R.V. COLLEGE OF ENGINEERING,
BANGALORE-560059
(Autonomous Institution Affiliated to VTU, Belgaum)

SELF STUDY REPORT ON

"Analysis of Text and Audio Reviews using ML Algorithm"

Submitted by

Name	USN
Ranjith G Hegde	1RV16CS112
Ujwalakavya J	1RV17CS435

Submitted to

Mrs. Sneha M	Ms. Veena Gadad	Ms. Pratiba D
Assistant professor	Assistant professor	Assistant professor

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
R.V.College of Engineering, Bangalore-59

R.V. COLLEGE OF ENGINEERING, BANGALORE - 560059

(Autonomous Institution Affiliated to VTU, Belgaum)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the Self Study work titled **"Analysis of Text and Audio Reviews using ML Algorithm"** is carried out by **RANJITH G HEGDE (1RV16CS112), UJWALAKAVYA J (1RV17CS435)** who is bonafide students of R.V College of Engineering, Bangalore, in partial fulfillment for the academic requirement of **Sixth semester** Bachelor of Engineering in Computer Science and Engineering of the Vishveshwarya Technological University, Belagavi during the year 2018-2019. It is certified that all corrections/suggestions indicated for the internal Assessment have been incorporated in the report. The Self Study report has been approved as it satisfies the academic requirements in respect of Self Study work prescribed by the institution for the said degree.

Mrs. Sneha M	Ms. Veena Gadad	Ms. Pratiba D
Assistant professor	Assistant professor	Assistant professor

TABLE OF CONTENTS

Chapter 1

REVIEW OF PROBLEM

Text and Audio reviews analysis using Deep learning based sentiment analysis model and speech recognition.

Sentiment analysis is contextual mining of text which identifies and extracts subjective information in source material and helping a business to understand the social sentiment of their brand, product or service by monitoring online reviews.

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format.

In the ever growing market of today customer feedback is very crucial in a company's survival. The gathering and accurate analysis of data is very hard which is the problem that needs to be solved in the industry. Customer feedback is an area in the current market which is not much focused on by many companies. So there is a huge scope for growth in this field which can be seized.

The objective of the project:

- To develop Android application that accepts customer feedback and reviews either as text or audio input.
- To build automatic speech recognition model that converts audio to text.
- Efficiently store and perform sentiment analysis on feedback to get market insights and trends that are prevailing.
- Provide the managers with easy to read dashboard and other information suggesting feedback given from customer how positive or negative they are which can help company's growth.

Chapter 2

METHODOLOGY

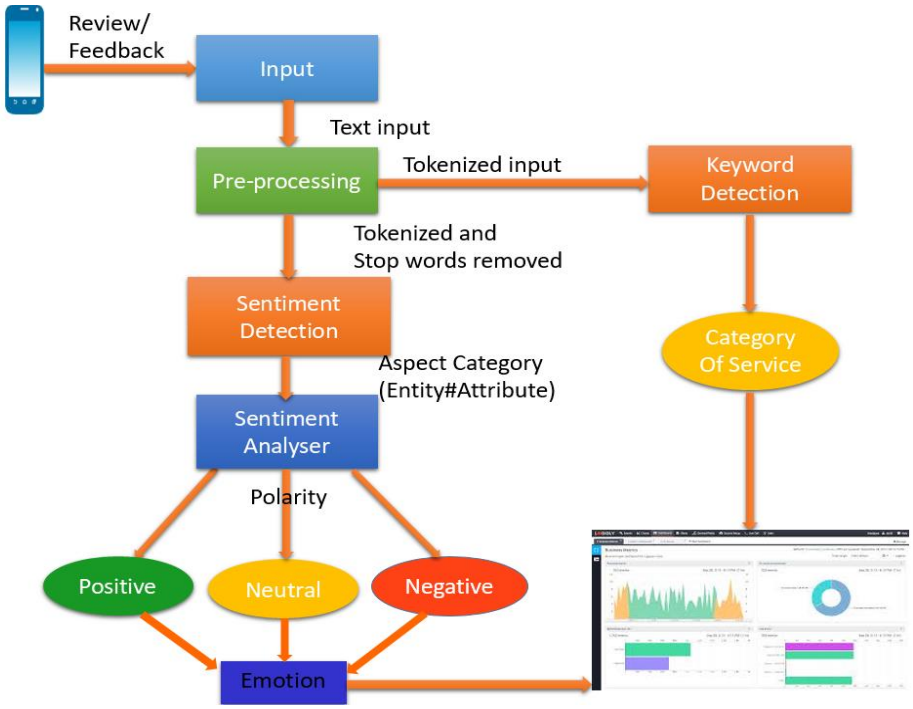

Fig 1. Flow Chart

- **Data Collection**

  Review is accepted from the customer through android application either in audio or text format.

- **Audio to Text Conversion**

  Using speech recognition model the audio is converted to text within the android application and sent for analysis.

- **Text Preparation**

  The extracted Text is filtered before analysis. It includes tokenizing, identifying and eliminating non-textual content (Stop words) and content that is irrelevant to the area of study from the data.

- **Sentiment Detection**

  At this stage, each sentence of the review and opinion is examined for subjectivity. Sentences with subjective expressions are retained and that which conveys objective expressions are discarded.

- **Sentiment Classification**

  Sentiments can be classified into groups, positive, negative and neutral. At this stage of analysis, each subjective sentence detected is classified into groups-positive, negative, neutral.

- **Topic Modelling**

  Once the sentiment is classified we get the keywords from sentence to know based on what was the review about.

- **Presentation of Output**

  After the completion of analysis, the text results are displayed on website for further analysis and future prediction.

Chapter 3

CORRELATION WITH SELF STUDY SUBJECTS

Computer Networks is used to analyse the transfer of data between the client and the server to appropriately set the server configuration for better performance and load analysis (Apache Tomcat).

Web Programming is used in the front end made in HTML, CSS, JavaScript embedded in the Java Server Pages with bootstrap used for delicate styling.

Compiler Design is very basic as Java uses both compiler and interpreter which makes it very robust.

Chapter 4

USE OF RAPID DEVELOPMENT TOOLS

ANDROID STUDIO

Android Studio is the official integrated development environment, built on JetBrains' IntelliJ IDEA software and designed specifically for Android development.

Eclipse

Eclipse SDK is free and open-source software, used in computer programming, and is the most widely used Java IDE.

The Eclipse Web Tools Platform (WTP) project provides **tools** for developing **Web** and Java EE applications.

Spring Framework

The Spring Framework is an application framework and inversion of control container for the Java platform. The framework's core features provides an extensions for building web applications on top of the Java EE platform.

Stanford CoreNLP

It provides a set of human language technology tools. It gives the base forms of words, their parts of speech, normalize dates, times, and numeric quantities, and syntactic dependencies.

Chapter 5

EXPERIMENTAL ANALYSIS AND PRESENTATION OF RESULT

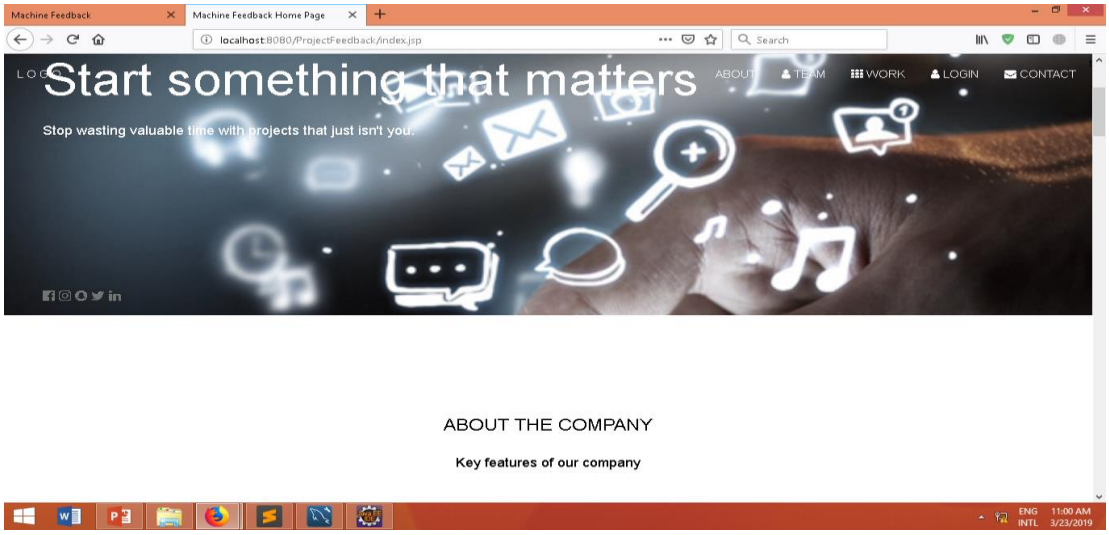

Fig 2. Home Page

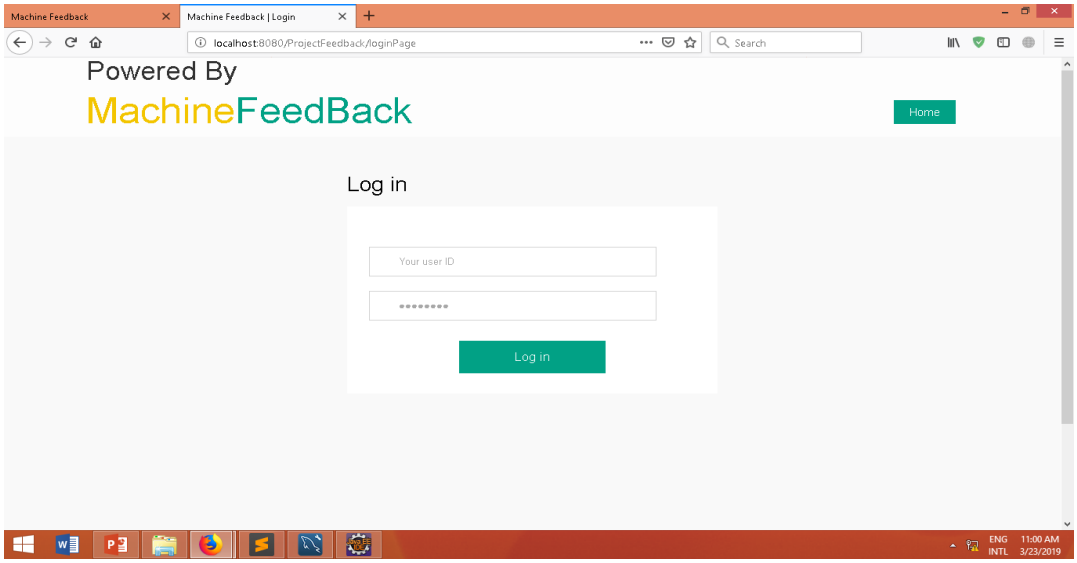

Fig 3. Login Page

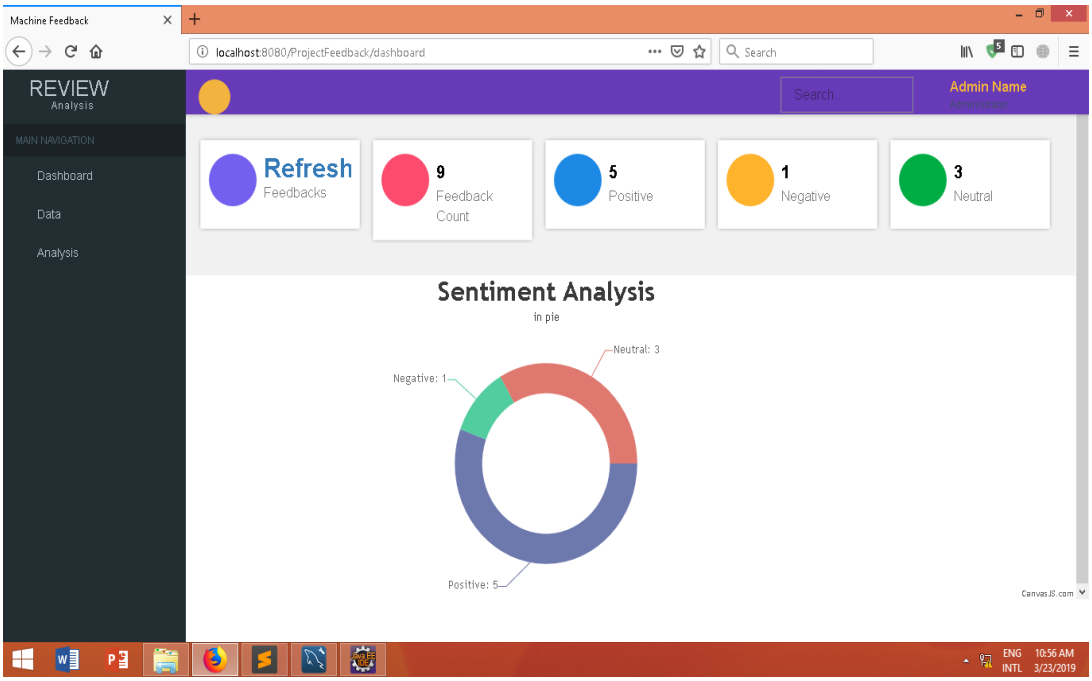

Fig 4. Dashboard and Sentiment Analysis Output

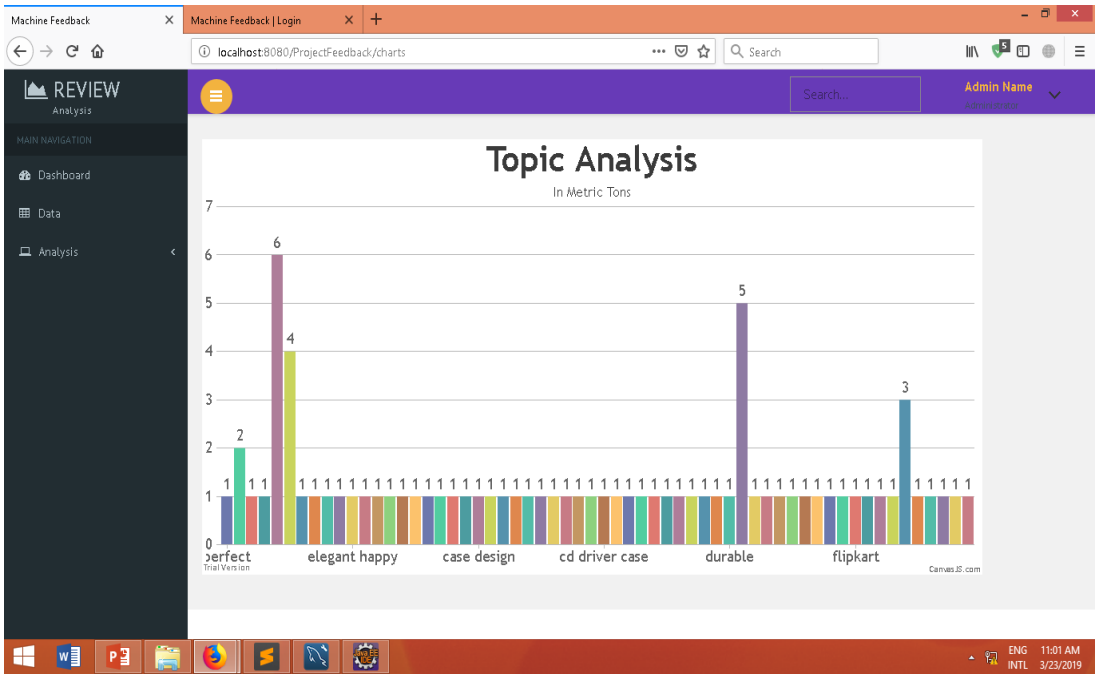

Fig 5. Topic Modelling Result

Chapter 6

CONCLUSION AND FUTURE WORK

A model was successfully built that is capable of analyzing the customer review and detect the topic of the review which can be leveraged to increase the customer base and the profit that is made by the company with its help can be over a specific service they provide or can be a very diverse range of products and services. This is a business tool with great commercial benefit.

The topic modelling will be improved in order to obtain better accuracy and specialization over a particular domain.

Chapter 7

REFERENCES

- https://developer.android.com/
- https://github.com/mozilla/DeepSpeech
- https://stanfordnlp.github.io/CoreNLP/
- https://spring.io/projects/spring-ws
- Dataset- http://en.wikipedia.org/wiki/List_of_blogs