

**UNDERSTANDING CLIMATE CHANGE CONCERN USING MACHINE
LEARNING TECHNIQUE**

BY

AFRIN SULTANA

ID: 211-25-931

**This Report Presented in Partial Fulfillment of the Requirements for the Degree of
Masters of Science in Computer Science and Engineering**

Supervised By

Dr. Sheak Rashed Haider Noori

Professor & Associate Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Co-Supervised By

Mr. Abdus Sattar

Assistant Professor & Coordinator M.Sc.

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

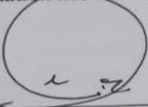
DHAKA, BANGLADESH

17 JANUARY 2023

APPROVAL

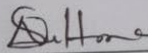
This Project titled “**Understanding Climate Change Concern Using Machine Learning Technique**”, submitted by **Afrin Sultana, ID No: 211-25-931** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17 January 2023

BOARD OF EXAMINERS



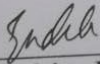
Dr. S M Aminul Haque, PhD
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



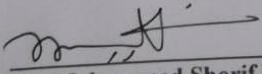
Ms. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



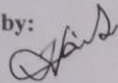
Dr. Mohammad Shorif Uddin, PhD
Professor
Department of Computer Science and Engineering
Jahangirnagar University

External Examiner

DECLARATION

We hereby declare that, this thesis has been done by **Afrin Sultana** under the supervision of **Dr. Sheak Rashed Haider Noori, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



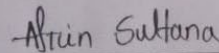
Dr. Sheak Rashed Haider Noori
Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Co-Supervised by:



Mr. Abdus Sattar
Assistant Professor & Coordinator M.Sc.
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Submitted by:



Afrin Sultana
ID: - 211-25-931
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty Allah for His divine blessing makes us possible to complete the final year thesis successfully.

We really grateful and wish our profound our indebtedness to **Dr. Sheak Rashed Haider Noori, Professor & Associate Head**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Data Mining” to carry out this thesis. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Mr. Abdus Sattar Assistant Professor & Coordinator M.Sc.** and **Professor Dr. Touhid Bhuiyan Head**, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Climate transformation has been identified as the principal environmental crisis of the 21st century and become a matter of considerable debate. Nowadays, the significance of climate change is extensively argued through various papers, newspapers, websites, and blogs. But that writing may lack accuracy, while the severity of results in scientific articles may be too unclear for the public to understand. With the rapid evolution of the internet and communication technology today, more and more people are interested to express their thoughts via social media and debate there a lot. Utilize the welfare of social media in this study, a gigantic dataset of tweets containing particular keywords connecting to climate change is analyzed using volume analysis and text mining methods such as case modeling and sentiment analysis and how ML can be a strong tool to predict accurately people's concerns and help society adapt to a changing climate. This study provides an explanation as well as a solution by classifying the tweet's natural, man-made, and neutral opinion reviews using the Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor, and Logistic Regression (LR). Among these algorithms, Support Vector Machine accuracy is the best accuracy and accuracy is 83.01%. The proposed model is made on Jupyter Notebook (a Python-based IDE) and trained on Kaggle's standard Twitter Climate Change Sentiment Dataset which has 43,743 records.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgments	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Background of the Research	1
1.2 Problem Statement	1
1.3 Motivation	1
1.4 Aim of the Research	3
1.5 Propose Solution	3
CHAPTER 2: LITERATURE REVIEW	4-7
2.1 Introduction	4
2.2 Related Work	4
2.3 Conclusion	6
CHAPTER 3: RESEARCH METHODOLOGY	8-20
3.1 Introduction	8
3.2 Research Subject and Instrumentation	8
3.3 Data Collection	9
3.4 Proposed Model	9
3.5 Data Transformation	10
3.6 Applied Mechanism	14
3.6.1 Classification Algorithms	14
3.7 Deep learning	19

3.8 Implementation Requirements	20
CHAPTER 4: RESULT & DISCUSSION	21-29
4.1 Introduction	21
4.2 Applied Machine Learning Algorithms	21
4.3 Preprocessing & Declaration	21
4.4 Data Visualization	21
4.5 Analysis of Performance Measured used	22
4.6 Limitations	29
4.7 Conclusion	29
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SWOT ANALYSIS	30-30
5.1 Impact on Society	30
5.2 Impact on Environment	30
5.3 SWOT Analysis	31
5.4 Conclusion	32
CHAPTER 6: CONCLUSION	33
6.1 Conclusion	33
6.2 Further Suggested Work	33
REFERENCES	34-35

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.3: Collected dataset sample	9
Figure 3.4: Proposed model	10
Figure 3.5.1: Data Transformation	11
Figure 3.5.2: Punctuation Character Removal	11
Figure 3.5.3: Tokenization	12
Figure 3.5.4: Stop Words Removal	12
Figure 3.5.5: Steaming	13
Figure 3.5.6: Vectorization	13
Figure 3.6.1: Logistic Regression	15
Figure 3.6.2: Decision Tree Classifier	16
Figure 3.6.3: Concept of Random Forest	17
Figure 3.6.4: support vector machine (SVM)	18
Figure 3.6.5: K-Nearest Neighbor (KNN)	18
Figure 3.7: Concept of deep learning	19
Figure 4.4: Data Visualization	22
Figure 4.5.1: Confusion matrix of Logistic Regression	26
Figure 4.5.2: Confusion matrix of Decision Tree Classifier	26
Figure 4.5.3: Confusion matrix of Support vector Machine	27
Figure 4.5.4: Confusion matrix of Random Forest Classifier	27
Figure 4.5.5: Confusion matrix of K-Nearest Neighbors	28
Figure 4.5.6: Comparison of Accuracy Result among DT, SVM, RF, KNN and LR	28

LIST OF TABLES

FIGURES	PAGE NO
Table 4.5.1: Accuracy rate of classification	23
Table 4.5.2: Precision rate of classification	24
Table 4.5.3: Recall rate of classification	24
Table 4.5.4: F1 Score of classification	25
Table 4.5.5: Basic Diagram of a Confusion Matrix	25
Table 4.5.6: Comparison of Accuracy among DT, SVM, RF, KNN and LR	29

CHAPTER 1

Introduction

1.1 Background of the Research

In current years, the issue of climate change has extended widespread concern. Climate change plays a vital role in temperature, people, environment as well as lifestyle. As a result, predicting accurately about climate change at an early stage is very important to conscious people correctly. Maximum accuracy of data helps to avoid confusion in public perception. With the evaluation of communication technology, more and more people tend to express their thoughts via social media. As millions of people are firing giant amounts of information day-to-day on social media, acquiring and exploring these data become quite challenging for everyone. However, social media can be used as a data source for mining public opinion on a variety of topics, including climate change, for the benefit of machine learning algorithms. A wide range of fields are using machine learning to address challenging problems.

1.2 Problem Statement

Climate change is one of the top reasons for natural catastrophes across the globe. Multiple actions have been brought to mitigate climate adaptation. People around the world don't still concern about climate change. Climate Change is the defining problem of our time and we are at a defining point. From shifting weather habits that endanger the food production to growing sea levels that raise the risk of disasters like flooding, the consequences of weather modification are international in scope and unusual in scale. Without powerful steps today, adjusting to these impacts tomorrow will be more complicated and pricier. Other hand any authorities can take the necessary steps to keep the climate normal if they can justify public concern accurately. The study is for catching people's sentiments accurately and taking necessary steps to reduce future disasters and issues.

1.3 Motivation

At long last, we should not overlook assessment appraisal is a NLP issue. It contacts each bit of NLP, co-reference targets, nullification managing, and word sense disambiguation,

which consolidates more troubles since these are not taken care of issues in NLP. In any case, it is besides helpful to grasp that feeling the appraisal is a fundamentally confined NLP issue because the framework doesn't need to thoroughly understand the semantics of each sentence or report yet basically needs to see a few pieces of it good or negative thoughts and their objective parts or subjects. In this sense, slant evaluation offers an uncommon stage for NLP analysts to increment liberal indications of ground on all fronts of NLP with the capacity of having a gigantic objective effect. In this report, I look at the speculation examination of the plane dataset and applied ML count for making assumption. I should utilize this report to draw in specialists from different ML computations of NLP to join to propel a cognizant undertaking to deal with the issue. Specialists before long in like way have a staggeringly improved view of the total extent of the issue, its structure, and center issues. Diverse new models and systems have been proposed. The examination has not exactly as of late developed now, moreover, stretched out on an essential level. Prior investigation in the field, by and large, in view of getting sorted out the possibility of subjectivity granted in reports or sentences, which is deficient for most genuine applications. Valuable applications routinely request progressively for what it's worth and fine-grained evaluation. Because of the improvement of the field, the report is also written in a created structure for twitter assessment examination using AI system. That helps with making a customized structure for twitter idea examination. Social affiliations or long reach relational correspondence help individuals with keeping related with their loved ones and are a direct strategy to discover what everybody is up to reliably in your get-together of partners. Easygoing affiliations can in like way be utilized to discover fun and charming things on the Internet since reliably your loved ones will share innumerable vague interests from you. Advertisers utilize social participation for developing brand confirmation and empowering brand responsibility. Since it makes an affiliation logically open to new clients and progressively undeniable for existing clients, long-broaden easygoing correspondence impels a brand's voice and substance. For instance, an interminable Twitter client may consider a relationship considering the way that through a news channel and choose to purchase a thing or association. The more acquainted individuals are with an affiliation's image, the more fundamental the affiliation's odds of finding and holding new clients.

1.4 Aim of the Research

This paper aims to analyze and research climate change and development matters more conveniently and stimulate debate between the various constituencies involved in these fields. To estimate the individual and integrated capabilities of a real-time machine learning model to predict climate changes sentiment with the help of social media data. Climate change can cause severe effects on adaptation and mitigation for poverty, equity, or social justice. Understanding the Earth's climate change, and its effects facing challenges have huge importance to society is very important. That is why we have chosen to study climate change so that, this research can propose a model to predict these problems early with the highest accuracy rate for contributing to myself and the climate analyzer

1.5 Propose Solution

The systematic model is based on data mining techniques applying machine learning algorithms that can easily classify climate changes.

CHAPTER 2

Literature Review

2.1 Introduction

Researchers in the data mining sector may use different machine learning algorithms with various practices to predict climate changes, which opens up a study area for them. There have been many studies on sentiment-based climate change classification in recent years. Due to our society's lack of awareness, our irresponsible lifestyle choices, and our incomplete grasp of its impacts, climate change is becoming one of the most serious issues facing modern society. Most people in today's world are unaware of how they are destroying nature and altering the climate. That is why many researchers decided to research climate change. However, predicting climate change accurately is the most challenging task. To tackle this problem, data mining combined with machine learning algorithms plays a crucial part in climate prediction.

2.2 Related Works

N. Mucha et al. [1] create a model Sentiment Study of Shared Tweets on International Heat Causes on Twitter with a Data Mining technique using Machine Learning and Naive Bayesian Approaches. On Twitter, tweets about global warming are gathered with the hashtag #turkish. They employed several Twitter APIs, which are separate from the most popular website used by the majority of users and are an independent media handled by the designers. From this forum, they gathered data in JSON format. The structure contains the upload date, details, text of the tweet contents as well as location data. Then they apply the Naive Bayesian method as the maximum outcome and the exactness rate was 66.9%.

Nabila Mohamad Sham et al.[2] developed a model for sentiment analysis related to climate change using machine learning and hybrid approaches. In this study, they raise 12 keywords to manage discussion threads connected to climate change on Twitter. This study uses data mining templates to aid the design of a model for forecasting climate change using massive data. They extracted tweets from April 01, 2019, to August 31, 2021, by scrape, using 12 keywords as the Harvard Dataverse, they used topic modeling utilizing

the Latent Dirichlet Allocation (LDA) model to summarize the topics that are discussed about climate change. The maximum result with an accuracy of 70.0%

This research used both quantitative and qualitative methods, by Tasha Erina Taufek et al. [3] One of the quantitative methods involved a sentiment analysis method using software called Azure Machine Learning to explore the polarity of the public sentiments on temperature transformation. They used 8 keywords for searching connected to this topic. The highest accurateness figure is 73.1%.

In this study, Mucha, N. et al. [4] used volume research and text mining techniques including topic modeling and sentiment analysis to examine a sizable dataset of geotagged tweets with specific keywords related to climate adaption. They used a dataset made up of 390,016 tweets that were gathered via the Twitter Stream Application Programming Interface between July 1, 2016, and February 28, 2018. (API). This study compares the results of volume research, sentiment analysis, and topic modeling performed on a set of tweets across space and time.

An artificial Neural Network algorithm-based system for climate change was presented by Patronella et al. [5]. supervised machine learning and unsupervised machine learning, the latter of which includes dictionary-based algorithms, are the two main techniques. large number of datasets. Their 60% accuracy rate.

Pudaruth et al [6] Used NVivo 11 for the analysis of the comments. NVivo 11 has a component for the automated tagging of sentiments to text. Sentiments can be coded as relatively positive, very positive, somewhat negative and very negative. NVivo supports individual lexicons for each of these classifications. They used a five-class prototype and machine learning classifiers to predict the personality of 20000 Facebook users. NVivo is a useful software which can be used to analyze unstructured data. The Overall accuracy of their proposed method was found to be 62%

Williams et al. [7] and Qiao, F. used the Latent Dirichlet Allocation (LDA) Model. This analysis uses latent Dirichlet allocation (LDA), topic modeling, and sentiment analysis to look at the subjects and sentiments of the Twitter conversation of global warming. There are two sections to this essay. The first section focuses on the subjects that received the

most attention throughout the study period as measured by LDA, and the second half explores the emotional polarities that people expressed in response to global warming.

Research on climate change and natural disasters is being done by Dzyuban, Y. et al. [8] with a focus on two cities: Phoenix, which has a Köppen-Geiger climate of BWh, and Singapore, which has an Af Köppen-Geiger climate (Kottek et al. 2006). Both cities have hot climates, however their levels of atmospheric moisture vary. They also forecast here how the people will react to actual weather facts.

JSietsmaMSca, A. et al. [9] systematically identified and mapped the scientific papers on weather change and healthiness published between January 1, 2013, and April 9, 2020 using supervised machine learning and other natural language processing approaches (topic modeling and geoparsing). Their findings included automated machine learning for implications on humans and climate change.

Lineman, M. et al. [10] Used tendency and opinion research as one technique whereby experimenters can determine modifications in the general perception that can be utilized to improve the growth of a social consciousness toward exact public welfare. The next study evaluated relative search volume (RSV) designs for global warming (GW) and also climate change (CC) to decide a general understanding and understanding of these words

Rohilla, R. and Zahoor, S. (2020) [18] used supervised machine learning and other natural language processing methods (topic modeling and geoparsing) to systematically determine and map the scientific journals on weather. Their conclusions were automated machine learning for climate change and as well as effects on human beings

Unsupervised Technique, which is a rule-based or lexical approach, was proposed by Yadav, S. et al. in 2022 [16]. It may be carried out using ready-made open-source libraries like TextBlob and VADER.

2.3 Conclusion

All of the analysis was done with the express goal of supporting climate analyzer and social media manager as well as business management sectors. A business is gradually grown up

according to the climate. Because people change their life style, food habit and interest s according to the climate adaption. Identifying and predicting this type of risk is a critical task. Also using this prediction applying in the real life is more challenging.

CHAPTER 3

Research Methodology

3.1 Introduction

There is no limitation to the extent of nuances brought by tweets and messages, consistently these little messages are used to give bits of knowledge and assessments that people have about what's going on in their overall environmental factors. Accordingly, the endeavor and the headway of a twitter incline corpus is made to propel research that will provoke an unrivaled perception of how assessment is passed on in tweets and messages. Working with these easygoing substance classes presents challenges for trademark language taking care of because of the language used is amazingly easygoing, with imaginative spelling and emphasis, off base spellings, the slang of word, new words, URLs, and kind unequivocal expressing and abbreviated structures. Another piece of electronic long range interpersonal communication data, for instance, Twitter messages is that it fuses rich coordinated information about the people drew in with the conversation.

3.2 Research Subject and Instrumentation

I suggested the subject name is “Understanding Climate Change Concern Using Machine Learning Technique”. This is a vital exploration district in NLP. I have assessed the course toward making assessment examination in English with the decided and theoretical technique first to now. A critical learning model necessities high structure pc and various instruments. Straightforwardly an idea analyzer have given under of the essential instrument for this model.

Hardware and Software:

- Intel core i3 7th generation with 8GB RAM.
- Hard Disk: 1 TB

Tools:

- Windows 10

- Numpy
- Python 3.7
- Seaborn
- Sklearn
- Pandas
- Jupyter Notebook
- NLTK

3.3 Data Collection

As a dataset, here utilized the Twitter Climate Change Sentiment Dataset that can be used to detect climate change opinions in a text from Kaggle containing 43,943 user records. The dataset is a collection of posts from the “climate” and “climate Changes” subreddits of the Reddit platform. This dataset aggregates tweets about climate adaptation contained between Apr 27, 2015, and Feb 21, 2018. In total, 43943 tweets were annotated. Per tweet is labeled independently by 3 reviewers. This dataset only includes tweets that all 5 reviewers agreed on.

	sentiment	message	tweetid
0	-1	@tiniebeany climate change is an interesting h...	7.929270e+17
1	1	RT @NatGeoChannel: Watch #BeforeTheFlood right...	7.931240e+17
2	1	Fabulous! Leonardo #DiCaprio's film on #climat...	7.931240e+17
3	1	RT @Mick_Fanning: Just watched this amazing do...	7.931250e+17
4	2	RT @cnalive: Pranita Biswasi, a Lutheran from ...	7.931250e+17

Figure 3.3: Collected dataset sample

3.4 Proposed Model

For any type of research, a structured dataset is needed to understand. In this research, we have collected a dataset from Kaggle. For using this research, we have to process the text and then convert it to numerical value across the individual text.

PROPOSED MODEL

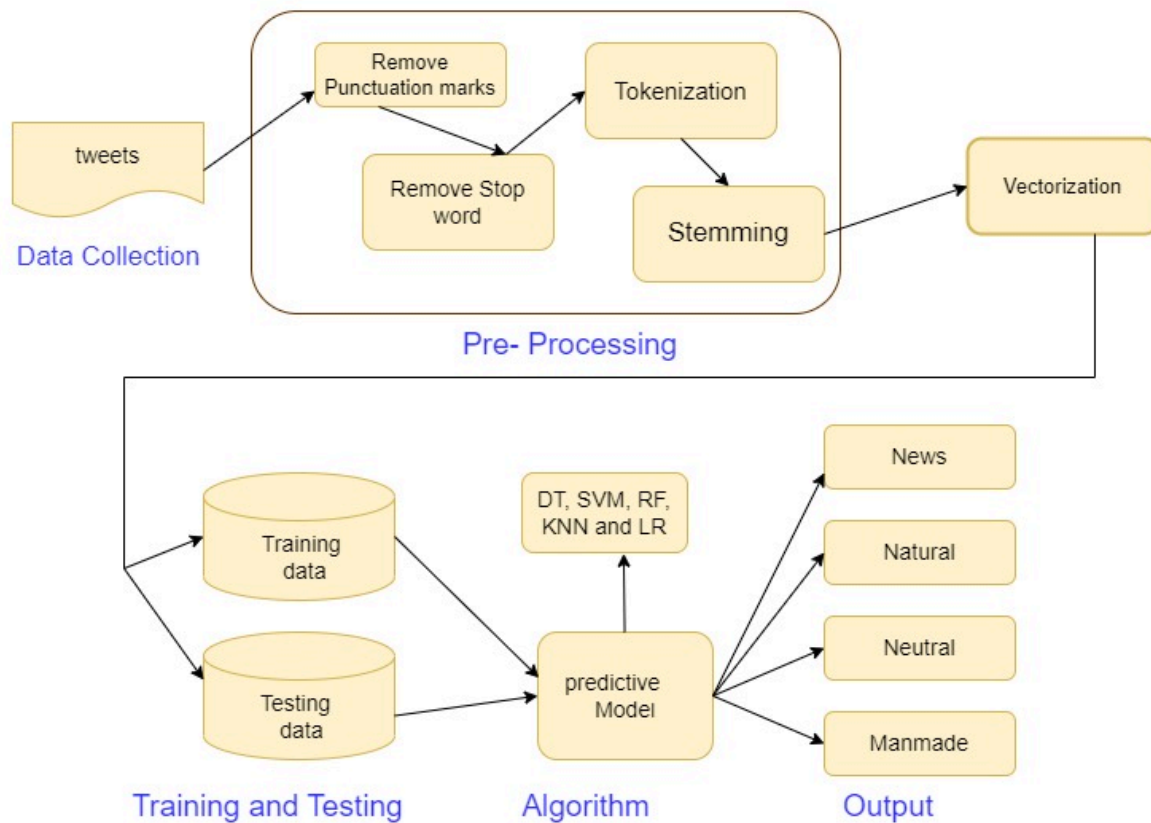


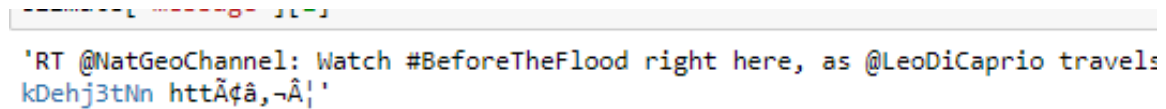
Figure 3.4: Proposed model

The summary of the procedure illustrated in this model also describes the overall idea of how the prediction will occur. Firstly, we collected data from Kaggle. After that, we develop a supervised machine learning which prepares the data and extracts information from the data. In our experimental case, we predict Climate Change. We furthermore use machine learning methods to train the learning. In this method, we predict climate change and then we discover the accuracy ratio. Each part of our proposed method is illustrated in the following section.

3.5 Data Transformation

Pre-processing is a method that is used to convert the raw data into a useful and efficient format before feeding it to the algorithm. To assure the quality of data, it can directly affect

the ability of our model to learn if not processed. Sometimes the reviews may include extreme and insignificant data for research. And hence they require some processing. The following example represents a snippet of a text word that consists of text along with punctuations, stop words, etc. we can take examples from our data. Our raw data was in this format in the figure 3.5.1 has shown. Here we shown one user tweet as an example:



```
'RT @NatGeoChannel: Watch #BeforeTheFlood right here, as @LeoDiCaprio travels kDehj3tNn httÃ¢â, -Ã¡'
```

Figure 3.5.1: Data Transformation

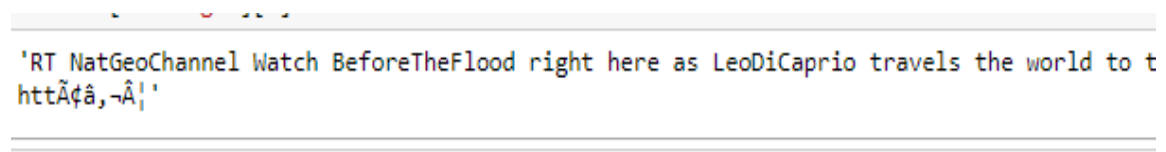
- **Remove Outliers from Dataset**

There have been seen many awkward and extra values for different attributes. It seems unrealistic to me. For gaining more satisfactory performance for the accurateness rate, those values are fired from different attributes of this dataset and assembled the dataset sensible to all by its containing values.

Dropped the column called 'tweetid: because it hasn't any practical role in this research

- **Punctuation Character Removal**

The English language utilizes many punctuation characters in texts which carries a little matter in sentiment polarity. Punctuation varies to get the correct impression from the text. So, a simple script was used to strip all punctuation characters clear from the data see the flowing example in figure below:



```
'RT NatGeoChannel Watch BeforeTheFlood right here as LeoDiCaprio travels the world to httÃ¢â, -Ã¡'
```

Figure 3.5.2: Punctuation Character Removal

- **Tokenization**

Tokenization directs to a procedure by which a portion of sensitive data is. Such as a David card number, which is replaced by a surrogate weight understood as a token. Text

segmentation or Tokenization is the method of separating the composed text into meaningful units, such as words, sentences, or topics. As an example, after applying Stop Words Removal and Tokenization to our dataset we see in the figure below:

```
tiniebeany, climate, change, is, an, interesting, hustle, as,  
➔rt, natgeochannel, watch, beforetheflood, right, here, as, le  
fabulous, leonardo, dicaprio, s, film, on, climate, change, i  
rt, mick_fanning, just, watched, this, amazing, documentary, |  
rt, cnalive, pranita, biswasi, a, lutheran, from, odisha, giv
```

Figure 3.5.3: Tokenization

▪ **Stop Words Removal**

A stop word is a word that seems usually in the dataset despite having no sentiment contradiction associated with it. This can be purified before or after the processing of natural language data. In sentiment analysis, many of the words in English like in the figure below

```
Sample sentence BEFORE removing stop words:  
['rt', 'natgeochannel', 'watch', 'beforetheflood', 'right', 'here', '  
le', 'climate', 'change', 'https', 't', 'co', 'lkdehj3tnn', 'httã', '  
  
Sample sentence AFTER removing stop words:  
['rt', 'natgeochannel', 'watch', 'beforetheflood', 'right', 'leodicap  
'https', 'co', 'lkdehj3tnn', 'httã']
```

Figure 3.5.4: Stop Words Removal

have least significance or do not convey any meaning that are filtered out from here. Which can discard from the input text seen in the flowing example. As the overall polarity of a study does not depend on those words.

▪ **Steaming**

Stemming is the method of creating morphological variants of a root/base word. Stemming programs are generally directed to as stemming algorithms or stemmers.

There are two types of steaming algorithm are there:

1. Porter stemming algorithm

2. Lancaster stemming algorithm

After using porter steaming algorithm figure 3.5.4

```
Sample sentence BEFORE stemming:
['rt', 'natgeochannel', 'watch', 'beforetheflood', 'right', 'leodicaprio',
'https', 'co', 'lkdehj3tnn', 'httã']

Sample sentence AFTER stemming:
rt natgeochannel watch beforetheflood right leodicaprio travel world tackl
```

Figure 3.5.5: Steaming

▪ Vectorization

In Machine Learning, vectorization is a term in element extraction. The concept is to obtain some distinct features out of the text for the model to train on, by transforming or converting text to numerical vectors. Let's look into 1 sample sentences to understand better what vectorization does this sentence has a couple of words in common - "climat", "chang"

```
#1 after vectorization:
(0, 9419)      1
(0, 12943)    1
(0, 13774)    1
(0, 14222)    1
(0, 27583)    1
(0, 27591)    1
(0, 35060)    1
(0, 35653)    1
(0, 40355)    1
(0, 49581)    1
(0, 50386)    1
```

Figure 3.5.6: Vectorization

Based on the column size of our vectorized data, we can notice there were 43943 tweets in the dataset and 67195 individual unique words.

3.6 Applied Mechanism

To implement a system for research, a qualitative dataset is needed. In prediction-based analysis, there are some independent factors and a target element. The target factor is dependent on independent elements. Independent factors should be a connection with the target factor. In this section, we will describe the visualization of various factors and also will describe the implementation of the procedure as methodology. Actually, from this chapter we can know the theoretical background of following terms which are used in this

3.6.1 Classification Algorithms

The classification algorithms are the supervised understanding method which will be utilized to determine the classification of new statements on the grounds of data which data to be trained. In classification, a program understands the provided dataset and then classifies the new statements into different categories or groups. Classification is the procedure of software understanding a dataset or observations and then classifying new statements into one of multiple categories or classes. In this study, I operated four types of classification algorithms among many machine learning algorithms founded on the previous records of their performance to get satisfactory results. They are the Logistic Regression (LR), and then Decision Tree Classifier, as well as Random Forest (RF), also the Support Vector Machine (SVM) and K-Neighbors Classifiers. The details of them are shown individually below. [18]

- **Logistic Regression**

A statistical method for describing a binary conditional variable in its simplest form using a logistic process. There are many developed versions, though. Regression analysis uses a technique called logistic regression to calculate the parameters of a logistic model (a form of binary regression).

Here they perfectly extract the complex sentences also semantic weight of every words, and the content of every tweet data. For that matter of their performance, they explore the text-based complex contents which are the tweets related to suicidal attempts.

Calculating probabilities utilizing a logistic regression equation is utilized in statistical software to understand clearly and perfectly where is the connection in the dependent variable as well as the additional distinct variables. This condition of analysis can help you in predicting the probabilities of a circumstance or a decision arising.

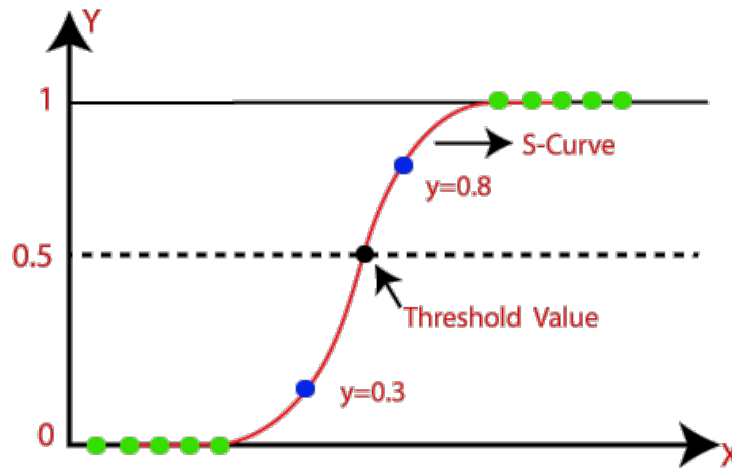


Figure 3.6.1: Logistic Regression [20]

An unconditional variable's outcome is indicated by operating logistic regression. And as an outcome, the result must be a discrete or absolute weight. It can be No or Yes, 1 or 0, False or True, and can be so on, but rather than giving detailed values it produces probability values. Instead of fitting a regression line, we suit a logistic operation in logistic regression, which predicts the two highest weights like (0 or 1). Finally, we can tell that this is a classification method that leverages the concept of predictive modeling as regression.

- **Decision Tree Classifier**

The decision tree classifier divides the activity set recursively until each segment contains only or mostly samples from one class. It is a class differentiator. The decision Tree algorithm is a close relative of the algorithms for supervised learning. The decision tree algorithm is a supervised learning technique that can be used to solve regression and classification problems. By understanding simple determination rules deduced from initial data, a Decision Tree is used to create a training model that may be used to identify the

class or value of the mark variable. In decision trees, we begin at the tree's base when assigning a class label to a record. The values of the root detail are then compared to the feature of the record. We then follow the branch associated with that weight and move on to the next node depending on the comparison. [17]

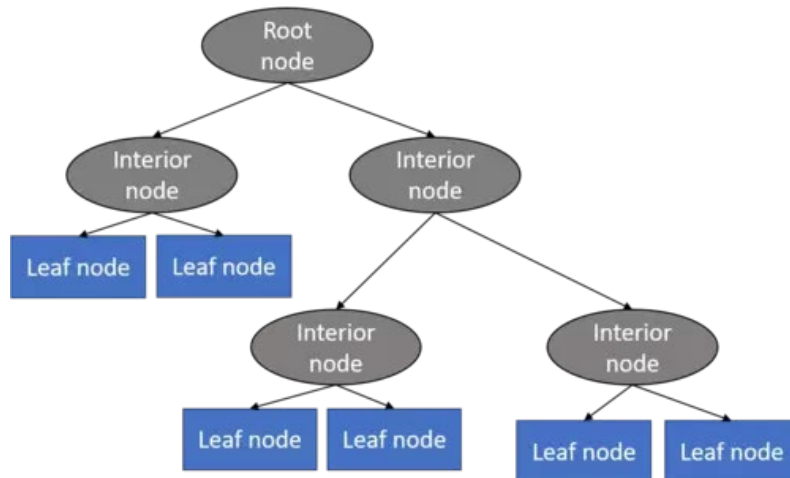


Figure 3.6.2: Decision Tree Classifier [22]

In the flowchart, the design of the internal nodes shows the difficulties or properties at a particular class or level. Every branch denotes a particular output, whereas the path from the leaf to the root defines categorization criteria. The most significant part of the learning algorithm based on numerous learning approaches is decision trees. They have improved the accuracy, stability, and readability of prediction models.

- **Concept of Random Forest (RF)**

Random forest (RF) is a powerful machine learning method developed by Leo Breiman and Adele Cutler that combines the outcome of numerous decision trees for creating a single outcome. Random forest is mainly a supervised easy-going technique and learning algorithm that is utilized for both classifications as well as regression. Random Forest is more effective and better than a single decision tree. The hidden reason is it reduces the over-fitting by averaging the result. We will be able to understand the functional stages of the Random Forest algorithm also the assistance of the subsequent stages given below:

First, let's begin by selecting randomly from a pre-provided dataset. Then, this RF algorithm will construct a decision tree for every sample. Then it will obtain the projection output from every decision tree. After that, voting will be conducted for every expected outcome. Lastly, it selects the tallest-voted projection outcome as the ultimate projection outcome.

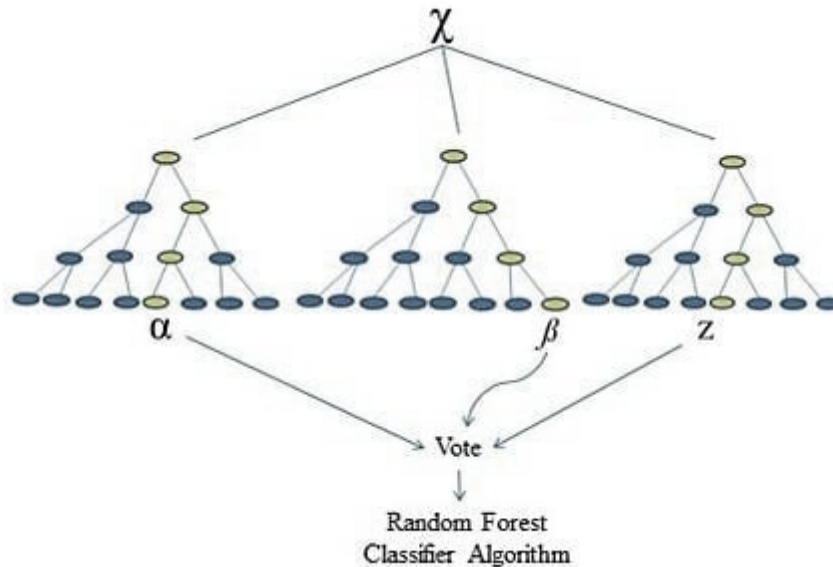


Figure 3.6.3: Concept of Random Forest [23]

The (RF) Random Forest algorithms are formed of a set of decision trees, and each tree in the ensemble is made up of a bootstrap model, which is a data model obtained from a training set with replacement. Almost One-third of the training selected sample is placed aside as test data, directed to the out-of-bag instance, which we'll consult later. Utilizing attribute bagging, the other instance of the random values needs to be injected into the dataset, enhancing the dataset's variety and reducing the correlation between decision trees. The forecast will be specified differently depending on the type of hardship. [17]

- **Support Vector Machine (SVM)**

The support vector machine is a supervised-based machine learning algorithm. Which can be utilized for meeting both classification and regression difficulties correctly. This is especially utilized in decoding classification difficulties. Through this support vector machine algorithm, we can plot each data item as a point in an n-dimensional margin by

using the weight of every element existing with certain lines. After that, we conduct the category by discovering the correct hyperplane that distinguishes the 2 styles of categories nicely. [17]

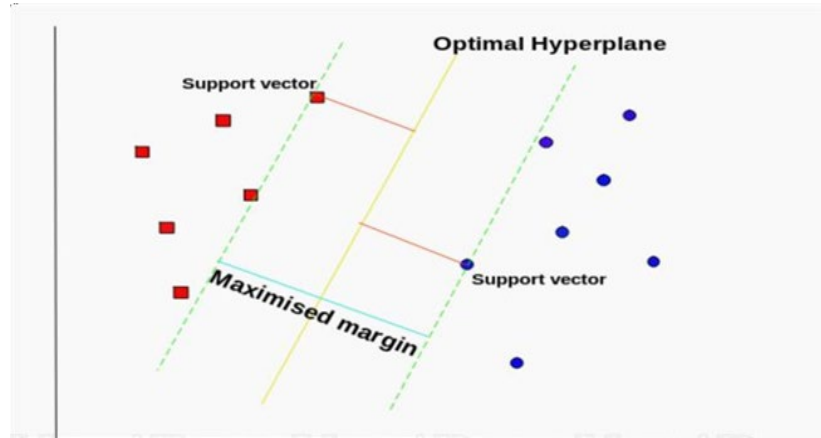


Figure 3.6.4: support vector machine (SVM) [21]

- **K-Nearest Neighbor (KNN)**

One of the simplest machine learning algorithms, based on the supervised learning methodology, is K-Nearest Neighbor. The K-NN algorithm takes into account how similar the new case/data is to unconstrained cases and classifies the most recent case into those categories. The k-NN technique provides all the data that is readily available and categorizes a new data point using the parallel. This shows that when new data is present, it can typically be classified using the K- NN algorithm into a suitable class. Although the K-NN technique can be used for both classification and regression, its primary application is for classification problems. Since K-NN is a non-parametric technique, it makes no assumptions about the underlying data. In the figure 3.6.5. [17]



Figure 3.6.5: K-Nearest Neighbor (KNN) [25]

3.7 Deep learning

Deep learning is called a subset of machine learning, which is a neural network with three or sometimes more extra layers. These neural networks attempt to affect or affect the conduct of the mortal brain albeit distant from reaching its capacity to allow it to “retain” enormous quantities of data. Deep learning doesn't indicate the machine learns more in-depth understanding. It means a machine utilizes various layers to understand the data. In deep learning, the learning step is done via a neural network. Where a neural network is an architecture where the layers are piled on top of each other. Most of the deep learning techniques utilize neural network architectures. For that reason, most deep learning pinnacles are usually directed to as deep neural networks. The phrase “deep” mainly directs to the number of hidden layers in the neural network. Standard neural networks just hold two to three secret layers, whereas deep networks can have as numerous as 150. Deep learning instances are prepared by utilizing extensive collections or groups of marked data and neural network architectures that understand elements straight from the data without the requirement for manual component extraction in the figure 3.7 has shown. [19]

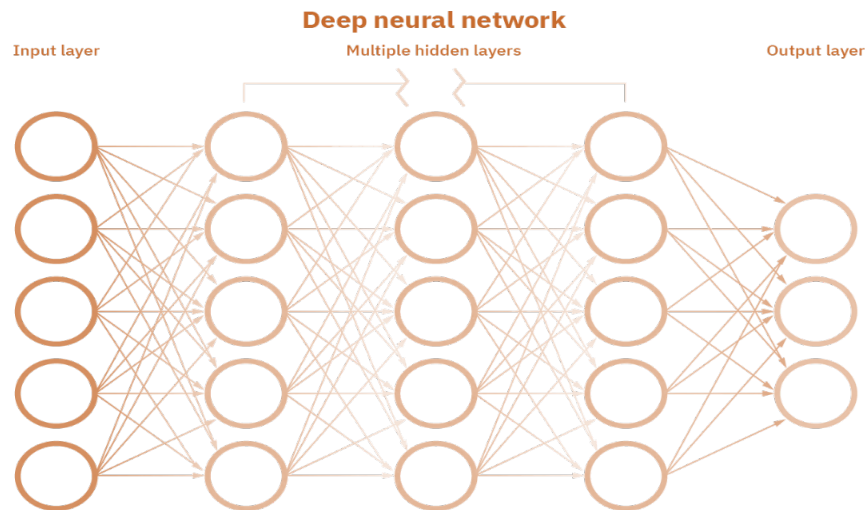


Figure 3.7: Concept of deep learning [24]

3.8 Implementation Requirements

For the implementation of algorithms, I used python Language as a programming language. Stacking the dataset pandas library is utilized and NLTK library for preprocessing. Additionally utilized sklearn library for those three machine learning algorithms. Absolute usage is work in a jupyter notebook.

CHAPTER 4

RESULT & DISCUSSION

4.1 Introduction

In this chapter, the process of finding out the ultimate result of this study will be discussed. For that cause, some measures (preprocessing & declaration) need to debate for using ML algorithms. After applying ML algorithms, the implementation measured parameters will be analyzed like accuracy, recall, F-Score, confusion matrix, ROC curve, and many more for estimating the best outcome of this study.

4.2 Applied Machine Learning Algorithms

Five ML algorithms are applied based on the previous research performance for classifying the target class (Suicide) to acquire a more sensible accuracy rate for the classification. The classification algorithms are Artificial Neural Network (ANN), Naive Bayes, Decision Tree Classifier, Support Vector Machine, Random Forest, and Logistic Regression.

4.3 Preprocessing & Declaration

Here, to apply classification ML algorithms, percentage division is used as a data mining method. The percentage split is a resampling method in which $n\%$ of the rows are set aside as the training dataset for building the model and $(n-100)\%$ of the rows are set aside as the trial dataset for testing the model. In contrast to the learned data, the target classifier is trained. The classification accuracy, on the other hand, is assessed using the trial dataset.

In this research we set the percentage split, 80% of the rows are utilized as the training dataset for constructing the model, and the remaining 20% as the trial dataset for testing the model purpose.

4.4 Data Visualization

Data visualization is an essential preprocessing job, which operated a graphical model to simplify and understand easily the overall status of complex data. Visualization methods have been newly used to visualize online learning factors. Instructors can use graphical

presentations to understand their learners nicely and become conscious of what is happening in distance lessons. This study visualizes the existing data set utilizing the Anaconda Navigator tool. As illustrated in Figure 4.1, the data set is pictured based on sentiment into 282 positive, 72 negative, and 40 neutral reviews. To understand the dataset for machine learning objectives, we have left as Natural as -1, Neutral as 0, Manmade as 1, and News as 2. Here, 80 % data was utilized for training, and 20% of the data was used for testing for every model this work

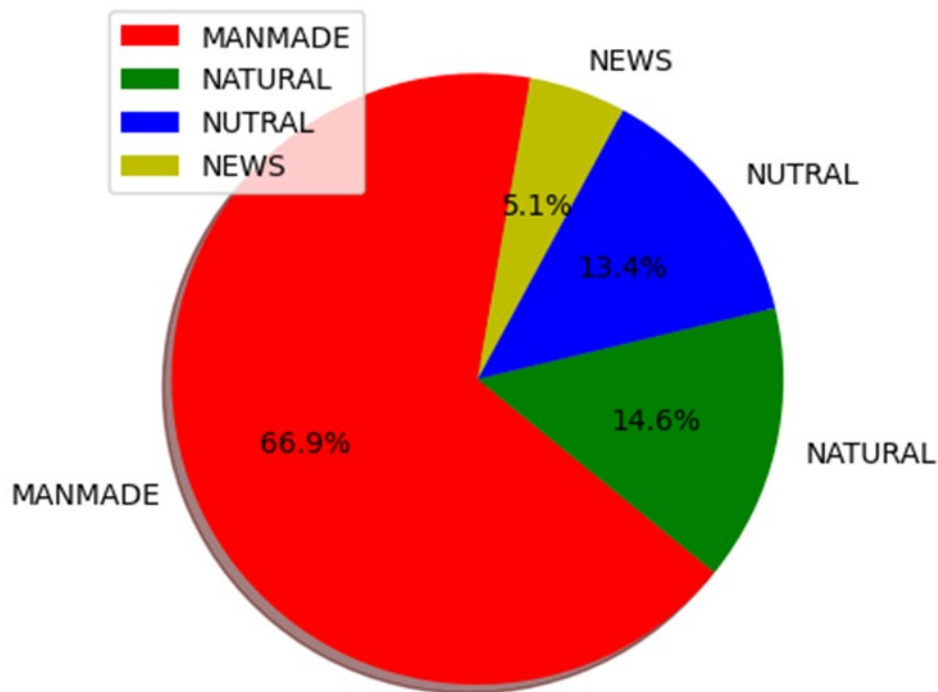


Figure 4.4: Data Visualization

4.5 Analysis of Performance Measured used

Used For measuring the performance applying to classification algorithms (Logistic Regression (LR), Decision Tree Classifier, Random Forest, and Support Vector Machine (SVM) and Ridge Classifiers (RC) of the declared model, performance measured parameters (accuracy, precision, recall, F-Score, confusion matrix) are used. The results for these performance-measured parameters are displayed below.

- **Accuracy Rate of Classification**

Here in this study, the most significant performance calculated parameter is the accuracy

rate. Truthfully, it is estimated by splitting the number of precisely categorized sample models by the entire number of samples multiplied by 100. The sum of True-Positive (TP) and True-Negative (TN) is the exact classified sample

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100$$

The obtained accuracy rates of these classification algorithms are given below:

Applied ML Algorithms	Training Accuracy	Test Accuracy
Logistic Regression (LR)	86.41%	82.59%
Random Forest (RF)	99.70%	82.79%
Decision Tree Classifier (DT)	99.70%	78.43%
Support Vector Machine (SVM)	88.58%	83.01%
KNN Classifiers	61.68%	57.33%

Table 4.5.1: Accuracy rate of classification

According to the table 5.3.1.1, it can declare that Logistic Regression Classifier got the highest accuracy rate for this proposed model.

- **Precision**

Precision is a performance metric utilized for pattern recognition and classification processes in machine learning. Precision is one of the measurements of a machine learning model's interpretation. The accuracy of a model's positive prediction. Precision is defined as the percentage of true-positive examples to predicted yes data, according to the Confusion Matrix

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

The obtained precision rates of these classification algorithms are given below:

Applied ML Algorithms	Precision
Logistic Regression (LR)	75%
Random Forest (RF)	83%
Decision Tree Classifier (DT)	79%
Support Vector Machine (SVM)	83%
KNN Classifiers	85%

Table 4.5.2: Precision rate of classification

- **Recall**

Recall is a parameter that considers how nicely a standard can detect positive models. Recall is another name for sensitiveness. Recall is described as the ratio of true-positive models to genuine yes samples, according to the Confusion Matrix

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

The obtained recall rates of these classification algorithms are given below:

Applied ML Algorithms	Recall
Logistic Regression (LR)	83%
Random Forest (RF)	83%
Decision Tree Classifier (DT)	78%
Support Vector Machine (SVM)	83%
KNN Classifiers	57%

Table 4.5.3: Recall rate of classification

- **F1-Score**

The F-Score is also understood as the F1-Score or the F-Measure. Using both recall and accuracy, the F-Score can supply a better useful measure of test performance. When the F-

Score score reaches 1, it means that both recall and accuracy are ideal

$$\text{F1 score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$

The obtained F1 scores of these classification algorithms are given below in the table 4.5.4:

Applied ML Algorithms	F1 Score
Logistic Regression (LR)	81%
Random Forest (RF)	82%
Decision Tree Classifier (DT)	78%
Support Vector Machine (SVM)	82%
KNN Classifiers	63%

Table 4.5.4: F1 Score of classification

- **Confusion Matrix**

The Confusion Matrix is a tool for displaying a model's performance or how a model generated a prediction in Machine Learning. The Confusion Matrix allows us to see where our model becomes confused while deciding between two classes. A 2×2 matrix may be used to visualize it, with the row representing the actual truth labels and the column representing the prediction labels

	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Table 4.5.5: Basic Diagram of a Confusion Matrix

Here, the Confusion Matrix is plotted below for all four ML algorithms. Hope, it will easy to understand after viewing the figure 4.6.1.

- **Confusion Matrix and Classification report for Logistic Regression**

```

Confusion matrix:
[[ 32  6 27  2]
 [  0 77 114 10]
 [  0 10 945 25]
 [  0  5  56 156]]

```

	precision	recall	f1-score	support
-1	1.00	0.48	0.65	67
0	0.79	0.38	0.52	201
1	0.83	0.96	0.89	980
2	0.81	0.72	0.76	217
accuracy			0.83	1465
macro avg	0.86	0.64	0.70	1465
weighted avg	0.83	0.83	0.81	1465

Figure 4.5.1: Confusion matrix of Logistic Regression

- **Confusion Matrix and Classification report for Decision Tree Classifier**

```

Confusion matrix:
[[ 38 13 14  2]
 [  4 113 73 11]
 [ 20  72 847 41]
 [  2 16  48 151]]

```

	precision	recall	f1-score	support
-1	0.59	0.57	0.58	67
0	0.53	0.56	0.54	201
1	0.86	0.86	0.86	980
2	0.74	0.70	0.72	217
accuracy			0.78	1465
macro avg	0.68	0.67	0.68	1465
weighted avg	0.79	0.78	0.78	1465

Figure 4.5.2: Confusion matrix of Decision Tree Classifier

- **Confusion Matrix and Classification report for Support Vector**

```

Confusion matrix:
[[ 33  5 27  2]
 [  0 87 98 16]
 [  3 19 927 31]
 [  0  7 41 169]]

```

	precision	recall	f1-score	support
-1	0.92	0.49	0.64	67
0	0.74	0.43	0.55	201
1	0.85	0.95	0.89	980
2	0.78	0.78	0.78	217
accuracy			0.83	1465
macro avg	0.82	0.66	0.71	1465
weighted avg	0.83	0.83	0.82	1465

Figure 4.5.3: Confusion matrix of Support vector Machine

- **Confusion Matrix and Classification report for Random Forest Classifier**

```

Confusion matrix:
[[ 34  8 25  0]
 [  1 99 97  4]
 [  2 34 931 13]
 [  0  2 66 149]]

```

	precision	recall	f1-score	support
-1	0.92	0.51	0.65	67
0	0.69	0.49	0.58	201
1	0.83	0.95	0.89	980
2	0.90	0.69	0.78	217
accuracy			0.83	1465
macro avg	0.84	0.66	0.72	1465
weighted avg	0.83	0.83	0.82	1465

Figure 4.5.4: Confusion matrix of Random Forest Classifier

- **Confusion Matrix and Classification report for K-Nearest Neighbors**

Confusion matrix:

```

[[ 30  37  0  0]
 [ 2 185  9  5]
 [ 1 455 518  6]
 [ 0  98 12 107]]

```

	precision	recall	f1-score	support
-1	0.91	0.45	0.60	67
0	0.24	0.92	0.38	201
1	0.96	0.53	0.68	980
2	0.91	0.49	0.64	217
accuracy			0.57	1465
macro avg	0.75	0.60	0.57	1465
weighted avg	0.85	0.57	0.63	1465

Figure 4.5.5: Confusion matrix of K-Nearest Neighbors

▪ **Final Result and analysis chart**

In this figure below accuracy and precision, recall and f1-score results of the best four machine learning techniques (DT, SVM, RF, KNN and LR) is given below:

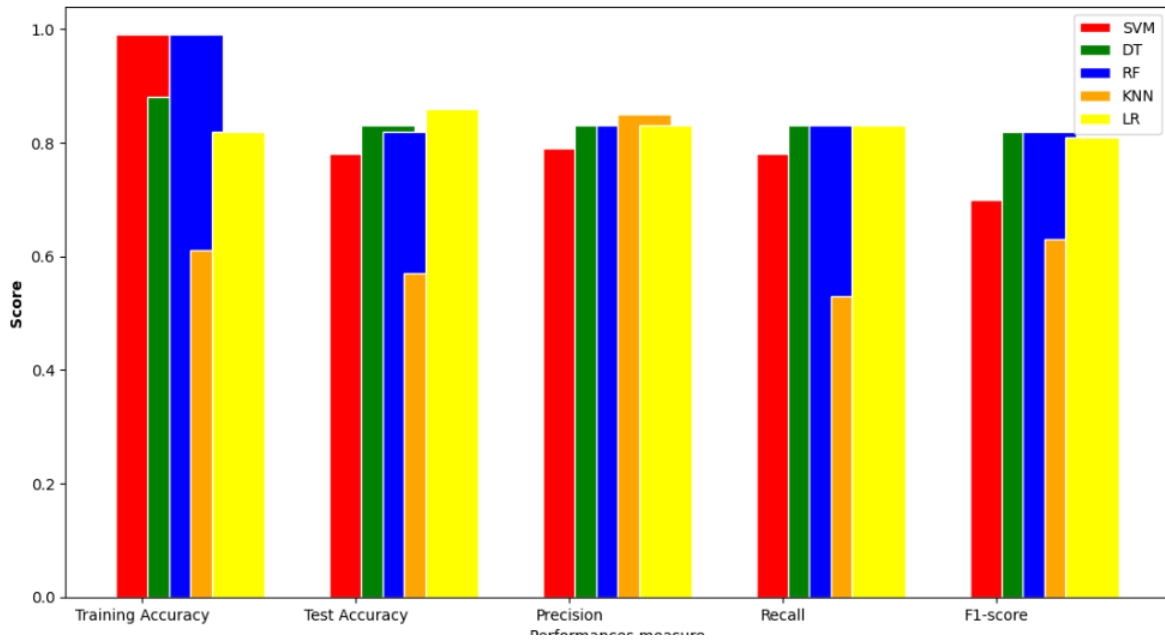


Figure 4.5.6: Comparison of Accuracy Result among DT, SVM, RF, KNN and LR

Using long short-term memory is training and test accuracy table: 4.5.6 is given below:

Measure	DT	SVM	RF	KNN	LR
Training Accuracy	99.70%	88.58 %	99.70 %	61.68%	82.88%
Test Accuracy	78.43%	83.01 %	82.79 %	57.33%	76.84%

Table 4.5.6: Comparison of Accuracy among DT, SVM, RF, KNN and LR

In Figure: 4.5.1: and Table 4.5.6: is the comparison of accuracy result among DT, SVM, RF, KNN and LR. Where SVM scored higher.

4.6 Limitations

In our investigation, we employed only 43,943 tweet data for content. This works, a small dataset uses this model, so accuracy and F1 score are not enough. It cannot determine exact reasons, such as, whether the climate change tweet contain how much positive and negative concern contains, etc. Other hand Support Vector machine, requires large data sets and is tremendously costly to train. These complex models can obtain weeks to train by using machines equipped with costly GPUs. Utilizing SVM for sentence classification is an interesting approach when one is working with short sentences, but as the size of sentences grows, recurrent networks should be a more adequate method.

4.7 Conclusion

In this chapter, it is marked that the Decision Tree Classifier & Random Forest both executed fantastically in almost all sectors (Accuracy, Precision, Recall, F-Score, Confusion Matrix & final chart curve). But, in the measure of precision rate, Random Forest functioned well than Decision Tree Classifier and acquired the tallest accuracy rate of 83.01% on this clustering dataset which is included in the experimental model chapter. Overall, the performance & result of this model fulfilled the requirements of this study.

CHAPTER 5

Impact on Society, Environment and SWOT Analysis

5.1 Impact on Society

Online life has a powerful influence on society, both positively and negatively. It provides people with a means of managing to stay in touch with others who are located far away. It enables people to share entertaining, captivating, and educational content. It provides associations with a method to control client attraction. Anyhow, one of the problems is that anyone can share anything, including information that might not be accurate. Genuine mischief is occasionally committed when people disseminate scathing, unverified, or outright false information. Private individuals may suffer as a result, such as when they are harassed online. It may also negatively affect society.

Data may now be spread quickly as a result of long reach relational correspondence. Things on Facebook and Twitter change so often that viewers rarely revisit what they've already seen. Additionally, a lot of information is disseminated via images, which can be based on large amounts of data. Different images are created to be enticing, seductive, or extravagant. In every case, we need the influence of others to persuade us. Undoubtedly, relationships with endorsed articles can also be misleading. It is made clear to convey that a significant portion of those who view a component or affiliate never read the complete article.

Fake news is typically spread in many places as false information. To acquire clicks and sell all of the further developing products, companies require people to tap on beautiful features. There are other goals that have a strong philosophical bent. These might be insanely leftist, plan, or preservationist areas. Many of these goals don't bother to spread "news" articles that support their viewpoint.

5.2 Impact on Environment

Whatever way you look at it, spreading false information through technological systems and media is what hurts people. If you own a website or business that provides a lot of

content, maybe with the aid of online life planning, you need exercise particular caution. To verify anything you observe via electronic framework organization media just requires a brief break or two. Consider the source of the narrative. Whether you've never thought about the broader picture, look it up on Google to see if it's appealing. It's best to ignore it if you don't have enough time, especially if it comes after anything that may be a satire, seductive substance, or conscious openness. By not spreading inaccurate information, you may help stop the spread of deceit and false news.

The ability allow information to reach innumerable people in an improbable amount of time is one of the fundamentally acceptable criteria of more than one electronic existence. Despite the fact that this can be seen as a huge advantage in a crisis situation, it can also be a huge barrier since unverified information can be spread quickly. This may result in out-of-control situations and certified deception.

5.3 SWOT Analysis

SWOT analysis is a strategic practice and strategic management technique that may be utilized to help an individual or institution discover its analysis planning powers, drawbacks, possibilities, and dangers. It's also understood as system analysis or situational evaluation.

- **Strength**

In this study, the primary strong point is the goal of this research outcome. Honestly, this analysis is founded on climate prediction opinions based on the concern for human beings. The suggested predictive measure benefits overcome this issue by detecting climate sentiment issues. This study also benefits social media engineers to use machine learning to predict the climate change analyzing people's opinions. This research also builds a connection between the environmental scientist sector and the Data Mining field focused on people concern on climate where the earth is pushing to rely on computer science day by day. In my opinion, it is the strength of my research work.

- **Drawback**

If I address the deficiency of this study work, then it will be the type of data that has been used finally after categorizing it as a dataset. The ultimately utilized dataset for creating this model is categorical. This instance can be executed when the input data would be separated into many categories for the features utilized in the dataset. In my point of view, it may be a drawback. But the most remarkable point is that as data mining researchers, on today's planet people can categorize any data like numerical or, distinct data. So, it cannot be a primary issue behind this study.

- **Opportunity**

This study has tremendous possibilities and opportunities. As today every people connected to social media, and they express their impressions. With this predictive model, social media engineers and the environmental scientist can use machine learning to predict the changes of climate and its effect on the society. Communities may provide additional helpful and practical advice to the users and estimate the marketing campaign. Also utilizing this predictive model, environmental scientist or, social media engineers can predict the concern about climate change and cause of changes, and the analyzer will be able to find out the exact number of people who are taking climate changes positively or negatively. In my point of view, it will be a wonderful option for both environmental science and data mining researchers as well as social media engineers.

- **Threat**

In my point of view, the most challenging threat is accurately detecting and predicting taking people concern. Because according to this model the highest accuracy people's sentiment is 83.01%. So, there has a 24.04 % probability of delivering the wrong detection to people. But it can be enriched or improved by further research shortly.

6.3 Conclusion

After the SWOT analysis of my study work, I can confidently state that this developed predictive sample can play a tremendous role in both the environment sector and social sector as well as the social commercial sector also.

CHAPTER 6

Conclusion and Future work

6.1 Conclusion

In this analysis, an innovative and creative ensemble approach is shown by integrating data mining methods. Evaluating classification approaches such as Logistic Regression (LR), Decision Tree Classifier (DT), Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) to define which is the most useful at accurately predicting climate changes. The absolute outcome illustrates that Support Vector Machine acquired the most increased accuracy rate of 83.01% founded on this proposed predictive model.

The accuracy, precision, recall, F-Score, and confusion matrix are among the six metrics utilized to evaluate the proposed model. The presented model's interpretation was compared to that of different existing models. Based on the results of the experimentation, we can finish that the proposed method improves climate changes accuracy. The proposed model was designed using Jupyter Notebook (a Python-based IDE) and trained to utilize Kaggle's standard Climate Change Sentiment Dataset, which has 43943 unique records.

In today's society, defining the reason for a problem like climate changes is critical. Estimating the climate changes the communities may provide additional helpful and practical advice to the users and estimate the marketing campaign.

6.2 Future Work

In the future, analysis can be performed to enhance prediction accurateness by integrating various algorithms. Besides, it will be able to concentrating on enhancing classification accurateness and will need to find out the most useful data mining technique utilizing several machine learning algorithms. To do so, our data set may be experimented in a variety of ways. For further research, more well-known suicide datasets may be selected to use. It also decided that it would like to utilize our research methods not just in the field of only climate change, but also in the field of other sentiment perspectives.

REFERENCE

- [1] N. Mucha, "Sentiment Analysis of Global Warming Using Twitter Data," library.ndsu.edu, 2018, Accessed: Dec. 07, 2022. [Online]. Available: <https://library.ndsu.edu/ir/handle/10365/28166>
- [2] Z. Shangguan, Z. Zheng, and L. Lin, "Trend and Thoughts: Understanding Climate Change Concern using Machine Learning and Social Media Data," arXiv:2111.14929 [cs], Nov. 2021, [Online]. Available: <https://arxiv.org/abs/2111.14929>
- [3] T. E. Taufek, N. F. M. Nor, A. Jaludin, S. Tiun, and L. K. Choy, "Public Perceptions on Climate Change: A Sentiment Analysis Approach," GEMA Online® Journal of Language Studies, 2021, doi: 10.17576/gema-2021-2104-11.
- [4] A. M. Patronella, "Covering Climate Change: A Sentiment Analysis of Major Newspaper Articles from 2010 - 2020," Inquiries Journal, vol. 13, no. 09, 2021, [Online]. Available: <http://www.inquiriesjournal.com/articles/1910/covering-climate-change-a-sentiment-analysis-of-major-newspaper-articles-from-2010--2020>
- [5] D. Tang, B. Qin, and T. Liu, "Deep learning for sentiment analysis: successful approaches and future challenges," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 5, no. 6, pp. 292–303, Oct. 2015, doi: 10.1002/widm.1171.
- [6] F. Qiao and J. Williams, "Topic Modelling and Sentiment Analysis of Global Warming Tweets," Journal of Organizational and End User Computing, vol. 34, no. 3, pp. 1–18, May 2022, doi: 10.4018/joeuc.294901.
- [7] Y. Dzyuban et al., "Sentiment Analysis of Weather-Related Tweets from Cities within Hot Climates," Weather, Climate, and Society, vol. 14, no. 4, pp. 1133–1145, Oct. 2022, doi: 10.1175/WCAS-D-21-0159.1.
- [8] L. Berrang-Ford et al., "Systematic mapping of global research on climate and health: a machine learning review," The Lancet Planetary Health, vol. 5, no. 8, pp. e514–e525, Aug. 2021, doi: 10.1016/s2542-5196(21)00179-0.
- [9] M. Lineman, Y. Do, J. Y. Kim, and G.-J. Joo, "Talking about Climate Change and Global Warming," PLOS ONE, vol. 10, no. 9, p. e0138996, Sep. 2015, doi: 10.1371/journal.pone.0138996.
- [10] N. Mohamad Sham and A. Mohamed, "Climate Change Sentiment Analysis Using Lexicon, Machine Learning and Hybrid Approaches," Sustainability, vol. 14, no. 8, p. 4723, Apr. 2022, doi: 10.3390/su14084723.
- [11] M. L. Loureiro and M. Alló, "Sensing climate change and energy issues: Sentiment and emotion analysis with social media in the U.K. and Spain," Energy Policy, vol. 143, p. 111490, Aug. 2020, doi: 10.1016/j.enpol.2020.111490.
- [12] M. L. Loureiro, M. Alló, and P. Coello, "Hot in Twitter: Assessing the emotional impacts of wildfires with sentiment analysis," Ecological Economics, vol. 200, p. 107502, Oct. 2022, doi: 10.1016/j.ecolecon.2022.107502.
- [13] H. T. P. Williams, J. R. McMurray, T. Kurz, and F. Hugo Lambert, "Network analysis reveals open forums and echo chambers in social media discussions of climate change," Global Environmental Change, vol. 32, pp. 126–138, May 2015, doi: 10.1016/j.gloenvcha.2015.03.006.

- [14] G. A. Ruz, P. A. Henríquez, and A. Mascareño, "Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers," *Future Generation Computer Systems*, vol. 106, pp. 92–104, May 2020, doi: 10.1016/j.future.2020.01.005.
- [15] S. Yadav, A. Kaushik, M. Sharma, and S. Sharma, "Disruptive Technologies in Smart Farming: An Expanded View with Sentiment Analysis," *AgriEngineering*, vol. 4, no. 2, pp. 424–460, May 2022, doi: 10.3390/agriengineering4020029.
- [16] S. Zahoor and R. Rohilla, "Twitter Sentiment Analysis Using Lexical or Rule Based Approach: A Case Study," *IEEE Xplore*, Jun. 01, 2020. <https://ieeexplore.ieee.org/abstract/document/9197910> (accessed Dec. 10, 2020).
- [17]"Data Mining vs Machine Learning - Javatpoint," www.javatpoint.com.
<https://www.javatpoint.com/data-mining-vs-machine-learning> (accessed Jan. 12, 2023).
- [18]"Classification Algorithm in Machine Learning - Javatpoint," www.javatpoint.com.
<https://www.javatpoint.com/classification-algorithm-in-machine-learning> (accessed Jan. 12, 2023).
- [19]"What is Deep Learning? | IBM," www.ibm.com. <https://www.ibm.com/topics/deep-learning#:~:text=the%20next%20step->
- [20]"Logistic Regression in Machine Learning - Javatpoint," www.javatpoint.com.
<https://www.javatpoint.com/logistic-regression-in-machine-learning>
- [21]javatpoint, "Support Vector Machine (SVM) Algorithm - Javatpoint," www.javatpoint.com.
<https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>
- [22]javaTpoint, "Machine Learning Decision Tree Classification Algorithm - Javatpoint," www.javatpoint.com. <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>
- [23]"Machine Learning Random Forest Algorithm - Javatpoint," www.javatpoint.com.
<https://www.javatpoint.com/machine-learning-random-forest-algorithm>
- [24]B. Butterfly, "Optical Neural Networks: The Future of Deep Learning?," *FindLight Blog*, Oct. 26, 2022. <https://www.findlight.net/blog/optical-neural-networks-the-future-of-deep-learning/> (accessed Jan. 16, 2023).
- [25]JavaTpoint, "K-Nearest Neighbor(KNN) Algorithm for Machine Learning - Javatpoint," www.javatpoint.com. <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>

ORIGINALITY REPORT

18%	10%	8%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	www.researchgate.net Internet Source	2%
3	Submitted to Liverpool John Moores University Student Paper	1%
4	Submitted to University of Northumbria at Newcastle Student Paper	1%
5	Submitted to Coventry University Student Paper	1%
6	Submitted to National Institute of Technology, Kurukshetra Student Paper	1%
7	academic-accelerator.com Internet Source	1%
8	Fang Qiao, Jago Williams. "Topic Modelling and Sentiment Analysis of Global Warming"	1%