REGION AND TIME WISE TWITTER BASED PCR ANALYSIS: A COMPARATIVE RESEARCH ON PEOPLE'S SERIOUSNESS TOWARDS COVID-19

By
Ghulam Musa Raz
2019-NUST-MS-CS-19 320112

Supervisor
Dr Abdul Wahid
Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Computer Science (MSCS)

In

School of Electrical Engineering and Computer Science,
National University of Sciences and Technology (NUST),
Islamabad, Pakistan.

(June 2021)

THESIS APPROVAL CERTIFICATE

Certified that final copy of MS Thesis written by NS **Ghulam Musa Raza** Registration No. **00000320112**, of **School of Electrical Engineering and Computer Science** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have been also incorporated in the said thesis.

Name of Supervisor**:** Dr. Abdul Wahid

Signature**:** _____

Date: _____**09-Jun-2021**_____

Committee Member 1: Dr. Muhammad Zeeshan

Signature**:** _____

Date**:** _____**10-Jun-2021**_____

Committee Member 2: Dr. Azizur Rahim

Signature**:** _____

Date**:** _____**10-Jun-2021**_____

DEDICATION

This thesis is dedicated to

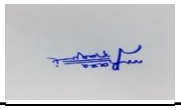
ALL THOSE who raised my morale and prayed for my success.

CERTIFICATE OF ORIGINALITY

I hereby declare that the thesis titled "***Region and Time Wise Twitter Based PCR Analysis: A Comparative Research on People's Seriousness Towards COVID-19***" my own work and to the best of my knowledge. It contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or any other education institute, except where due acknowledgment, is made in the thesis. Any contribution made to the research by others, with whom I have worked at SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic is acknowledged. I also verified the originality of contents through plagiarism software.

Author Name: Ghulam Musa Raza

Signature: 

ACKNOWLEDGEMENT

All praises be to ALLAH;Al-Muizz, Al-Kabeer, Al-Hadi and Al-Fattah

The successful completion of this thesis is accomplished by the devoted participation and cooperation of all guidance committee members. With gratitude and affection, I acknowledge active and guided guidance of my honorable supervisor Ass prof. Dr. Abdul Wahid. He supported me in hours of need and channelized my way in hard times. His motivation, guidance and supervision acted as the driving force that has enabled me to achieve my objective. I also admire and value participation of my respectable committee members; Ass prof DR. Mohammed Zeeshan and Ass prof Dr. Azizur Rahim for their time and advice. I am very grateful for the teachings of faculty members of Computer Science Department that has fueled my sense of continued determination over the years. I appreciate efforts of my all-family members, friends and class fellows who raised my morale and their motivation opened new ways for me. Their prayers and ALLAH's help had enabled me to be the best version of myself.

Contents

ABBREVIATIONS

DEFINATION	ABBREVIATIONS
Coronavirus Disease 2019	Covid19
Polymerase Chain Reaction	PCR
Support Vector Machine	SVM
Naive Bayes	BN
Single Layer Perceptron	SLP
Multilayer Perceptron	MLP
Logistic Regression	LR
bidirectional LSTM	BiLSTM
Conditional Random Fields	CRF
Enhanced Language Representation with Informative Entities	ERNIE
Bidirectional Encoder Representations from Transformers	BERT
Long Short-Term Memory	LSTM
Natural Language Processing	NLP
Deep Convolutional Neural Network	DCNN
Recurrent Convolutional Neural Network	RCNN
Determiner	DT
Random forest	RF

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

As we all are living in pandemic covid19, which is affecting every perspective of daily life. Every class is badly suffering from this disease. It has been a subject of discussion since 2019 due to the increased prevalence of social media and its extensive use and it has been a source of tension, fear, and disappointment for people all over the world. In this research, we have taken the covid19 tweets data of ten different regions including United States of America, Pakistan, India, China, South Africa, Philippines, United Kingdom, Switzerland, Philippines, and Ireland with timespan of 25th July of 2020 to 29th august of 2020. Using the common word embedding technique count-vectorizer, we experimented with different classifiers on data to train deep neural networks to improve the accuracy rate for predicting emotions. We assigned sentiments with highest accuracy rate. After mining the opinions of these regions about covid19, we have collected the PCR results of these regions. we have compared the percentage of opinion in form of positive or negative responses with percentage of per million PCR covid results of these regions. After this region and time wise twitter-based PCR analysis, we came to know that how much these regions are serious for pandemic. Also, we figured out a real time international measure to detect these region's seriousness for any future pandemic. This research can help the administrations of different regions for taking wise and suitable steps for controlling spread of any outgoing and upcoming pandemic.

Chapter 1 Introduction

In recent two years, COVID has embraced whole world into a pandemic. Every country is suffering from this virus and its effects. Both developed and developing countries have had the deadly effects of the disease. Everyday life is being severely affected by Corona. The corona virus is in its third wave, once again on the rise around the world, with SOPs being reintroduced in the United States, Europe and in Asian countries and lockdowns being imposed in many places disrupting normal life. Private and government educational institutions across many countries have been closed for certain period now. A series of smart lockdowns has been started in the affected regions and SOPs have been issued for social events.

While appreciating the government's initiative to control pandemic, it is important to know what the public opinion is about covid19 and whether SOPs issued by government are being taken seriously or not. A person's opinion can be used to gauge how serious a person is about particular thing. While living in this pandemic, people have become more active on social media. People use different platforms to express themselves and their mindsets. If their opinions are positive, then it means that they are serious about that thing. If opinion is negative, it means that they do not bring this thing to mind. In the same way, we can determine how much people care about Corona by taking their opinions about covid19.

In the write up below we present an idea of taking people's opinion of a region about covid19 and then comparing it with their PCR analysis. It would find out that what people thought about corona and what is the result of PCR test for corona virus. We have used the most popular social media platform twitter to get people's opinions.

Twitter has been known for its popularity in information flow of most talked about trends. So, covid related tweets will be used to get sentiments of people about covid. It is an established source for forecasting many big events, such as the opening of film box offices and general elections.

The tenacity of sentiment analysis is to evaluate the expressive orientation of user feedback. In this regard, Machine Learning and Artificial Intelligence has played incredible role in relevant areas and there has been a growing interest in the use of Deep Learning techniques to accomplish sentiment analysis and emotional intelligence.

Our focus is to determine people's seriousness towards covid19 with respect to their regions and time by analyzing the PCR results and covid related tweets. This study helps to provide an international measure to detect nation's seriousness about pandemic. Data of ten countries including United States of America, Pakistan, India, China, South Africa, Philippines, United Kingdom, Switzerland, Philippines, and Ireland is used for experimentation and for analyzing results.

1.1 Motivation

A lot of research has been done on different aspects of corona virus. People have worked in many features on the problems caused by Covid19. However, there are many gaps in this domain which can be covered. Knowing the seriousness of the people about corona is one of them, also most important one. If we find out that the people of this area are not taking corona seriously and the covid PCR test of the people of that area also has positive ratio then we can highlight those areas.

We can inform and help the administration of that area in keeping check on the people. Similarly, if people of an area are taking covid seriously and ratio of PCR results is lower, then we can appreciate people and relevant authorities of that area.

In this study, we compare the seriousness of countries regarding corona virus and following SOPs corresponding to the PCR results in that region for a specific time interval. We have predicted the effectiveness of the analysis of PCR results and people's seriousness in helping the governments to make better decisions to prevent of spread of pandemic in specific region. Covid19 is a pandemic that we all are going through. Region wise sentiment analysis provided insights to seriousness of people towards it. We can make predictions about different aspects which were affected by covid19, after proposed research like controlling misinformation, preventing negative propaganda, following SOPs. We can highlight hotspots of pandemic spread for coming waves.

1.2 Problem Statement and Objectives

In COVID situation when there's lockdown and less physical interaction between people, to deal with any of any epidemic in a positive way, it is very important to know that what is the awareness of the people about this epidemic and how many people have been affected by this epidemic. When we know these two things then we can formulate a best strategy to deal with that epidemic.

Taking COVID'19 into consideration we don't have a measure which can facilitate us to detect any particular region's real time seriousness for covid19 based on their covid tweets sentiments and PCR results.

So far, there's no research technique which can guide us to know the real time difference between region's positive or negative sentiment percentage and PCR positive or negative percentage of a particular region in particular time interval.

Objectives of the thesis include:

- To track users' behavior towards pandemic and provide a platform to take necessary steps.
- Make easier for government and authorities to figure out in what areas they need to take sensitive measure (resource allocation, implementing urgent lockdowns, region specific awareness campaigns etc.) for COVID or any future pandemic.
- Prediction of emotions on highest possible accuracy.

1.3 Thesis Contribution and Outline

Natural language processing (NLP) based tasks are performed for assigning sentiments to tweets. Data set of ten countries including United States of America, Pakistan, India, China, South Africa, Philippines, United Kingdom, Switzerland, Philippines, and Ireland taking the time of 25th July 2020 to 28th Aug 2020 for covid tweets is taken from opened source platform.

After assigning the sentiments, five Deep Learning classifiers. Those are State Vector machine, Bernoulli Naive Bayes, Single Layer Perceptron, Logistic Regression and Multi-Layer Perceptron have been implemented to test and train the data.

Classifier SVM with highest accuracy rate with 93.78% is submitted to model. After NLP and AI tasks, PCR for Covid19 test results are taken from an opened source platform. Region and time wise twitter-based PCR analysis is done on above mentioned countries and results are mentioned in results section.

Thesis report is divided into chapters listed below:-

- **Chapter 1: Introduction**
- **Chapter 2: Literature Review**
- **Chapter 3: Methodology**
- **Chapter 4: Results**
- **Chapter 5: Discussion**
- **Chapter 6: Conclusion**
- **Chapter 7: Recommendation**
- **Chapter 8: References**
- **Chapter 9: Index (Optional)**

Chapter 2 Literature Review

Nature of thesis is consisted upon two major factors which are assigning emotions to the text and preprocessing of data which would help us to better usage of data to get results.

Literature Review is described into two sections.

2.1 Assigning Emotions to Data

An immersive multitasking learning system [1] for text sentiment classification of Chinese text is suggested in this article. The Traditional BiLSTM, attention and CRF model is used here to take full advantage of the relationship of interaction between tasks and solve the two tasks of relational dictionary extension and description of emotions at the same time. The suggested approach distinguishes textual emotions classification and the extension of the emotional dictionary into main activity and subtask and adopts the ERNIE paradigm as the main activity's learning model of textual representation. Then, the text sentiment classification task is accomplished to the highest grouping stage and the totally related level. Meanwhile, to remove emotional terms from the text in the subtask, the classical BiLSTM, attention and CRF model is used. Furthermore, the interaction process of multitasking information is used and the predictive information about the autonomous subtask is fed back into the future representation of the two operations. The efficiency of the two tasks is further optimized after iterative training. As an entity to shape the experimental dataset, microblogs with COVID-19 are used here. The findings show the supremacy of the system suggested over other methods and further check the superiority of ERNIE for the classification of Chinese textual sentiment over BERT, RoBERTa and XLNet.

It is totally experiment based research [2]in which authors have implemented Short long-term memory (LSTM) model for sentiment analysis. This article is representing a complete solution related to hardware design for memristorbased LSTM. Both internal and externals structures are considered and implemented in making of design by memristor crossbar. Goal of minimum hardware design for sentiment analysis is also accomplished and it is plus point. Sigmoid and tanh functions are involved. Proposed model is experimented on IMBD and SemEval dataset. Total 65

memristive crossbar are required to achieve the goal of sentiment analysis. Results show that proposed model have achieved higher accuracy and greater performance as compared to CMOS and FGPA based methods.

This article [3] is focusing on enhancement of aspect-based sentiment analysis with the help of BERT. BERT is famous language model for NLP. Purpose of this article is to predict an opinion related to specific aspect. Ancillary sentence is created from aspect then sentence pair classification is obtained by converting ABSA with the help of BERT. Authors have fine-tuned the BERT model with the small changing in softmax function and addition of classification layer. They have divided their results in two experiments. Results have been compared with existing state of the art results. Results show that aspect classification is much batter then Dmu-entnet however accuracy is small then Dmuentnet and LSTM with contrast of 3.8 and 5.5 sequentially.

Purpose of this article [4] is to implement different artificial techniques on twitter data set and compare the obtained results. After comparing different results, they came to know that DCNN is most suitable for better accuracy for prediction. Authors have classified overall feature representations into three further features before going into depth of proposed model.

1. Baseline features – unigram and bigram models are used.
2. Word sentiment polarity score feature-PMI method is used.
3. Word representation featureGlove model is used for this for representation of this feature.
4.

Naïve-Bayes, Max Entropy, SVM Support vector machine then finally DCNN is used. Results show that for base line and for Glove model, RCNN provide more accuracy.

Most of the methods used for social media sentiment classification just focus on the sentiment polarity given by text however ignore other information. In this article, [5]authors have included human behavior which is found in text. To obtain this purpose, convolutional neural networks is used as base model. Weka library is playing key role to implement moel.40 different features are used to train the model. Different classifiers are also used to compare results obtained by props0ed model and other classifiers. Two different datasets are used to get results. Both of data set are obtained by twitter API. Results show that proposed model not only performs well on unbalance dataset also can be train with low data set.

Predicting ordinal regression [6]is always interesting in sentiment analysis. Different approaches have been introduced till now to predict ordinal regression. Proposed model is drawing out sentiment analysis of twitter with the help of constructing a balancing and scoring model. After that, Different classifiers are used to classification of ordinal regression. Proposed model consists of four main modules. MLG, SVM, DT and RF are used as classifiers to get and compare results. Data set is obtained by NLTK corpora collection. After getting results, it is clear that SVM and RF has same accuracy, which is more than MLG. 91.81% accuracy is achieved with the help of Decision trees. Also, mean absolute error and Mean squared error is also considered to measure the performance.

Real rime sentiment analysis [7] is now hot topic for research and technical perspective. Many people want it in daily life like businessmen want it in their business development. Politicians want it to analysis their followers etc. this article is representing a straight way of implementing sentiment analysis with the help opened source library KERAS. Whole representation of data processing is described in article. Data is collected by authorized API. Data is cleaned after getting it. A lot of garbage is removed for data like numbers, special characters; slang used in tweets etc. cleaned data is converted into vectors to fed DL models. Finally, neural network is traind and results have been produced. Word2V and KERAS are used to do achieve goal of run time sentiment analysis.

For twitter sentiment analysis, the [8]authors suggested a deep learning framework. The key purpose of this work was to initialize the weight of the parameters of the convolutionary neural network and to correctly train the model to prevent the need to incorporate new functionality. A traditional neural network is used to further optimize integration into a broad supervised corpus. Words and parameters previously incorporated for the configuration of the network were used, with the same design and the creation of a Semeval 2015 supervised corpus. Activations, grouping of sentence matrices, softmax and convolutional layers are the components used in the proposed work. The non-convex function optimization and stochastic gradient descent (SGD) algorithms were used to train the network and to calculate the gradient back propagation algorithm. To boost the regularization of neural networks, the technique of abandonment was used. In two tasks, the deep learning paradigm is applied: message level activity and Semeval2015 sentence level activity to forecast polarity and produce high performance. The suggested model ranks first with regard to consistency when conducting six sets of evaluations.

It is a review paper [9]based on different deep learning techniques used in sentiment analysis. Authors have put efforts to highlight taxonomy of sentiment analysis. According to them core sentiment analysis tasks are document-level sentiment classifications, sentence-level sentiment classifications, aspect-level sentiment classifications. And there are others minor sentiment analysis task which are called sub-task. The different DL models have been used for the sentiment analysis like convolutional neural networks, word embedding, RecNNs, RNN, LSTM, GRUs, DBNs. Authors have also compared these models based on applications, performance, execution time Drawbacks and benefits of using these models are also discussed in paper.

Authors	Models	Accuracy Rate.
[5]	CNN	89%
[2]	LSTM	84.3
[10]	Naive Bayes Support Vector Machine Multinomial Naive Bayes	75% 78% 86%
[6]	Decisions Trees	91.81%
[1]	BiLSTM + attention + CRF	85.49%

Table 2-1: Related work for assigning emotion.

2.2 Preprocessing of Data

The authors of [11] investigate the impact of preprocessing on twitter data to improve sentiment rankings. They focus on tweets that contain a lot of phrases, abbreviations, folklore, and unspecified words. URLs, hashtags, user references,

punctuation and non-significant words are removed and slang relevance and spell check are highlighted. They employ a Support Vector Machine (SVM) classifier.

[12] investigates the impact of preprocessing on film ratings. They use preprocessing methods including extension of abbreviations, removal of non-alphabetic signs, removal of non-significant words, handling of negation with the prefix "NO" and derivation. They often employ an SVM classifier, with the number of characteristics related to their accuracy. They show that the use of appropriate text preprocessing techniques, such as data transformation and filtering, will greatly improve the efficiency of the classifier. To produce new data, Zou [13] used some simple data alteration techniques (EDA), such as synonymous terms, random change, random insertion and random deletion. Since these methods are easy to apply and require no additional resources, they greatly increase efficiency.

Technique	LM	Ex Dat.
Trans. data aug.	yes	yes
Back-translation	yes	yes
Back-translation	yes	yes
Noising	yes	no
Back-translation	yes	no
LM + SR	yes	no
Contextual aug.	yes	no
SR - kNN	no	no
EDA	**no**	**no**

Table 2-2: Zuo's work for data preprocessing

According to [14] the cumulative percentage of noise in a data set will exceed 40%, which creates problems for machine learning algorithms. Spelling and typing errors, as well as the use of abbreviations and jargons, are common among twitter users.

They can also use exclamation marks and other punctuation marks to highlight their feelings. In most cases, it is not necessary to use all words in the initial form of a text in the machine learning stage, and some of them may be omitted, replaced, or combined with others. As a result, data needs to be cleaned and normalized, as data consistency is a critical element in the machine learning performance that follows preprocessing.

In two twitter datasets, [15]compares 15 preprocessing techniques widely used in an experiment. We use three different machine learning algorithms, Linear SVC, Bernoulli Nave Bayes, and Logistic Regression, and report the accuracy of the classification and the number of features produced for each preprocessing technique.

Number	**Pre-processing Technique**	**Number**	**Pre-processing Technique**
0	Basic (Remove Unicode strings and noise)	8	Replace negations with antonyms
1	Remove Numbers	9	Handling Negations
2	Replace Repetitions of Punctuation	10	Remove Stopwords
3	Handling Capitalized Words	11	Stemming
4	Lowercase	12	Lemmatizing
5	Replace Slang and Abbreviations	13	Other (Replace urls and user mentions)
6	Replace Elongated Words	14	Spelling Correction
7	Replace Contractions	15	Remove Punctuation

Table 2-3:Different data preprocessing techniques

Finally, they rank these strategies based on their effectiveness based on our results. They found that while some methods, such as derivation, number deletion, and long sentence replacement, increase accuracy, others, such as punctuation elimination, do not.

Performance	**Description**	**Techniques**
Best	High accuracy in all classifiers and all datasets	1,2,11
High	High accuracy in most classifiers and all datasets	9,12,13
Poor	Low accuracy in most classifiers and all datasets	3,5,8,10,14
Worst	Lowest accuracy in all classifiers and all datasets	15
Varying	High or poor accuracy in most classifiers depending on the dataset	4,6,7

Table 2-4: Performance wise ordering of techniques

Specially for twitter sentiment analysis [16]has proposed a model, which consists upon 3 phases. Architectural diagrams of proposed model is given below.

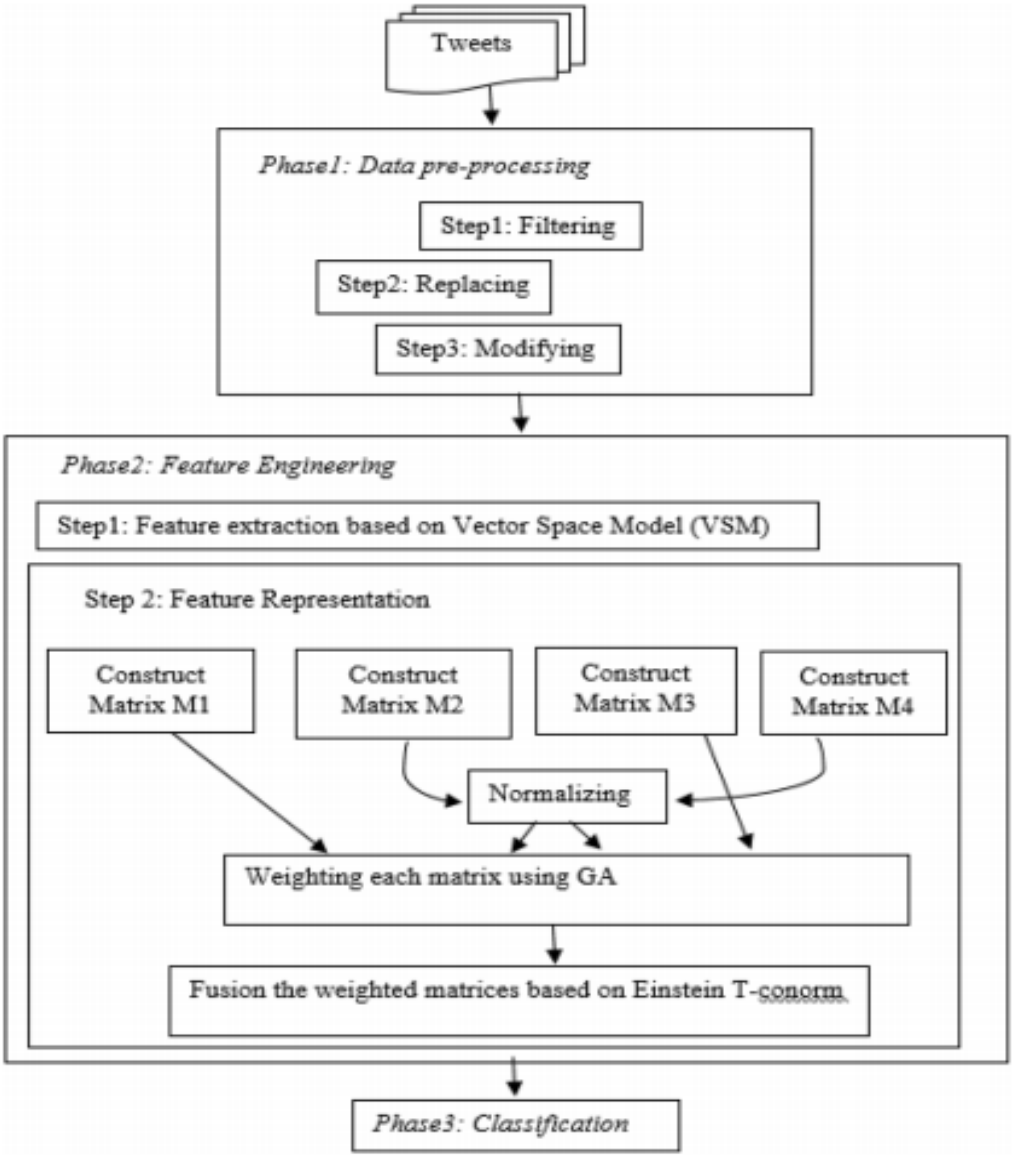


Figure 2-1: Data Preprocessing Model

The developers of the approach suggested in [17] took the tweets and converted them into a series of terms, which they then interpreted using a sentiment evaluation module based on the SWN lexicon. Then, for rule inference, each of the represented tweets is evaluated using algorithms based on approximate set theory (RST).

The relational increase was suggested by Kobayashi [18]. They stochastically replace terms with other words predicted by a bidirectional language model in word

positions, and the language models have been developed with a conditional tag design that allows the model to integrate sentences without compromising tag compatibility.

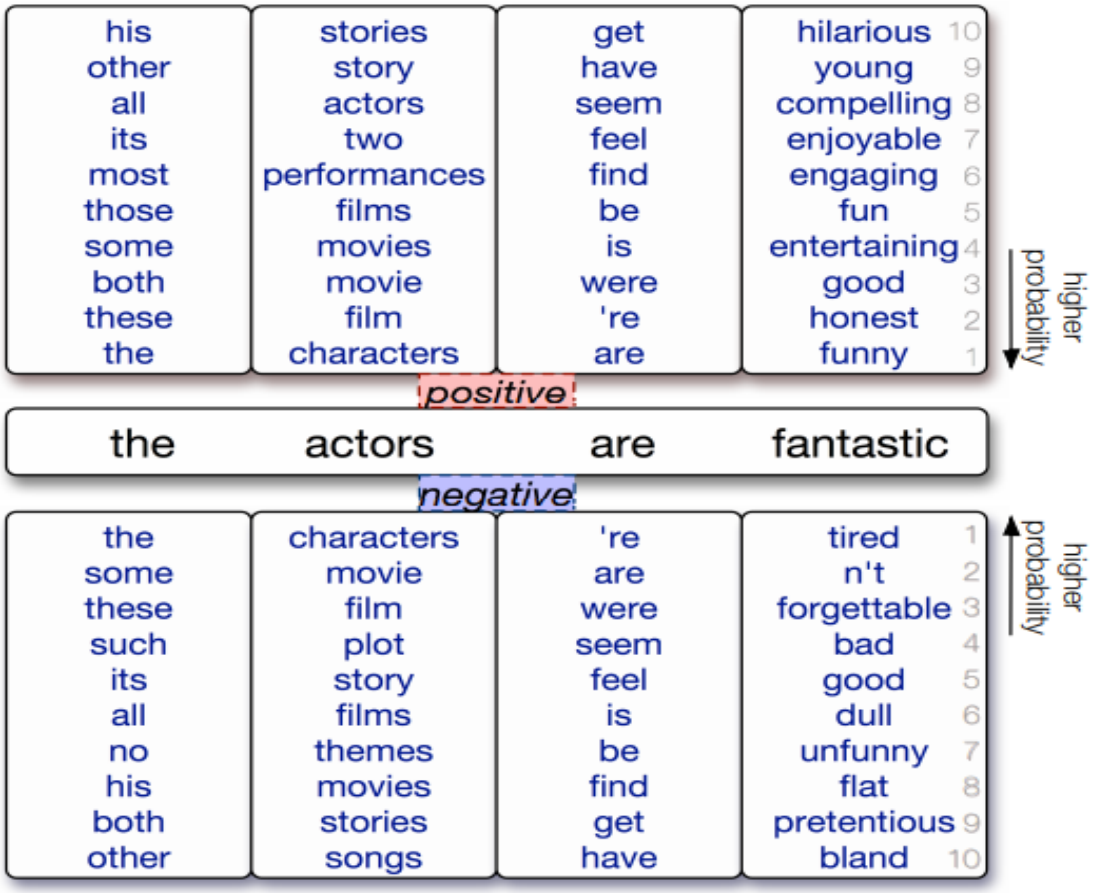


Figure-2-2: Kobayashi's model for relational increase

Each of these approaches is discussed in the work of A. Balahur [19], which addresses the question of the classification of twitter messages, which are short sentences that belong to a single topic. It explains how to use a variety of interesting preprocessing modules (emoticon replacement, tokenization, punctuation marks, z-word, etc.) and how to use them. Fortunately, these approaches are collected prior to data classification, and his dissertation does not focus on why or how one of these modules helps to increase the accuracy of the classifier.

Inline texts, as seen in [20], generally contain a lot of noise and non-informative elements, such as HTML tags. This increases the complexity of the problem and complicates the classification process. The most commonly used algorithms to polish and prepare twitter message data are: Tokenization, derivation, and noise words are all methods of removing punctuation and symbols as we can see an example in [21]

Chapter 3 Methodology

Overall implementation is divided into subparts. Implementation flow diagram is given below.

3.1 Data Collection

In this section, we have satisfactorily described the collection of the data and the different NLP based tasks which are performed on it. First data is collected then NlP

based task are implemented. NLP based tasks include data processing and assigning emotions to data.

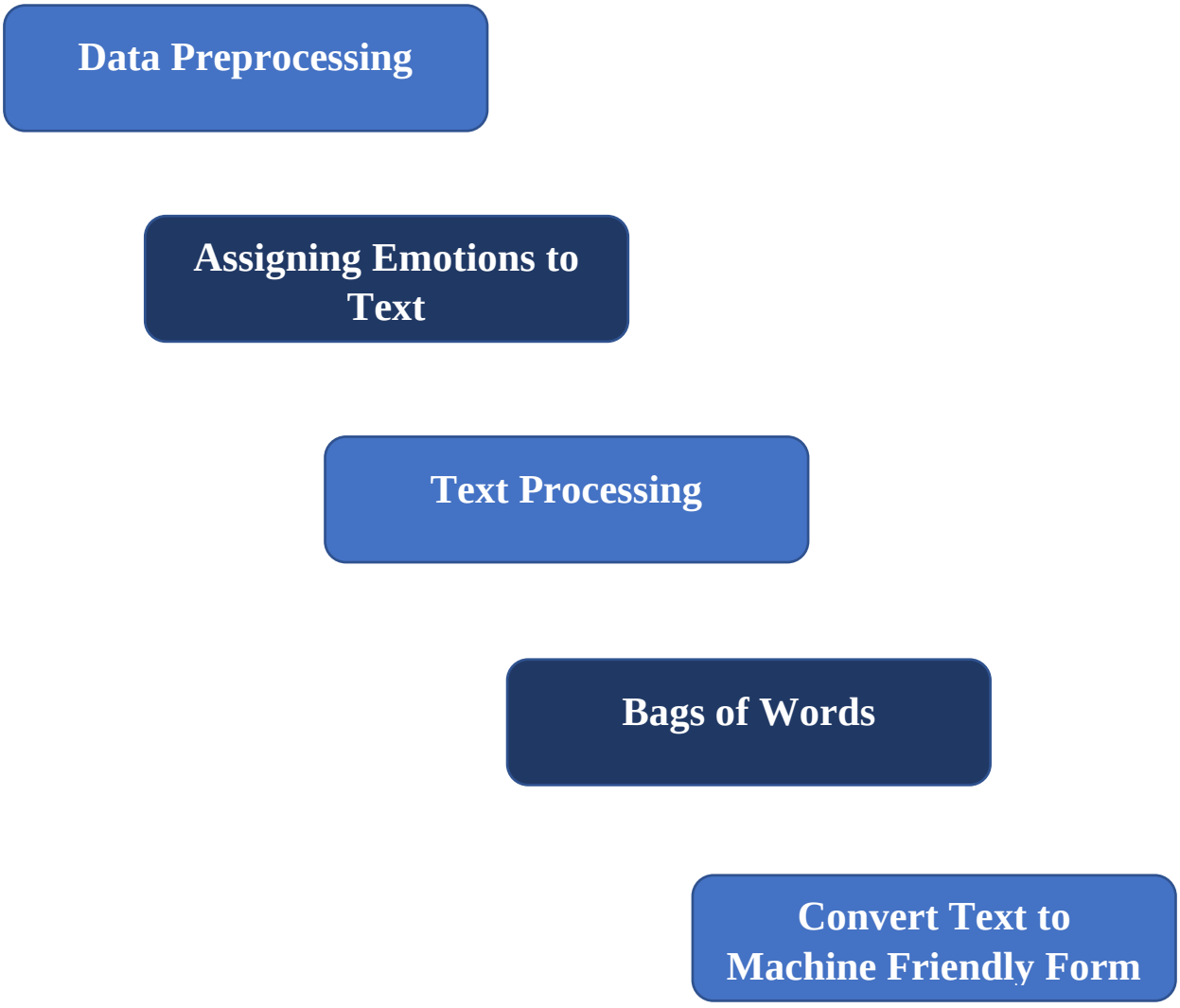
Collecting data allows an individual or business to answer relevant questions, evaluate results, and predict potential probabilities and patterns. Maintaining the accuracy of science, making informed business decisions, and ensuring quality assurance require accurate data collection. In the proposed model, we used two datasets to provide a measure. One for the covid19 tweets from a particular region and the second for the covid PCR results from that specific region.

Numerous datasets have been developed for the classification of supervised texts on twitter. Each is made up of tweets that have been manually tagged into a single opinion group by human experts. Positive and negative are the most general, although some data sets often have numerical labels related to the strength of emotions. An opensource platform [22]presents twitter sentiment ranking datasets that are used regularly. This dataset contains more than one lac tweet from around the world. On the other hand, covid19 PCR results are obtained from [23], which is also an unprofitable opensource platform. This platform has a short record of even just one day from each country. How many PCR test samples were collected in which country, on what day, at what time and what result was obtained in relation to these tests? This platform has very considerable statistics related to covid19.

3.2 NLP Tasks

Natural language processing, or NLP for short, is an area of research that focuses on the relationships between human language and computers. Computer technology,

artificial intelligence and computational linguistics converge in this area. This is an NLP-based business flow chart. Researchers can use natural language processing (NLP) to organize and structure information for tasks such as advanced summary, interpretation, named object recognition, relationship detection, emotion evaluation, speech recognition, and subject segmentation.



3.2.1 Data Preprocessing

Data preprocessing is critical in data mining operations because it directly affects the success rate of the project. Since the natural world data is not clean, this diminishes the sophistication of the data under study. If there are incomplete attributes, attribute values, noise, or outliers, and redundant or incorrect data, the data is considered impure. If all of these are present, the accuracy of the results will suffer. Use of systems such as abbreviations, dialects, exclamation points, words with recurring letters and incorrect grammar. In addition to some words from social networks, such as URL, hashtag etc., twitter users are used to using emoticons and emojis. So, in this search, first, the unimportant columns were removed from the tweet dataset. Then the following NLP-based preprocessing techniques were implemented.

Numbers excluded: Removing numbers from text is a standard technique as they do not convey any emotion. Some scholars argue that keeping the numbers can increase the accuracy of the classification.

Lowercase: All lowercase words is one of the most popular preprocessing strategies. All capital words are converted to lower case. As a result, many terms are combined and the dimensionality of the problem is minimized.

User Mention and URL Removal: Each sentence in a twitter text is a query that contains a mention of the user and a URL. Including them does not imply sentiment, however one solution is to remove them with preprocessing tags, as it did.

Stop word Removal: Word breaks are important words that appear frequently in all sentences. It is believed that there is no need to analyze them because they do not provide valuable material. The list of these terms is not completely predefined and, depending on the application, can be changed by removing it or adding others.

Contraction Replacement: In preprocessing, one technique that can be used is the use of contractions, as phrases like "I don't want" after contraction replacement it will be “I do not want”.

3.2.2 Assigning Emotions to Text

Interpreting a body of text to explain the opinion it conveys is a central feature of sentiment analysis. We usually assign a positive or negative meaning to this emotion, known as polarity. The polarity score sign is often used to infer whether the general mood is positive, neutral, or negative. In many cases, sentiment analysis for text data can be done in many stages, including individual sentences, paragraphs, and the entire document. Sentiment is often calculated in the text as a whole, or some aggregations are made after sentiment has been calculated for particular phrases. Sentiment research works well when there is contextual meaning to the document, rather than when there is only an empirical context. Without projecting feelings, thoughts or moods, the target text typically represents any common statement or information. Subjective text is text that is normally transmitted by a person with a particular mood, feeling or thought.

We have used NLP library named ‘Textblob’ that has a method ‘Sentiment’ with a property ‘Polarity’. Value of polarity is a float type that lies between -1 and 1 [26] where -1 refers to negative and 1 refers to positive sentiment.

Text	Polarity Score	Sentiment.
image doesn’t list source d careful overall risk dying statistics related.	negative	-0.033
kolar need blood type b positive jalappa hospital blood component need plasma b ve covid19 recover	positive	0.227273

coronavirus covid19 deaths continue rise s almost bad politicians businesses want	negative	-0.7
Smelled scent hand sanitizers today someone past think intoxicated.	negative	-0.25

Table 3-1A: Polarity Table

3.2.3 Text Processing

So far, we're done with data preprocessing and motion mapping to text. However, we need to reduce the dimensionality of the data. To do this, we use two main concepts.

• Stemming

• Lemmatization

Both lemmatization and stemming are intended to reduce the inflected and often derivational forms associated with a word to a common base form. Acceptability is exemplified by stemming and derivation. They choose a canonical representative for a group of closely related word types. Furthermore, there are notable differences in them.

Stemming: Stemming is the process of removing the suffix from a word and reducing it to its basic form. Stemming is a natural language processing normalization strategy that reduces the amount of computation involved. We can make NLP derivations using libraries like Porter Stemming, Snowball Stemmer, etc. Derivation is primarily used to minimize the number of dimensions in records. In other words, if there are words like walk, run, wait, wait, they are contextually different but identical. By removing the suffixes from all sentences, we can get to the root of the word "walk".

Stemming: Let's try to reduce a given word to its root using stemming, which is one of the most common text preprocessing strategies used in natural language processing (NLP) and machine learning in general. In the process of derivation, the root of the word is called the root and in the process of lemmatization, a lemma.

Consequently, a stemming algorithm would recognize that the best comes from good, and therefore the lemme is good. A bypass algorithm, on the other hand, will not be able to do this. Over or under response can occur, and the best word can be simplified for gambling or gambling, or just kept better. However, there is no way to minimize it to its positive root word by derivation.

Stemming vs Lemmatization

1. The lemmatization reduces the forms of words to linguistically correct headwords, while the derivation reduces them to (pseudo) roots. This distinction is noticeable in languages with more complex morphology. However, may not be important in many IR implementations.

2. Derivation can also deal with derived variance, while stemming can only deal with inflectional variance.

3. Stemming is more complicated to apply (especially for morphologically complex languages) and usually requires the use of vocabulary. Successful derivation, on the other hand, can be achieved using relatively simple rule-based methods.

We used the derivation in our implementation because of its advantage over the slogan mentioned above.

3.2.3 Bag of Words

A bag of words model, or Bow for short, is a method of extracting features from text for use in modeling, such as machine learning algorithms. The method is simple and adaptable and can be used to remove features from records in a variety of ways. A word bag is a representation of text that represents how often words appear in a document. It consists of two steps:

1-A lexicon of known words.

2-A metric to determine the presence of known terms.

It is called a "bundle" of terms because all the details about the word order or the composition of the document are discarded. The model only cares whether recognized terms appear in the text or not, not where they appear. The assumption is that records with similar material are equivalent. Furthermore, we can deduce a lot about the context of the document solely from its text. You can make the sack of words as simple or complicated as you like. The difficulty arises from the choice of how to create a vocabulary of known words (or cards), as well as how to evaluate the inclusion of known words.

3.2.4 Words Vector

Until now, all major related activities are performed with data. All we need now is to convert the words to vector form. Machines only understand vectors. Vectorization is a mechanism by which text data is translated into a machine-readable format. So, to achieve this, we are using a technique called Count Vectorization. The count vectorizer helps us both to create word bags and to convert them later into vector form. The result is a coded vector of the length of the entire

vocabulary and an integer count of how many times each word appears in the text. We call these scattered vectors since they can have several zeros. Python's scipy. sparse module simplifies working with sparse vectors.

The vectors returned by transform () are scattered vectors, and you can use the toarray () function to convert them back to numpy arrays to investigate and better understand what's going on.

To do this, we use the "Count Vectorizer" tool from the sklearn library. Since the amount of vocabulary is so high, it is important to keep the scale of function vectors to a minimum. The 700 most common terms(features) are used in this initiative. It is also worth noting that we set min df = 2 and ngram range = 0. (1,3). min gl = 2 indicates that a word must appear in at least two texts for the language to be used in the array. The term ngram interval refers to the number of ngrams used to cut a sentence. Let's say we have a sentence; I am a child. If we cut the sentence by digraph (ngram = 2), the sentence will be cut as ["I am", "am a", "a child"].

3.3 Implementing AI Classifiers

NLP is created by taking artificial intelligence and focusing it on human linguistics. "NLP allows humans to communicate with machines:" This subset of artificial intelligence allows computers to understand, translate and control human language. NLP, like machine learning and deep learning, is a subclass of AI.

To grasp the context of text documents, machine learning (ML) for natural language processing (NLP) and text analysis uses "restricted" machine learning and artificial intelligence (AI) algorithms. Social media articles, web comments, survey results, and even banking, medical, legal and regulatory documents are examples of documents that contain text. In natural language processing and text analysis, the job of machine learning and artificial intelligence is to develop, accelerate, and automate

the underlying text analysis functions and NLP functions that transform unstructured text into data and useful information.

A batch of text documents is tagged or annotated with descriptions of what the algorithm can look for and what that element of supervised machine learning should look like. These documents are used to "practice" a mathematical model, which is then received with unlabeled text for investigation. As the algorithm learns more about the documents it parses, it can retrain it for larger or more robust data sets. For example, supervised learning can be used to train a model to rate movie reviews and then to rate the critic's star rating. If you come across these words, what you need to know is that they refer to a series of data scientist-driven machine learning algorithms. Lexalytics requires supervised machine learning to develop and expand its text analysis and natural language processing skills.

In this study, I applied various machine learning methods to natural language processing.

Those AI classifiers are mentioned below:

- Support vector Machine.
- Bernoulli Naive Bayes.
- Logistic Regression.
- Single Perceptron.
- Multilayer perceptron.

3.3.1 Support Vector Machine

Over the past two decades, the extensive development of machine learning has focused on improving the efficiency of classifiers, leading to a new wave of next

generation classifiers, such as carrier vector machines. SVM is a kind of meaningful classifier: it's a vector space-based machine learning tool for finding a decision boundary between two groups as far away as possible from any point in the training data (any point can be discarded as outliers or noise).

We used Support Vector Machine (SVM) to classify the data. The SVM class in the Sklearn library has a function LinearSVC that performs linear support vector classification. We choose Hinge as the default loss function.

I set class weight to "balanced" by defaulting to the penalty term L2, which will automatically change weights that are inversely proportional to frequencies against and class from the fed results. After shuffling on random state = 1800 on 5 k folds, we train our data and then forecast on test data. On test data, SVM achieves a 93.78 percent accuracy, which is a decent starting point for experimenting with other classifiers.

3.3.2 Bernoulli Naive Bayes

The Nave Bayes (NB) classifier is a community of basic probabilistic classifiers that is often used as the basis for classifying text. It is based on the premise that all functions are independent of each other as long as the category variable. Bernoulli's theorem is a version of Naive Bayes. A probabilistic classifier, the Naive Bayes classifier estimates the probability of an entry being evaluated for all groups that receive an entry. Conditional possibility is another name for it. The Bernoulli model uses the knowledge of binary events to identify a test document, ignoring the number of occurrences, while the multinomial model keeps track of the different occurrences.

Since we need to categorize tweets into negative and positive emotions, the Bernoulli Nave Bayes classifier is a good choice because it works well for discrete results. We use the Sklearn library's naive bayes class, which has a BernoulliNB method, to implement it (). After training our data on 0.03 alpha, we discovered that it has a test accuracy of 90.65%.

You don't need too much data for education. It is capable of handling both continuous and discrete data. It can handle a large number of predictors and data points. It is fast and can be used to make real-time predictions. Since only the probability of each class and the probability of each class need to be determined given different input values (x), training is fast. The optimization methods do not have to match any coefficient.

3.3.3 Logistic Regression

The logistic regression classification algorithm is used to attribute observations to a different group of groups. Spam or unsolicited email, fraud or non-fraud in online transactions, and malignant or benign tumors are some examples of labeling problems.

The logistic regression classifier passes the weighted combination of the input characteristics to a sigmoid function. Any real number can be converted to a number between 0 and 1 using the sigmoid algorithm. In the text data, the logistic regression was found to be very accurate and the underlying algorithm is also reasonably simple to understand. More specifically, in the field of natural language processing, logistic regression is widely considered a good first algorithm for classifying text.

On test results, I achieve 93.31 percent accuracy using logistic regression as our linear model on 5 folds with penalty term l2. The match and predict properties of GridSearchCV (scikit-learn.org n.d.) is used to train and test our models.

3.3.4 Single Layer Perceptron

The first neural model suggested was the single-layer perceptron. A vector of weights constitutes the content of the neuron's local memory. A single-layer perceptron is calculated by calculating the number of input vectors, each with the value multiplied by the corresponding part of the weight vector. A feedback network based on a threshold transfer characteristic is known as a single-level perceptron (SLP). SLPs are the most basic artificial neural networks and can only distinguish cases that are linearly separable and have a binary target (1, 0).

In this study Single Layer Perceptron is implemented. Surprisingly, tuning hidden layer did not increase sentiment prediction accuracy on test results, so we achieve 54.68 percent accuracy on a single hidden layer.

3.3.5 Multi-Layer Perceptron

A feedback artificial neural network called multilayer perceptron (MLP) is a type of feedback artificial neural network (ANN). The term MLP is ambiguous; it can refer to any feedforward ANN, or it can refer specifically to networks composed of different perceptron layers. In supervised learning problems, multi-layered perceptron are often used. They learn to model the association (or incompatibilities) between inputs and outputs by training in a series of input-output pairs. MLPs are suitable for classification prediction problems where inputs are classified or tagged. They are also useful for regression prediction problems where a quantity of actual value is predicted from a series of inputs. On our data collection, we have used Multiplayer perceptron, which provided us with 93.73 percent accuracy using five hidden layers and Relu as the triggering feature with 0.3 alpha. Finally, we train

logistic regression as a linear model on five folds with the penalty term l2 and achieve a test data accuracy of 93.31 percent. GridSearchCV (scikit-learn.org, n.d.) was used to train and evaluate our models, and its match and forecast properties.

Below table is describing overall view of implemented Classifiers.

AI Classifiers	Hyperparameter
SVM	param_grid = lr2_param, verbose = 1, cv = kfold, n_jobs = -1, scoring = roc_auc lr2_param=(dual= True, C=0.05,class weight=balance,loss=hinge)
Bernoulli Naive Bayes	mlp_param_grid=[alpha=0.03,binarize=0.001], cv = kfold, scoring = roc_auc, n_jobs= 1, verbose = 1
Single Layer Perceptron	hidden_layer_sizes=1,activation=logistic,solver=sgd, alpha=0.1,learning_rate=constant max_iter=1000
Logistic Regression	param_grid = [lr2_param], cv = kfold, s coring = 'roc_auc', n_jobs = 1, verbose = 1],lr2_param=penalty=l2,dual=False C=0.05,class_weight=balanced
Multilayers Perceptron	mlp_param_grid = [hidden_layer_sizes=5, activation=relu, solver=adam, alpha=0.3, learning_rate=constant,max_iter=1000], param_grid = mlp_param_grid, cv = kfold, scoring = roc_auc, n_jobs= -1, verbose = 1

Table 3-2B: Classifiers vs Hyperparameter

3.4 Submit Best Classifier

After implementing all AI classifiers on dataset with respect to their best hyperparameter we came to know that SVM classifier has achieved highest accuracy rate with 93.78% with elapsed time of 1.2s. here we have results of all implemented classifiers

Classifiers	Accuracy Rate	Elapsed Time
SVM	93.78%	1.2s
Bernoulli Naive Bayes	90%	1.0s
Single Layer Perceptron	54%	12.9s
Logistic Regression	93.21%	2.4s
Multi-Layer Perceptron	93.73%	60s

Table 3-3C: Accuracy vs Elapsed Time

3.5 Feature Extraction

Since there are 700 features, it is inconceivable to look at all the coefficients at the same time. As a result, we can sort them and look at the ones with the highest coefficients. The bar chart below shows the 30 largest and 30 smallest coefficients in the linear SVM model, and the bars indicate the size of each coefficient.

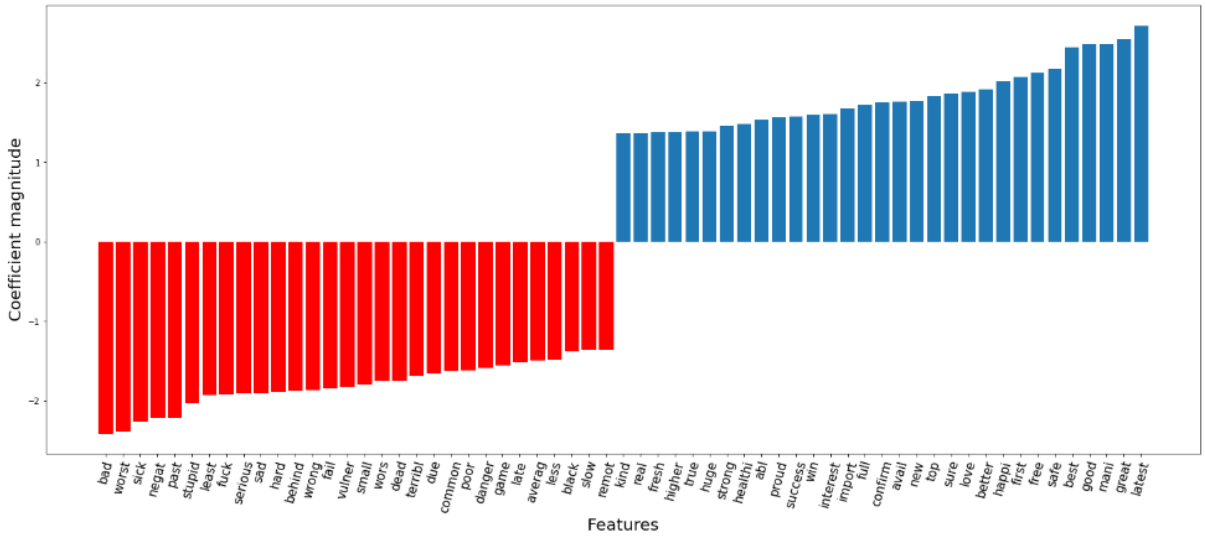


Figure 3-1:Most Significant Feature Representation

Smallest thirty indexed features: array(['bad', 'worst', 'sick', 'negat', 'past', 'stupid', 'least', 'fuck','serious', 'sad', 'hard', 'behind','wrong', 'fail', 'vulner','small', 'wors', 'dead', 'terribl', 'due', 'common', 'poor','danger','game', 'late', 'averag', 'less', 'black', 'slow','remot'] Largest thirty indexedfeatures: array (['kind', 'real', 'fresh', 'higher', 'true', 'huge', 'strong','healthi', 'abl', 'proud', 'success', 'win','interest', 'import','full', 'confirm', 'avail', 'new', 'top', 'sure', 'love', 'better', 'happi', 'first', 'free', 'safe', 'best', 'good', 'mani', 'great','latest']. Here common suffix are dropped because of stemming.

3.6 Region & Timewise Division

In this section, data is analyzed according to specific regions and time interval. Ten different countries or regions have been selected for Region and Timewise Division. Names of those countries are united states of America, Pakistan, India, China, South Africa, Philippines, United Kingdom, Switzerland, Philippines, Ireland. Time interval of 10 countries is analyzed from 25/07/2020 to 29/08/2020.

In this time interval these 10 countries have different tweets related to covid19. We have thoroughly analyzed the texts and came to know the percentage of positive and

negative tweets by these regions. Here positive tweets are referring to tweets which are referring that people are serious about covid and negative tweets which are referring that there is nothing like covid or covid is nor harmful etc. These are the details of analysis including the percentage of positive tweets and negative tweets of related region.

1-USA

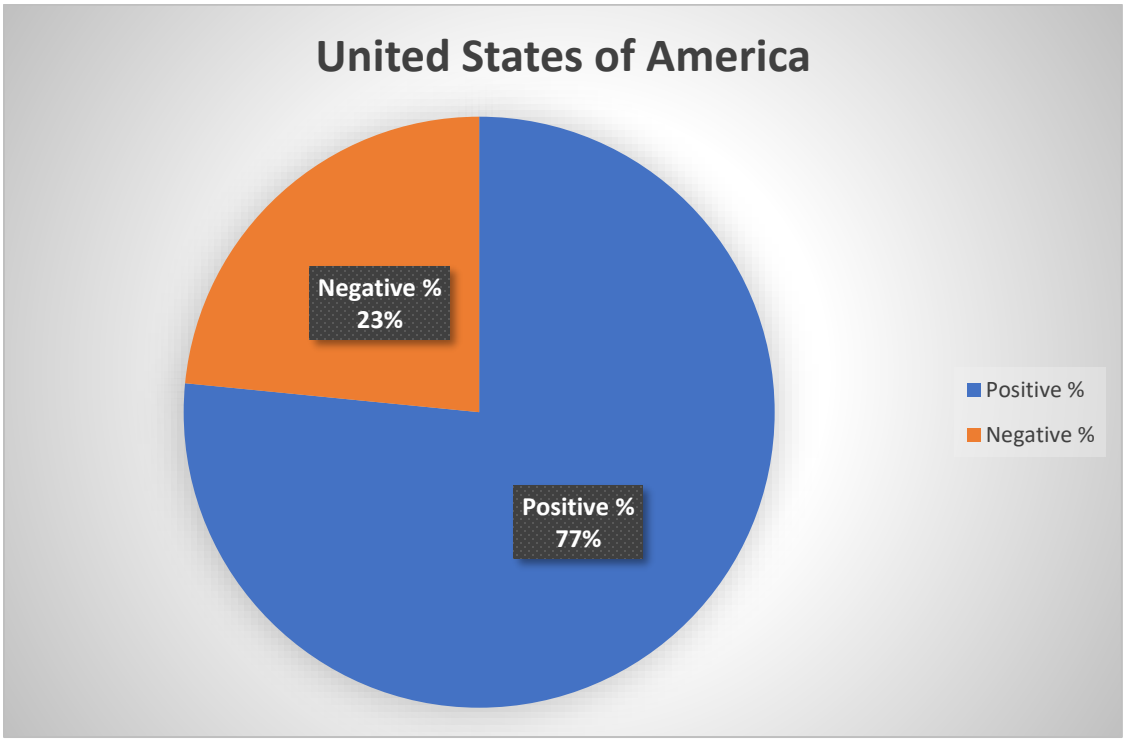


Figure 3-2:USA Region Opinion About Covid19

2-India

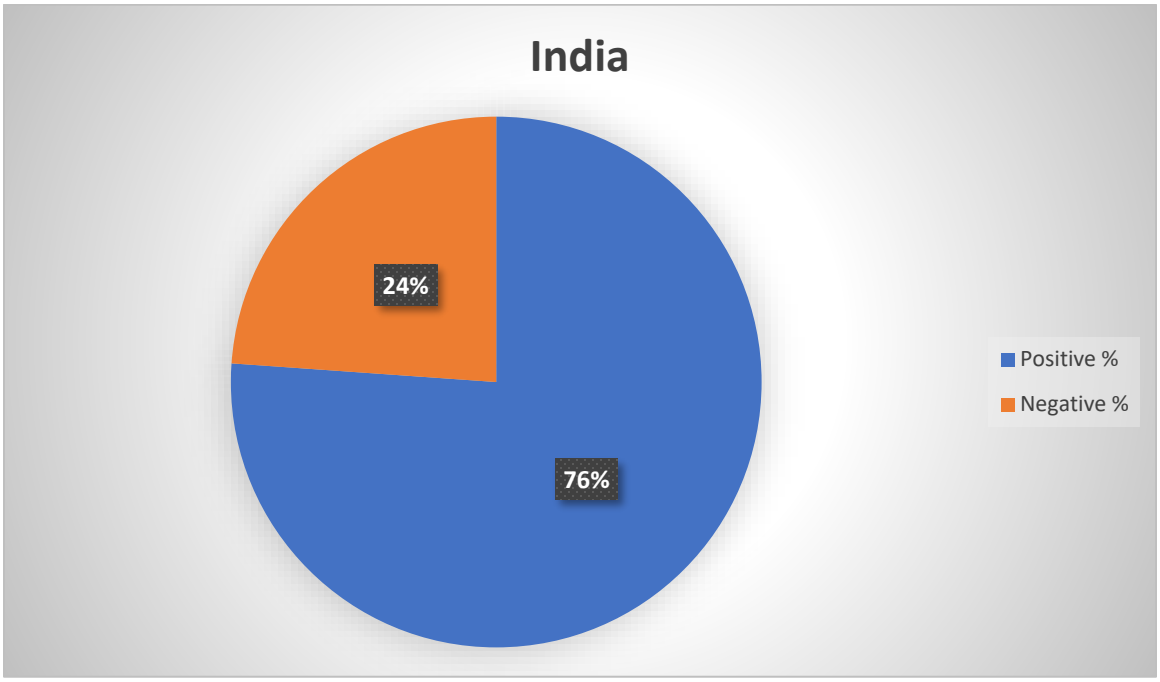


Figure 3-3: India Region Opinion About Covid19

3- China

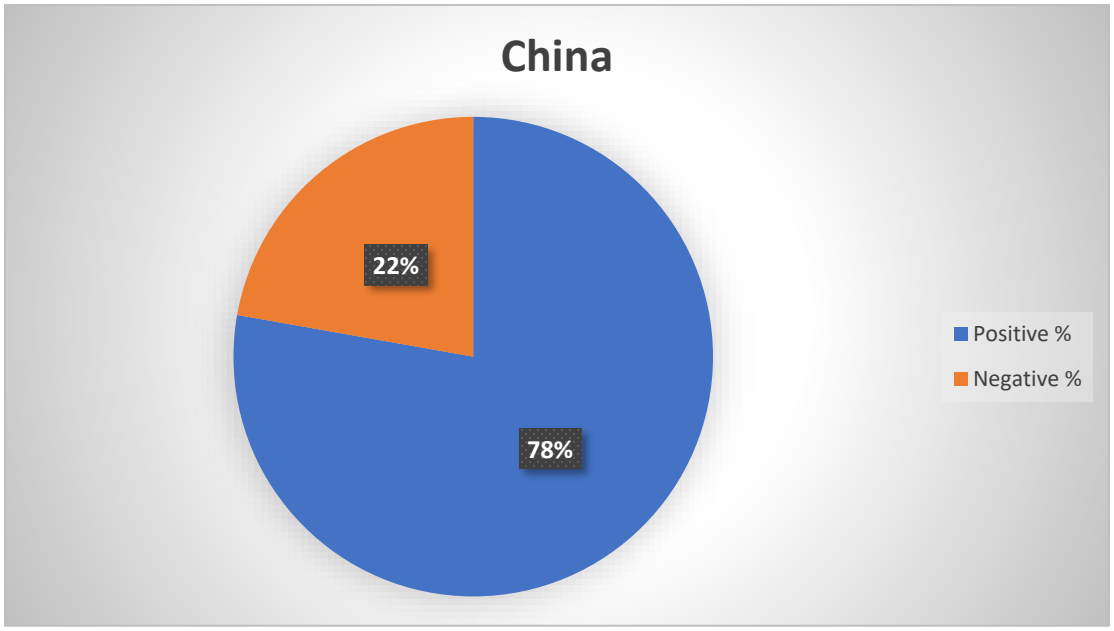


Figure 3-4:China Region Opinion About Covid19

4- Australia

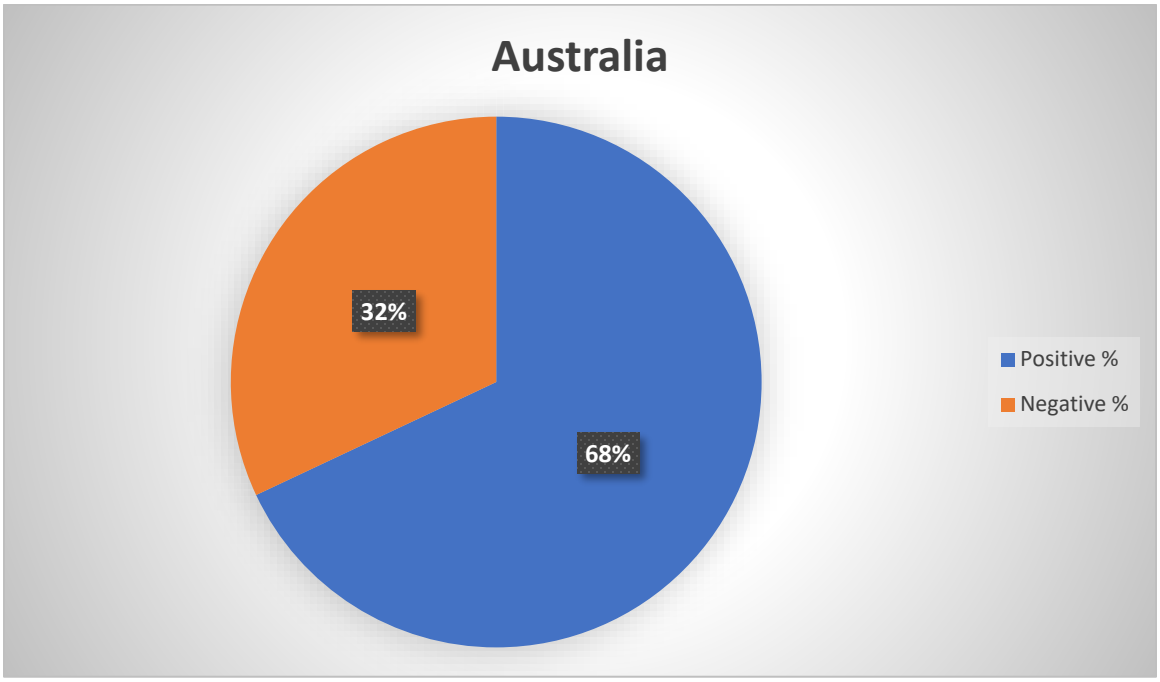


Figure3-5:Australia Region Opinion About Covid19

5- Ireland

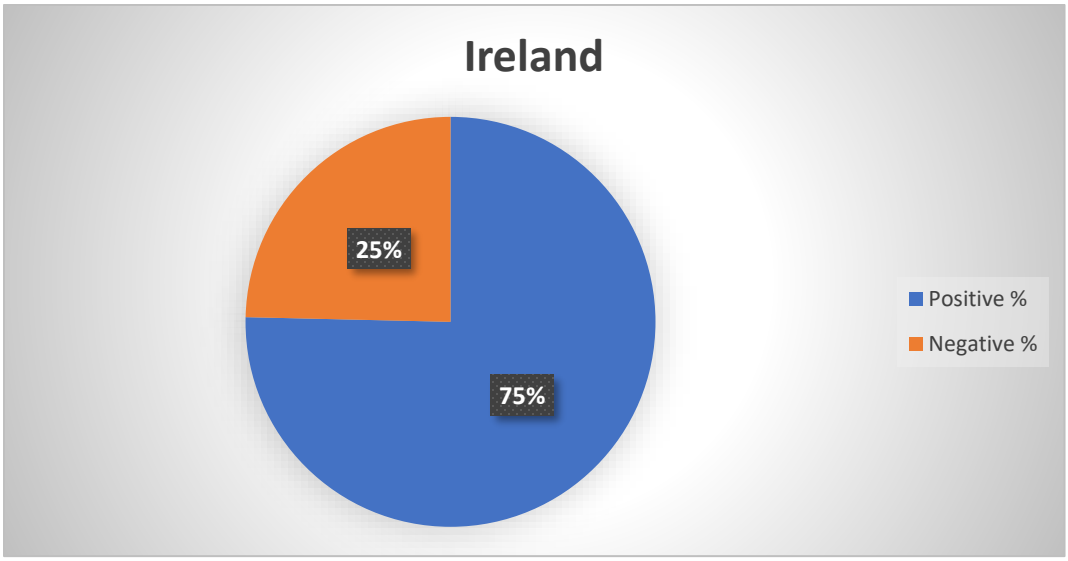


Figure 3-6:Ireland Region Opinion About Covid19

6- Switzerland

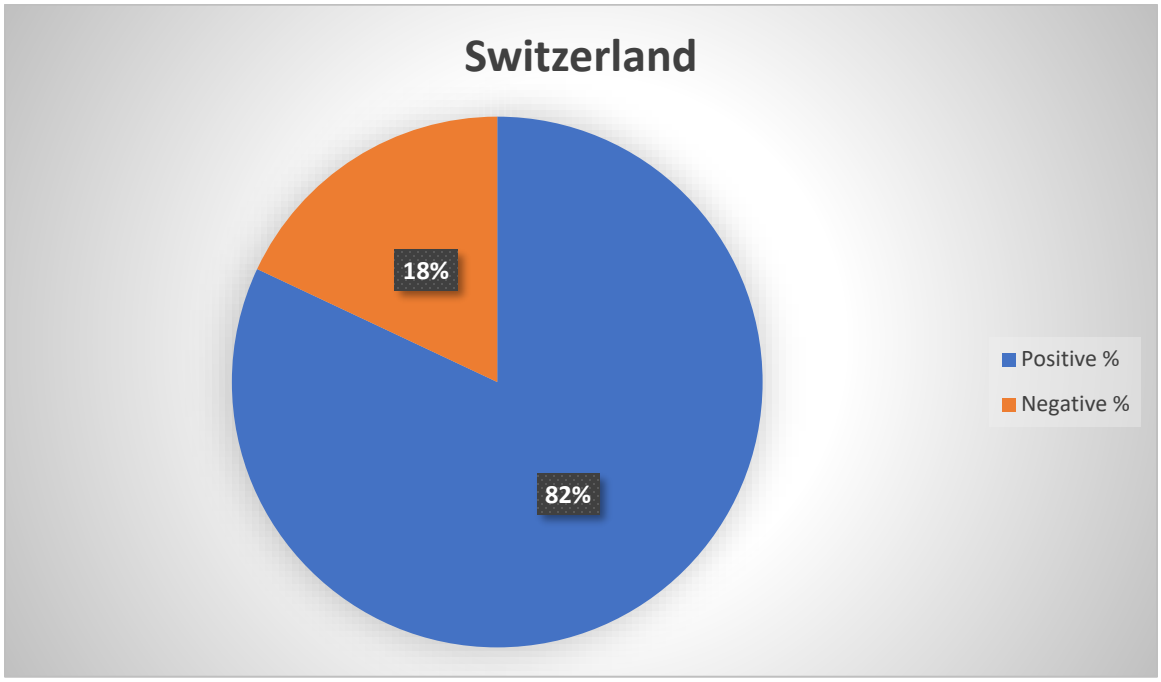


Figure 3-7: Switzerland Region Opinion About Covid19

7- South Africa

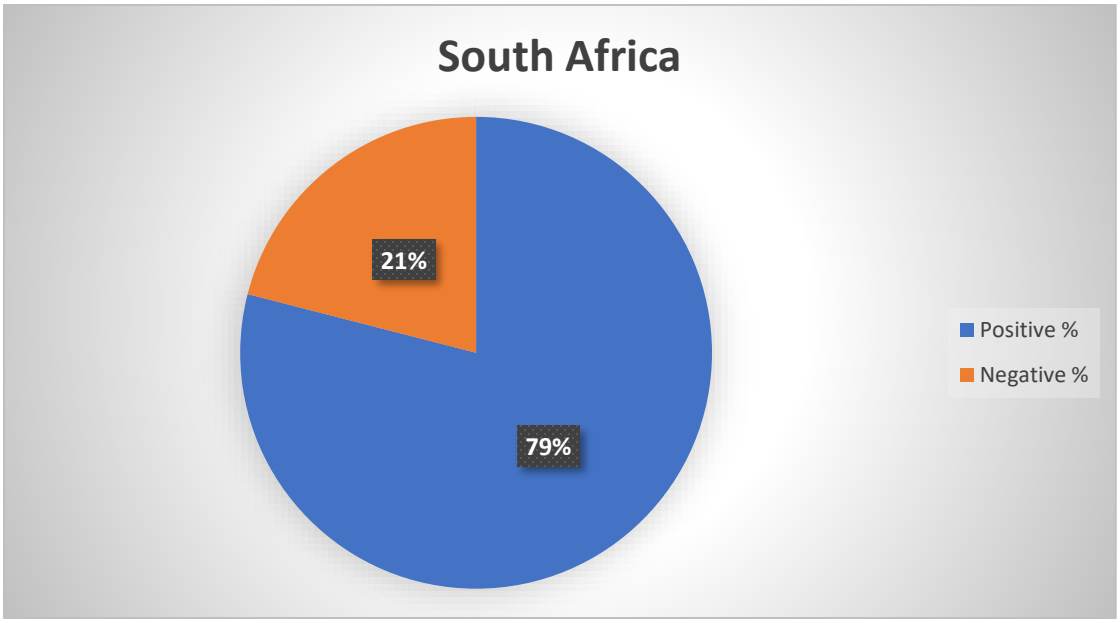


Figure-3-8: SA Region Opinion About Covid19

8- Pakistan

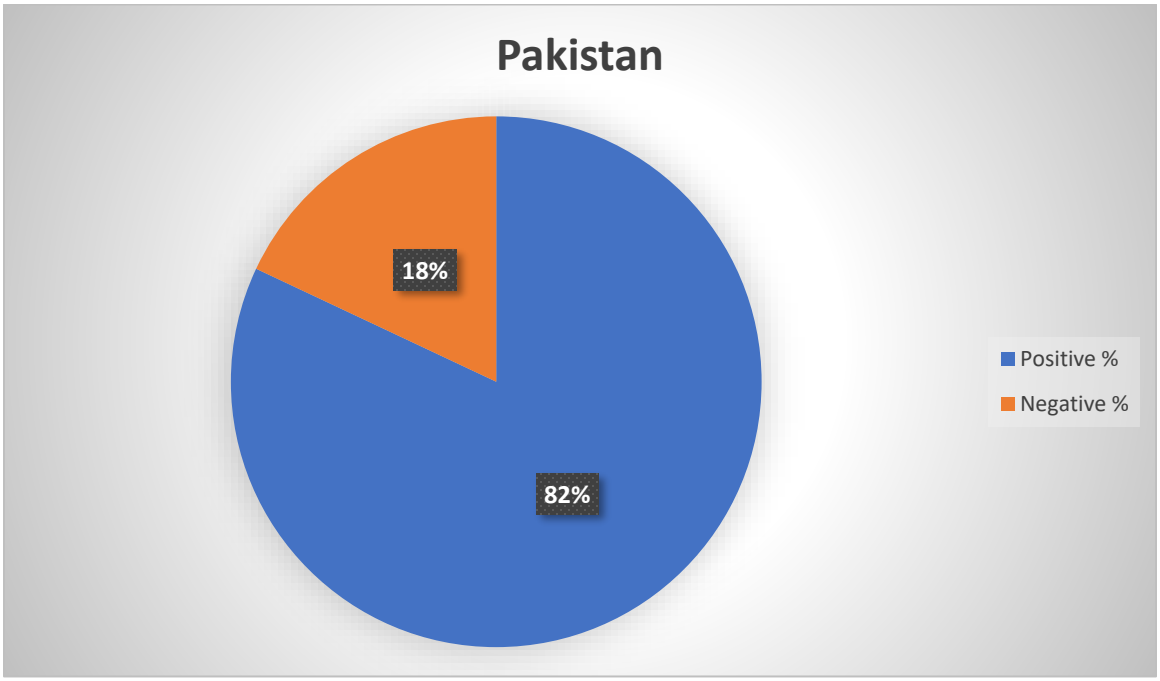


Figure 3-9: Pakistan Region Opinion About Covid19

9- United Kingdom

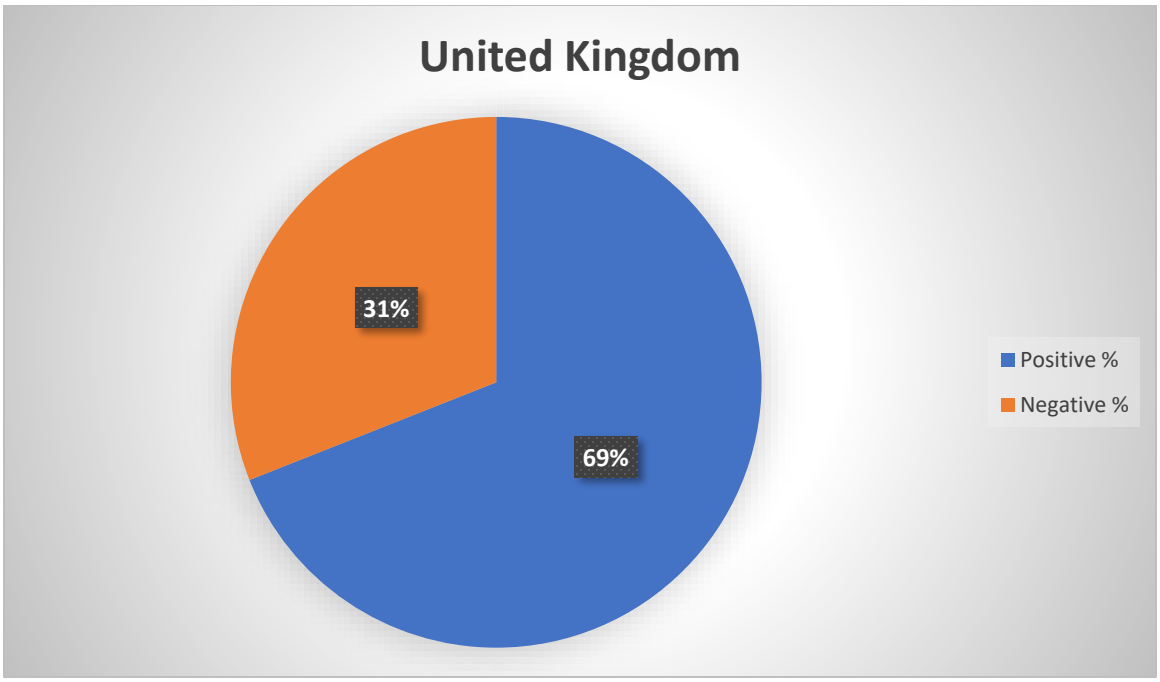


Figure 3-10: UK Region Opinion About Covid19

10- Philippines

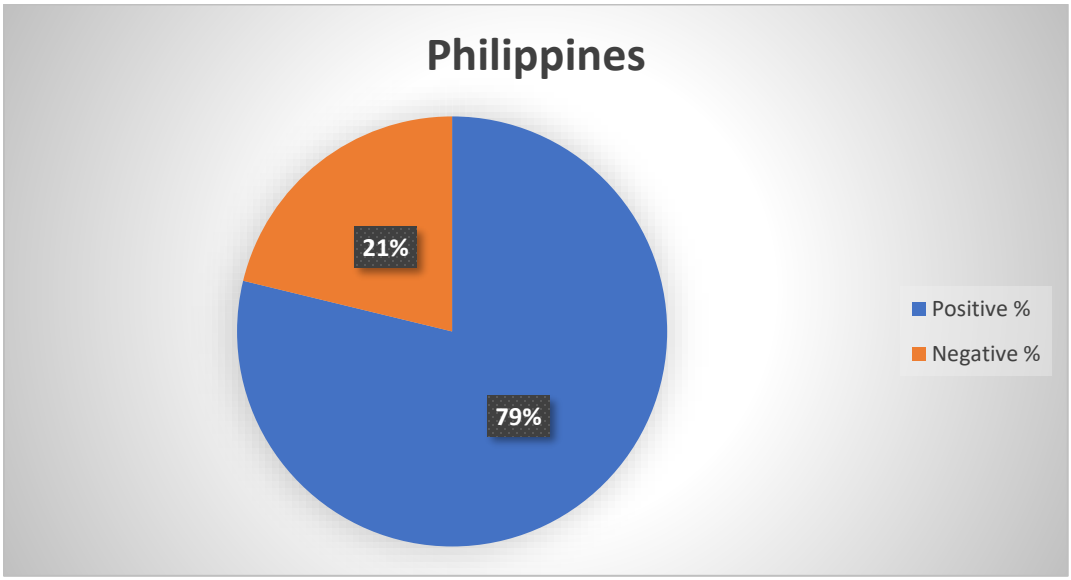


Figure 3-11: Philippines Region Opinion About Covid19

Chapter 4 Results

Results section is divided into three subsections:

- Getting PCR Observations
- Assigning Sentiments to tweets with high accuracy rate.
- Time and region wise twitter-based PCR analysis.

4.1 Getting PCR Observations

The COVID-19 nose swab PCR procedure is the most precise and safe way to diagnose the virus. A positive result indicates that you have COVID-19. A negative result indicates that you were not infected with COVID-19 at the time of the test. To diagnose genetic material from a single organism, such as a virus, a polymerase chain reaction (PCR) test is used. If you are sick at the time of the test, the test will detect the existence of a virus. Even if you are no longer infected, the test can detect virus fragments.

The accuracy and reliability of the COVID-19 PCR test are its key advantages. It is the most reliable COVID-19 detection test possible.

This section describes coronavirus PCR test results in ten different regions. Results are gained from an opened source platform [23]. This platform tells that how many covid19 PCR tests are taken in ten different regions at which day and how many test results are positive or negative from 25th of July to 29th of august.

If we talk about covid 19 PCR test in united states of America per million stats.

Date	Total Tests Per Million	Positive Results
25/07/2020	2,905	197.98
25/07/2020	2,856	195.68

..	..	..
..	..	..
28/08/2020	2,549	125.91
29/08/2020	2,549	125.74

Table 4-4-1: USA PCR Result

If we divide number of total positive to total number of test then multiply it with 100. We will come to know that USA had 6.13% ratio of positive Cases. In the same way, we can extract required positive percentage of required region.

India had 8.99% of positive cases in specific time interval 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	2500	31.87
25/07/2020	259	32.86
..	..	..
..	..	..
28/08/2020	622	50.55
29/08/2020	612	51.53

Table 4-4-2: India PCR Results

China had 0.57% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	2497	14.57
25/07/2020	2570	15.06

..	..	..
..	..	..
28/08/2020	2594	5.29
29/08/2020	2590	4.81

Table 4-4-3: China PCR results

Ireland had 1.13% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	1454	3.44286
25/07/2020	1437	3.50071
..	..	..
..	..	..
..	..	..
28/08/2020	1570	23.81043
29/08/2020	1662	23.49229

Table 4-4-4: Ireland PCR Results

Switzerland had 2.25% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	586	13.37029
25/07/2020	587	13.55186
..	..	..

..	..	..
..	..	..
28/08/2020	1019	29.92629
29/08/2020	1074	31.11471

Table 4-4-5: Switzerland PCR Results

South Africa had 21.94% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	742	200.69586
25/07/2020	727	195.35814
..	..	..
..	..	..
..	..	..
28/08/2020	308	40.45186
29/08/2020	284	37.34957

Table 4-4-6:SA PCR Results

Australia had 0.56% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	2508	16.78443
25/07/2020	2491	18.53243
..	..	..

..	..	..
..	..	..
28/08/2020	2594	5.294
29/08/2020	2590	4.81229

Table 4-4-7: Australia PCR Results

Pakistan had 2.89% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	94	6.21943
25/07/2020	95	5.19314
..	..	..
..	..	..
..	..	..
28/08/2020	110	2.241
29/08/2020	108	2.06829

Table 4-4-8:Pakistan PCR Results

United Kingdoms had 0.90% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	1906	9.76
25/07/2020	1913	9.81057

..	..	..
..	..	..
..	..	..
28/08/2020	2717	17.97771
29/08/2020	2723	17.51057

Table 4-4-9:UK PCR Results

Philippines had 10.50% of positive cases in specific time interval from 25th of July to 29th of august.

Date	Total Tests Per Million	Positive Results
25/07/2020	258	17.08829
25/07/2020	265	16.93714
..	..	..
..	..	..
..	..	..
28/08/2020	309	35.43243
29/08/2020	309	33.74157

Table 4-4-10: Philippines PCR Results

4.2 Assigning Sentiments to Tweets with High Accuracy Rate.

We also compare our achieved highest accuracy with some of existing work. Results are shown below in table 4.1:

Authors	Models	Accuracy
[5]	CNN	89%
[2]	LSTM	84.3%
[10]	Naive Bayes Support Vector Machine Multinomial Naive Bayes	75% 78% 86%
[6]	Decision Tree	91.81%
[1]	BiLSTM + attention + CRF	85.49%
Proposed Model	**SVM**	**93.98%**

Table 4-11: Proposed Model Vs LR

So, it is seen that with hyperparameter mentioned in methodology section, SVM has given us highest accuracy rate with 93.98% which is more than many publications discussed in literature review.

4.3 Time and Region Wise Twitter-Based PCR Analysis

Table 4.2 is showing the trend of time and region wise PCR analysis. In above graph, x-axis is for number of PCR percentage and y-axis is for regions.

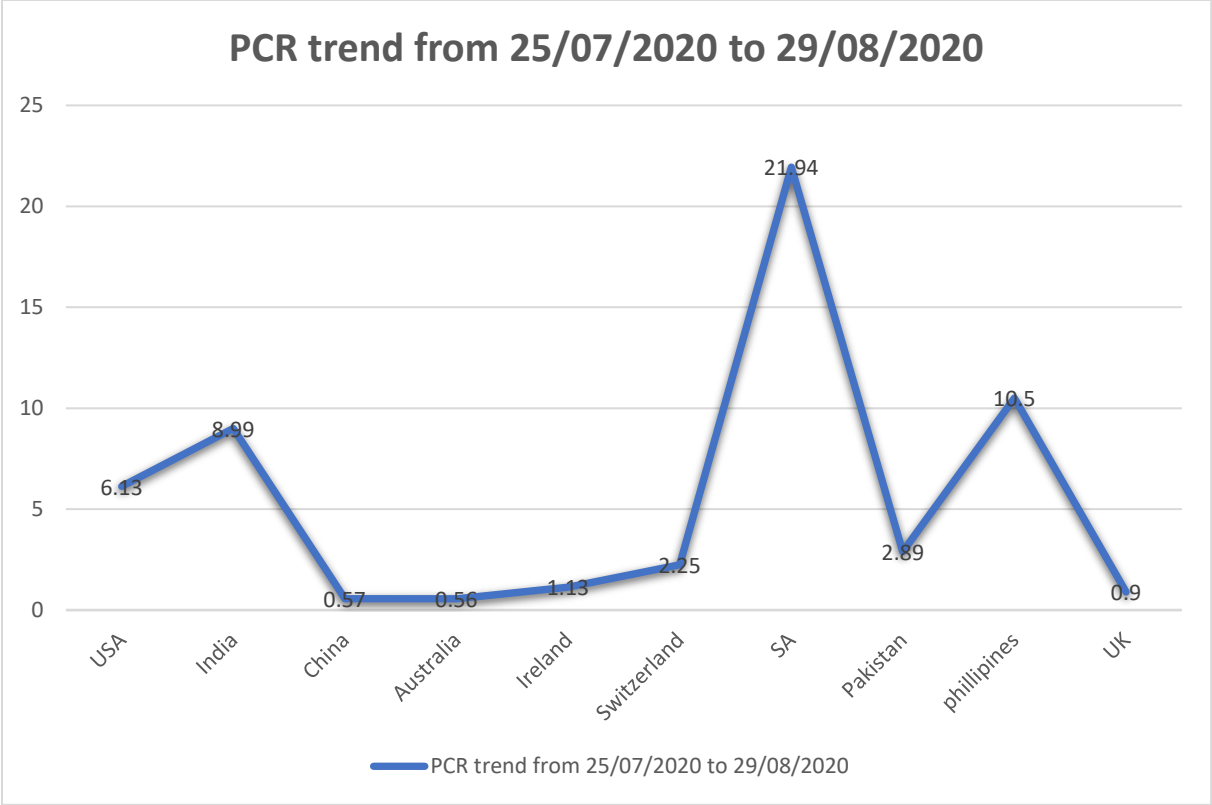


Table 4-12: PCR Trend

As we can see from table 4.2 that South Africa had highest positive PCR ratio with 21.94% from our data set. After that Philippines has highest positive PCR ratio with 10.5%. After Philippines, it is shown that India has highest positive PCR ratio with 8.99%. Then USA is leading with 6.13% of positive PCR ratio from available data set of 10 countries. Pakistan is at number five with 2.89% of positive PCR in list of ten countries. Switzerland is at number six with 2.25%. Ireland has 1.13% of positive PCR. China is at 2nd last position with low positive PCR ratio of 0.57%. Australia has lowest positive PCR percentage with 0.56. It comes at the last number in our dataset.

Table 4.3 is showing the trend of twitter-based PCR ration with respect to USA, India, China, Australia and Ireland.

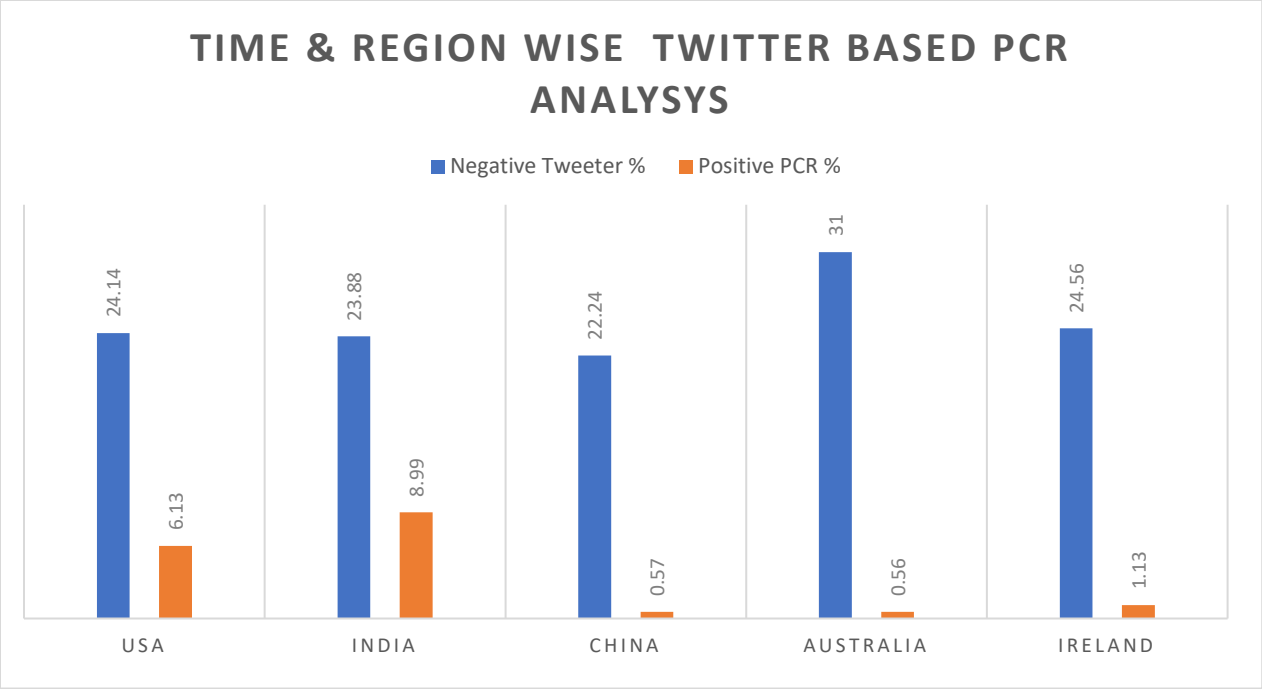


Table 4-13: Time & Region wise Twitter Based PCR Analysis of 5 Regions.

Table 4.4 is showing the trend of twitter-based PCR ration with respect to Switzerland, South Africa, Pakistan, United Kingdom, Philippines.

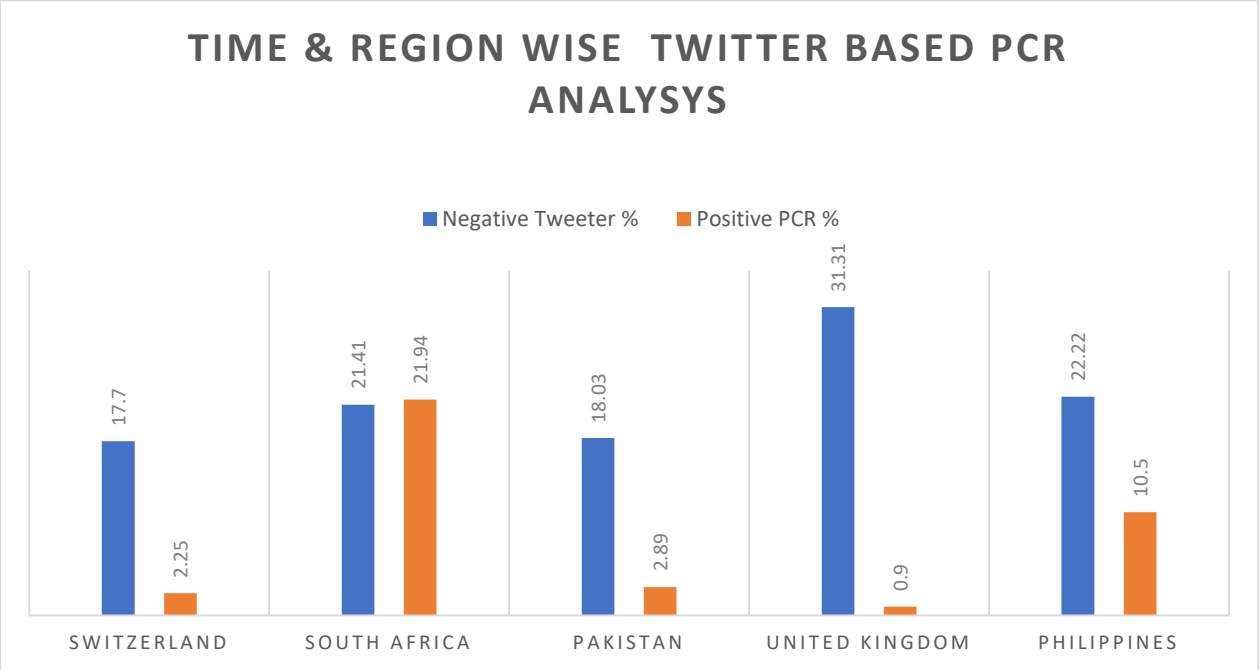


Table 4-14: Time & Region wise Twitter Based PCR Analysis of Remaining 5 Regions.

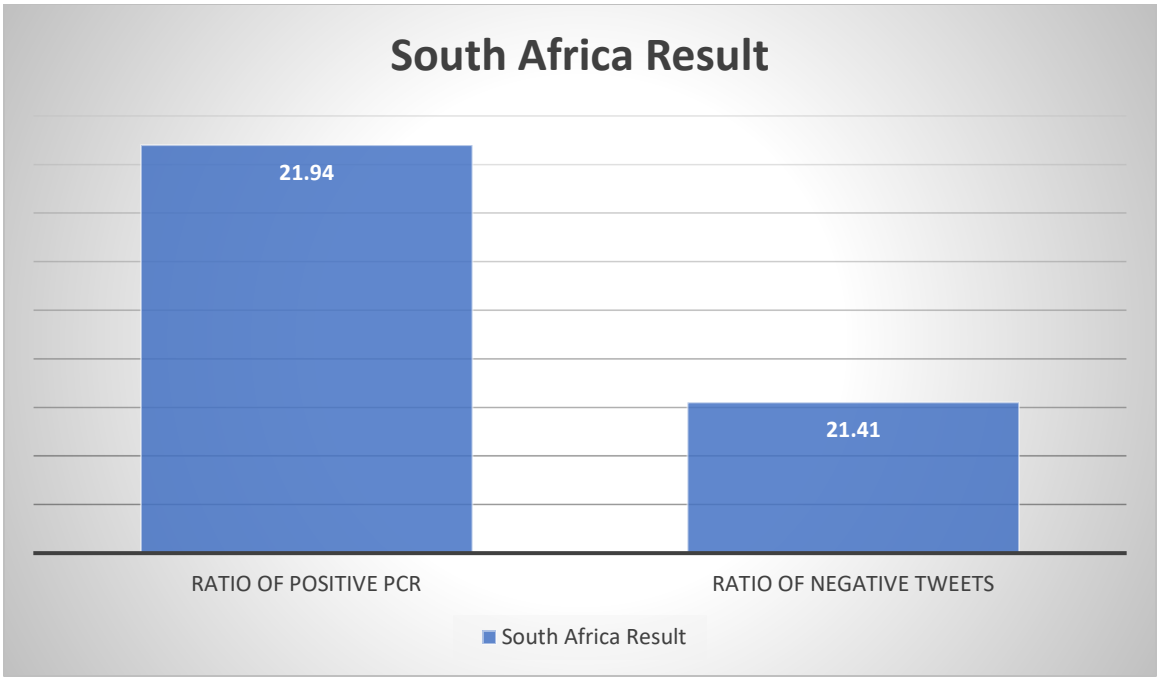


Figure 4-1: South Africa Results

In above table fig 4-5, example of South Africa is given with same ratio. In fig 4-5, it is clearly mentioned that this region has almost equal ratio between positive PCR results and negative tweets. So, if any pandemic or another wave of covid comes in future then admin of South Arica already will have stats of region. They can estimate easily that effected ratio of region by mining people's opinion about that wave of covid or future pandemic. Both ratios will almost be same.

On the other hands countries like Australia is also in list with least effected ratio but high negative tweets.

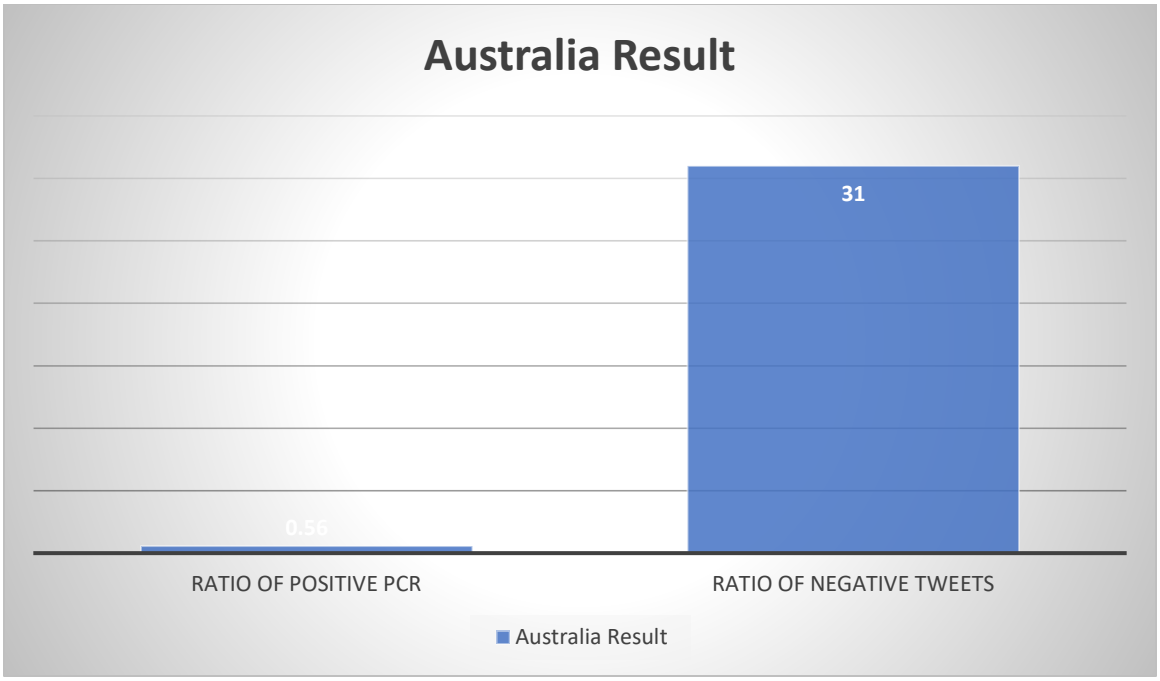


Figure 4-2:Australia Results

In fig 4.2, it is seen that there is large contradiction between the ratio of people's negativity about covid19 and PCR positive ratio. Reason is that government of Australia took strict action to control it. For future, Australian admin has stats of region. They know better that their people showed carelessness towards covid in their opinion, but result was opposite. So, they can repeat their policies and strategies for controlling covid. But still, they will know that this region had not positive thoughts about covid.

Chapter 5 Discussion

In above section, we worked on region and timewise twitter-based PCR analysis. We took data of ten countries. We have seen that the opinion of these countries regarding the corona and the PCR results of covid19. In this section we will look at how much contradiction or similarity has been found in the region's opinion about covid19 and PCR results of the covid19 of these regions. When we have learnt to find the difference of contradiction and similarity in opinions and PCR results. Then we can provide an international measure related to experimented regions. With the help of this measure, administration of these regions can take more wise steps or make more suitable strategies for safety of their people from third or fourth wave or any future pandemic.

For example, if we talk about USA then we would come to know that the contradiction ratio between negative opinions and positive PCR is 18% because 24% of USA region gave their negative comments about covid (like they don't believe in covid) and 6% of region was covid positive. So, contradiction ratio is 18%. So, if admin of a country knows his region's contradiction ratio, he can take necessary steps, accordingly in present of in future. For example, in future if any other pandemic comes and admin of any region don't have the kits to praise the positivity or negativity of the deceases related to that pandemic then only thing any admin would have, it will be the people's opinion about it. So, at that time, admin of that region can corelate this research experiment with going on pandemic. This research can help the administrations in taking many decisions in effective way.

Some necessary decision related facts are given below:

- source allocation
- implementing urgent lockdown
- region specific awareness campaign
- apprehend people spreading propaganda.
- region specific relief fund
- Travelling restrictions
- admiring the regions for their behave.

5.1 Source Allocation

The method of distributing and handling assets in a way that supports an organization's strategic interests is known as resource allocation.

Managing physical assets like hardware to allow the most use of softer assets like human resources is part of resource utilization. In order to optimize the optimal utilization of available resources and achieve the highest return on investment, resource selection entails juggling conflicting interests and goals and deciding the most effective course of action.

When we talk about covid19, governments, foreign bodies, and insurance services all have a responsibility to provide quality health coverage to all citizens to the best of their abilities. However, during a pandemic, where all resource is expected to be scarce, this might not be feasible. Rationing resources in this sense entails making tragic decisions, however these decisions should be morally justifiable. This is why

resource allocation exists. Links to doctors, ventilators, vaccines, and medications may be among the resources allocated. In such situations, it is important resource allocations are ethically justified. According to our research, if we take examples of South Africa region then we will come to know that there is 0% of contradiction in people's opinion and positive PCR. SA has 21% of negative opinion of total count, and same 21% has positive PCR. So, we can say that in source allocation, SA regions more attention and priority then other regions like China who has difference of 21% in opinion and PCR results.

5.2 implementing urgent lockdown.

A lockdown is a technique used while the building's inhabitants are in imminent danger. In the case of a lockdown, teachers, faculty, and staff will be told to stay in their rooms and not leave until the situation has been resolved. This helps first officials to protect the students and staff in place, deal with the imminent threat, and relocate any innocent civilians to a haven. lockdown includes:

- Barricade your door and stay in your bed or office.
- Continue to be silent.
- If you want to leave the building or space, you will be arrested.
- Wait before the ambulance crews say, "all clear!"

If we talk about these coronaviruses, then in many countries had strict and smart lockdown. Every region which is highest ration of positive PCR, government can impose lockdown at that regions.

5.3 Region Specific Awareness Campaign

It's critical to raise public visibility to boost excitement and patronage, encourage self-mobilization and initiative, and mobilize local information and resources. Raising political consciousness is critical because policymakers and lawmakers are central players in the adaptation policy process. To achieve the desired result, successful communication techniques are required for raising awareness. The term " region specific awareness campaign " refers to the combination of these marketing techniques with a specific audience over a set period of time. The goal of awareness-raising campaigns varies depending on the situation, it usually involves raising awareness, educating the target group, projecting an optimistic image, and attempting to improve their behaviors. If we see towards our research, then SA deserves this campaign badly because this region has zero difference in people negative opinion and positive PCR. So, all regions who have behaviors like this, then government of that region can arrange region specific awareness campaign.

5.4 Apprehend People Spreading Propaganda.

Rumors are a deadly weapon that has a negative impact on people's confidence. Law enforcement agents have the authority under the law to pursue disciplinary charges against someone who spreads news about the virus and induces widespread hysteria.

Although online disinformation and misinformation about the coronavirus are distinct, the former is the deliberate dissemination of inaccurate or misleading facts, while the latter is the unintended dissemination of the same, both pose a significant health risk. Like some specific individuals who have a number of following, also said that there is nothing like corona etc. When admin of a region would come to know that this region has high ratio of negative opinion then admin can find out the source of this misinformation and take action against them.

5.5 Region Specific Relief Fund.

Even if some people have learnt to adapt their daily lives to the new pandemic, millions of people are also unable to cope with the consequences of COVID-19. And if they now receive federal assistance, it is insufficient to obtain them by themselves. Millions of people also struggle to make ends meet, including rent, electricity, and food. With fears of poverty, deprivation, and a sense of hopelessness growing in an already disadvantaged demographic, we need to help them more than ever before. So, in this case, governments have ethical obligations to arrange a funds for those regions which are badly affected like South Africa, India, Philippines etc.

5.6 Travelling Restrictions.

Governments can ban resident to travel those countries who has high positive PCR ratio. Government cannot allow the residents of major affected countries to come in their countries. Like people of Pakistan and India are banned to land in Australia because Pakistan and India's health situation is getting worst day by day.

5.7 Admiring The Regions for Their Behave.

This research also tells us about the regions who showed minor negative opinions about covid or who have positive PCC ratio less than one percent like Australia, United Kingdom etc. So, governments should also encourage or appreciate these regions for their efforts like they maintained physical distancing, following SOPs etc.

There can be many other region-specific policies which can be made by international platform or local governments to stopping the spread of covid 19.

Chapter 6 Conclusion

In this research, we took covid19 tweets of ten different regions. Regions are United States of America, Pakistan, India, China, South Africa, Philippines, United Kingdom, Switzerland, Philippines and Ireland. Time span of these tweets lies between 25th of July,2020 to 29th of august,2020. We applied different classifiers on COVID related tweets to train deep neural network to enhance the accuracy rate after implementing word embedding and cleaning techniques using Natural Language Processing. Afterwards, we share results of each implemented classifier and pick the highest accuracy model (SVM) to investigate impactful word that are true representatives of the user sentiments shared by tweets. After assigning the emotions to tweets, we found the positive and negative percentage of opinions of these regions towards covid19. We took the PCR results of these regions and again found the positive and negative percentage of covid19 PCR results per million of same regions. We compared these percentages and provided an international measure which will help governments to deal with any future pandemic on the basis of opinions related to pandemic. This research related to a global pandemic is of huge importance as we are still experiencing COVID and lockdowns in many regions of the world. This research can help governments to take wise and suitable decisions to control the spread of any future pandemic or any another wave of covid19. Further, dynamic measures should be taken to provide concentrated comfort to social media users despite of negativity being shared every minute.

6.1 Future Work

People on social media like to share their emotions by emoticons and different abbreviations, such as "awsl" (English for "that's awesome, I'm blown away"). As a result, in future experiments, hyper parameters can be finetune and adding more functionalities to increase the model's efficiency. There is more work required on whether users' anonymity is abused when we gather these texts from the Internet for emotion analysis testing, which is an important thing to think about.

Chapter 6 Recommendations

This model can be more significant if we take PCR results and tweets of same people instead of taking PCR results and tweets based on regions. Sentiments is assigned through library of python which can be neutral due to many cases. A proper algorithm can be designed for dealing with sentiments which can ensure that sentiments are assigned perfectly.

Chapter 7 Bibliography

[1] S. S. J. L. Y. G. H Zhang, "Sentiment Classification for Chines Text Based on Interective Multitask Learning.," 2020.

[2] S. Wen, "Memristive LSTM Network for Sentiment Analysis," in *IEEE Transactions on System*, 2019.

[3] L. H. a. X. Q. C. Sun, "Utilizing BERT for AspectBased Sentiment Analysis via Constructing Auxiliary Sentence.," in *Conf. North Amer pp. 1–6.*, 2019.

[4] H. S. P. R. Y. Sailaja Thota, "OPINION MINING OF TWITTER DATA USING MACHINE LEARNING," in *USING MACHINE LEARNING." International Journal of Advanced Research in Computer Science. ISSN No. 0976-5697 .* , 2020.

[5] E. d. D. 2. hmed Sulaiman M. Alharbi, "Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information,Cognitive Systems.," in *ResearchVolume , ISSN 50- 61*, 2019.

[6] S. S. a. J. Yang, "Twitter Sentiment Analysis Based on Ordinal Regression.," in *IEEE Access vol. 7, pp. 163677-163685.*, 2019.

[7] S. G. A. S. S. K. a. N. Y. R. S. Kathuria, "Real Time Sentiment Analysis On Twitter Data Using Deep Learning(Keras).," in *International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, India, 2019.

[8] A. S. a. A. Moschitti, "Twitter Sentiment Analysis with Deep Convolutional Neural Networks.," in *SIGIR pp. 959962.*, 2015.

[9] A. V. D. adav, "Sentiment analysis using deep learning architectures: a review.," in *Artif Intell Rev 4335–4385* ., 2020.

[10] T. M. K. J. P. a. M. C. B. Heredia, "CrossDomain Sentiment Analysis: An Empirical Investigation.," in *IEEE 17th Int. Conf.* , 2016.

[11] T. K. M. Singh, "Role of text pre-processing in twitter sentiment analysis.," in *Proc. Comput. Sci. 89, 549–554 http://www.sciencedirect.com/science/ article/pii/S1877050916311607*, 2016.

[12] Y. X. Y. W. P. T. Y. L. J. B. D. C. Z. H.-V. K. G. L. H. P. Y. Y. L. (. Shi, "Proceedings of the First International Conference on Information Technology and Quantitative Management," in *, Procedia Computer Science, vol. 17. Elsevier (2013). http://www.sciencedirect.com/science/journal/ 18770509/17*, Dushu Lake Hotel, Sushou, China, 2013.

[13] Z. K. Wei J, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification,," in *ICLR 2019-7th International Conference on Learning Representations.*, 2019.

[14] U. P.-S. G. U. R. Fayyad, " Summary from the KDD03 panel: data mining: the next 10 years.," in *SIGKDD Explor. 5(2), 191–196 (2003). doi:10.1145/980972.981004*, 2003.

[15] S. S. a. S. S. Dimitrios E, "A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis," in *{dimievfr,ssymeoni,avi}@ee.duth.gr http://www.nonrelevant.net*.

[16] F. S. Fatemeh Zarisfi, "Solving the twitter sentiment analysis problem based on a machine learning-based approach," in *Evolutionary Intelligence 13(6):1-18, DOI: 10.1007/s12065-019-00301-x.*.

[17] A. M.-S. Keshavarz H, "ALGA: adaptive lexicon learning using genetic algorithm for sentiment analysis of microblogs.," in *Knowl-Based Syst 122:1–16*, 2017.

[18] K. S., "Contextual Augmentation: Data Augmentation by Words with Paradigmatic Relations,," in *in Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), New Orleans. 2018; pp. 452–457*, 2018.

[19] A. Balahur, "Sentiment Analysis in Social Media Texts," in *European Commission Joint Research Center.*, Varese, Italy, 2013.

[20] E. L. X. S. Y. Haddi, "The role of text pre-processing in sentiment analysis," in *Procedia Computer Science 17, 26–32*, 2013.

[21] B. Z. Y. Duncan, "Neural networks for sentiment analysis on twitter. In: Cognitive Informatics & Cognitive Computing (ICCI* CC),," in *IEEE 14th International Conference on. pp. 275–278. IEEE 2015.*, 2015.

[22] G. Preda, 2020. [Online]. Available: www.kaggle.com. https://www.kaggle.com/gpreda/covid19-tweets.

[23] "ourworldindata,," 2020. [Online]. Available: https://ourworldindata.org/coronavirus-testing.