RATING PREDICTION USING REVIEWS: CLOUD COMPUTING PROJECT

Group No. : 8

Project Group Members:

1. Janvi Pangoriya (Sec-D // Roll No. 26 // Univ Roll No. 181500292)
2. Nidhi Gupta (Sec- D // Roll No. 36 // Univ Roll No. 181500422)

Project Supervisor: Mr. Saurabh Singhal, Assistant Professor, Department of Computer Science and Engineering, GLA University

About the Project:

The aim of the project is to create a machine learning model which helps us to analyse 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 and this will open the gateways to other text classification model and real time applications.

Motivation:

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 classification that can be added to add to the facilities of the customer.

Project Planning:

Below, is a rough Gantt Chart that shows phase and approximate time to complete those phases where,

Planned **Completed** **Ongoing**

S. no	Work Elements	Status	February		March				April			
			Wk3	Wk4	Wk1	Wk2	Wk3	Wk4	Wk1	Wk2	Wk3	Wk4
1.	Understanding the problem statement	Planned										
		Actual										
2.	Finding, understanding and loading the dataset	Planned										
		Actual										
3.	Feature Extraction and pre-processing	Planned										
		Actual										
4.	Data Cleaning(tokenisation, removing punctuation and stop words)	Planned										
		Actual										
5.	Data Visualization and some ML Algorithm.	Planned										
		Actual										
6.	Applying other machine Learning Algorithm and deployment	Planned										
		Actual										
7.	Deployment and documentation	Planned										
		Actual										

Tools required:

- **Hardware Requirements:**
  - Processor: intel core i5
  - Operating System :Windows 10
  - RAM : 4GB (minimum)

- **Software Requirements:**
  - Software Used: Anaconda
  - Deployed on: AWS or Heroku

Signature of Project Guide: _____