ESSAY TONE DETECTOR

A

Project Report

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology in Computer Science& Engineering *By*

Shivam Garg (161298)

Under the supervision of

Dr. Rajni Mohana

Jaypee University of Information Technology Waknaghat, Solan – 173234 Himachal Pradesh, India

June-2020

Candidate's Declaration

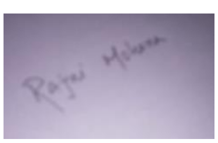
I hereby declare that the work presented in this report entitled **"Essay Tone Detector"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science &Engineering** submitted in the department of Computer Science & Engineering and Information Technology, Jaypee University of Information Technology Waknaghat, Solan is an authentic record of my own work carried out over a period From May 2020 to June 2020 under the supervision of **Dr. Rajni Mohana**(Associate Professor, Department of CSE & IT).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature)

Shivam Garg, 161298

This is to certify that the above statement made by the candidate is true to the best of my knowledge.



(Supervisor Signature) Dr. Rajni Mohana Associate Professor Department of Computer Science& Engineering

Dated 23/June/2020

Acknowledgement

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. I am grateful to my project guide **Dr. Rajni Mohana** for the guidance, inspiration and constructive suggestions that helped me in the preparation of the project.

I also thank our colleagues who have helped me in successful completion of the project.

Shivam Garg

Table of Contents

1

List of Tables

List of Figures

Abstract

Identifying the emotion or the feeling in a written essay or in a message is much harder than it is in a conversation. For one thing, there are no facial expressions or body language to hint at a writer's emotional state. Also there's usually no single attribute that is responsible for the tone of a message. This project Essay tone detector as the name suggests detects the tone of the essay or the message written by the writer. This is an extrapolation of the popularly used "Sentiment Analysis Tool" to determine the tone or the emotion of the writer in his/her literary works/social media/ essay writing competitions etc. This project Essay Tone Detector depends on a mix of rules and ml (machine learning) to identify various hints or flags in a piece of writing which adds to its general tone or the emotion. This Essay tone detector can tell you how your message is probably going to sound to somebody reading it. The goal of this project is to detect the tone or the emotion of the essay or the articles or the reviews and various other literary works of the writer.

CHAPTER 1

INTRODUCTION:

1.1 Background

Emotion is one type of affect, other type of being mood, temperament and sensation. Emotions have been widely studied in psychology and in behavior sciences, as they are an important element of human nature. Nowadays they have also attracted the attention of researchers of computer science, especially in the field of artificial intelligence. With recent advances in the field of textual analysis, the area of emotion detection has become a favorite in computational linguistic. Since emotion detection is the newer area of textual analysis, it has weaker standard methods. Emotion can be expressed as happiness, sadness, anger, disgust, fear, surprise and so forth.

What is a tone of writing?

Tone is the mood ones writing shows. As tone of voice, tone in writing gives much more meaning beyond the words used. It can tell us the intention of the writer or can hide it. The award winning writing coach Adair Lara has said, “Tone is what the dog hears.”

Suppose, for instance, one’s manager sent him/her a message that stated, “Do you have a moment to talk?” he/she might think, “Oh no, what’s wrong?” If the manager rephrased it as, “Got time to chat real quickly?” one may be less scared. There is nothing so negative or scary about that line but it is because of its tone.

And many a times, the intention or the meaning of the message is taken in a wrong way even if the sender doesn’t mean that. Many times it is taken in a wrong way and thus creates the problem for the sender, therefore understanding the tone or emotion of the message is very important and checks whether the message or essay written is in right tone or emotion with that of the writer.

Different Types of Tones.

Tone and emotions are full of diversity and there are various emotions and tones. It conveys various emotions such as sad, happy, negative, and positive and many others.

Here are some of the tones and emotions with some examples:

Appreciative: Thanks for inviting me! .

Joyful: Yes! My heart is leaping with joy!

Informal: Yeah, see ya at the party

Formal: This is to inform you that I'm going to join you at the party.

Confused: I have no clue.

Skeptical: Have you really thought this through?

Regretful: It's a pity I can't go.

Neutral: kk

1.2 Introducing Domain

We have mainly used features which can be divided into 2 groups. These features are used for making different patterns and used to classify. These groups are: 1. Formal Language Based and 2. Blogging. Language based features are the features which are related to formal linguistics and former knowledge of different words and phrases is part of polarity and speech tagging of sentences. Prior tone schism means that there are various words and phrases which usually have a natural trend to express specific and specific tone. Let's take for instance, the word "outstanding" shows a strong positive opinion, and the word "corrupt" shows a strong negative opinion. So, when a word is conveying positive attitude in a sentence the whole sentence is likely to convey a positive tone. Using parts of the sentences to detect the tone of the phrase is a good approach to detect the tone. There are various online platforms where people tend to write for example whatsapp, twitter, email and many more. All these online sites provide various features which informally allow people to express their emotions or tone through various emojis, hashtags, wordcapitalisation, Internet emoticons & Internet slang.

There are widely two Classification methods and they are: Supervised versus Unsupervised and non-compatible versus adaptable or reinforcement methods. The first approach i.e. supervised approach comes with data labels & these labels will be used to train the classifier.

We also have two other methods which can also be called as adaptive methods. These are: **passive & active**. Passive Methods are the methods that only take in consideration and use feedback to know and learn about the environment.

There are various methods and matrices which have been suggested and used for calculating & comparing the outcomes of the experiments. Some most commonly used matrices are: Precision, Accuracy, F1 Measure, True Rate and False Alarm Rate. It is shown here an example of how to calculate the metric we need.

	Machine says yes	**Machine says no**
Human says yes	**tp**	**fn**
Human says no	**fp**	**tn**

Table 1: A Typical 2x2 Confusion Matrix

- **Precision(P)** = $tptp+fp$
- **Recall(R)** = $tptp+fn$
- **Accuracy(A)** = $tp+tntp+tn+f+fp+fn$
- **F1** = $2.P.RP+R$
- **True Rate(T)** = $tptp+fn$
- **False-alarm Rate(F)** = $fptp+fn$

1.3 Objectives

The main aim of the project is to develop a system which is able to detect or tell about the tone or the emotion of the essay or the phrase. The project is designed to fulfill following features as listed below:

- To detect the tone of the essay.
- To extract the features needed by the code or the algorithm from the essay.
- To classify the features as needed to detect the tone or the emotion of the passage.

1.4 Scope of the Project

We will design a system which will compromise of two modules. The first one will help us to clean the data using various techniques such as tokenization, stemming, removing stop words in order even the data. The second model comprises of the algorithm written using machine learning that will help in detecting the tone or the emotion of the essay or the phrase written as input to the program.

Chapter 2

Literature Review

2.1 Limitations To Prior:

Conclusion investigation is totally new research theme in field of micro blogging, so there is a lot of space for additional exploration here. A limited quantity of significant earlier work has been done in supposition investigation of client surveys, records, web online journals articles and general expression level notion examination. The better outcomes engaged with the conclusion arrangement are the employments of administered learning procedures such as innocent bays and SVM (Support Vector Machines), yet manual naming required for the directed methodology is exorbitant. A portion of the work was done on the unaided and semi managed approaches and there is incredible breadth for development. Different analysts testing new highlights and order strategies have looked at the results to baseline-line execution. There is need of better more formal examinations b/w the outcomes through numerous particular highlights and grouping techniques to choose best highlights and most proficient order strategies for the particular applications.

2.2 Related Work:

The sack of words models is one of the most extensively used component models for a basically all book gathering endeavors, with its ease and extraordinary execution. The model suggests message appointed the pack or arrangement of individual words with associations or conditions to a word, which implies it thoroughly, dismisses the language structure and solicitation of words in the substance. This model also amazingly well known in inclines examination and has been used by various experts. The least mind boggling way to deal with unites this model into or request is to use unigram a segment. Generally speaking, n gram is the [6]progression of "n" words in our substance, absolutely self-governing of some other word or gram on the substance.

In this way, we acknowledge that the substance to be named a unigram is only a combination of individual words, and the proximity or nonappearance off another word in a book doesn't impact the likelihood of a word occurring. It is extremely essential yet has been seemed to give incredible performance. The direct way of using unigram is to consign them the fixed pre-polarization and to average the general furthest point of the text, where the general limit of the text is dictated by including the different before polarities of the individual unigrams. If the word is used as a picture of motivation, the earlier furthest point of the word is sure, for e.g. "Good". However, the word is commonly negative in case it is connected with negative implications, for example "wicked". There may furthermore be limit in the model, which infers how unequivocal the term is to a particular class.

"Overpowering" likely has a strong passionate furthest point, while "extraordinary" has a strong positive limit, yet perhaps with weak autonomy.

There are three distinct approaches to use the pre-furthest point of words to depict. An independent essential technique is to use openly available online word references, which map a word to its past furthest point. Multi-Perspective-Questioning Answering (MPQA) is a subjectivity dictionary that maps a full scale ofl 4,850 words whether "positive" or "negative" and whether they are "strong" or "weak" themes. SentiWordNet 3.0 is another benefit that gives the opportunity of each word in the positive, negative and fair-minded classes.

The resulting strategy is to construct a positive pre furthest point word reference according to the occasion of leach word in each particular class from our planning data. For example, if a particular word in our arrangement dataset (stood out from other classes) occurs more sometimes in determinedly checked articulations, we find out the probability that solitary word has a spot with the positive class rather than the diverse class. This system has been seemed to give better performance, as the earlier polarization ofl words is progressively proper and fitted with a particular kind ofl text, and isn't as typical as the past strategy.

In any case, the latter is the overseen approach in light of the fact that the planning data must be set apart to the suitable class before it is possible to register the relative occasion ofl a word in each ofl the class. Execution diminishes were recognized by Kouloumpis et al. using word reference word features with custom n gram word features worked from the readiness data, rather than using n grams alone.

The third procedure is the mediation between the two approaches. In this, we develop our own polarization anyway not so much from our arrangement data, so we don't need to name getting ready data. One way to deal with do this is proposed by Turnty et al. The pre semantic bearing (furthest point) ofl a word or articulation is controlled by reacting the information with "wonderful" and removing the result from "poor" with the coordinated effort of that word or articulation. They used the amount of results from the huge request question's online web crawler to learn normal information.

Here hits (express “shocking”) are the number of documents given by the web file (whose limit must be resolved) and "incredible" happen together. Hits (“wonderful”) infer the numbers of documents that contain "excellent" co-occur. Prabowo et al. thought of this idea and used 120 positive words and 120 negative seeds of data to conduct an internet search. Along these lines, the general semantic bearing of the word suitable can be found by averaging the word's region to each word's seed words.

Another graphical system for estimating the furthest point of adjectives is discussed by Hatzivasiloglou et al. The strategy incorporates first perceiving all the unmistakable word blends from the corpus and a while later arranging each pair of descriptors using an algorithm (supervised). A graph is worked in which center point engaging words and associations address the comparable or different semantic heading. Finally, the batching estimation is executed, which isolates the outline into two subsets, which suggests that the center points in the subset have basically a comparative bearing joins, and the associations between the two subsets have in a general sense different headings. Most subsets have positive descriptors and various have negatives.

Various experts in this field have quite recently used word references publicly open inclination, while others have also examined creating their own pre-polar word references.

The focal issue with the philosophy of prior furthest point perceived by Wilson et al. He perceives prior limit and intelligent furthest point. They in like manner express that the prior limit of a word may truly be not exactly equivalent to the word used in a particular setting.

In this model, the four underlined words "Trust", "well", "cause" and "fitting" are certain references when seen without reference to the articulation, anyway are not used here to pass on a positive inclination. This prompts the end that "Trust" may regularly be used in positive sentences; anyway this doesn't block the probability that it is also found in non-positive sentences.

Prior polarization of individual words (whether or not words are usually positive or negative wisdom) isn't the principle issue. In light of the significant limit of the articulation, examines some various features, including semantic and syntactic associations between words to improve their request.

The introduction of sentiment assessment may be solidly related to communicate level supposition examination. In 2005, Wilson and others presented a key paper on express level appraisal examination. It perceives another approach to manage the issue by first gathering the articulations according to subjectivity (polar) and the objectivity (neutral) and requesting the theoretical obvious articulations as positive or negative. Various passionate articulations use prepositional verbalization, which explicitly adds to the course of action of theoretical articulations. If we use an essential request we acknowledge that the important limit of the term is proportional to its previous furthest point, the result is around 38%. The epic portrayal structure propose contains an overview of general features that

Contain information about the significant furthest point, achieving a critical improvement in the introduction (in wording of accuracy) of the game plan process. The eventual outcomes of this paper are presented in the going with table:

Features	Accuracy	Subjective F.	Objective F.
Word tokens	73.6	55.7	81.2
Words + prior polarity	74.2	60.6	80.7
28 features	75.9	63.6	82.1

Table 2: Step 1 results for Objective / Subjective Classification in [16]

Features	Accuracy	Positive F.	Negative F.	Both F.	Objective F.
Word tokens	61.7	61.2	73.1	14.6	37.7
Word + prior	63.0	61.6	75.5	14.6	40.7
10 features	65.7	65.1	77.2	16.1	46.2

Table 3: Step 2 results for Polarity Classification in [16]

One way to deal with decline the opportunity condition and solidify fragmented references into our word model is to use bigrams and trigrams as well as unigrams. Bigram is a collection of two on very basic level irrelevant words in a book, and trigram is a variety of three consecutive words. Subsequently, we can calculate the previous furthest point or the probability of the bigram/trigram of the explicit class – rather of prior limit of disengaged class. Various experts have investigated various roads in regards to them, saying that in case we have to use one of them, unigram perform better, and a bit of the unigram with bigram can give better results. Trigrams generally have poor performance. Execution decline using trigrams considering the way that there is an exchange off between capturing continuously complex models and word incorporation when moving to higher numbered grams. A couple of researchers have endeavored to recall disclaimers for the Unigram word model. Throb et al. Furthermore, used a model in which prior polarity was changed to the word, which means renouncing (like "no", "not", "don't, etc.). In this manner, some appropriate information is associated with the word model.

Syntactic features, (for instance, Parts of speech marking" or POS naming) are similarly normally used around there. The thought of tagging each word of a tweet concerning any part of speech is: thing, pronoun, activity word, descriptor, modifier, power, etc. The thought is to perceive and use models reliant on this POS which we can use in the gathering system. For example, it has been represented that target tweets have more run of the mill things and outcast activity words than enthusiastic tweets, so if a tweet is masterminded, the more conspicuous use of nonexclusive things and activity words is generally speaking as an untouchable glancing in, that tweet is objective (according to this component). Likewise, passionate tweets contain more qualifiers, graphic words, and increases. These associations are developed in the figures underneath:

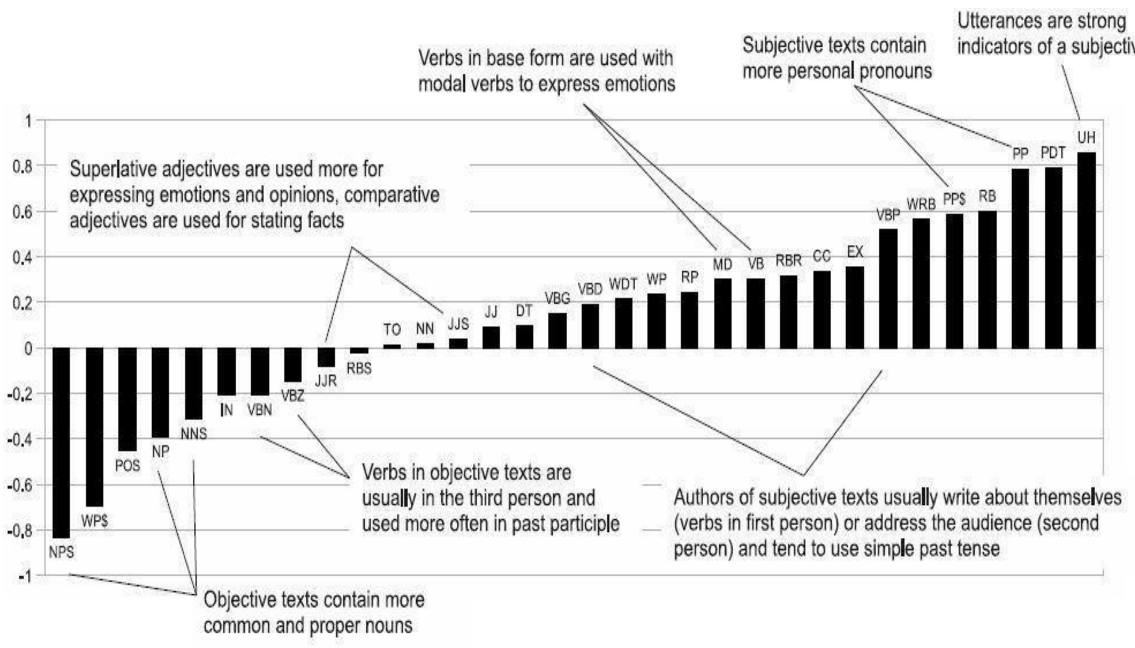


Figure 1: Using POS Tagging as features for objectivity/subjectivity classification

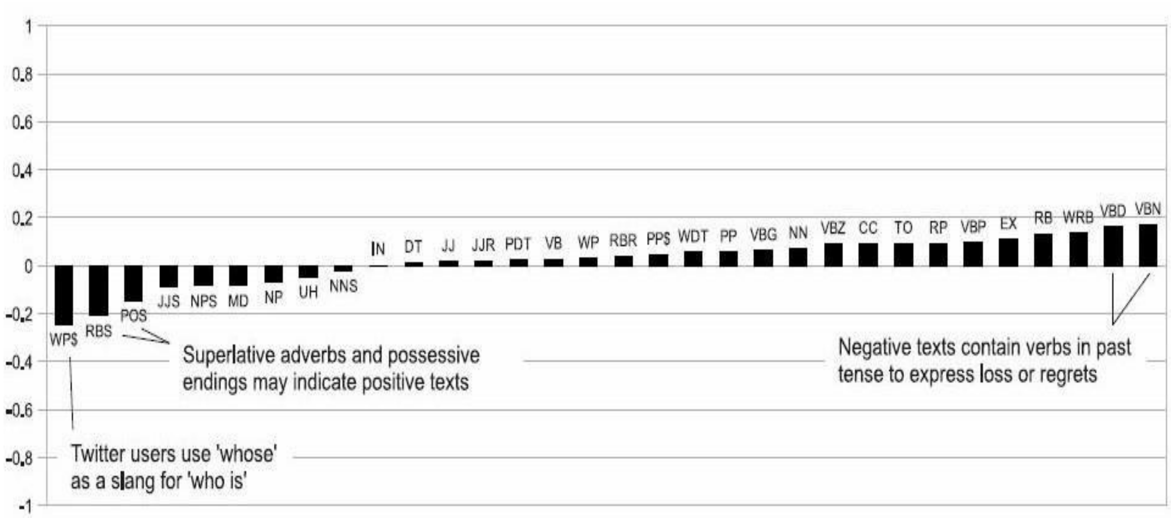


Figure 2: Using POS Tagging as features in positive/negative classification

In any case, there is still conversation about whether part of speech is a useful component of suspicion gathering. A couple of experts battle for better POS features while others don't recommend them.

Beside these, there is a ton of work to be done in filtering for a component class that is only sensible for the micro blogging territory. The closeness of URLs in a tweet and the number of capitalized letter/letters in a tweet were noted by Koulompis et al. likewise, Barbosa et al. Koulmpis addresses positive results for the use of features, for instance, emoticons and Internet slang terms. Brady et al concentrates on the reaching out of words as a picture of subjectivity in a tweet. The paper reports positive results of their study, suggesting that if a word is progressively visit; the term is seen as a strong sign of subjectivity.

Gullible bayes classifier and state vector machines are the most consistently used portrayal techniques. A couple of authorities, for instance, Barbosa et al. circulate extraordinary results for help vector machines, while Pak et al. reinforce Naive Bayes.

Disregarding the way that it continues adding logically named essays to planning data, it has been seen that having a greater getting ready test is somewhat payoff to a particular degree, by then the precision of the request is for all intents and purposes consistent. Barbosa et al. for getting ready classifiers used essays named by Internet sources as opposed to hand naming. Simultaneously the precision of the stamped models is lost (shown as the development in uproar) anyway if the accuracy of the readiness name outperforms a large portion of, the higher the name, the higher the course of action result precision. In this way, if there are a gigantic number thus, by then our imprints will make racket, will be mistaken and can be compensated for the misunderstanding. On the other hand Pak et al And Go et al see the proximity of positive or negative emoticons to give out names to tweets. Like the above case, they used a tremendous number of essays to diminish the impact of noise on their arrangement data.

CHAPTER-3

SYSTEM DEVELOPMENT

3.1 Introduction

The tool made as a bit of this arrangement program engages a customer to move a book record and predicts the tone of the moved content with about 85% precision. The model trains at 50,000 unpredictable sentences out of the... sentences in the dataset obtained from Kaggle. The estimation used to envision the sentiment of the sentences in the data record is Logical Regression. After viably choosing the sentiment of each sentence, the instrument reestablishes the most a great part of the time happening feeling in all the sentences. It furthermore forms a pie chart outlining the general scattering of various sentiments in the information record.

3.2 Model Implementation

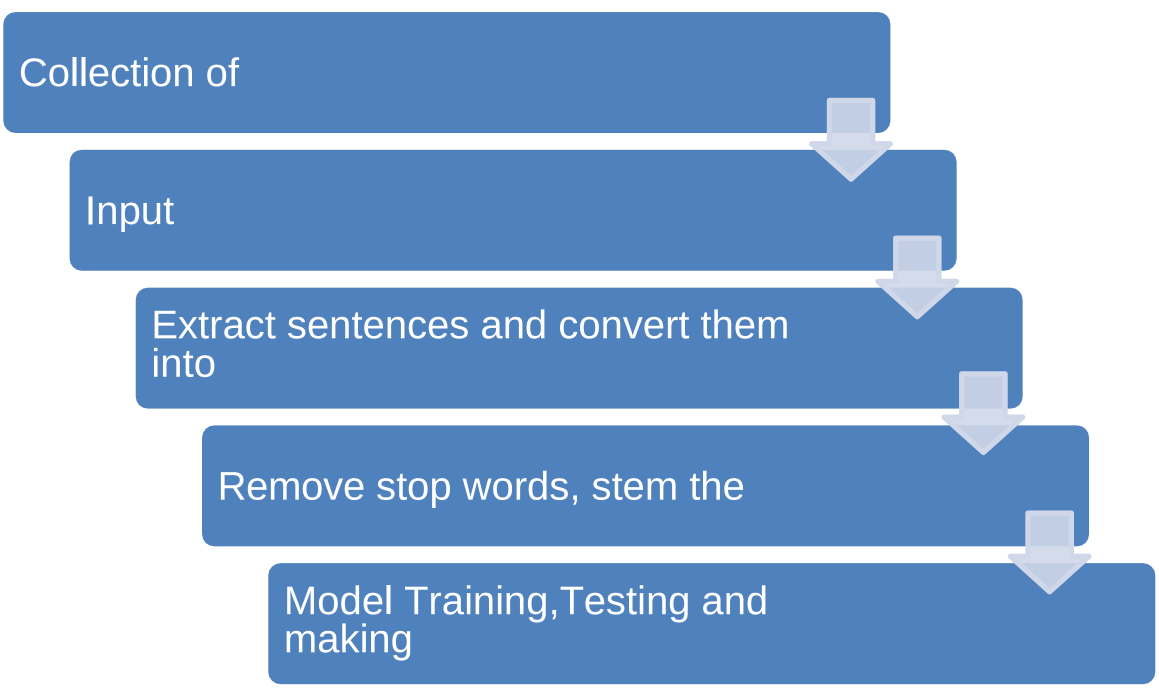


Figure 3: Model Implementation.

3.3 Software Used

This system is built with the help of number of tools which provide us platform to run our algorithms or store data or the functions for the front end services some of the main tools are as follows:

NLTK (Natural Language Toolkit):

The Natural Language Toolkit (NLTK) is a phase used for building Python programs that work with human language data for applying in genuine standard language taking care of (NLP).

It contains text getting ready libraries for tokenization, parsing, request, stemming, marking and semantic reasoning. It moreover fuses graphical presentations and test instructive lists similarly as joined by a cook book and a book which explains the measures behind the key language dealing with assignments that NLTK supports.

The Natural Language Toolkit is an open source library for the Python programming language at first created by Steven Bird, Edward Lopper and Ewan Klein for use being created and guidance.

It goes with a hands-on control that presents focuses in computational phonetics similarly as programming basics for Python which makes it proper for etymologists who have no significant data in programming, pros and researchers that need to plunge into computational historical underpinnings, understudies and instructors.

Microsoft Excel:

This is a kind of spreadsheet that is made by the Microsoft for a wide scope of working systems. The standard features of MS Excel is that it contains diverse logical limits that can help us in calculations and there are a piece of instruments that can be used to plot diagrams in different structures that urges us to separate data and various instruments, for instance, turn tables, programming language named Visual stray pieces for various applications.

Spreadsheets will provide you with the characteristics arranged in rows and columns that can be changed deductively using both basic and complex number shuffling exercises. Despite the standard spreadsheet features, Excel offers programming support by methods for Microsoft's Visual Basic for Applications (VBA), the ability to get to data from outside sources through Microsoft's Dynamic Data Exchange (DDE). Microsoft Excel is an Electronic Spreadsheet Computer Program.

3.4 Preprocessing Data:

Cleaning of the data that we are going to use is very important to highlight the important points that are necessary for our machine learning framework to pick. This includes various steps:

1. Eliminate Punctuation:

Punctuations like comma, full stop, question mark etc. These give meaning to the sentence. But in vectorizer that will count the no of words and thus these punctuations does not add a value and thus these are removed. For instance: why did you do this? > Why did you do this.

In [7]:
```python
import string
string.punctuation
```

Out[7]: `'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'`

In [8]:
```python
#Function to remove Punctuation
def remove_punct(text):
    text_nopunct = "".join([char for char in text if char not in string.pun
    return text_nopunct

data['body_text_clean'] = data['body_text'].apply(lambda x: remove_punct(x)

data.head()
```

Out[8]:

	label	body_text	body_text_clean
0	ham	I've been searching for the right words to tha...	Ive been searching for the right words to than...
1	spam	Free entry in 2 a wkly comp to win FA Cup fina...	Free entry in 2 a wkly comp to win FA Cup fina...
2	ham	Nah I don't think he goes to usf, he lives aro...	Nah I dont think he goes to usf he lives aroun...
3	ham	Even my brother is not like to speak with me. ...	Even my brother is not like to speak with me T...
4	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	I HAVE A DATE ON SUNDAY WITH WILL

Figure 4.1: Eliminate Punctuation.

2. Tokenization:

In tokenization every sentence is read line by line and is converted into individual words.

For instance, I am going out. It is expressed as [I, am, going, out]. This process is called as tokenization.

In [9]:
```python
import re

# Function to Tokenize words
def tokenize(text):
    tokens = re.split('\W+', text) #W+ means that either a word character (A-Za-z0-9_) or a dash (-) can go there.
    return tokens

data['body_text_tokenized'] = data['body_text_clean'].apply(lambda x: tokenize(x.lower()))
#We convert to lower as Python is case-sensitive.

data.head()
```

Out[9]:

	label	body_text	body_text_clean	body_text_tokenized
0	ham	I've been searching for the right words to tha...	Ive been searching for the right words to than...	[ive, been, searching, for, the, right, words, ...
1	spam	Free entry in 2 a wkly comp to win FA Cup fina...	Free entry in 2 a wkly comp to win FA Cup fina...	[free, entry, in, 2, a, wkly, comp, to, win, f...
2	ham	Nah I don't think he goes to usf, he lives aro...	Nah I dont think he goes to usf he lives aroun...	[nah, i, dont, think, he, goes, to, usf, he, l...
3	ham	Even my brother is not like to speak with me. ...	Even my brother is not like to speak with me T...	[even, my, brother, is, not, like, to, speak, ...
4	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	I HAVE A DATE ON SUNDAY WITH WILL	[i, have, a, date, on, sunday, with, will]

Figure 4.2: Tokenization

3. Stopwords:

Stopwords are the words that are common in the sentence and do not add to the emotion of the passage and thus we remove them. Some Stopwords are is, am, are, this, a, an, the.

In [10]:

```
import nltk

stopword = nltk.corpus.stopwords.words('english')# All English Stopwords
```

In [11]:

```
# Function to remove Stopwords
def remove_stopwords(tokenized_list):
    text = [word for word in tokenized_list if word not in stopword]# To remove all stopwords
    return text

data['body_text_nostop'] = data['body_text_tokenized'].apply(lambda x: remove_stopwords(x))

data.head()
```

Out[11]:

	label	body_text	body_text_clean	body_text_tokenized	body_text_nostop
0	ham	I've been searching for the right words to tha...	Ive been searching for the right words to than...	[ive, been, searching, for, the, right, words,...	[ive, searching, right, words, thank, breather...
1	spam	Free entry in 2 a wkly comp to win FA Cup fina...	Free entry in 2 a wkly comp to win FA Cup fina...	[free, entry, in, 2, a, wkly, comp, to, win, f...	[free, entry, 2, wkly, comp, win, fa, cup, fin...
2	ham	Nah I don't think he goes to usf, he lives aro...	Nah I dont think he goes to usf he lives aroun...	[nah, i, dont, think, he, goes, to, usf, he, l...	[nah, dont, think, goes, usf, lives, around, t...
3	ham	Even my brother is not like to speak with me. ...	Even my brother is not like to speak with me T...	[even, my, brother, is, not, like, to, speak, ...	[even, brother, like, speak, treat, like, aids...
4	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	I HAVE A DATE ON SUNDAY WITH WILL	[i, have, a, date, on, sunday, with, will]	[date, sunday]

Figure 4.3: Stopwords

4. Stemming:

Stemming is the process in which word is written into its root form. There are various forms of a single word only and by reducing the word to its root form we can reduce the noise in the data and avoid using the same word again and again. For instance words connections, connecting connected are all derived from the same word connect.

In [12]:

```python
ps = nltk.PorterStemmer()

def stemming(tokenized_text):
    text = [ps.stem(word) for word in tokenized_text]
    return text

data['body_text_stemmed'] = data['body_text_nostop'].apply(lambda x: stemming(x))

data.head()
```

Out[12]:

	label	body_text	body_text_clean	body_text_tokenized	body_text_nostop	body_text_stemmed
0	ham	I've been searching for the right words to tha...	Ive been searching for the right words to than...	[Ive, been, searching, for, the, right, words,...	[Ive, searching, right, words, thank, breather...	[ive, search, right, word, thank, breather, pr...
1	spam	Free entry in 2 a wkly comp to win FA Cup fina...	Free entry in 2 a wkly comp to win FA Cup fina...	[free, entry, in, 2, a, wkly, comp, to, win, f...	[free, entry, 2, wkly, comp, win, fa, cup, fin...	[free, entri, 2, wkli, comp, win, fa, cup, fin...
2	ham	Nah I don't think he goes to usf, he lives aro...	Nah I dont think he goes to usf he lives aroun...	[nah, i, dont, think, he, goes, to, usf, he, l...	[nah, dont, think, goes, usf, lives, around, t...	[nah, dont, think, goe, usf, live, around, tho...
3	ham	Even my brother is not like to speak with me. ...	Even my brother is not like to speak with me T...	[even, my, brother, is, not, like, to, speak, ...	[even, brother, like, speak, treat, like, aids...	[even, brother, like, speak, treat, like, aid,...
4	ham	I HAVE A DATE ON SUNDAY WITH WILL!!	I HAVE A DATE ON SUNDAY WITH WILL	[i, have, a, date, on, sunday, with, will]	[date, sunday]	[date, sunday]

Figure 4.4: Stemming

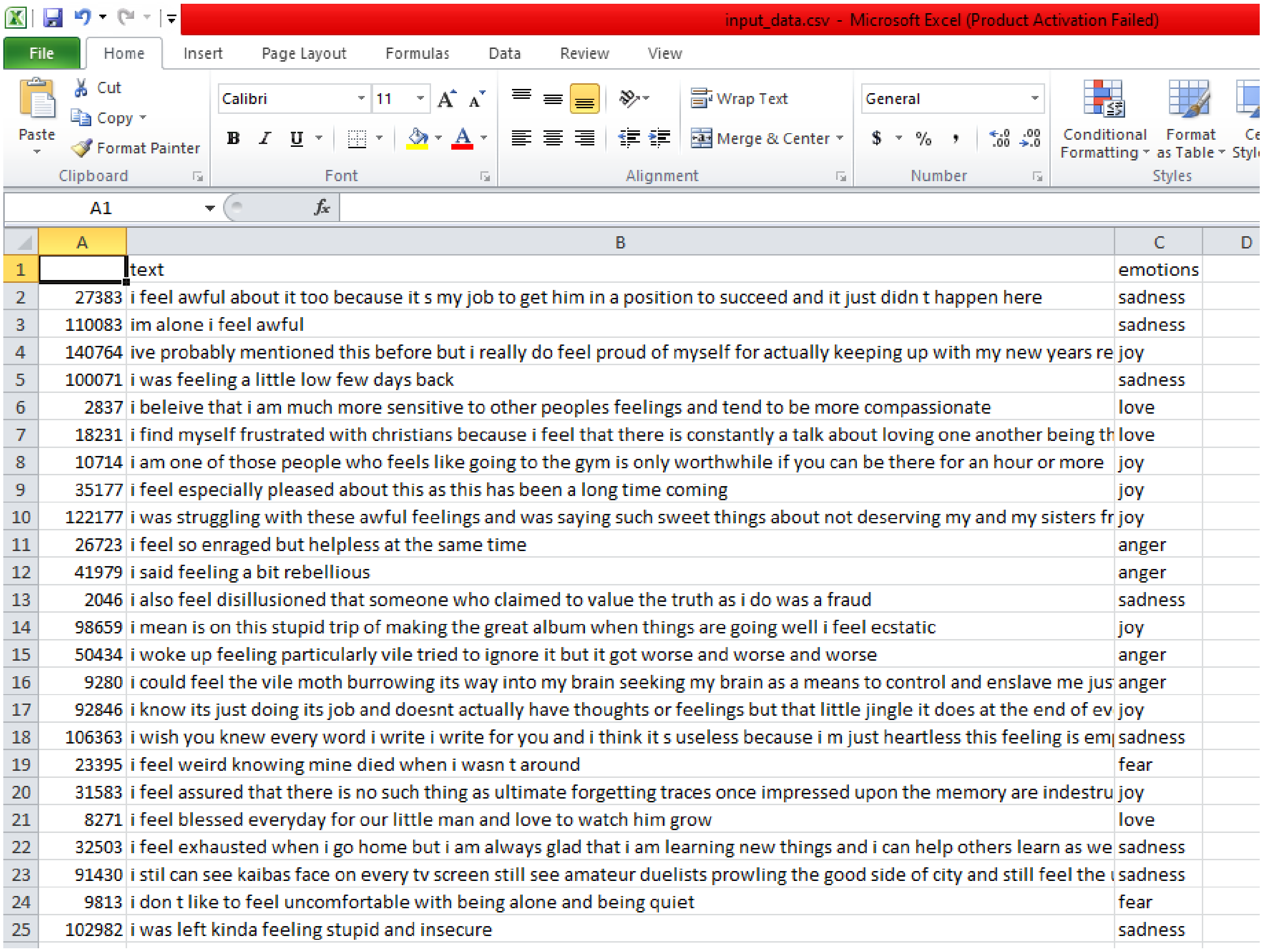
Chapter 4

Code Implementation

This module consists of different codes for various functioning of the project and also the outputs figures captured in running environment.

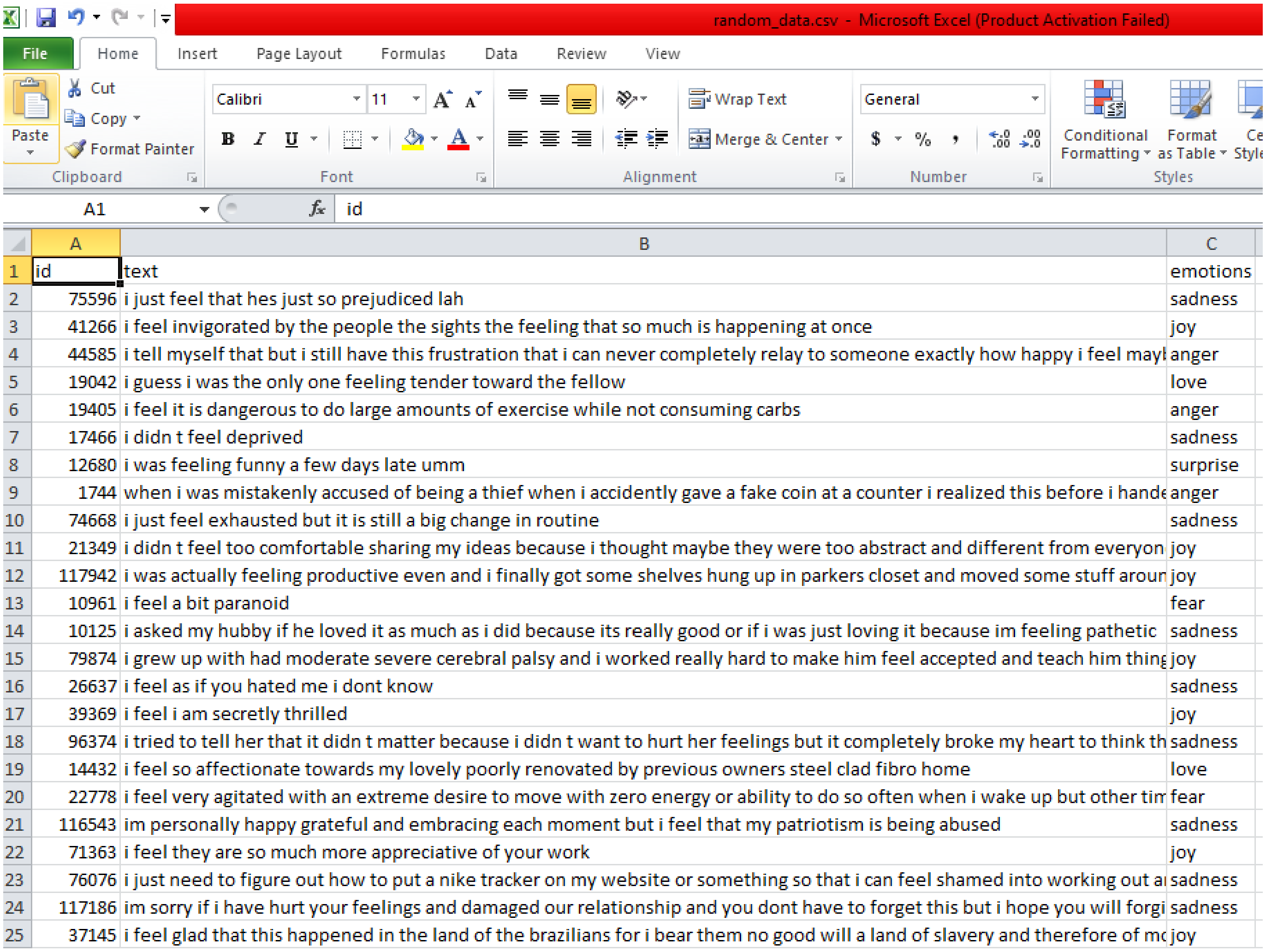
4.1 DATASET:

The dataset consists of 4, 15, 809 sentences with labeled emotion. The data is collected from Kaggle. The screenshot of the dataset is:



	A	B	C
1		text	emotions
2	27383	i feel awful about it too because it s my job to get him in a position to succeed and it just didn t happen here	sadness
3	110083	im alone i feel awful	sadness
4	140764	ive probably mentioned this before but i really do feel proud of myself for actually keeping up with my new years re	joy
5	100071	i was feeling a little low few days back	sadness
6	2837	i beleive that i am much more sensitive to other peoples feelings and tend to be more compassionate	love
7	18231	i find myself frustrated with christians because i feel that there is constantly a talk about loving one another being th	love
8	10714	i am one of those people who feels like going to the gym is only worthwhile if you can be there for an hour or more	joy
9	35177	i feel especially pleased about this as this has been a long time coming	joy
10	122177	i was struggling with these awful feelings and was saying such sweet things about not deserving my and my sisters fr	joy
11	26723	i feel so enraged but helpless at the same time	anger
12	41979	i said feeling a bit rebellious	anger
13	2046	i also feel disillusioned that someone who claimed to value the truth as i do was a fraud	sadness
14	98659	i mean is on this stupid trip of making the great album when things are going well i feel ecstatic	joy
15	50434	i woke up feeling particularly vile tried to ignore it but it got worse and worse and worse	anger
16	9280	i could feel the vile moth burrowing its way into my brain seeking my brain as a means to control and enslave me jus	anger
17	92846	i know its just doing its job and doesnt actually have thoughts or feelings but that little jingle it does at the end of ev	joy
18	106363	i wish you knew every word i write i write for you and i think it s useless because i m just heartless this feeling is em	sadness
19	23395	i feel weird knowing mine died when i wasn t around	fear
20	31583	i feel assured that there is no such thing as ultimate forgetting traces once impressed upon the memory are indestru	joy
21	8271	i feel blessed everyday for our little man and love to watch him grow	love
22	32503	i feel exhausted when i go home but i am always glad that i am learning new things and i can help others learn as we	sadness
23	91430	i stil can see kaibas face on every tv screen still see amateur duelists prowling the good side of city and still feel the	sadness
24	9813	i don t like to feel uncomfortable with being alone and being quiet	fear
25	102982	i was left kinda feeling stupid and insecure	sadness

Figure 5.1: Dataset



id	text	emotions
75596	i just feel that hes just so prejudiced lah	sadness
41266	i feel invigorated by the people the sights the feeling that so much is happening at once	joy
44585	i tell myself that but i still have this frustration that i can never completely relay to someone exactly how happy i feel mayl	anger
19042	i guess i was the only one feeling tender toward the fellow	love
19405	i feel it is dangerous to do large amounts of exercise while not consuming carbs	anger
17466	i didn t feel deprived	sadness
12680	i was feeling funny a few days late umm	surprise
1744	when i was mistakenly accused of being a thief when i accidently gave a fake coin at a counter i realized this before i hande	anger
74668	i just feel exhausted but it is still a big change in routine	sadness
21349	i didn t feel too comfortable sharing my ideas because i thought maybe they were too abstract and different from everyon	joy
117942	i was actually feeling productive even and i finally got some shelves hung up in parkers closet and moved some stuff arour	joy
10961	i feel a bit paranoid	fear
10125	i asked my hubby if he loved it as much as i did because its really good or if i was just loving it because im feeling pathetic	sadness
79874	i grew up with had moderate severe cerebral palsy and i worked really hard to make him feel accepted and teach him thing	joy
26637	i feel as if you hated me i dont know	sadness
39369	i feel i am secretly thrilled	joy
96374	i tried to tell her that it didn t matter because i didn t want to hurt her feelings but it completely broke my heart to think th	sadness
14432	i feel so affectionate towards my lovely poorly renovated by previous owners steel clad fibro home	love
22778	i feel very agitated with an extreme desire to move with zero energy or ability to do so often when i wake up but other tir	fear
116543	im personally happy grateful and embracing each moment but i feel that my patriotism is being abused	sadness
71363	i feel they are so much more appreciative of your work	joy
76076	i just need to figure out how to put a nike tracker on my website or something so that i can feel shamed into working out a	sadness
117186	im sorry if i have hurt your feelings and damaged our relationship and you dont have to forget this but i hope you will forgi	sadness
37145	i feel glad that this happened in the land of the brazilians for i bear them no good will a land of slavery and therefore of mc	joy

Figure 5.2: Dataset

This dataset is obtained from Kaggle. Our module learns on this dataset. The tool learns on 50, 000 randomly picked sentences out of whole dataset. This data is then transformed into the matrix and our model is trained on this data.

The algorithm used is Logical Regression.

4.2 Code:

pro(3).py - F:\essay tone detector\Essay-Tone-Detector-master\pro(3).py (3.8.3)

File Edit Format Run Options Window Help

```
from tkinter import *
from tkinter import ttk as ttk
import tkinter as tk
import tkinter.filedialog as fd
import tkinter.messagebox as mb
import pandas as pd
import numpy as np
import re
from nltk.tokenize import sent_tokenize
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn import metrics
from nltk.stem.porter import PorterStemmer
import nltk
nltk.download('stopwords')
#nltk.download('all')
nltk.download('punkt')
import random
from collections import Counter
import matplotlib.pyplot as plt

ps = PorterStemmer()
stop_words=set(stopwords.words("english"))
cv = CountVectorizer(max_features = 4000) #to select top 4000 words most used
reg=LogisticRegression(solver='lbfgs',multi_class='auto',max_iter=1001)
lab=LabelEncoder()


def info():
   # mb.showinfo("Info","Please browse a file first")
   pass

def openfile():
    filename=fd.askopenfilename()
    el.insert(0,filename)
```

Figure 6.1: Code

pro(3).py - F:\essay tone detector\Essay-Tone-Detector-master\pro(3).py (3.8.3)

File Edit Format Run Options Window Help

```
def pie():
    plt.show()


def model():
    fh=open("input_data.csv")
    #fh2=open("random_data.csv",encoding='utf-8')
    fh2=open("random_data.csv","w+")
    fh2.write("id,text,emotions\n")
    contents=[]
    for line in fh:
        contents.append(line)
    for i in range(0,50000):
        i=random.randint(1,416809)
        fh2.write(contents[i])
    fh.close()
    fh2.close()
    dataset=pd.read_csv("random_data.csv",encoding='cp1252')
    processed_list = []


    for i in range(50000):
        contents=re.sub('@[\w]*',' ',dataset['text'][i])
        contents = re.sub('[^a-zA-Z]', ' ', contents)
        contents = contents.lower()
        contents = contents.split()
        filtered_sent=[]
        for w in contents:
            if w not in stop_words:
                filtered_sent.append(ps.stem(w))

        filtered_sent = ' '.join(filtered_sent)
        processed_list.append(filtered_sent)


    X = cv.fit_transform(processed_list) #convert it in string and store data in X


    y=dataset["emotions"]
    y=y[0:50000]
```

Figure 6.2: Code

pro(3).py - F:\essay tone detector\Essay-Tone-Detector-master\pro(3).py (3.8.3)

File Edit Format Run Options Window Help

```
    y=dataset["emotions"]
    y=y[0:50000]

    y=lab.fit_transform(y) #to  make y as interger type label
    reg.fit(X,y)

def DisplayOnGUI(tone):
    ta=Text(root,height=1,width=40,bg="slategray1")
    ta.insert(tk.END,tone)
    ta.place(x=100,y=250)
    bt3=Button(root,text="Details",fg="white",bg="SteelBlue",width=10,font="Arial 10 bold",command=pie)
    bt3.place(x=330,y=272)

def display_result(result):
    result_list=result.tolist()
    result2=Counter(result_list)
    count=0
    for ele in result_list:
        curr_freq=result_list.count(ele)
        if curr_freq>count:
            count=curr_freq
            label=ele
    tone="The tone of the essay is: "+str(label)
    DisplayOnGUI(tone)

    unique_label=[]
    sizes=[]

    for ele in result_list:
        if ele not in unique_label:
            unique_label.append(ele)
            sizes.append(result2[ele])

    plt.pie(sizes,labels=unique_label, autopct='%1.1f%%',shadow=True, startangle=90)
    plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.


def train(X_test):
    y_pred=reg.predict(X_test)
    result=lab.inverse_transform(y_pred)
    display_result(result)
```

Figure 6.3: Code

pro(3).py - F:\essay tone detector\Essay-Tone-Detector-master\pro(3).py (3.8.3)

File Edit Format Run Options Window Help

```python
def convert_into_words(contents):

    tokenized_text=sent_tokenize(contents)

    processed_list=[]
    for i in tokenized_text:
        con=re.sub('@[\w]*',' ',i)
        con = re.sub('[^a-zA-Z]', ' ', con)
        con = con.lower()
        con = con.split()
        filtered_sent=[]
        for w in con:
            if w not in stop_words:
                filtered_sent.append(ps.stem(w))

        filtered_sent = ' '.join(filtered_sent)
        processed_list.append(filtered_sent)

    X_test = cv.transform(processed_list) #convert it in string and store data in X
    return X_test


def textmining(event):
    filename=str(el.get())
    if filename =="":
        info()
    else:
        result=re.search(r'\.([A-Za-z0-9]+$)',filename)
        if result:
            if str(result.group(1))!="txt":

                el.delete(0,'end')
                mb.showerror("Error","Only .txt files supported!")
                root.destroy()
            else:
                el.delete(0,'end')
                fh=open(filename,"r")
                contents=fh.read()

                model()
                X_test=convert_into_words(contents)
```

Figure 6.4: Code

pro(3).py - F:\essay tone detector\Essay-Tone-Detector-master\pro(3).py (3.8.3)

File Edit Format Run Options Window Help

```
        result=re.search(r"\.([A-Za-z0-9]+)",filename)
        if result:
            if str(result.group(1))!="txt":

                e1.delete(0,'end')
                mb.showerror("Error","Only .txt files supported!")
                root.destroy()
            else:
                e1.delete(0,'end')
                fh=open(filename,"r")
                contents=fh.read()

                model()
                X_test=convert_into_words(contents)
                #print("Filterd Sentence:",X_test)
                train(X_test)


root=Tk()
root.title("Essay Tone Detector")
root.geometry("500x500")
root.geometry("500x500+100+100")
root.resizable(False,False)
root.config(background="slategray1")
logo = tk.PhotoImage(file="logo.png")

w1 = tk.Label(root, image=logo)
w1.place(x=405,y=20)

l1=Label(root,text="Essay Tone Detector",fg="white",bg="Skyblue4",font="Calibri 20 bold",relief=RIDGE,padx=10)
l1.place(x=100,y=40)

bt=Button(root,text="Click here to browse File",fg="white",bg="SteelBlue",width=20,font="Arial 10 bold",command=openfile)
bt.place(x=20,y=150)

e1=Entry(root,width=45)
e1.place(x=210,y=155)

bt2=Button(root,text="Upload",fg="white",bg="SteelBlue",width=10,font="Arial 10 bold")
bt2.place(x=180,y=210)
bt2.bind('<Button>',textmining)
```

Figure 6.5: Code

4.3OUTPUT:

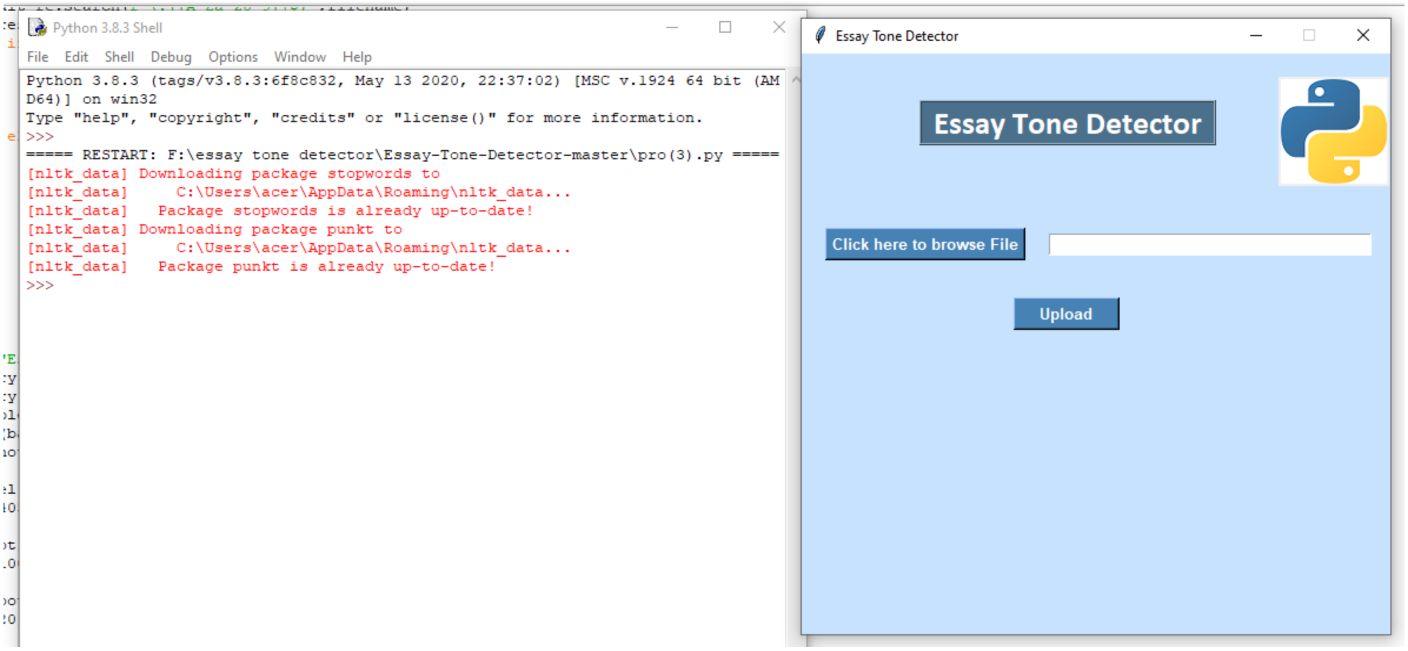
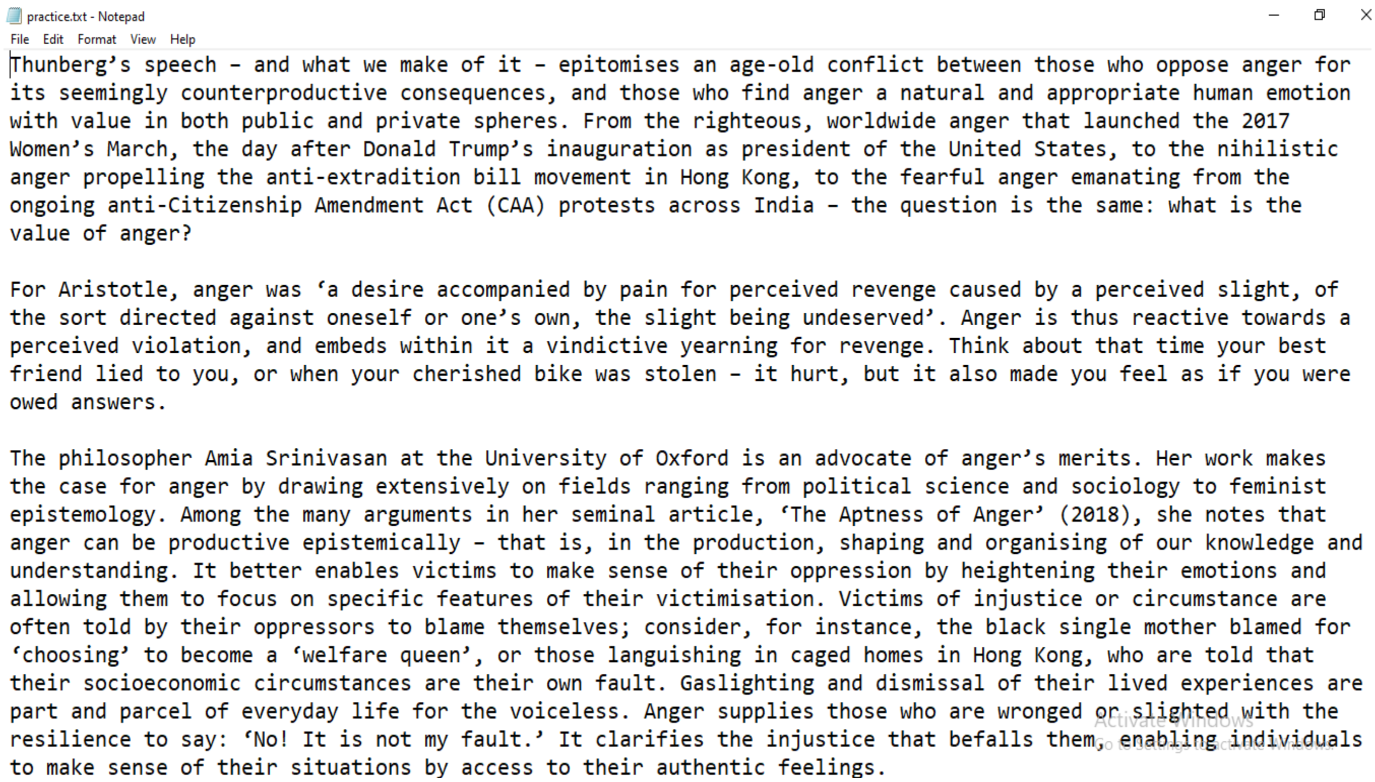


Figure 7.1: OUTPUT WINDOW WHEN PROGRAM IS RUN

4.4 INPUT TEXT FILE:


practice.txt - Notepad

File Edit Format View Help

Thunberg’s speech – and what we make of it – epitomises an age-old conflict between those who oppose anger for its seemingly counterproductive consequences, and those who find anger a natural and appropriate human emotion with value in both public and private spheres. From the righteous, worldwide anger that launched the 2017 Women’s March, the day after Donald Trump’s inauguration as president of the United States, to the nihilistic anger propelling the anti-extradition bill movement in Hong Kong, to the fearful anger emanating from the ongoing anti-Citizenship Amendment Act (CAA) protests across India – the question is the same: what is the value of anger?

For Aristotle, anger was ‘a desire accompanied by pain for perceived revenge caused by a perceived slight, of the sort directed against oneself or one’s own, the slight being undeserved’. Anger is thus reactive towards a perceived violation, and embeds within it a vindictive yearning for revenge. Think about that time your best friend lied to you, or when your cherished bike was stolen – it hurt, but it also made you feel as if you were owed answers.

The philosopher Amia Srinivasan at the University of Oxford is an advocate of anger’s merits. Her work makes the case for anger by drawing extensively on fields ranging from political science and sociology to feminist epistemology. Among the many arguments in her seminal article, ‘The Aptness of Anger’ (2018), she notes that anger can be productive epistemically – that is, in the production, shaping and organising of our knowledge and understanding. It better enables victims to make sense of their oppression by heightening their emotions and allowing them to focus on specific features of their victimisation. Victims of injustice or circumstance are often told by their oppressors to blame themselves; consider, for instance, the black single mother blamed for ‘choosing’ to become a ‘welfare queen’, or those languishing in caged homes in Hong Kong, who are told that their socioeconomic circumstances are their own fault. Gaslighting and dismissal of their lived experiences are part and parcel of everyday life for the voiceless. Anger supplies those who are wronged or slighted with the resilience to say: ‘No! It is not my fault.’ It clarifies the injustice that befalls them, enabling individuals to make sense of their situations by access to their authentic feelings.

Figure 7.2: Input Text File

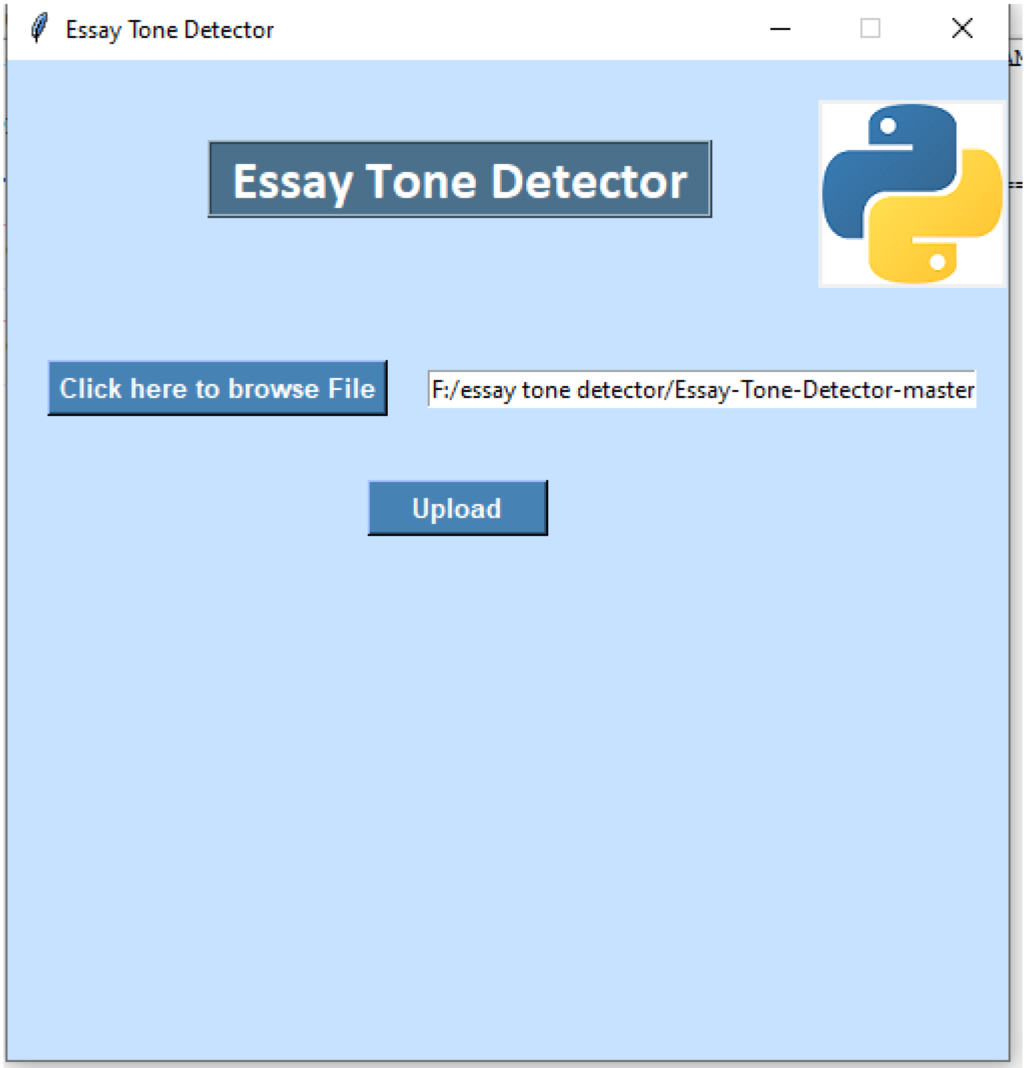


Figure 7.3: UPLOADING THE INPUT TEXT FILE.

4.5 FINAL OUTPUT:

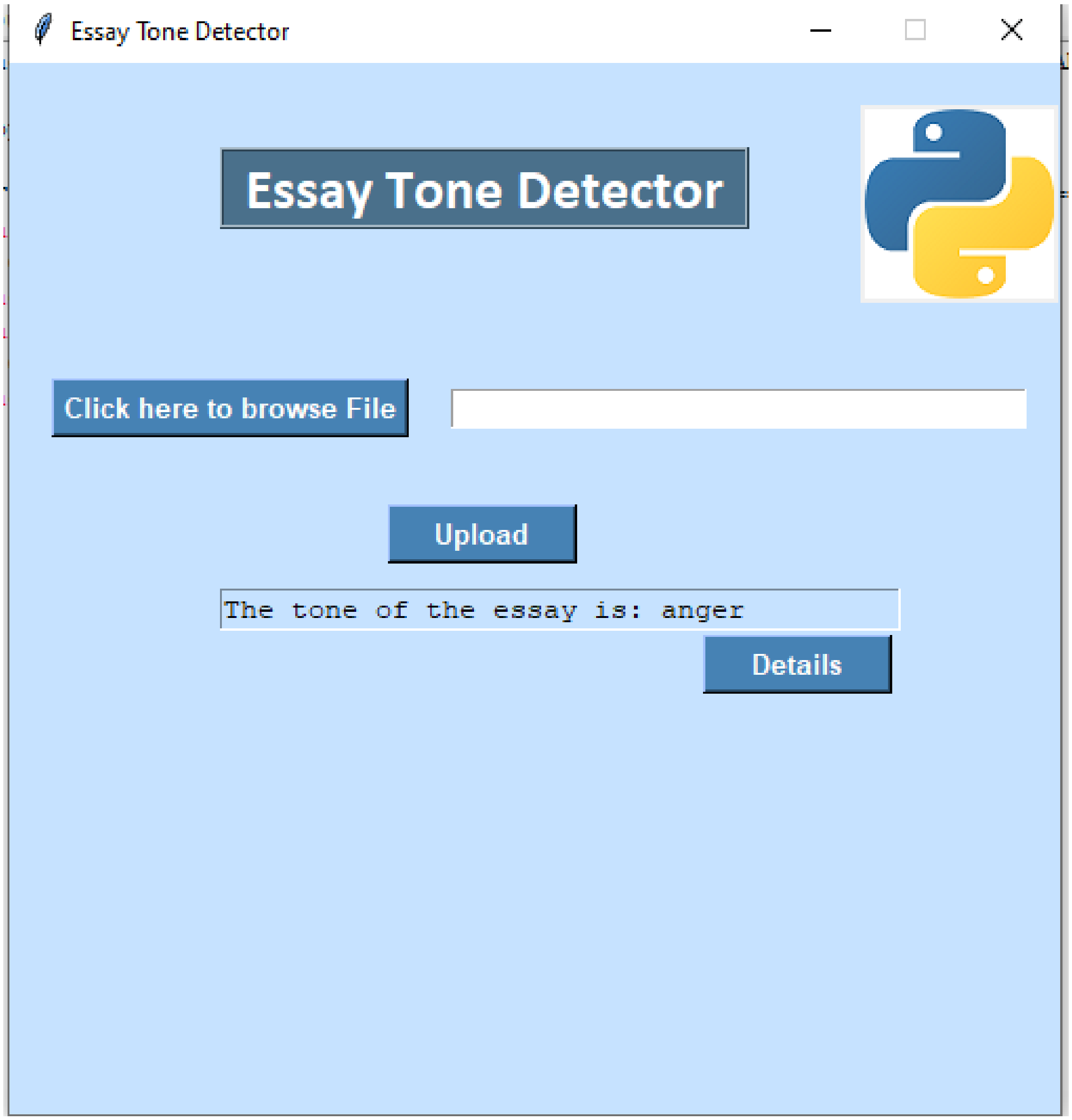


Figure 7.4: The figure shows the tone of the essay.

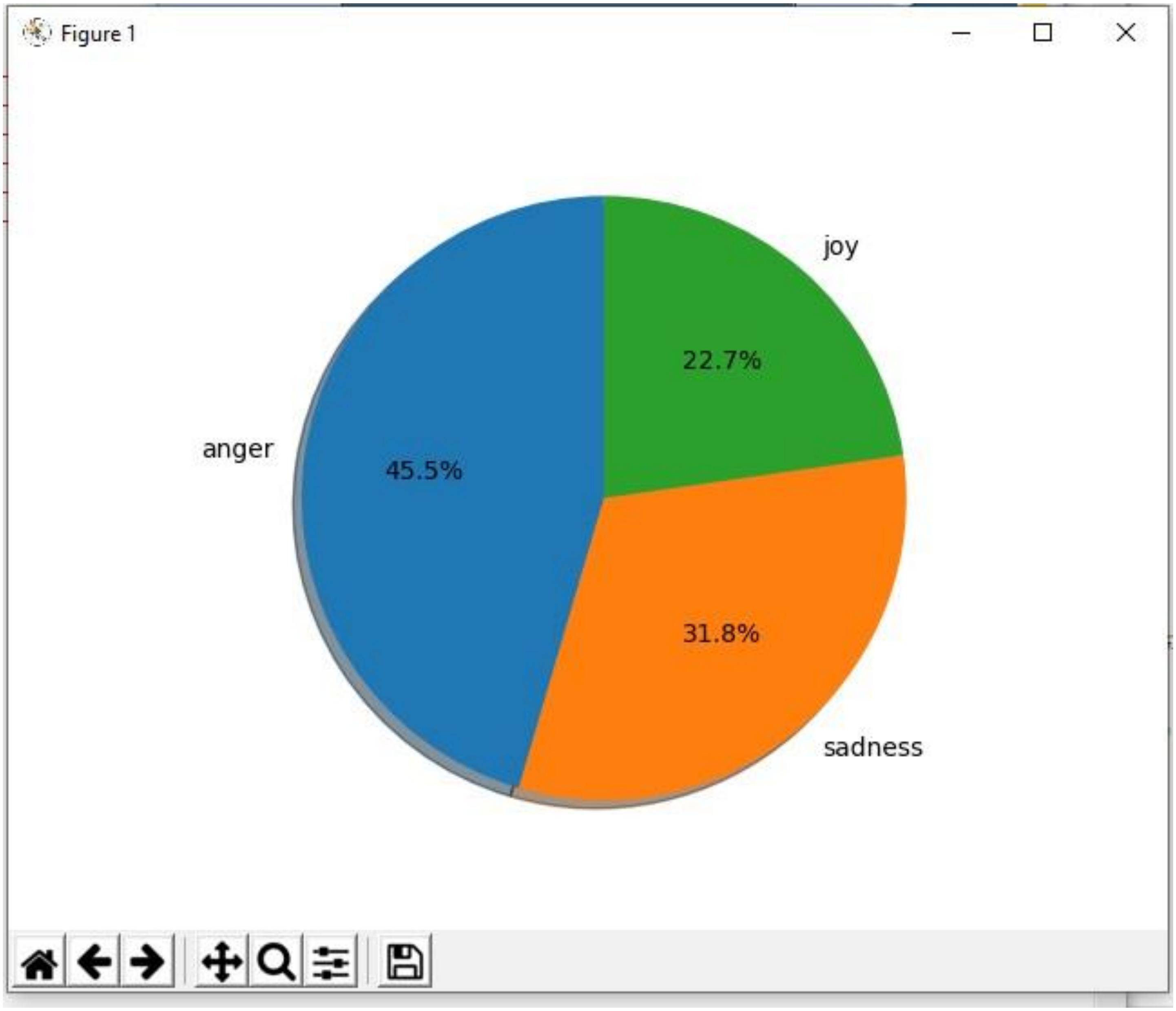


Figure 7.5: Figure depicting details of various emotions via pie chart.

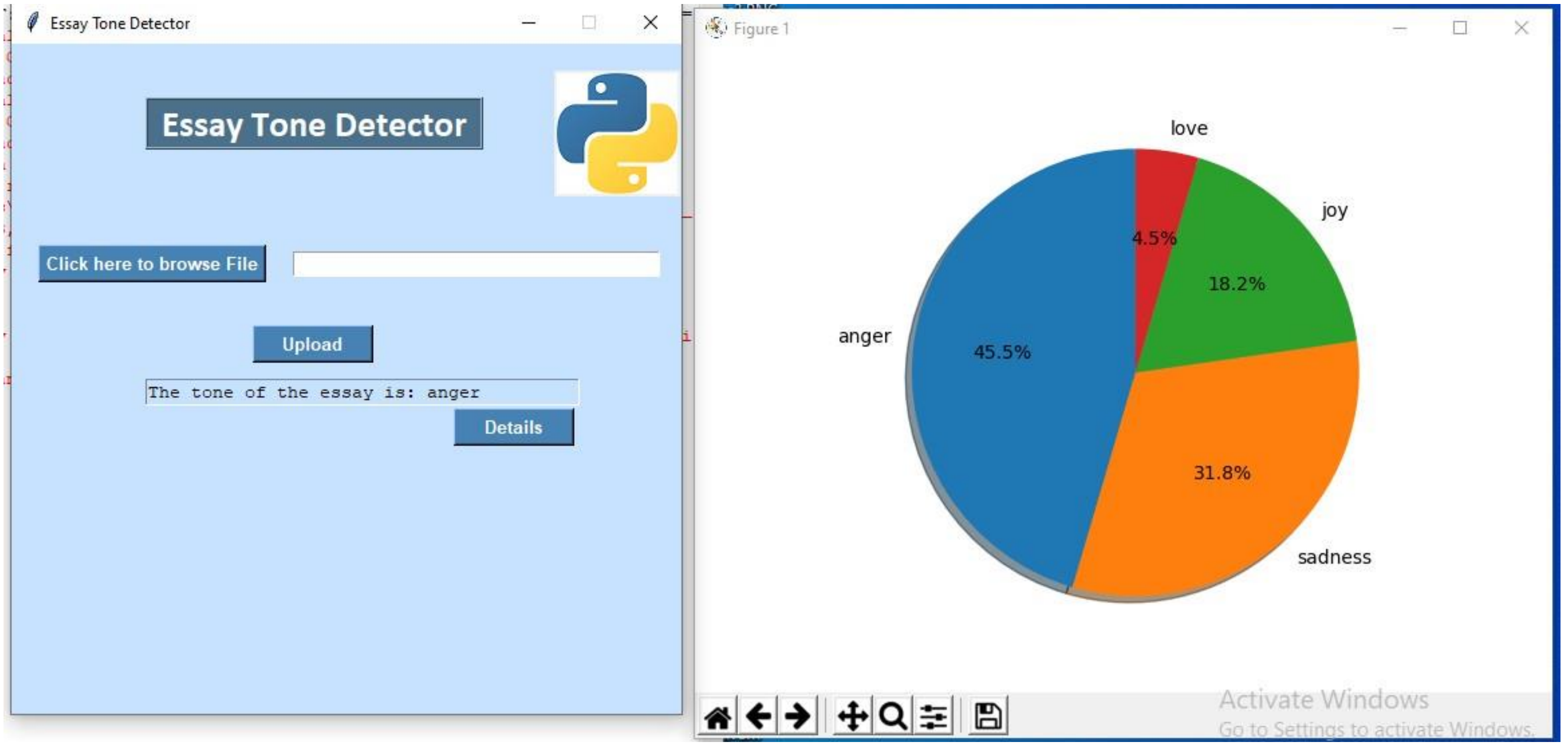


Figure 7.6: Combined image.

Chapter 5
Conclusion

In this chapter, we observed how machine learning can be used to implement essay tone detection. Various other machine learning algorithms can be used but we have used logistic regression.

To conclude, we have discussed the whole process of development of this system and we can rely on this system.

Future Scope

This chapter discusses the future scope or the implementation of this robot. To increase the scope of this device we can add some new features. As technology is becoming more advance it will be mandatory to change the structure some day with better replacement and sometimes based on customer requirements. The results that we achieved are as follows:

- The model can be used by a non-IT technician.
- The model is ready to be of commercial use.
- The system has the capacity to carry up a large amount of textual data.
- The system can serve as much people as they want within anorganization.

References

i. . www.kaggle.com/datasets.

ii. https://www.datacamp.com/community/tutorials/text-analytics-beginners-nltk.

iii. https://www.grammarly.com/blog/tone-and-emotions/.

iv. Abdul Hanan et al. Emotion Detection of Text, International Journal of Engineering Research and Development e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com Volume 11, Issue 07 (July 2015), PP.23-34.

v. https://aeon.co/essays/anger-is-a-valuable-emotion-driving-private-and-public-good.

final project

ORIGINALITY REPORT

16%	14%	4%	7%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	docplayer.net Internet Source	8%
2	www.ijitee.org Internet Source	3%
3	www.ijerd.com Internet Source	2%
4	Submitted to The British College Student Paper	1%
5	Submitted to National Institute of Technology Delhi Student Paper	1%
6	Submitted to University of KwaZulu-Natal Student Paper	<1%
7	www.greycampus.com Internet Source	<1%
8	Submitted to Bharati Vidyapeeth Deemed University College Of Engineering Student Paper	<1%

9	Submitted to CITY College, Affiliated Institute of the University of Sheffield Student Paper	<1%
10	Submitted to University of Sheffield Student Paper	<1%
11	Submitted to ASA Institute Student Paper	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: 16-07-2020

Type of Document (Tick): PhD Thesis | M.Tech Dissertation/ Report | B.Tech Project Report ✓ | Paper

Name: SHIVAM GARG **Department:** CSE **Enrolment No** 161298

Contact No. 9805057113 **E-mail.** SHIVAMSIDGARG@GMAIL.COM

Name of the Supervisor: DR. RAJNI MOHANA

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): ESSAY TONE DETECTOR

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages = 45
- Total No. of Preliminary pages = 08
- Total No. of pages accommodate bibliography/references = 03

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at 16 (%). Therefore, we are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

Rajni Mohana

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Generated Plagiarism Report Details (Title, Abstract & Chapters)	
	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String		Word Counts	
Report Generated on			Character Counts	
		Submission ID	Total Pages Scanned	
			File Size	

Checked by
Name & Signature **Librarian**

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com