Sentiment Analysis
Final NLP project

December 10, 2021

Overview

What is sentiment analysis?

Sentiment analysis is the process of reading tons of product reviews automatically and extracting useful and meaningful information to discover if the customers are really satisfied with your product or not.

A person's feedback is more subjective rather than factual. Feedbacks can be negative, positive, or neutral.

Sentiment analysis applies Natural Language Processing (NLP) and Text Analysis techniques to highlight the subjective information from the text.

Performing the sentiment analysis on the customer reviews can help identify what is lacking and therefore guide you to improvement.

You can review customer feedback and responses, and thus identify the negative comments and reasons why the customers have issues with your product or service.

Goals

To determine the sentiment of a given review. (Negative/ Neutral/ Positive)

Specifications

For data, we used “Amazon Fine Food Reviews”

Source: https://www.kaggle.com/snap/amazon-fine-food-reviews

Description

This dataset consists of reviews of fine foods from amazon.

The data span a period of more than 10 years, including all ~500,000 reviews up to October 2012.

Reviews include product and user information, ratings, and a plain text review. It also includes reviews from all other Amazon categories.

Data includes :

- Reviews from Oct 1999 - Oct 2012

- 568,454 reviews

- 256,059 users

- 74,258 products

Attributes :

- ProductId : Unique identifier for the product

- UserId : Unqiue identifier for the user

- ProfileName : Name of customer

- HelpfulnessNumerator : Number of users who found the review helpful

- HelpfulnessDenominator : Number of users who indicated whether they found the review helpful or not

- Score : Rating between 1 and 5

- Time : Timestamp for the review

- Summary : Brief summary of the review

- Text : Review text

Github:

https://github.com/saifeldeenmohamed/NLP-sentiment-analysis

https://github.com/JOoAlashmawy/ITI/tree/main/NLP/Project

Procedures:

1-Imports

2-Reading the Dataframe

3-Exploratory Data Analysis (EDA)

Check null values

Removing nulls

Removing Duplication

Removing outliers

Create target column

4-Preprocessing

Handling imbalance

Downsampling

Stopwords

Clear Reviews

WordCloud

5-Modeling

Vectorization

5

Bag of Words Vectorizer

TF-IDF

Label Encoder

Training

Logistic Regression with bag of words

Logistic Regression with Tf-Idf

Prediction on a review

Testing

Conclusion:

Tf-Idf has got the best accuracy.

Developed by:

Youssef Mohamed Alashmawy

Saif Eldeen Mohamed Abdelbasit

Group:

Group 1 Alexandria