MANIPAL
INTERNATIONAL UNIVERSITY

Smart Internz

Analysis Of Amazon Cell Phone Reviews Using IBM Watson Services

Student Name : Puvanaah A/P Manokaran

ID : 1106191007

Program : Artificial Intelligence Externship Program

Subject : Natural Language Processing

Date of Submission : 19.09.2021

TABLE OF CONTENTS

I. INTRODUCTION

Mobile phones have revolutionized the way that we purchase products online, making all the information available at our fingertips. Reviews and ratings submitted by consumers became an integral part of the customer's buying decision process. The review and rating platform provided by eCommerce players creates a transparent system for consumers to take decisions and feel confident about it.

II. PROBLEM STATEMENT

The problem statement of this project is that it is difficult to read all the feedback for a particular item especially for the popular items with many comments.

III. PROJECT OBJECTIVES

The objectives that will be achieved by the end of this project are:

- Knowing fundamental concepts and techniques of natural language processing (NLP).
- Gaining a broad understanding of text data.
- Knowing how to pre-process/clean the data using different data pre-processing techniques.
- Knowing how to build a web application using Flask framework.

IV. PROJECT FLOW

This project contains the following activities such as:

- Data Collection.

  ML depends heavily on data, without data, it is impossible for an "AI" to learn. It is the most crucial aspect that makes algorithm training possible. In Machine Learning projects, a training data set is needed. It is the actual data set used to train the model for performing various actions. The dataset used for this project was obtained from Kaggle. The data collected from Amazon it is about unlocked mobile phones. It consists of 67986 reviews and it contains 8 columns (67986 rows × 8 columns). The columns present in this dataset are:

- Asin,
- Name
- Rating
- Date
- Verified
- Title
- Body
- Helpful Votes

- Feature Selection

  Feature Selection is the process where we automatically or manually select those features which contribute most to the prediction variable or output in which that are needed. Having irrelevant features in our data can decrease the accuracy of the models and make the model learn based on irrelevant features.

- Text Pre-processing.

  In natural language processing, text pre-processing is the practice of cleaning and preparing text data. NLTK and re are common Python libraries used to handle many texts pre-processing tasks. Text Pre-processing includes the following main tasks:

  - Importing the Libraries.
  - Importing the dataset.
  - Removing Punctuations.
  - Converting each word into a lower case.
  - Stemming.
  - Splitting Data into Train and Test.

- Model Building

  Now that the data is cleaned and ready for model building, let's use the pre-processed data to train the model. The model building tasks includes:

  - Importing the model building Libraries
  - Initializing the model
  - Adding Input Layer
  - Adding Hidden Layer
  - Adding Output Layer
  - Configuring the Learning Process
  - Training and testing the model
  - Optimizing the Model
  - Saving the Model

- Application Building

  Flask Frame Work with Machine Learning Model. In this section, we will be building a web application that is integrated into the model we built. A UI is provided for the user where they have to enter the values for predictions. The entered values are given to the saved model and the prediction is showcased on the UI. To build a web application the basics of "HTML, CSS, Bootstrap, flask framework, and python" are needed. The created project folder should contain:

  - A python file called app.py.
  - Bag of word or Count Vector file (count_vec.pkl).
  - Trained model file (phone.h5)
  - Templates folder which contains the index.HTML file.
  - Static folder which contains styles.css(index12.css).

- Deploying Application on IBM

  In this milestone, we will be deploying our flask app in IBM Cloud as a cloud foundry application.

V. SOLUTION

In this project, the factors that contribute to classifying reviews as positive or negative will be attempted to understood with the help of Natural language processing to analyse the sentiment (positive or a negative) of the given review. Besides, a sample web application is integrated to the model built. Below in figure1, the flow of the project is described in pictorial format.

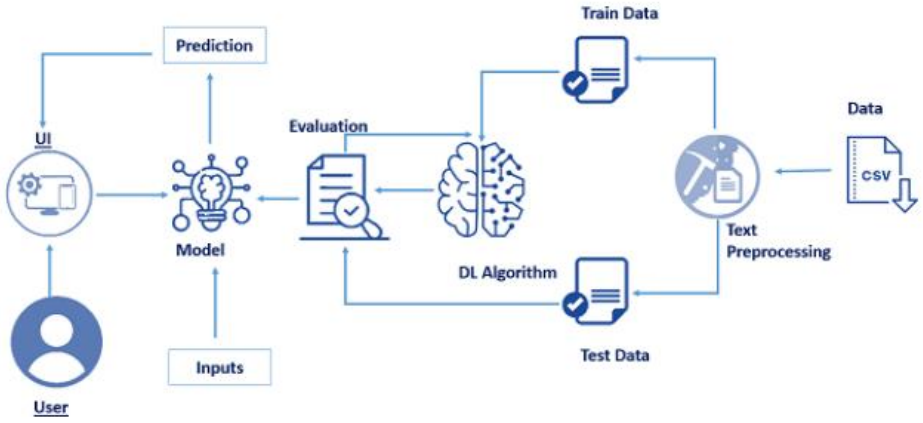

Figure 1: Technical Architecture

Figure 2: Positive Prediction for Positive Review

Figure 3: Negative Prediction for Negative Review

```
(1, 2000)

Prediction is [[0.0284515]]
127.0.0.1 - - [14/Sep/2021 19:06:34] "←[37mPOST / HTTP/1.1←[0m" 200 -
127.0.0.1 - - [14/Sep/2021 19:06:34] "←[33mGET /static/style/index12.css HTTP/1.1←[0m" 404 -
Hey Def not best, but not worst I had the Samsung

(1, 2000)

Prediction is [[0.01574755]]
127.0.0.1 - - [14/Sep/2021 19:07:09] "←[37mPOST / HTTP/1.1←[0m" 200 -
127.0.0.1 - - [14/Sep/2021 19:07:09] "←[33mGET /static/style/index12.css HTTP/1.1←[0m" 404 -
Hey Simply Amazing!
```

Figure 4: Prediction Level

VI. FUTURE IMPROVEMENTS

- **Evolving from human-computer interaction to human-computer conversation.**

  It is feasible to employ a single medium for interaction alone, which might be any spoken or nonverbal communication. However, it is necessary to employ both verbal and nonverbal communication mediums in tandem for effective communication. Though it is believed that as Natural Language Processing and Biometrics advance, technologies such as humanoid robots will be able to read facial expressions, body language, and speech. To accomplish it, there is a necessity of the integration of many modern-day technologies such as recognition of Human users, Sentiment analysis, recommendation analysis and techniques with the engagement in conversations is possible in a dynamic manner. To do so, various modern-day technologies such as human user recognition, sentiment analysis, suggestion analysis, and approaches with dynamic conversation interaction are required.

- **Biometrics**

  Body language, touch, gestures, and facial emotions are all examples of nonverbal communication. Biometrics like as facial recognition, fingerprint scanners, and retina scanners are required to incorporate nonverbal communication into the game. Nowadays, biometrics are becoming a prominent element in the field of providing security on computers, tablets, and even smartphones, leading to the assumption that biometrics may be used to identify patterns in human facial expressions and so understand feelings and emotions.

- **Humanoid Robotics**

  To express itself, every soul needs a body. Similarly, a physical unit is required to express NLP progress in a formal and commercial setting. Thought gadgets like as iPads, interactive television, and specialist conversational devices (such as Siri and Google Home) have begun to cover this area, although they are still only scratching the surface due to constraints such as the utilisation of a limited range of senses (hear, speak, to an extent see but cannot feel touch). This engagement should be bidirectional and involve the fourth sense (touch), like in a face-to-face conversation.

VII. CONCLUSION

In conclusion, this project was a great exposure to Natural Language Processing. NLP will undoubtedly gain in popularity as the amount of available data continues to rise, and algorithms get more complex and accurate. Human-machine interaction is changing as a result of it. Unstructured information makes up as much as 80% of what we encounter. As a result, NLP is one of the most important disciplines of data science. Organizing this data is a significant problem that numerous academics face on a daily basis. NLP is advancing at a rapid pace, and we may anticipate it to impact more and more parts of our life in the future.