PDPM Indian Institute of Information Technology,
Design and Manufacturing, Jabalpur-482005

CS314b

SENTIMENT ANALYSIS

Machine Learning Project Report

Submitted To:

Dr Kusum Kumar Bharti

Submitted By:

Kanika Dhiman (2016115)

Kushdeep Mittal (2016135)

Sanjeev Singh (2016226)

Abstract

Emotions form a very important and basic aspect of our lives. Whatever we do, whatever we say, somehow does reflect some of our emotions, though may not be direct. To understand the very fundamental behaviour of a human, we need to analyze these emotions through some emotional data. This data can be text, voice, facial expressions, etc.

In this project, we have implemented emotion analysis for two broader categories i.e. positive, negative and neutral using **Linear SVM.**

Our results show that it is possible to perform sentiment analysis on spontaneous sentences, whether the sentence is positive, neutral or negative.

Introduction

Sentiment analysis, also called opinion mining, is the field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. Sentiment analysis and opinion mining mainly focus on opinions which express or imply positive or negative sentiments. Sentiment analysis is perhaps one of the most popular applications of **NLP** (Natural Language Processing).

Applications:

Enhancing the Customer Experience through Sentiment Analysis in Business:

A business breathes on the gratification of its customers. The experience of the customers can either be positive, negative or neutral. Owing to the internet savvy era, this experience becomes the text of their social posting and online feedback. The tone and temperament of this data can be detected and then categorized according to the sentiments attached. This helps to know what is being properly implemented with regards to products, services and customer support and what needs improvement.

Sentiment Analysis in Business for Brand Brisking:

A brand is not defined by the product it manufactures or the services it provides. The name and fame that build a brand majorly depend on their online marketing, social campaigning, content marketing and customer support services. Sentiment analysis in business helps in quantifying the perception of the present and the potential customers regarding all these factors. Keeping the negative sentiments in knowledge, you can develop more appealing branding techniques and marketing strategies to switch from torpid to terrific brand status. Sentiment analysis in business can majorly help you to make a quick transition.

Background and Problem Description

One subproblem of NLP is sentiment analysis, i.e classifying a statement as **positive, neutral or negative**. For example, on any commercial website, it's users can leave a comment about a product stating whether it was good, bad or it could even be neutral. Now, using a human to read all the comments and obtaining the overall customer feedback on the product would be expensive and time-consuming. The machine learning model can churn through a vast amount of data, making inferences and classifying the comment. Using this ML model, the website can better its products through the customer reviews which would bring in more revenue for the company.

Level of Analysis:

Sentiment analysis has been investigated mainly at three levels:

a) **Document-level**: The task at this level is to classify whether a whole opinion document expresses a positive or negative sentiment
b) **Sentence level**: The task at this level goes to the sentences and determines whether each sentence expressed a positive, negative, or neutral opinion. Neutral usually means no opinion. This level of analysis is closely related to subjectivity classification.
c) **Entity and Aspect level**: Both the document level and the sentence level analyses do not discover what exactly people liked and did not like.

In this project, we have focused upon the **sentence level analysis.**

We have taken up the tweets of different airlines and trained our dataset on the same.

Related Work

A) **Expression of sentiments** - Providing the flavours of polarity by identifying if the positive, neutral or negative sentiment.

B) **Datasets** - We are going to look at tweets from different airlines to build our sentiment analysis model.

C) **Emotion models** - The major step in a machine learning text classifier is to transform the text into a numerical representation, usually a vector or matrix. Usually, each component of the vector represents a word or expression in a predefined dictionary. This process is known as feature extraction or text vectorization.

D) **Computational approaches** - There exist many algorithms that can be used for sentiment analysis. The classification step usually involves a statistical model like Naïve Bayes, Logistic Regression, Support Vector Machines, or Neural Networks. In this project, **Linear SVM** has been used for the implementation for the same.

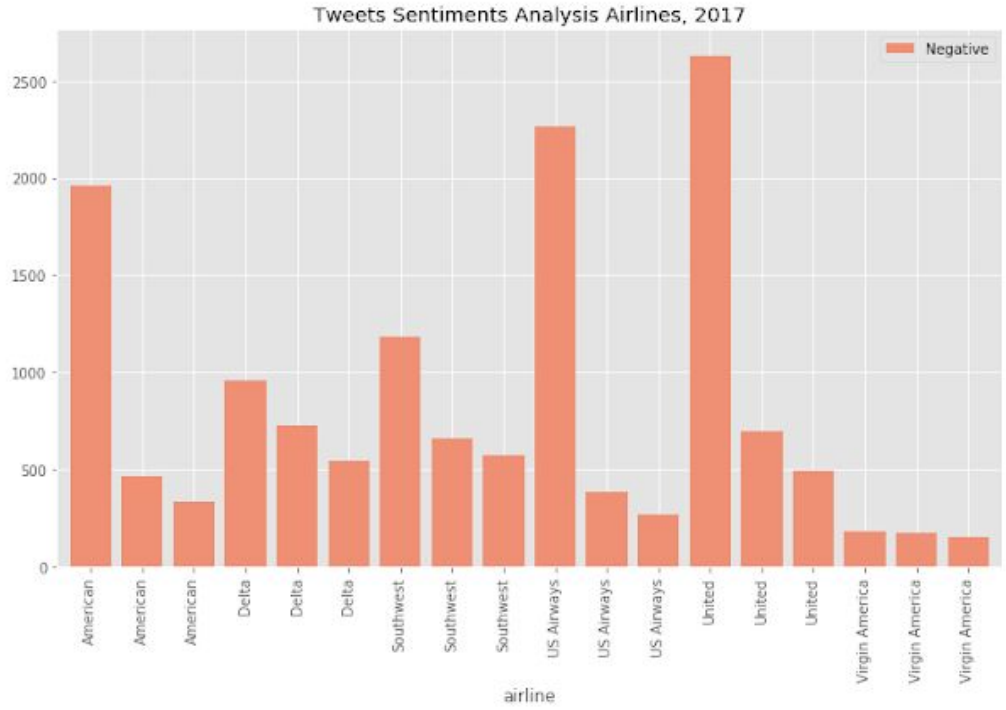
WHY SVM? Shows up a pretty good accuracy with SVM bigrams.

Proposed Solution

1. **Loading and analysing dataset:**

   The dataset is being loaded using the pandas library.

   The number of positives, negatives and neutrals are being observed for different airlines.



2. **Dataset Operations:**

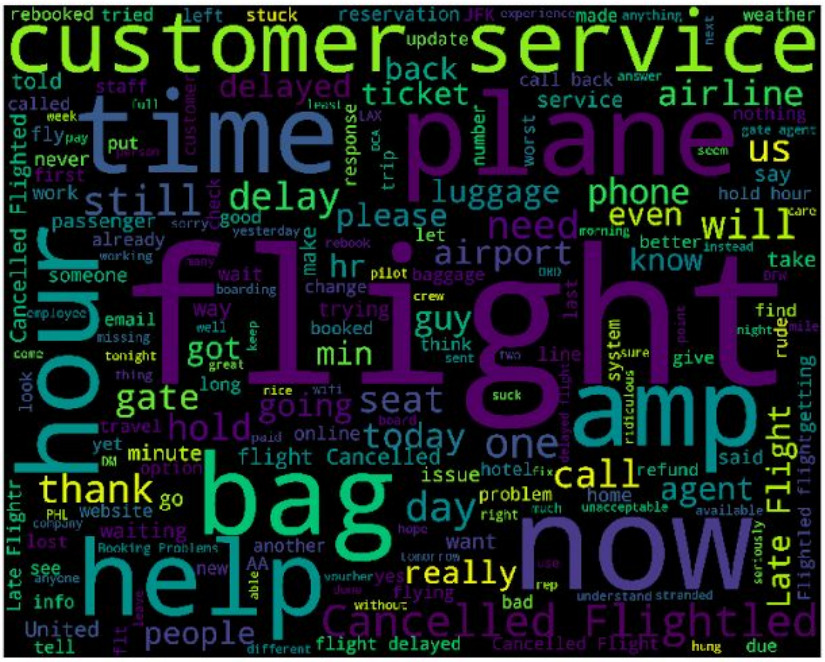
   Total Tweets: 14641

   Lemmatizer has been used to convert words into meaningful ones.

   Extracting the bigrams and trigrams out of each tweet

   Splitting the dataset into 60% of training data, 40% of test data

Negative words



Positive words

Word cloud of negative and positive words

3. Training using linear SVM

Training data (normalized bigrams and trigrams)is used for training the model using Linear SVM.

4. Checking on Test data

Testing on 40% of the test data. The accuracy achieved: 78.51%

```
In [40]: from sklearn import svm
         from sklearn.multiclass import OneVsRestClassifier
         from sklearn.model_selection import cross_val_score

         clf = OneVsRestClassifier(svm.SVC(gamma=0.01, C=100., probability=True, class_w
         # scores = cross_val_score(clf, indexed_data, targets, cv=1)
         # scores
         clf_output = clf.fit(data_train, targets_train).decision_function(data_test)

In [56]: clf.score(data_test, targets_test)

Out[56]: 0.7851775956284153
```

5. Predicting on new sentences

Using the trained model, probability distribution on negative, positive and neutral can be shown.

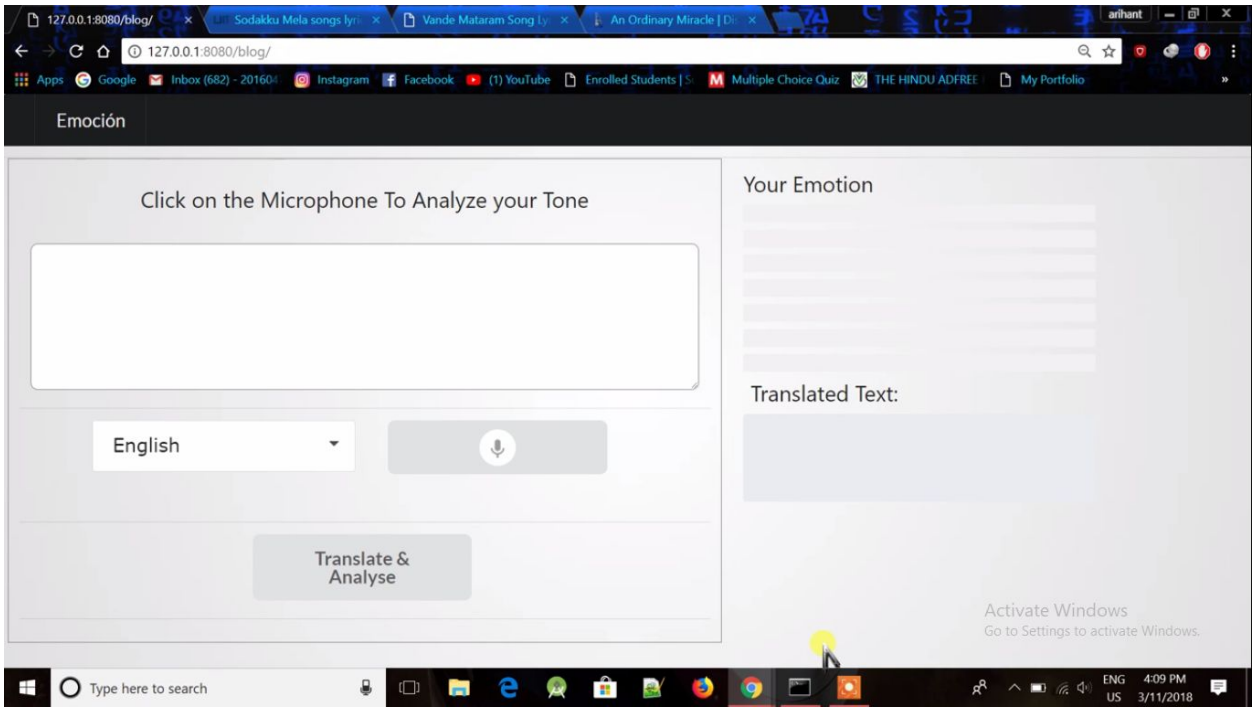
```
In [61]: sentences = count_vectorizer.transform([
             "What a great airline, the trip was a pleasure!",
             "My issue was quickly resolved after calling customer support. Thanks!",
             "What the hell! My flight was cancelled again. This sucks!",
             "Service was awful. I'll never fly with you again.",
             "You fuckers lost my luggage. Never again!",
             "I have mixed feelings about airlines. I don't know what I think.",
             ""
         ])
         clf.predict_proba(sentences)

Out[61]: array([[0.20722808, 0.05923751, 0.73353441],
                [0.14036472, 0.07158684, 0.78804844],
                [0.94281112, 0.04041069, 0.01677819],
                [0.89040329, 0.07339116, 0.03620555],
                [0.9735437 , 0.01760042, 0.00885589],
                [0.46770982, 0.50194496, 0.03034521],
                [0.26527808, 0.52255681, 0.21216511]])
```

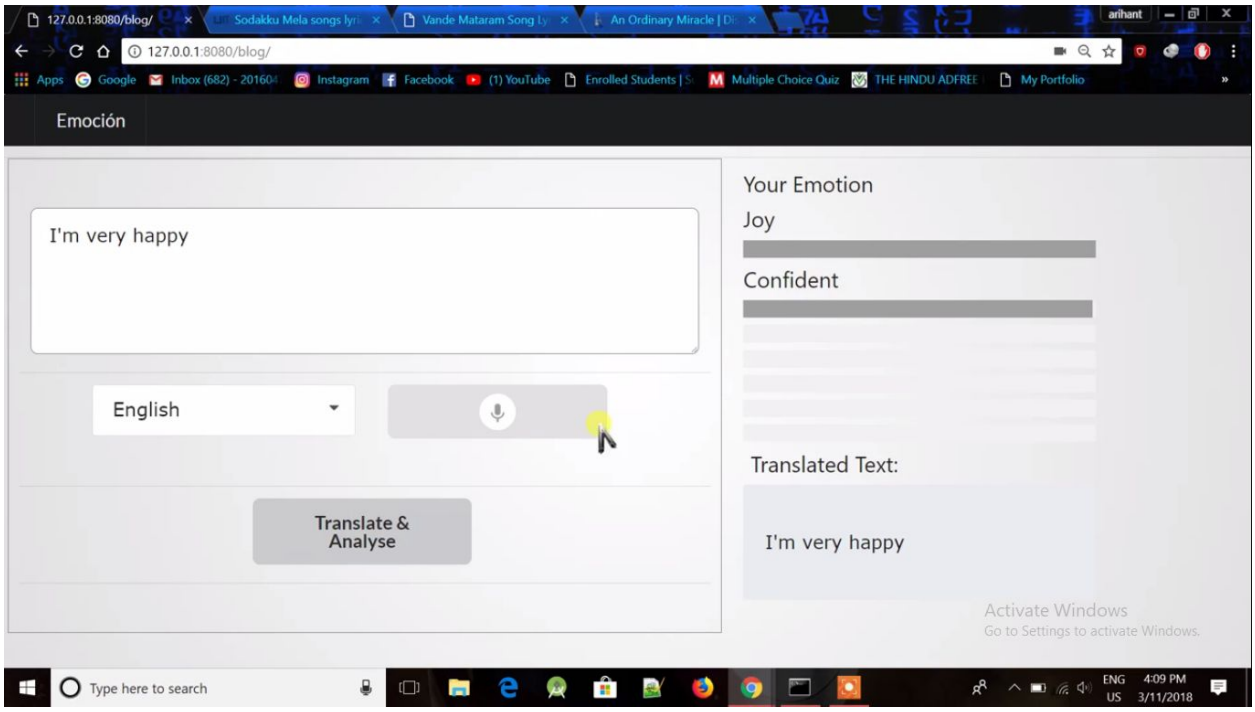
Giving probability distribution on negativity, neutral and positivity.

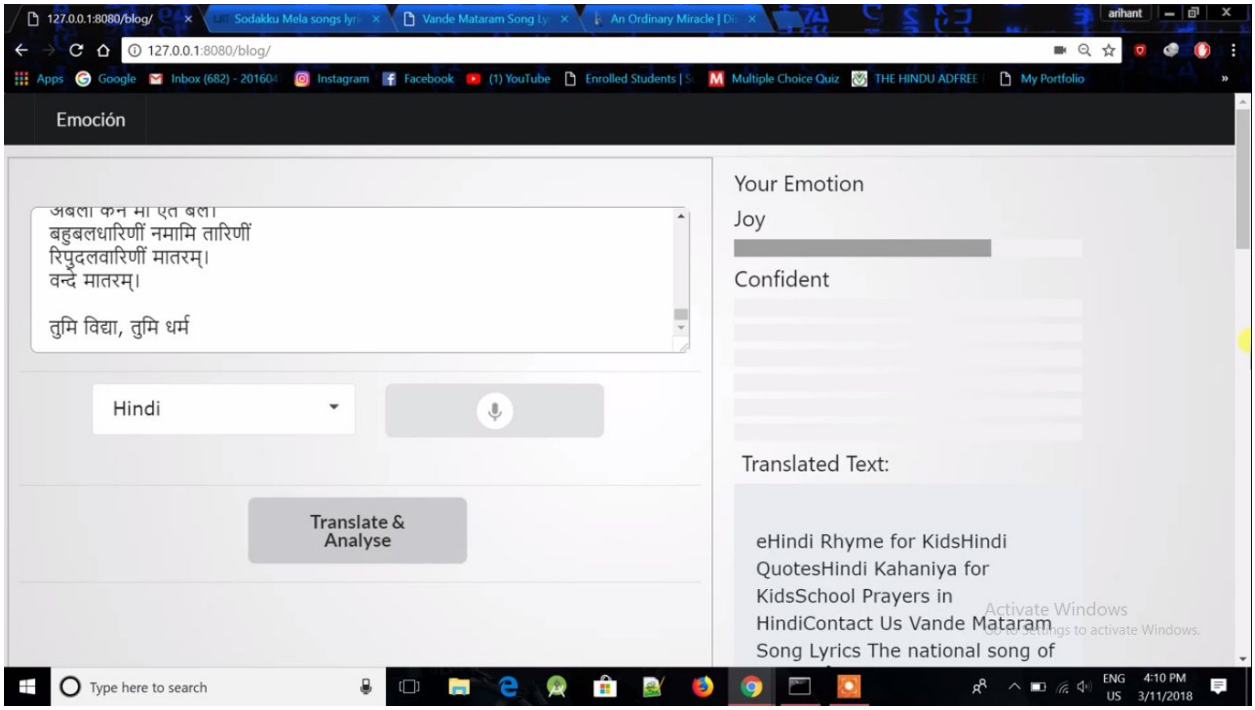
User Interface

For User Interface Technologies used are :

- Django (Python framework)
- Ajax (Javascript Library)
- Javascript




127.0.0.1:8080/blog/
Emoción
I'm very happy
English
Translate & Analyse
Your Emotion
Joy
Confident
Translated Text:
I'm very happy
Activate Windows
Go to Settings to activate Windows.
Type here to search
ENG US
4:09 PM
3/11/2018


127.0.0.1:8080/blog/
Emoción
बहुबलधारिणीं नमामि तारिणीं
रिपुदलवारिणीं मातरम्।
वन्दे मातरम्।
तुमि विद्या, तुमि धर्म
Hindi
Translate & Analyse
Your Emotion
Joy
Confident
Translated Text:
eHindi Rhyme for KidsHindi QuotesHindi Kahaniya for KidsSchool Prayers in HindiContact Us Vande Mataram Song Lyrics The national song of
Activate Windows
Type here to search
ENG US
4:10 PM
3/11/2018

Evaluation Study and Results

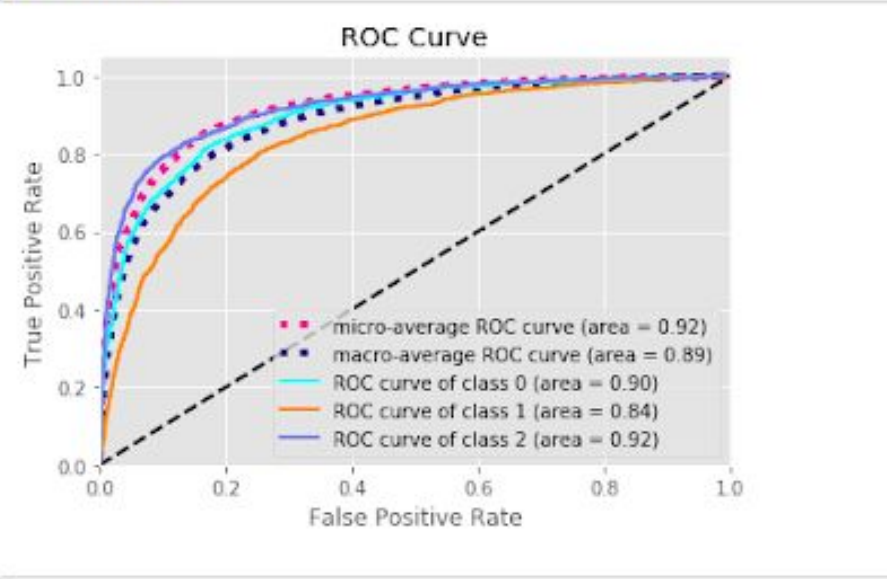
Data set size: 14641 tweets

Test data set: 40%

i) The accuracy achieved on test data: 78.51%

```
In [40]: from sklearn import svm
         from sklearn.multiclass import OneVsRestClassifier
         from sklearn.model_selection import cross_val_score

         clf = OneVsRestClassifier(svm.SVC(gamma=0.01, C=100., probability=True, class_w
         # scores = cross_val_score(clf, indexed_data, targets, cv=1)
         # scores
         clf_output = clf.fit(data_train, targets_train).decision_function(data_test)

In [56]: clf.score(data_test, targets_test)

Out[56]: 0.7851775956284153
```

ii) ROC curve



Class 0: Negative, Class 1: Neutral, Class 2: Positive

Conclusion

This project has discussed a study of the mechanisms and methodology in sentiment analysis. The implementation of the same using Linear Support Vector Machine has been successfully done with a decent accuracy of 78.51%.

Future Scope

There is a lot of scope in analyzing the video and images on the web. Nowadays, with the advent of Facebook, Instagram and Video vines people are expressing their thoughts with pictures and videos along with the text. Sentiment analysis will have to pace up with this change. (Speech -> Text -> Sentiment Analysis)

The most important bit for sentiment analysis in the future has less to do with improving the accuracy of the algorithms but instead lies in the area of determining where you can correlate sentiment with behaviour.

Despite all the challenges and potential problems that threaten sentiment analysis, one cannot ignore the value that it adds to the industry. Because sentiment analysis bases its results on factors that are so inherently humane, it is bound to become one of the major drivers of many business decisions in future. Improved accuracy and consistency in text mining techniques can help overcome some current problems faced in sentiment analysis.

References

[1] Emotion analysis: A survey

https://www.researchgate.net/publication/319362855_Emotion_analysis_A_survey

[2] Sentiment Analysis and Opinion Mining

https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf

[3] Sentiment Extraction from Natural Audio Streams

http://www.utd.edu/~john.hansen/Publications/CP-ICASSP13-KaushikSangwanHansen-Sentiment-0008485.pdf

[4] A Review on Sentiment Analysis and Text-To-Speech

https://pdfs.semanticscholar.org/22a2/080cfe7dd3d2e29ecd8e333bedaf7cbeb629.pdf

[5] Sentiment Analysis

https://monkeylearn.com/sentiment-analysis/