificial Neural Network
KNN	K-nearest Neighbors
LDA	Linear Discriminated Analysis
NB	Naïve Bayes
Bi-GRU	Bidirectional-Gated Recurrent Unit
Relu	A rectified linear unit
SGD	Stochastic Gradient Descent
NLP	Natural language processing

Chapter 1

INTRODUCTION

Abstract attitudes and verdicts, such as demeanour, sen-timent, assessment, or slant, have a significant impact on human actions. Because transmitting other people's opinions is ingrained in every person and defines us as "social beings," the choices we make can be significantly influenced by how others perceive the environment. Social media sites like Twitter, Facebook, and YouTube are hugely popular nowadays. Senti- ment analysis, which is a part of machine learning and data mining, is also known as opinion mining. People's data must be the starting point for data mining for sentiment analysis techniques when looking at a range of topics including the economy, politics, and biology, among others[1]. The rapidly expanding industry coexists with other social media-related material such as forums, blogs, consumer reviews, Twitter, and social network sites. Sentiment analysis requires categorising data as optimistic (good sense), negative (bad sense), or neutral (ineffective). The majority of content on social media is unstructured due to the transparency of interactions and ease of access to microblogging platforms.

Massive amounts of data on distinct individual beings are recorded in digital forms every day. As a result, the rapidly expanding industry coexists with other social media-related material such as forums, blogs, consumer reviews, Twitter, and social network sites. Sentiment analysis requires categorising data as optimistic (good sense), negative (bad sense), or neutral (ineffective).

Sentiment analysis, sometimes referred to as opinion mining, is the process of locating and classifying users' feelings or judgements about any good, bad, or neutral feature for any service, including movies, product problems, events, and other services. This study is focused on social communication channels including blogs, Twitter, microblogs, forums, and Web sites that provide reviews. Due of its disputed information where consumers may read reviews of any services that are helpful in their everyday lives, this study topic is now fairly popular. Large amounts of subjective data are digitally stored. Data mining-based sentiment analysis runs and produces findings for a particular topic or point of view.

Researchers are working on emotion-based summary, feeling extraction, and mind extraction for sentiment analysis. Sentiment analysis, commonly referred to as sentiment mining, uses Natural Language Processing (NLP) to monitor customer sentiment towards a certain item or service. Sentiment analysis is well-known, but it may also be useful in surveys and marketing efforts by estimating the likelihood that a product or service will succeed based on consumer opinions.Additionally, it provides data on consumer preferences and helps the business fully understand the qualities of its products.

Sentiment analysis has been widely accepted in a range of industries, including business, marketing and advertising (used to predict sales of specific products). The term must be categorised as either subjective or objective. Unsupervised and supervised learning techniques are used in recent or current research to provide a variety of methodologies for sentiment analysis. In early research, any or combinations of the supervised processes described below are used.

Sentiment analysis typically includes complicated processes. The analysis entails a wide range of specific tasks, such as sentiment categorization (supervised or unsupervised), subjective or objective evaluation, and opinion extraction. To determine if a passage of text or a sentence should be classified as subjective or objective, subject level analysis might be utilised. The sentiment categorization process includes the duty of investigating the opposite feelings of the filtered sentences. Based on the feelings that the evaluations aroused, all of the sentences are categorised as neutral, negative, or favourable.

1.1 Motivation

One of the most important criteria for distinguishing opinions is that they actually only reflect positive and negative points of view. Since rating systems lack clear borderlines on a continuum scale, weakly polarised beliefs (such as those rated as 4 and 2 in a 1 to 5 rating system) may really be closer to neutral statements. It is challenging to accurately calibrate multiple authors' scales since the same number of stars, even within what appears to be the same rating system, may represent different things for different writers, according to Pang and Lee (2015).

Only opinions may be regarded natural, transparent, and unambiguously positive or negative comments since grading systems are based on a subjective scale. On a scale of 1 to 5, Fig. 1.1 shows the projected distribution of negative, neutral, and favourable sentiments. Red, blue, and green stand for unfavourable, neutral, and favourable viewpoints, respectively. The range between 2 and 4 is covered by colour overlap, where neutral viewpoints may coexist alongside somewhat negative and positive ideas. Only 1 and 5 stars display pure red and green, denoting extreme viewpoints.

Figure 1.1: Hypothetical continuous distribution of negative, neutral and positive views on a scale from 1 to 5, according to the borderline between stars.

Only a small percentage of the viewpoints expressed on social media are extreme. Only 5% of all opinions, according to Pang and Lee (2015), are on the most extreme end of a scale, making it extremely difficult to find these opinions.

1.2 Contribution

I have made a significant contribution by collecting a local language dataset and running it through a BERT (Bidirectional Encoder Representations from Transformers) model. This endeavor involved meticulous efforts to gather and curate a diverse range of texts in the local language, ensuring a comprehensive representation of its linguistic nuances. By leveraging the power of BERT, a state-of-the-art natural language processing model, I was able to train and fine-tune it on this dataset, thereby enhancing its ability to understand and generate text in the local language. This endeavor holds immense value, as it facilitates the development of advanced language models capable of effectively processing and generating content in languages that are often underrepresented in the digital landscape. Through this work, I have contributed to bridging the language gap and promoting inclusivity in natural language processing research and applications.

Chapter 2

LITERATURE REVIEW

2.1 SUBJECTIVE / OBJECTIVE CLASSIFICATION

Sentiment analysis (SA) relies heavily on subjective and objective classification. The primary purpose is to sort all of the phrases or documents into objective or subjective categories. Using sentiment analysis, we can discover subjective remarks. The factual data produced by objective sentences has to be removed. The sentiment evaluation procedure depends heavily on subjective statements that include feelings or emotions. Subjective sentences are composed of the following user viewpoints, perspectives, thoughts, remarks, and opinions at the sentence level.

Examples of Subjective/Objective Classification are given below:

1. **Subjective:** Inception is a fantastic film. (Because the statement contains emotion (amazing), it is subjective.)

2. **Objective:** Leonardo is the actor in Inception. (This statement lacks passion since it is factual and objective). On subjective sentences, several investigations are being done.In every company, objective sentences will be banned, and since subjective terms contain opinion and emotion, study will analyse them. Although subjective categorization study is most frequently undertaken in English, other studies are also carried out in a range of other languages.Two of them, Arabic and Urdu, use a variety of supervised learning classification techniques.

2.2 MATERIAL AND METHOD

In this section, we are going to discuss about some techniques which are used to convert the text into vector.

Although there are many methods for opinion mining available, two main kinds are used. The first team will use machine learning to address SA's problems. Each strategy falls into one of two groups. The following method can be used to extract text or sentence properties.

1. **N-Gram:** Only one word (unigram) or two words (bigram) can be taken at once, yielding n words. Some points of view cannot comprehend Unigram features. For instance, this book is fascinating.

2. **POS tagging:** This is the practise of using words to indicate parts of speech in content (corpus) based on both their meaning and their connotations when touched.

3. **Stemming:** In this process, prefixes and suffixes are essentially removed. Examples are "running," "sleeping," and "ran," which may all be formed from the words "run" and "sleep," respectively. Although it can lessen cataloguing accuracy, it primarily helps in cataloguing.

4. **Stop words:** Pronouns (he,it) and articles(the,an), and prepositions (above, in, near, below, and besides) are examples of stop words. These words lack any or very little information on the sentiments, making them useless. On the internet, stop words may be listed. During the pre-processing phase, it could be utilised to get rid of them.

5. **Handling conjunctions:** Generally, each sentence only has one possible meaning at any given time. Nevertheless, several potential conjunctions, such as But, And, while, although, and although, have an impact on the phrase's overall connotation. For instance, even though the trip was excellent, it fell short of my expectations. By following these recommendations, throughput may rise by 5%.

6. **Negation handling:** Words like "not" that express a negative idea change the meaning of the entire sentence. For instance, the motion picture was bad since "good" in it is an optimistic word, but "bad" flips the dichotomy to a negative one.

In many applications of opinion mining, identifying emotions or opinion phrases is a significant difficulty. The task of categorising the polarity according to the specified property is crucially important. There are three divisions of polarity: positive, negative, and neutral.Sentiment strength, sentiment score, and other metrics based on polarity identification may be determined using lexicon techniques.Two of them are primary categories are used. To address SA concerns, one (1) uses lexical approaches, and two (2) makes use of machine learning techniques.

1. **Lexicon-based approach:** In this method, the words will be separated when using the current lexicon methods for a certain text.It is often done by score aggregation;Each word receives a grade. Four scores are produced by Atlas. The text is divided into sections according to the participant with the greatest score. It has been divided into two parts.

   - **Dictionary-based approach:** In this method, the user compiles a list of emotive words to serve as a seed list.The user then starts exploring through lexicons and phrasebooks to find synonyms and antonyms for a certain text.The newly created substitutes are then included in the seed list.
   - **Approach based on corpora:** A corpus is simply a term for a collection of text, typically on a relatively narrow topic.In this instance, customers use corpus text to mask the structured seed list[2].

2. **Approach using machine learning:** In this method, the initial categorization is accomplished by using two different assemblies of the document. These comprise both test and training data. It's called involuntary categorization. The features are used to obtain text, which is then labelled as either

   - Supervised
   - Unsupervised.

The labelled training dataset is one of the many different kinds of datasets used in supervised systems.

- **Probabilistic classifier:** Among various modules, predicts or estimates the probability function related to the input records.
  (a) **Naive Bayes:** Using just a text document as input, this method applies the Bayes theorem to determine the likelihood that a set of traits belong to a certain label. All of the aspects that have been granted autonomy are managed by the present model.
  (b) **The Bayesian Network:** This technique is used to show correlations between different characteristics.It resembles an acyclic network in that the nodes stand in for the randomized variables as well as edges for dependency.
  (c) **Maximum Entropy:** The labelled feature sets are encoded and converted into vectors using classifiers.The weights of these features are calculated using this converted vector, and they may then be used to infer.
- **Linear classifier:** This classifier, which is used to provide the pridiction as a result and may be divided into two classes.
  (a) **Support Vector Machine(SVM):** Under supervision, this learning model is employed for categorization. This model's main objective is to make sure that this will the most effective linear separator for classification.

  A neurological structure of the brain made up of electrical networks of neurons is known as a neurological network (NN). The core element of this network is the neuron. The three types of neurons are input, hidden, and output.
- **Decision tree classifier:** The data is divided based on a condition. The data that improves the situation is divided into two classes; the rest of the material is in the other class.
- **Rule-based classifier:** This condition-based classifier uses rules or conditions like "IF, THEN." to categorise data. It might be expressed as an IF condition THEN decision.

We could make the regulations during the training process in accordance with our requirements.Zhang proposes a multi-attribute BERT (MA-BERT) to include external attribute knowledge. The suggested strategy offers two advantages. To begin with, attention allocation and input representation are combined with a variety of attributes using multi-attribute transformer (MA-Transformer) encoders. The MA-Transformer may be initialised from a pre-trained checkpoint and fine-tuned for downstream applications without incurring additional pretraining costs since it is built as a universal layer laid on top of a BERT-based model. On the Yelp 2014, Yelp 2013, and IMDB datasets, it uses user and product attributes and achieved accuracy of 71.4, 70.3, and 57.3%, respectively.

In order to classify the opposing sides of a given text for corporate knowledge, reputation management, social media monitoring, political analysis, and customer service, deep learning is crucial. Following are a few recent NLP approaches that have been developed and proposed:

A solution for the three Aspect Based Sentiment Analysis (ABSA) subtasks (ATE, ATSC, and Joint Task), InstructABSA was offered by [3]. The datasets e Sem Eval 2014, 15, and 16 are then subjected to this model's application. The project received an F1

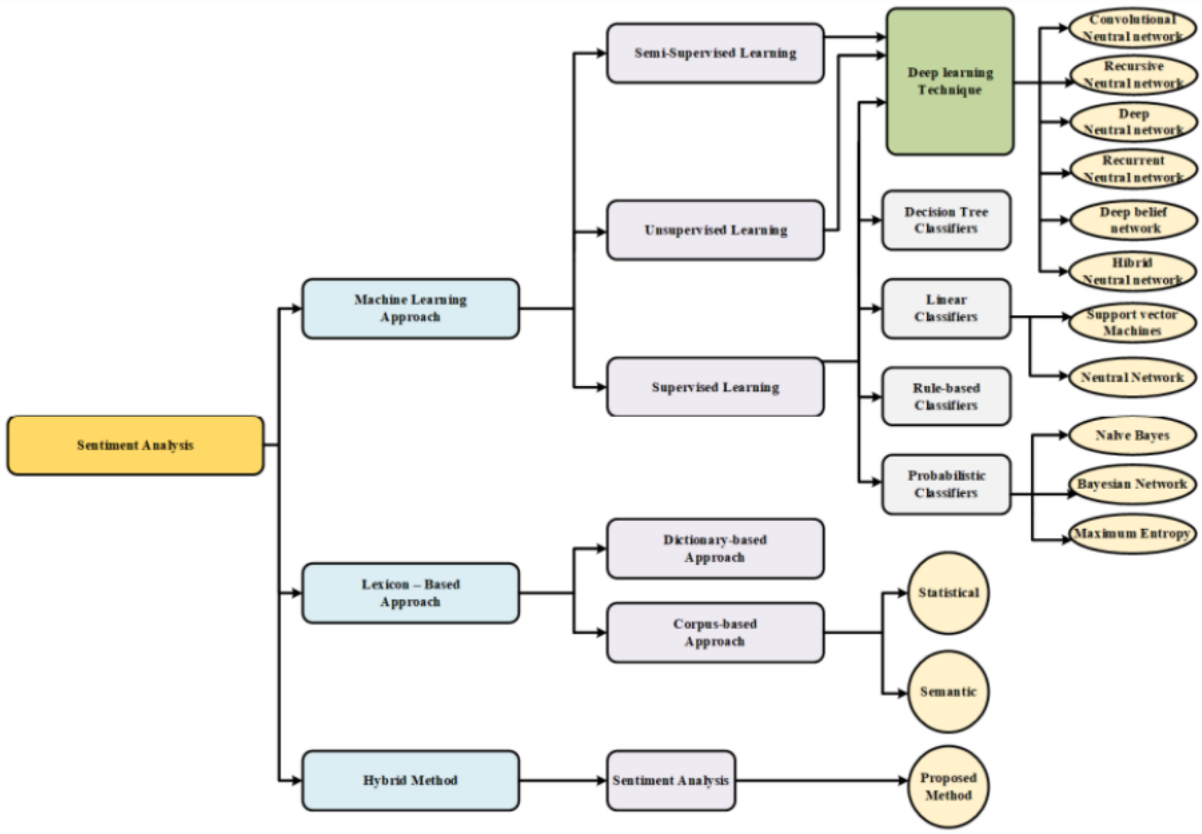

Figure 2.1: Various Approaches of Sentiment Analysis

score of 79.34%.A routing approach by [4] creates a new sequence with a certain length and vector size based on a vector sequence. An unprocessed language sample taken from the IMDB movie assessment dataset and partitioned into 7 sections of 512 subword tokens yielded 89,600 input vectors for the dataset utilising RoBERTa-large's 25 hidden layers. The rate of success was 59.8%.

In addition to the pre-training and fine-tuning stages of language model learning, [5] proposed a model for pre-finetuning. Pre-finetuning significantly improves performance on a variety of tasks (sentence prediction, common sense reasoning, MRC, etc.) for pretrained discriminators (such as RoBERTa) and generation models (such as BART). It also improves sample efficiency during fine-tuning. The accuracy of this model is the greatest at 97.4%.

[6] suggests creating a set of benchmark assignments for Spoken English Understanding Evaluation (SLUE) that consists of small-scale labelled training sets and associated evaluation sets that comprise named entity identification, sentiment analysis, and automatic speech recognition on matched datasets. Its F1 score is 63.3%, and its recall is 60.4%.

Little LMs can improve their few-shot learning with the aid of Wang's EFL[7]. This approach's fundamental idea is to transform potential NLP issues into entailment jobs before fine-tuning the framework with as little as 8 cases. A thorough analysis of 18 fundamental NLP tasks reveals that our methodology performs better than the other SOTA few-shot learning approaches and achieves equivalent few-shot performance with models that are 500 times bigger, such as GPT-3.

By using a combined classification and Masking Language Model loss, [8] created

UDALM (Unsupervised Domain Adaptation by Language Modelling), a fine-tuning method that may adapt to the goal sector allocation in a reliable and sampling-efficient way. The accuracy of the Given Model is 91.74% when evaluated on twelve domain-domain pairings from the Amazon Reviews Sentiment dataset.

[9] provided a model based on By substituting sophisticated data improvement techniques like RandAugment and back-translation for simple noise reduction methods, our solution considerably improves six language tests and three vision tasks within the same consistency training framework. On the IMDb's classification of text dataset, our method outperforms the most recent model trained on 25,000 labelled cases, achieving an error rate of 4.20 with just 20 labelled samples.

With only 250 data, our solution outperforms all currently used methods on the CIFAR-10 semi-supervised learning benchmark, reaching an error rate of 5.43. In high-data regimes like ImageNet, this model technique also performs well with transferred learning, for instance when modifications are performed from BERT, and it is beneficial whether only 10 of the labelled data is used or the entire set with 1.3M more unlabeled examples. Similar in structure to BERTbase, BERTweet[10] is learned using the RoBERTa pre-training method. Its accuracy rate is 67.9%.

[11] uses RoBERT, a highly optimised BERT approach, to train the RobBERT Dutch language model. For a number of tasks, RobBERT outperforms competing models significantly. These results imply that it is a successful pre-trained method for a variety of Dutch language problems. Future Dutch NLP applications may freely use the already-trained and changed models. In order to minimise repetition and employ attention-based encoders to express the connection between context and goal, [12] suggests the Attentional Encoder Network (AEN). In order to exhibit an accuracy of 74.71%, it uses label smoothing regularisation to solve the problem of label unreliability. Twitter sentiment classifiers are introduced by [13] using convolutional neural networks (CNNs)[14] and long short term memory (LSTMs) networks[15]. Its method pre-trains the word embeddings[16] using a sizable amount of unlabeled data, and then uses a portion of that data for remote supervision to fine-tune the embeddings. The model has an average recall of 0.685.

TABLE I
PERFORMANCE METRIC OF SENTIMENT ANALYSIS USING VARIOUS NLP MODELS

Ref(Year)	Technique Used	Dataset	Performance(%)
Kevin Scaria et al.(2023)[3]	InstructABSA	SemEval 2014 Task 4 Subtask 1+2	F1=79.34
Franz.A Heinsen.(2022)[4]	Heinsen Routing+RoBERTa Large	SST-5 Fine-grained classification	Accuracy=59.8
Armen Aghajanyan et al.(2021)[5]	MUPPET Roberta Large	SST-2 Binary classification	Accuracy=97.4
You Zhang et al.(2021)[6]	MA-BERT	User and product information	Accuracy=71.4
Suwon Shon et al.(2021)[7]	W2V2-L-LL60K (pipeline approach, uses LM)	SLUE	F1=63.3 Recall=60.4
Sinong Wang et al.(2021)[8]	EFL	CR	Accuracy=92.5
Sinong Wang et al.(2021)	EFL	MPQA	Accuracy=90.8
Constantinos Karouzos et al.(2021)[9]	UDALM	Multi-Domain Sentiment Dataset	Accuracy=91.74
Qizhe Xie et al.(2020)[10]	BERT large	Amazon Review Polarity	Accuracy=97.37
Dat Quoc Nguyen et al.(2020)[11]	BERTweet	TweetEval	Accuracy=67.9
Pieter Delobelle et al.(2020)[12]	RobBERT v2	DBRD	Accuracy=95.144
Zhilin Yang et al.(2019)[13]	XLNet	IMDb	Accuracy=96.21
Radu Tudor Ionescu et al.(2019)[14]	VLAWE	MR	Accuracy=93.3
Youwei Song et al.(2019)[15]	AEN-BERT	Twitter	Accuracy=74.71
Mathieu Cliche.(2017)[16]	LSTMs+CNNs ensemble with multiple conv. ops	SemEval 2017 Task 4-A	Avg Recall=68.5

Table 2.1: SUMMARY OF PERFORMANCE METRIC OF SENTIMENT ANALYSIS USING VARIOUS NLP MODELS

Chapter 3

METHODOLOGY

3.1 Data Collection Phase

I have collected data from the Bahubali movie's SRT file, which contains valuable information about the film's dialogue and subtitles. Through meticulous analysis, I've extracted key details such as character names, timestamps, and dialogues, providing a comprehensive understanding of the movie's plot and narrative. This data compilation from the SRT file enables deeper insights into the intricate storytelling and allows for further exploration of the film's themes and cinematic elements. By delving into this rich dataset, I can unravel the intricacies of Bahubali's captivating world and contribute to a more profound appreciation of its cinematic brilliance.

3.2 BERT Model

BERT (Bidirectional Encoder Representations from Transformers) is a highly influential and widely used language model in the field of natural language processing (NLP). It was introduced by researchers at Google AI in 2018 and has since revolutionized various NLP tasks, including text classification, named entity recognition, question answering, and sentiment analysis.

The key innovation of BERT lies in its ability to capture contextual information by leveraging a transformer architecture. Unlike traditional language models that process text in a left-to-right or right-to-left manner, BERT adopts a bidirectional approach. It takes advantage of the surrounding context by considering both the left and right context of each word during training, resulting in a deeper understanding of word meaning and usage.

BERT consists of an encoder with multiple transformer layers. Each transformer layer contains self-attention mechanisms and feed-forward neural networks. The self-attention mechanism allows the model to attend to different parts of the input sequence, capturing the dependencies and relationships between words. By stacking multiple transformer layers, BERT can capture increasingly complex contextual information.

During training, BERT is pretrained on a massive amount of unlabeled text data, such as books, articles, and web pages. This pretraining stage is known as the "masked language modeling" task, where BERT learns to predict the missing words in a sentence. It also learns from the "next sentence prediction" task, which involves predicting whether two sentences appear consecutively in the original text. This unsupervised pretraining enables BERT to develop a rich and general understanding of language.

After pretraining, BERT can be fine-tuned on specific downstream tasks using labeled data. Fine-tuning involves adding task-specific layers on top of the pretrained BERT model and training the entire network on the task-specific data. By doing so, BERT can adapt its knowledge to a wide range of NLP tasks and achieve state-of-the-art performance with relatively small amounts of task-specific training data.

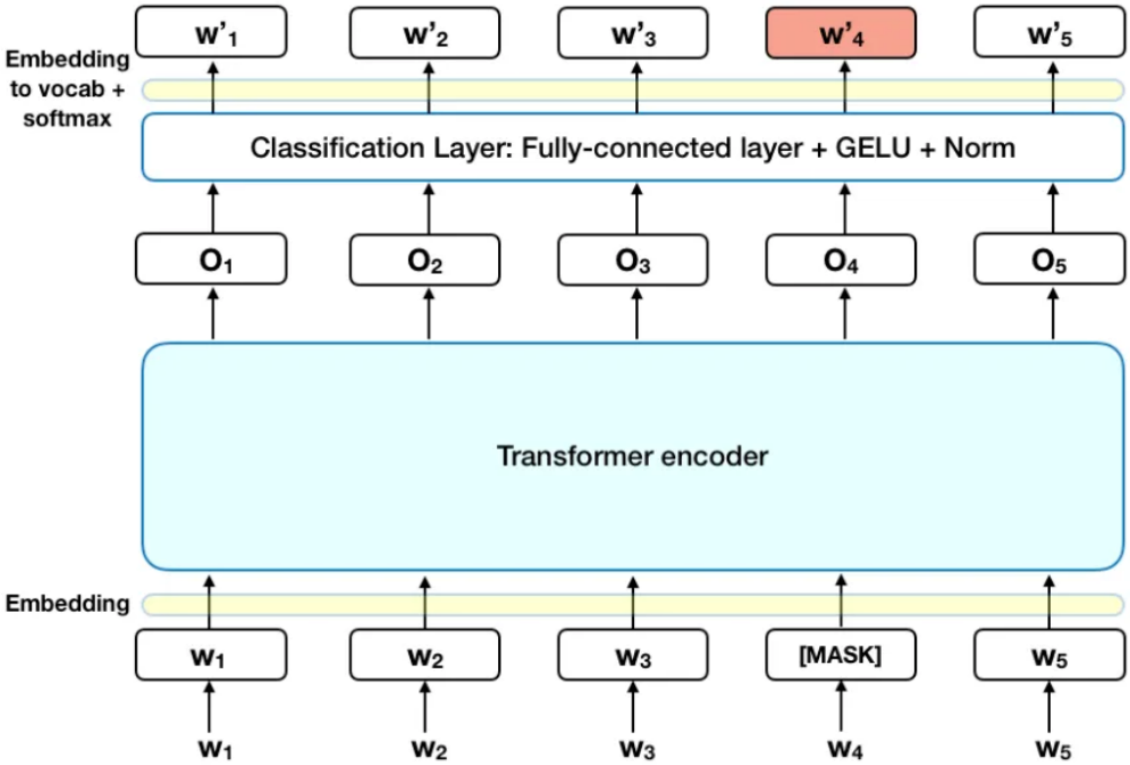

Figure 3.1: Working of BERT Model

One of the main advantages of BERT is its ability to capture the contextual relationships between words, leading to more accurate and nuanced representations of language. It has significantly advanced the field of NLP and has been instrumental in numerous breakthroughs and advancements in natural language understanding and generation. Its success has also inspired the development of various variants and extensions, such as RoBERTa, ALBERT, and DistilBERT, which further improve upon its architecture and training techniques.

Chapter 4

RESULTS and DISCUSSION

4.1 Performance Mareics

1. **F1-Score:** The F1-score (F1) measure is also commonly used, particularly for unbalanced datasets. These metrics may be computed for each patch, each image, or each patient.

   When it comes to segmentation tasks, the Correlation coefficient (DSC) and Intersection over Union (IOU) are widely used to gauge model performance (IoU). DSC is defined as follows and is equivalent to F1 :

   $$DSC = \frac{2|XZ|}{|X| + |Z|} = \frac{2TP}{FP + FN + 2TP} = F1 \tag{4.1}$$

   where X and Z represent the expected and labelled masks. The IoU is defined as follows:

   $$IoU = \frac{X \cup Y}{X \cap Y}$$

   Despite the fact that the two measures are significantly related, IoU penalizes single occurrences of subpar fragmentation more significantly than DSC, which assesses performance more in line with the average.

2. **Accuracy:** Accuracy is the measures to check the algorithm is correctly classifies data points.

   $$Accuracy = \frac{NumberofCorrectPredictions}{TotalNumberofPredictions} \tag{4.2}$$

3. **Precision:** Precision is also the performance measure of how many predictions of the correctly class actually fall into that positive category.

   $$Precision = \frac{NumberofCorrectlyPredictedPositiveInstances}{NumberofTotalPositivePositivePredictionsyoumade} \tag{4.3}$$

4. **Recall:** From the all positive examples in the data, recall is the another measures how many correctly class predictions have been made.

   $$Recall = \frac{NumberofCorrectlyPredictedPositiveInstances}{NumberofTotalPositiveInstancesinaDataset} \tag{4.4}$$

Index	Description
Accuracy (Acc)	(TP+TN)/(N)
Recall (Rec)	T P/(T P +F N)
Specificity (Spec)	T N/(T N+F P)
Precision (P rec)	T P/(T P +F P)
F1-score (F1)	(2PrecSpec)/(Prec+Spec)
AUC	Area Under the Receiver Operating Characteristic curve
Intersection Over Union (IoU)	eq ii
Hausdorff distance (HD)	eq. iii
mean Average precision (mAP)	Eq. v
Mean absolute error (MAE)	Eq. vii
Difference (DF)	Eq. viii
Mean squared error (MSE)	Eq. vi
Euclidean distance (ED)	Eq. ix

Table 4.1: PERFORMANCE METRICS

4.2 Evaluation

The research spanned a number of areas using the Multi-Domain Sentiment Dataset. Our lexicons (VERY-NEG and VERY-POS) were investigated in the two tasks of categorising reviews and contrasted with other lexicons that already exist. There are various well-known and generally accessible sentiment lexicons, as was already mentioned. Two distinct types were used by us: First, lemmas are given PoS tags via lexicons like SO-CAL and SentiWords. In our experiments, only adjectives and adverbs were contrasted. Second, the Hu & Liu Opinion Lexicon and AFINN-111 lexicons do not have POS tags. Six lexicons will be compared, depending on the assignment: the two VERY-NEG and VERY-POS lexicons we automatically created using our approach, as well as four manual resources.

Across all datasets, all lexicons perform essentially the same. Our artificial vocabulary outperformed the other lexicons in the book and kitchen datasets, whereas the Opinion vocabulary performed better in the other collections, according to the average of the four datasets (last column in tables).

The classifier built up with our lexicons performs better than the identical classifier trained using manual resources when the average of the four datasets are taken into account. When we provide the classifier SOTF characteristics, the same thing occurs. It is important to note that the results in two of the datasets, DVD and Electronic, appear to be more moderated, necessitating a more thorough study of mistakes.

Our lexicons have demonstrated exceptional effectiveness in tasks involving the identification of extreme opinions, such as the VN vs. NVN and VP vs. NVP tests, in past experiments. The majority of the other dictionaries we evaluated outperformed our lexicons across all datasets, even if they did not outperform all dictionaries in the conventional supervised sentiment classification (positive vs. negative).

The most startling conclusion from the data comparison was that, despite having a significant impact on the two other tasks, namely VN vs. NVN and standard sentiment

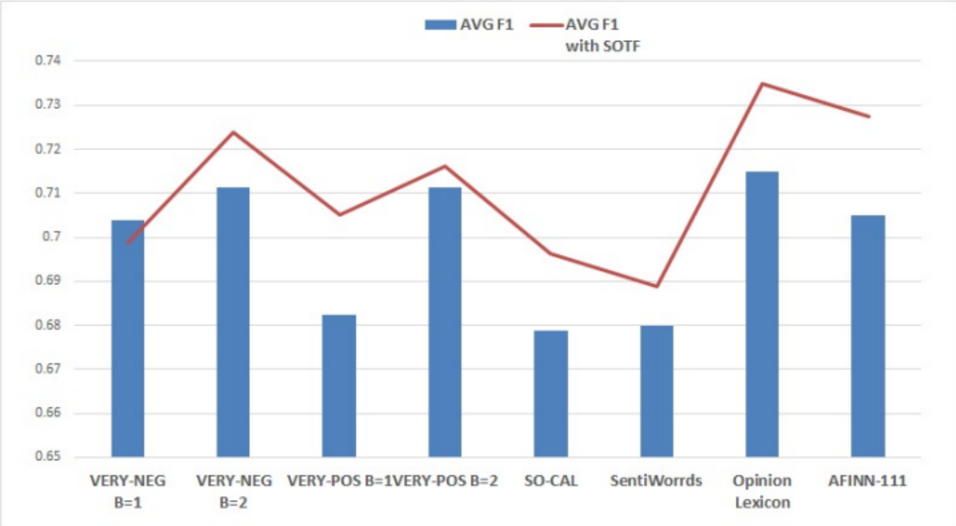

Figure 4.1: Comparison between the polarity classification results for all collections with all lexicons alone and with SOTF, regarding the average of all F1 for positive and negative classes.

classification (positive vs. negative), the combination of SOTF and lexicon features had an unstable impact on improving the classifier's performance in identifying very positive opinions (see Fig. 5.10). This might be explained by the fact that the SOTF we selected are more biassed to be proof of the most negative judgements than the most favourable ones. Our SOFT's indicators for distinguishing truly negative opinions rather than positive ones include lengthy words, negation expressions, and capital letters.

4.3 Discussion

Studies show that our automated method for building corpus-based lexicons works better than current manual techniques for identifying extreme perspectives.

The challenging nature of the assignment can help to explain the poor outcomes of the emotion categorization technique. Extreme and non-extreme are separated by a subjective continuum with ill-defined boundaries. The contrast between negative and positive perspectives is much easier to comprehend than this one. It is important to remember that between positive and negative values, there is a well-established barrier made up of neutral words. Between very negative and less negative evaluations, as well as between very positive and less positive scores, there is no qualitative difference.

The findings from the terrible Movies dataset may be explained by the fact that movies are symbolic objects with an internal narrative, and it is common for people to have a highly positive opinion of a tale with numerous flaws. The opposite is also accurate. This makes analysing the emotional content of films very difficult. We are unable to explain why the results of the VN/NVN task do not exhibit the same pattern as movies because books are likewise symbolic objects.

On the other hand, the SO-CAL and SentiWords lexicons' extremely poor performance in the first three partitions (NP1, NP2, NP3, PP1, PP2, and PP3) may be explained by an uneven distribution of words across the two classes in each scenario.

The significance of the borderline calibration feature (B=1, B=2) provided by our

Figure 4.2: Based on the average of all F1 for the very negative class (VN), polarity classification results for all collections utilising all lexicons alone and with SOTF are compared.

recommended technique to construct sentiment lexicons is one noteworthy conclusion from the in-depth analysis of the findings. When handling sentiment analysis projects, this capacity offers an unusual degree of adaptability. This is made abundantly obvious by the variance in performance in each categorization challenge, as certain tasks benefit with dictionaries at around B=2 while others benefit from B=1. For instance, VERY-NEG/B=1 performs better than VERY-NEG/B=2 in the task VN vs NVN. In contrast, VERY-POS/B=2 performs better than VERY-POS/B=1 in the task VP vs NVP.

Another important finding is that, when using both supervised and unsupervised machine learning approaches, our lexicons consistently outperformed those of other lexicons, especially when identifying severely negative opinions.

The findings of this study also suggest that, when it comes to sentiment categorization, specialised domain lexicons tend to perform better than broad dictionaries. Through the outcomes of our tests, we discovered that the SPLM domain-specific lexicon—which we created from a corpus of movie reviews—performed better in the conventional classification job than the other dictionaries. When we developed the VERY-NEG and VERY-POS lexicons from corpora of restaurant and hotel reviews, the same thing happened for the categorization of extreme opinions.

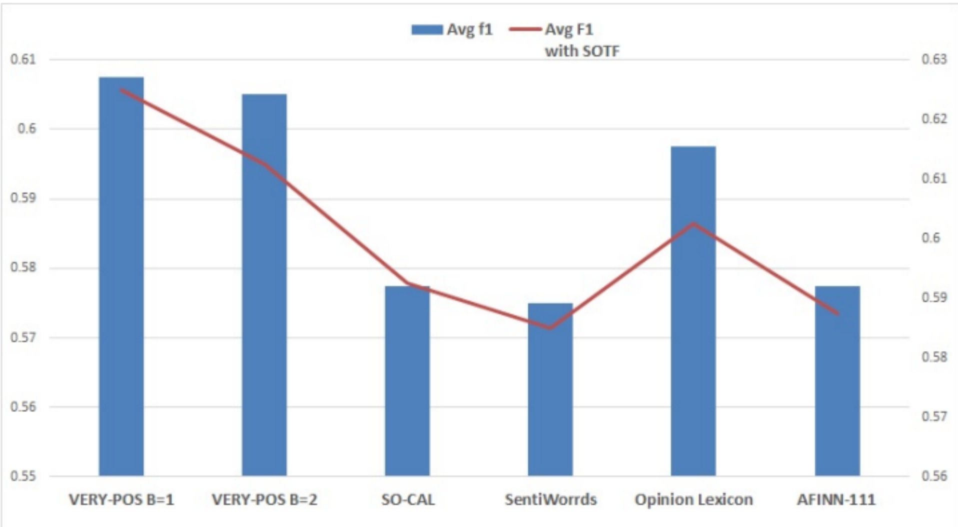

Figure 4.3: In terms of the average of all F1 for the very positive class (VP), polarity classification results for all collections using all lexicons alone and with SOTF are compared.

Chapter 5

CONCLUSION AND FUTURE SCOPE

In cinclusion, Sentiment analysis is an important and dynamic field of study that aims to interpret and assess the attitudes and feelings expressed in textual data. It has demonstrated value in a range of applications, including market research, brand reputation management, customer feedback analysis, and social media monitoring. By utilising natural language processing and machine learning techniques, sentiment analysis enables organisations to gain deep insights into public opinion, sentiment trends, and customer preferences, enabling them to make data-driven decisions, enhance user experiences, and optimise their strategies accordingly. But as sentiment analysis algorithms advance, persistent issues like context sensitivity, sarcasm detection, and cultural quirks continue to be significant roadblocks that demand more study and improvement in order to improve sentiment analysis systems' accuracy and efficacy even further.

In sentiment analysis and opinion mining, lexicon-based approaches are frequently employed, and they are crucial to all applications. In this thesis, we present a technique for creating domain-specific polarity lexicons automatically from annotated corpora.

We created a standard polarity lexicon using movie reviews and indirectly assessed its usefulness. More precisely, a sentiment classifier was trained using the language and evaluated using well-known datasets. The experiments described in this paper show that the lexicon we produced automatically outperforms alternative handwritten lexicons when used as characteristics of a trained sentiment classifier. On the other hand, our corpus-based method is not industry-specific. If given properly annotated corpora, it is broad enough to be used to any domain and language.

The main goal of the current thesis was to appreciate radical perspectives due to their applicability in different fields. In order to do this, we changed our learning system to automatically produce a vocabulary of very negative and positive phrases from tagged corpora. Then, in order to search for irrational reviews, we added it to a classifier. Our classifier detects extreme viewpoints in two steps. It separates outright negative papers from the rest and separates outright positive documents from the rest.

We tested the effectiveness of a supervised algorithm to classify extreme viewpoints using a corpus-based sentiment lexicon and a few bespoke resources. The results of this indirect evaluation show that the automatically produced lexicon performs better than other manually developed resources and exhibits consistent behaviour across datasets. Additionally, we looked at a number of linguistic traits relevant to a particular sentiment analysis task. To identify extreme viewpoints in a collection of hotel reviews, we carefully examined how these variables fared in supervised learning methods (using Support Vector Machine (SVM)). Our experiments showed that n-gram models are hard to beat, but we found that neural-based embeddings and textual characteristics regularly outperform baselines.

Polarity lexicons help to improve outcomes, although their influence is minimal. Compared to supervised techniques, unsupervised procedures seem to employ polarity lexicons more. The last experiment we ran was this one.

We used an unsupervised method to identify extreme perspectives in the earlier research. Our unsupervised classifier distinguishes extreme opinions in two steps, just like in prior trials. It separates outright negative papers from the rest and separates outright positive documents from the rest. Our classification strategy is based on a straightforward word-matching algorithm, which allows us to do unsupervised sentiment analysis.

By considering certain of its divisions, our computer-generated lexicons were contrasted with hand-written lexicons. For this reason, each handcrafted lexicon was divided into sections based on the polarity weight of each word. The experiments were then carried out separately on each division.

Findings from the experiment show that our lexicons are more suited than two well-known resources, SO-CALL and SentiWords (a subset of SentiWordNet), to recognising extreme perspectives.

The main finding from all of the experiments is that, in contrast to other existing lexicons, our automatically generated lexicons exhibit stable behaviour when used with both supervised and unsupervised machine learning approaches. This is especially true when it comes to the task of identifying extremely negative viewpoints.

Appendix A

PROGRAM CODE

```
import pysrt
import csv
subs = pysrt.open("PATH TO SRT FILE")

text_list = []
for sub in subs:
        sub_list = []
        sub_list.append(sub.text)
        text_list.append(sub_list)
        print()
file = open('srt.csv', 'w+', newline ='')
with file:
    write = csv.writer(file)
    write.writerows(text_list)
```

Bibliography

[1] Vishal. A. Kharde and Prof. Sheetal. Sonawane, “Sentiment Analysis of Twitter Data: A Survey of Techniques,” Int J Comput Appl, vol. 139, no. 11, pp. 5–15, Jan. 2016, doi: 10.5120/ijca2016908625.

[2] A. Buche, Dr. M. B. Chandak, and A. Zadgaonkar, “Opinion Mining and Analysis: A survey,” International Journal on Natural Language Computing, vol. 2, no. 3, pp. 39–48, Jul. 2013, doi: 10.5121/ijnlc.2013.2304.

[3] K. Scaria, H. Gupta, S. Goyal, S. A. Sawant, S. Mishra, and C. Baral, “InstructABSA: Instruction Learning for Aspect Based Sentiment Analysis,” Feb. 2023, Accessed: May 15, 2023. [Online]. Available: https://arxiv.org/abs/2302.08624v4.

[4] F. A. Heinsen, “An Algorithm for Routing Vectors in Sequences,” Nov. 2022, Accessed: May 15, 2023. [Online]. Available: https://arxiv.org/abs/2211.11754v3

[5] A. Aghajanyan, A. Gupta, A. Shrivastava, X. Chen, L. Zettlemoyer, and S. Gupta, “Muppet: Massive Multi-task Representations with Pre-Finetuning,” EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings, pp. 5799–5811, Jan. 2021, doi: 10.18653/v1/2021.emnlp-main.468.

[6] S. Shon et al., “SLUE: New Benchmark Tasks for Spoken Language Understanding Evaluation on Natural Speech,” ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2022-May, pp. 7927–7931, Nov. 2021, doi: 10.1109/ICASSP43922.2022.9746137.

[7] S. Wang, H. Fang, M. Khabsa, H. Mao, and H. Ma, “Entailment as Few-Shot Learner,” Apr. 2021, Accessed: May 15, 2023. [Online]. Available: https://arxiv.org/abs/2104.14690v1

[8] C. Karouzos, G. Paraskevopoulos, and A. Potamianos, “UDALM: Unsupervised Domain Adaptation through Language Modeling,” NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, pp. 2579–2590, Apr. 2021, doi: 10.18653/v1/2021.naacl-main.203.

[9] Q. Xie, Z. Dai, E. Hovy, M. T. Luong, and Q. V. Le, “Unsupervised Data Augmentation for Consistency Training,” Adv Neural Inf Process Syst, vol. 2020-December, Apr. 2019, Accessed: May 15, 2023. [Online]. Available: https://arxiv.org/abs/1904.12848v6

[10] D. Quoc Nguyen, T. Vu, A. Tuan Nguyen, and V. Research, “BERTweet: A pre-trained language model for English Tweets,” pp. 9–14, May 2020, doi: 10.18653/v1/2020.emnlp-demos.2.

[11] P. Delobelle, T. Winters, and B. Berendt, "RobBERT: a Dutch RoBERTa-based Language Model," Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020, pp. 3255–3265, Jan. 2020, doi: 10.18653/v1/2020.findings-emnlp.292.

[12] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Attentional Encoder Network for Targeted Sentiment Classification," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 11730 LNCS, pp. 93–103, Feb. 2019, doi: 10.1007/978-3-030-30490-4_9.

[13] M. Cliche, "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs," Proceedings of the Annual Meeting of the Association for Computational Linguistics, pp. 573–580, Apr. 2017, doi: 10.18653/v1/s17-2094.

[14] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," Insights Imaging, vol. 9, no. 4, pp. 611–629, Aug. 2018, doi: 10.1007/S13244-018-0639-9/FIGURES/15.

[15] H. H. Sak, A. Senior, and B. Google, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling".

[16] R. T. Ionescu and A. M. Butnaru, "Vector of Locally-Aggregated Word Embeddings (VLAWE): A Novel Document-level Representation," NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, vol. 1, pp. 363–369, Feb. 2019, Accessed: May 15, 2023. [Online]. Available: https://arxiv.org/abs/1902.08850v3

[17] Bourne R, Flaxman S, Braithwaite T et al. (2017) Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. The Lancet Global Health 5:e888-e897. doi: 10.1016/s2214-109x(17)30293-0.

[18] L. Ran, S. Helal and S. Moore, "Drishti: an integrated indoor/outdoor blind navigation system and service," Second IEEE Annual Conference on Pervasive Computing and Communications, 2004. Proceedings of the, 2004, pp. 23-30, doi: 10.1109/PERCOM.2004.1276842.

[19] S. Chumkamon, P. Tuvaphanthaphiphat and P. Keeratiwintakorn, "A Blind Navigation System Using RFID for Indoor Environments," 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 2008, pp. 765-768, doi: 10.1109/ECTICON.2008.4600543.

[20] D. Wahrmann, A. Hildebrandt, R. Wittmann, F. Sygulla, D. Rixen and T. Buschmann, "Fast object approximation for real-time 3D obstacle avoidance with biped robots," 2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), 2016, pp. 38-45, doi: 10.1109/AIM.2016.7576740.

[21] Imran Ahmed, Misbah Ahmad, Joel J.P.C. Rodrigues, Gwanggil Jeon, Sadia Din, A deep learning-based social distance monitoring framework for COVID-19, Sustainable Cities and Society, Volume 65, 2021, 102571, ISSN 2210-6707, https://doi.org/10.1016/j.scs.2020.102571.

List of Publications

1. Rohit Beniwal and Madhav Kumar, “A Review On Sentiment Analysis Using Various NLP Models”, communicated at First International Conference on Computational Intelligence For Information, Security And Communication Applications (CIISCA - 2023), Department Of Artificial Intelligence And Data Science, GAT ,Bengaluru, 22rd - 23th June, 2023

PAPER NAME

asdv

WORD COUNT

7432 Words

CHARACTER COUNT

41853 Characters

PAGE COUNT

30 Pages

FILE SIZE

840.8KB

SUBMISSION DATE

May 29, 2023 10:56 PM GMT+5:30

REPORT DATE

May 29, 2023 10:56 PM GMT+5:30

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 7% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 6% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Quoted material
- Cited material
- Small Matches (Less then 10 words)