Report On

Prediction of Suicidal Sentiment

Submitted in partial fulfillment of the requirements of the Course project in Semester VIII of Final Year Computer Engineering

by
Mrunmai S. Kore (Roll No. 60)
Ishika K. Koytekar (Roll No. 61)
Bhakti P. Mhatre (Roll No. 64)

Mentor
Dr. Tatwadarshi P. N.

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

(A.Y. 2021-22)

Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

CERTIFICATE

This is to certify that the Mini Project entitled **"Prediction of Suicidal Sentiment"** is a bonafide work of **Mrunmai S. Kore(Roll No. 60), Ishika K. Koytekar(Roll No. 61) and Bhatre P. Mhatre(Roll No. 64)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in Semester VIII of Final Year **"Computer Engineering" .**

_____
Dr. Tatwadarshi P. N.
Mentor

_____
Dr Megha Trivedi
Head of Department

_____
Dr. H.V. Vankudre
Principal

Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

Course Project Approval

This Mini Project entitled "**Prediction Of Suicidal Sentiment"** by **Mrunmai S. Kore(Roll No. 60), Ishika K. Koytekar(Roll No. 61) and Bhatre P. Mhatre(Roll No. 64)**is approved for the degree of **Bachelor of Engineering** in Semester VIII of Final Year **Computer Engineering.**

Examiners

1……
(Internal Examiner Name & Sign)

2……
(External Examiner name & Sign)

Date:

Place:

Contents

Abstract

Natural language processing, or NLP, is a type of artificial intelligence that deals with analyzing, understanding, and generating natural human languages so that computers can process written and spoken human language without using computer-driven language. Natural language processing, sometimes also called “computational linguistics,” uses both semantics and syntax to help computers understand how humans talk or write and how to derive meaning from what they say. This field combines the power of artificial intelligence and computer programming into an understanding so powerful that programs can even translate one language into another reasonably accurately. This field also includes voice recognition, the ability of a computer to understand what you say well enough to respond appropriately.

Acknowledgements

It is said that “learning is a never-ending process.” While working on the project we have undergone the same experience of learning new things as we proceeded in our goal of building a Toxic Chat Detection System.

Working on the project was a new experience for us. As it opened a new gateway wherein we had an opportunity to work on a totally new concept as far as the engineering syllabus is concerned where most of the concepts are to be learned by rote.

The joy of working in a new domain and learning new things was a welcome experience for the four of us and all we have to say is that we have cherished all the moments as they came by, right from working on a project to making this report.

We would like to thank our Principal **Dr. Harish Vankudre** for constant motivation and support to excel and for having faith in our ability. We would also like to thank our professor **Dr. Megha Trivedi** (Head - Department of Computer Engineering) for providing her views on the subject.

We would like to thank **Dr. Tatwadarshi P. N.,** who guided us and shared their knowledge & invaluable experience about the topic, and gave their precious time towards solving our difficulties. We would also like to thank our college management for providing us with the facilities and infrastructure for working on the project.

_ _ _ _ _ _ _ _ _ _ _ _ _ _

Mrunmai S. Kore (60)

_ _ _ _ _ _ _ _ _ _ _ _ _

Ishika K. Koytekar (61)

_ _ _ _ _ _ _ _ _ _ _ _ _ _ _

Bhakti P. Mhatre(64)

Date :

1. Introduction

1.1 Introduction

Suicide is the 17th leading cause of death in the world and accounted for 1.4% of all deaths worldwide. The Majority of these deaths are occurring in low and middle-income countries. The best way to decrease or prevent suicide is to do treatments for mental health disorders because 90% of people who die by suicide are suffering from major mental health illnesses and the most common is Depression. Many people die by suicide because they did not receive effective treatment. To cure depression, the best way is to talk about their feelings with someone they trusted and release the pain inside of them. So we've developed an application that will listen to them and provide help from mental health professionals. In this application, we're using natural language processing to train the chatbot to perform sentiment analysis. Various other features like emergency contact, and calling an ambulance within seconds are available in this application as well.

1.2 Problem Statement and Objectives

Suicide is a critical issue in modern society. The main reason for suicide is untreated

depression. Depression is a mood disorder that causes changes in the thinking of humans. A depressed person starts thinking negatively and overthinks about the bad events of life.
Our project aims to predict the sentiment of suicidal notes using techniques of Natural Language Processing and predicting using Machine Learning.

1.3 Scope

Much of the current work is focused on two major directions:

- To implement NLP preprocessing to identify text
- To use the Logistic Regression network to train and identify toxic chat.

This project will make the prediction of suicidal notes easier. Users can use this system to detect and possibly remove suicidal thought.

2. Proposed System

2.1 Introduction

Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others' opinions. Second, it presents many challenging research problems, which had never been attempted before the year 2000. Part of the reason for the lack of study before was that there was little opinionated text in digital forms. It is thus no surprise that the inception and the rapid growth of the field coincide with those of the social media on the Web. In fact, the research has also spread outside of computer science to management sciences and social sciences due to its importance to business and society as a whole. In this talk, I will start with the discussion of the mainstream sentiment analysis research and then move on to describe some recent work on modeling comments, discussions, and debates, which represents another kind of analysis of sentiments and opinions.

Sentiment classification is a way to analyze the subjective information in the text and then mine the opinion. Sentiment analysis is the procedure by which information is extracted from the opinions, appraisals, and emotions of people in regards to entities, events, and their attributes. In decision making, the opinions of others have a significant effect on customers' ease, in making choices with regard to online shopping, choosing events, products, and entities. The approaches of text sentiment analysis typically work at a particular level like phrase, sentence, or document level. This paper aims at analyzing a solution for the sentiment classification at a fine-grained level, namely the sentence level in which the polarity of the sentence can be given by three categories as positive. negative and neutral.

2.2 Block Diagram

Figure 1: Flowchart for data preprocessing

2.3 Algorithm and Process Design

- The Dataset is cleaned and Loaded in the CSV file
- The data is preprocessed by using Stopword Removal from the NLTK English stopword repository.
- The data is then tokenized using the same library.
- then Stemming is performed on each word using the same library.
- The tokenized data is then converted to array form using Vectorization algorithms.
- The array is then passed to a Logistic Regression model to train.
- The trained model is evaluated and Loaded.
- The test Data is predicted on the loaded Model

2.4 DETAILS OF HARDWARE AND SOFTWARE

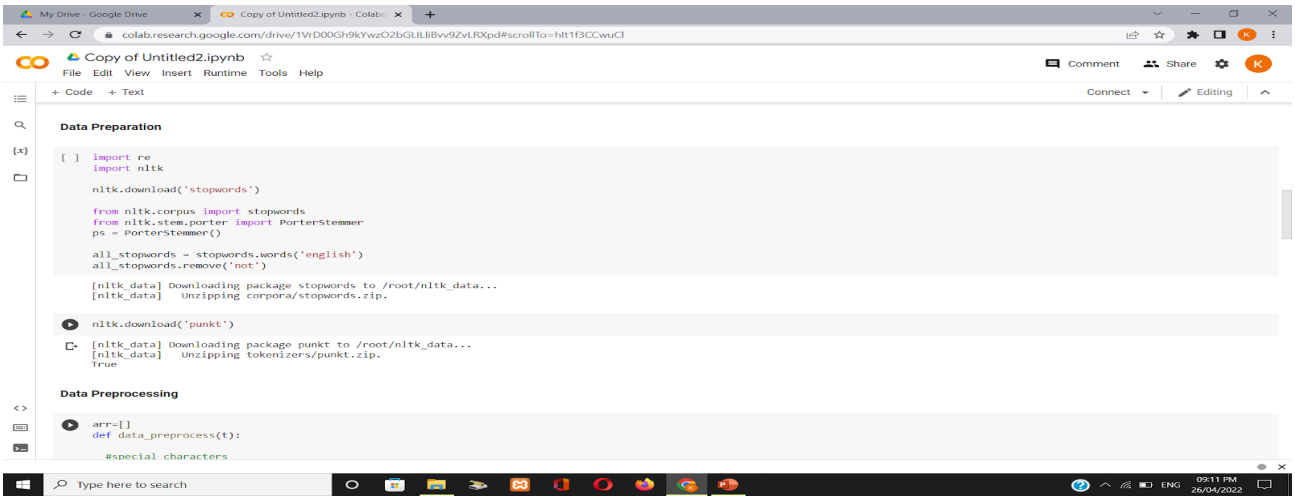
Hardware

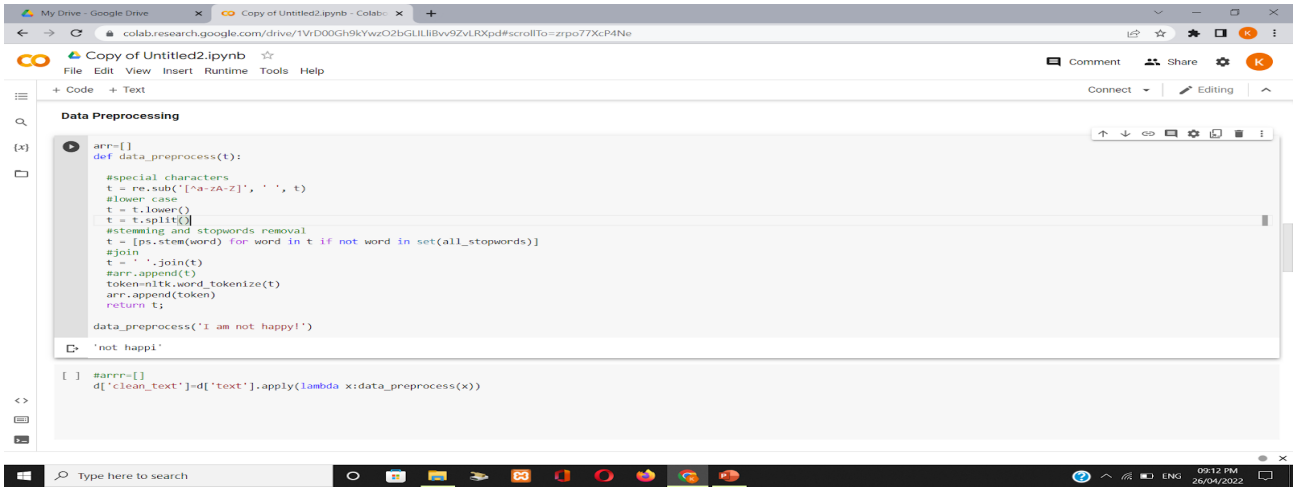
- Intel i5 processor
- RAM – 8GB
- Hard disk – 10GB
- Internet Connection

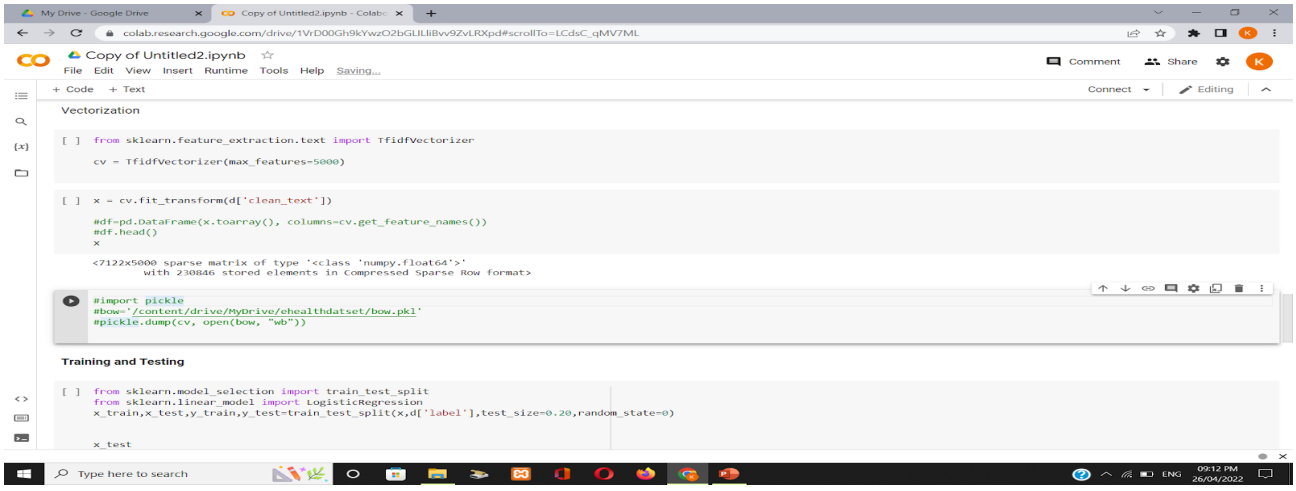
Software

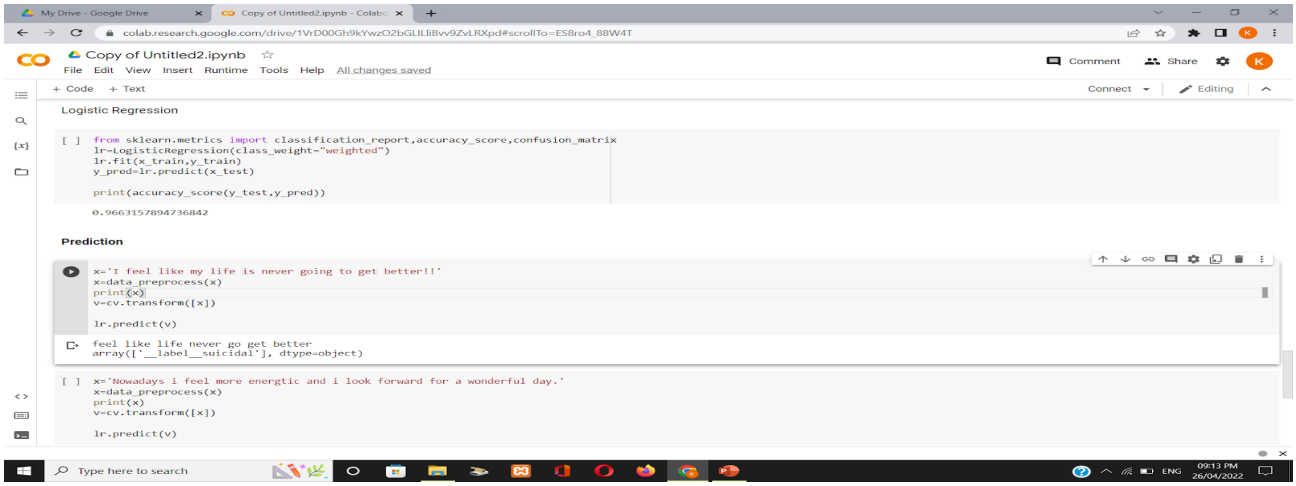
- Python
- Scikit
- Pandas
- Numpy
- Google Colab
- Windows

2.5 Experiment and Results for Validation and Verification









2.6 Conclusion and Future Scope

We considered three statistical models for solving our problem Logistic Regression, Bag Of Words.

Logistic Regression is mainly based on the independence assumption Training is very easy and fast In this approach each attribute in each class is considered separately. Testing is straightforward, calculating the conditional probabilities from the data available. One of the major tasks is to find the sentiment polarities which is very important in this approach to obtain the desired output. In this approach, we only considered the words that are available in our dataset and calculated their conditional probabilities. we have obtained successful results after applying this approach to our problem. In supervised learning methods next, we adopted a bag of words. This approach assumes that every single word in the test data is repeated at least once, which eliminates the zero probability problem. After applying this approach, the results are obtained correctly and their execution is also very fast.

However, we were successful at predicting the sentiment of Suicidal notes on a small scale using Logistic Regression and also gained a lot of information in machine learning.

2.7 References

- https://www.kaggle.com/
- www.geeksforgeeks.org/