RATING PREDICTION USING REVIEWS

SOFTWARE REQUIREMENT SPECIFICATION

VERSION 1.0

SUPERVISED BY:

MR. SAURABH SINGHAL

EFFORTS BY:-

(GROUP 8)

JANVI PANGORIYA(181500292)

NIDHI GUPTA(181500422)

Table of Contents

Revision History

Date	Version	Description	Author
13/04/2021	1.0	Hotels play a crucial role in travelling and with the increased access to information new pathways of selecting the best ones emerged. This aim of this project is to develop a machine learning model named “Rating Prediction using Reviews” that will be deployed on cloud and be available as a web based application The dataset we are using, consisting of 20k reviews crawled from Trip Advisor,which can explore what makes a great hotel and maybe even use this model in various travelling blogs and websites.	Nidhi Gupta Janvi Pangoriya

Introduction

1.1 Purpose

Guest reviews are becoming a prominent factor affecting people's bookings/purchases. When we look for a place to stay for a vacation on Expedia/Booking/Trip Advisor, we scroll the screen to check on the reviews In other words, guest reviews clearly influence people's booking decision, which means, the social media sites that are related to travelling and travelling blogs better pay attention to what people are saying about various hotels. Reviews can tell you if hotels and restraunts are keeping up with customers' expectations and help you learn the most about customers, which is crucial for developing marketing strategies based on the personas of your customers.

The aim of this project is to develop a Machine Learning Model which predicts the Rating out of the Reviews that are being collected and this can be used as a real world application by various Travel blogs and sites to increase their effectiveness and utility.

1.2 Intended Audience and Reading Suggestions

The document is intended for developers and marketing managers as this type of real world application can be used by various travelling blogs and websites and can be embedded to calculate the rating out of the reviews that are collected on daily basis on their site and hence can easily determine the customer demand and can be beneficial for the tourism industry and Hotel Industry as well.

To get the summary of this project, one can proceed to the use case diagram to understand the problem statement properly and hence read the implementation details to know how the model is developed and deployed over cloud. The Future Scope mentioned extends the scope of the project to various other problems that can be solved using the same approach.

1.3 Product Scope

In the last decades, travelling has changed dramatically due to the evolution and popularisation of information and communication technologies as well as mobile devices, namely, smart phones. Not only travellers are continuously sharing on-line information regarding their travel experiences through ratings, reviews, comments, photos or videos, but the Web became the main source of tourism information. Increasingly, tourists search on

websites, wikis or social networks for information and rely on ratings, reviews or posts of tourists, for decision making. Therefore ratings play a very important role in picture and hence the model we are developing as this model predicts the rating out of the reviews with the help of the dataset and hence can be embedded on any travel blog to check the rating of a review.

1.4 References

The Machine Learning Model that we are developing uses a dataset that contains 20k Reviews taken from Kaggle https://www.kaggle.com/andrewmvd/trip-advisor-hotel-reviewsand whole of the project is available at Github:https://github.com/nidhi2611/Rating_Prediction_Using_Reviews

Overall Description

2.1Product Perspective

This Machine Learning Application is a self-contained product. On the interior it contains a machine learning model that predicts out the Rating out of the Reviews and on the part of User Interface it is displayed as a Web Based Application where you could enter the Reviews and predict out the Rating. The major components of the system are displayed as follows.

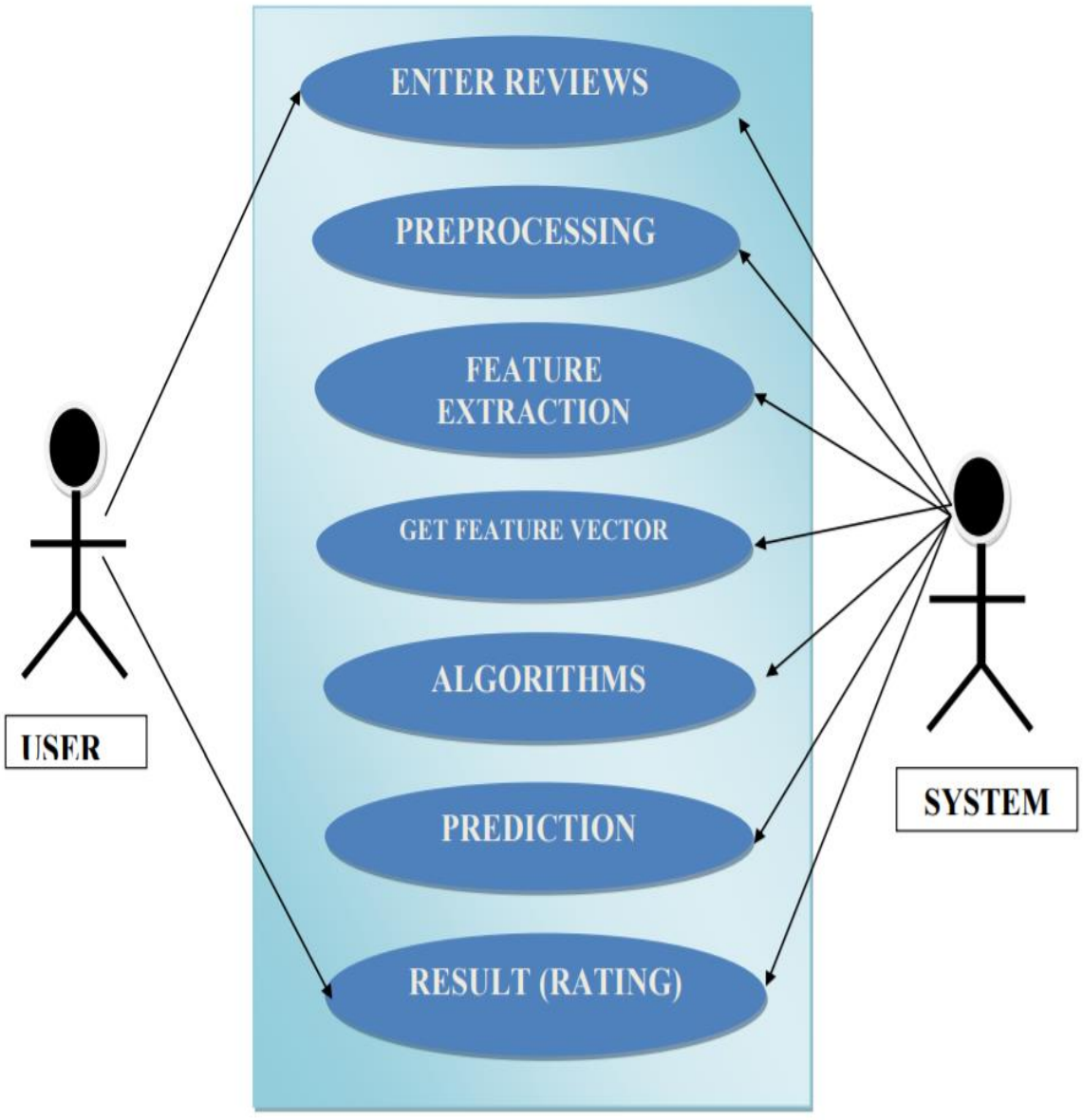

Figure 1: Use Case Diagram

2.2Product Functions

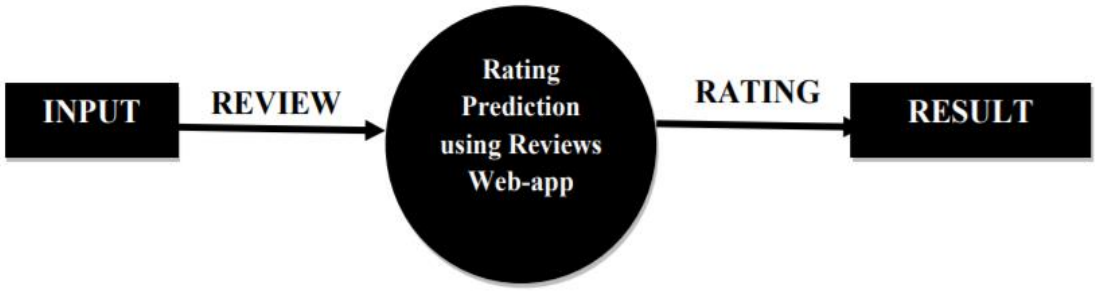

Figure 2: Level -0 DFD

The major function of this product is it provides the user with a web based app which takes the input from the user in the form of the Reviews and the output is the Rating predicted by the machine learning model.

This has got a really wide scope for the Hotel Industry and Tourism industry as everyone who wishes to visit the place first check up on the reviews given by the other people on a particular destination. Above the data flow diagram which further explains the functions of the product.

2.3 User Classes and Characteristics

Although the product holds a single functionality but there are the various users that will use this real life application for their benefit and will be effective in count of every penny the spend using it. Some of the distinguished users that count in are:

- Bloggers: Travel Bloggers or various types of tech writers whose whole of the study is based on the reviews that are given up by the people can make use of this model to choose the place they focus to highlight on.
- General Public: For anyone who has not visited a particular hotel, will definitely look up for reviews on the various sites and hence a good or bad rating might be helpful to make a choice.
- Stake Holders of Hotel Industry: Checking the rating predicted by this model, the stake holders can work upon the feedbacks and requirements of the customer.

2.4 Operating Environment

Once the model is being developed and deployed on any of the cloud platform like Heroku or AWS then all of the computations will take place on the cloud only and the user will be provided with Live Link effectively running on any browser like Chrome / Mozilla Firefox.

2.5 Design and Implementation Constraints

Hardware Requirement:-

- Processor: intel core i5
- Operating System: Windows10
- RAM: 4GB (minimum)

Software Requirement

- Software used :Anaconda
- Deployed :Heroku or AWS

2.6 User Documentation

Along with product there are three documents we are providing

1. Synopsis of the project
2. Software Requirement Specification
3. Final Report of the Project
4. Github repository for the Implementation Details

External Interface Requirements

3.1 User Interfaces

For the User Interface we will be presenting a web based application and that will be containing two WebPages .The first page will be marked for the Input that will be taken from the user. There will be text box where the user will be supposed to enter the review and below it will be action button that will take the user to the result page where the output will be there as predicted by the Machine learning model running in background that will predict the rating from the review entered 1 will be the most depreciated rating value and 5 will be considered as the highest value that can be crawled from the review. For designing the Webpages we have used the various front-end languages like HTML and CSS hence collecting input from the web pages (users) and feeding to the machine learning algorithm and again collecting output from them (the predicted value) and displaying it to the user and hence providing an efficient User Interface. Level I DFD shows user Interactions Interface

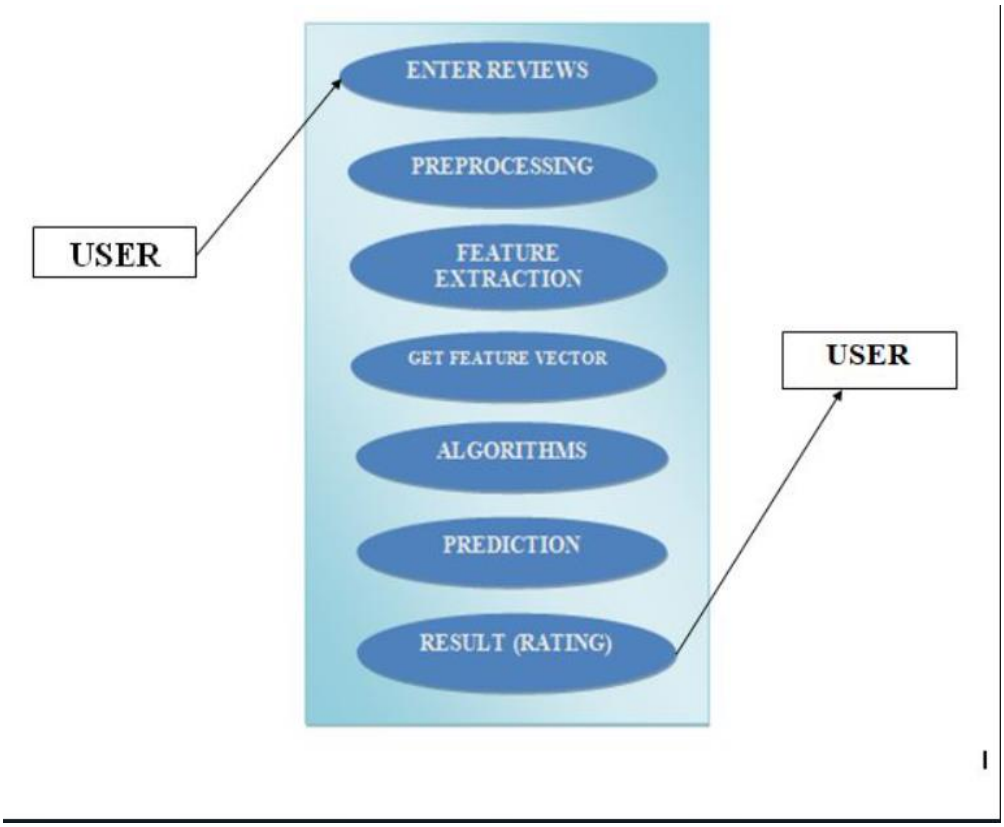

Figure 3: Level -1 DFD

3.2 Hardware Interfaces

Whole of the project is made on softwares,tools, libraries and in such a case no particular hardware is required in particular except for the system we are developing on which requires an intel core i5 processor and a minimum of 4GB RAM.

3.3 Software Interfaces

For developing this product, the machine learning model is developed using the Python language on a software called Anaconda in particular the Jupyter Notebook which holds the implementation of machine learning algorithms and after that the model is deployed on a cloud platform the Heroku App.

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more. Heroku is a cloud platform as a service supporting several programming languages. One of the first cloud platforms, Heroku has been in development since June 2007, when it supported only the Ruby programming language, but now supports Java, Node.js, Scala, Closure, Python, PHP, and Go.

3.4Communications Interfaces

When the user will be interacting with the machine learning model by means of Web Based Application we will be providing the user with the Live link and while predicting the rating the computations will be taking place on the cloud so we need the connection between the web based application and the model deployed on the cloud that can be provided by Internet. So Internet stands as one of the communication interface required.

System Features

To the era we are living in, our choices are made virtually, in an online mode rather than actually visiting and making choices so the only way left with us to verify our choices is the review we get from the other people who have been in the same place as we are in .So having a numerical scale on the review might be beneficial for both on the part of customer to know about the choice he/she is making and the owner to serve the customer better. This is a text classification project which classifies the text review of the customer on a numerical scale and has scope of various other text classifications that can be added to add to the facilities of the customer.

4.1Rating Protection Out of Reviews

4.1.1 Description and Priority

A machine learning model which helps us to analyze the reviews of various customers and to predict the rating using the text classification techniques and applying various algorithms for the same. At the end we would be comparing the results of the algorithms and selecting the one with the maximum accuracy and will be deploying the model on cloud using either the AWS or the Heroku so as to ensure all the computations will be taking place on the cloud and hence this model can be practically implemented as a real time application in the field of text classification and an application of cloud computing as well. The dataset which we will be using will contain reviews from various travelling websites and hence the implementation of this model will help these website to understand and serve their customer well There will be text box where the user will be supposed to enter the review and below it will be action button that will take the user to the result page where the output will be there as predicted by the Machine learning model running in background that will predict the rating from the review entered 1 will be the most depreciated rating value and 5 will be considered as the highest value that can be crawled from the review.

4.1.2 Stimulus/Response Sequences

The response of the application will be displayed on the web based application that the user is interacting with .It will be displaying the rating for which 5 stands as the highest compliment in rating decreasing the value of the comment to 1 which stands for the most underrated comment.

4.1.3 Functional Requirements

The flowchart below explains the various Functional requirements for the model.

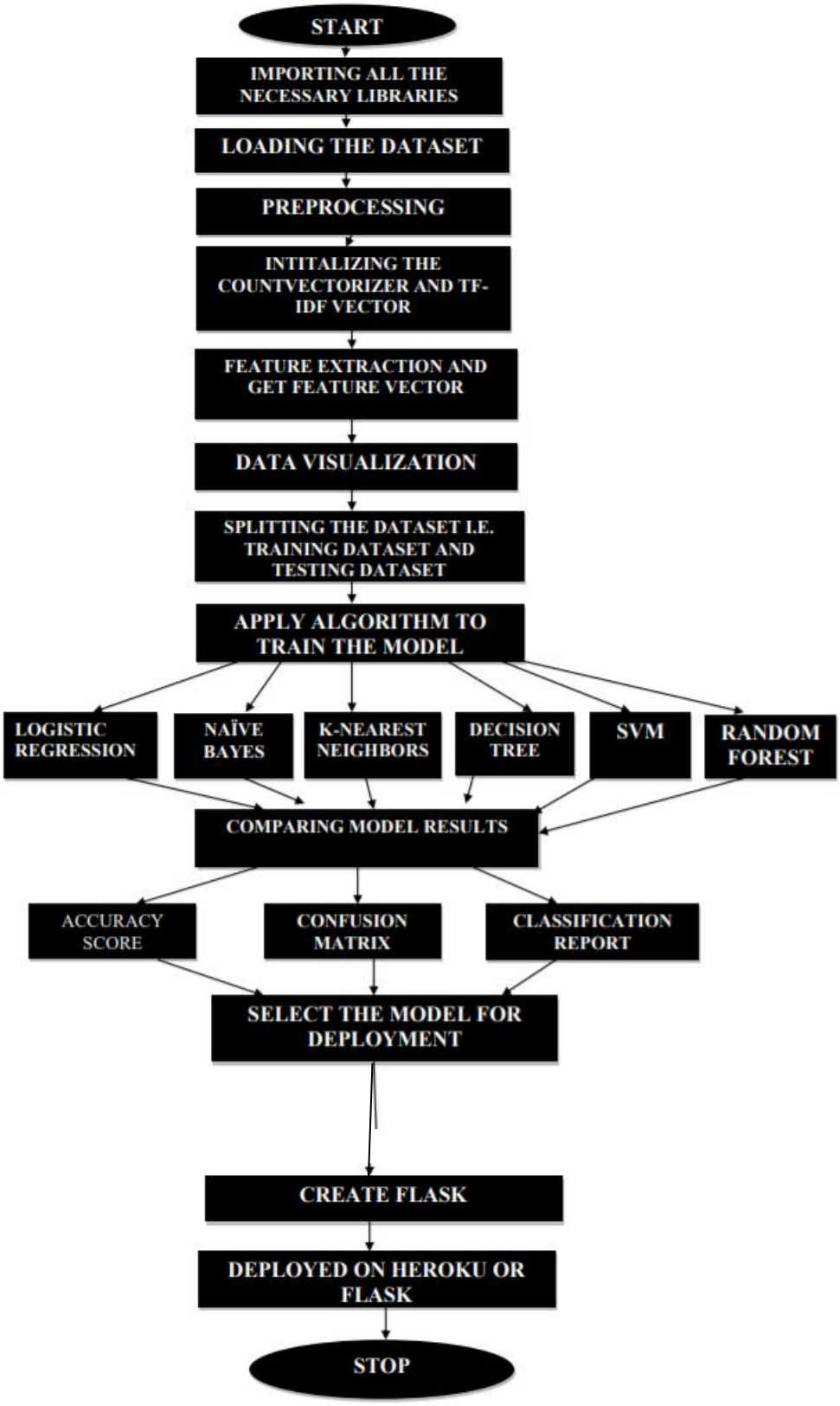

Figure 4: Flow Chart

At each of the step the functional requirements are fulfilled by creating functions**.**