**Sentiment Analysis and Recommendation system on Amazon and Walmart Product Reviews (GROUP 14)**

**Natural Language Processing (NLP)**

**CSCE 5290 - (Spring 2023 Section 003)**

**Github -** [**https://github.com/Sreedanthojue2/NLP-project.git**](https://github.com/Sreedanthojue2/NLP-project.git)

**Instructor** **Information**

**Instructor** **Name** **-** Dr.Zeenat Tariq

**Email**: [Zeenat.Tariq@unt.edu](about:blank)

**Group Information:**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | NAME | UNT ID | Contact |
| 1 | Srikanth Reddy Dwarakapalli | 11650259 | [SrikanthReddyDwarakapalli@my.unt.edu](http://SrikanthReddyDwarakapalli@my.unt.edu) |
| 2 | Sreevani Danthojue | 11643429 | [sreevanidanthojue@my.unt.edu](mailto:sreevanidanthojue@my.unt.edu) |
| 3 | Kishore Sai Lakshman Rangisetti | 11666403 | [KishoreSaiLakshmanRangisetti@my.unt.edu](http://KishoreSaiLakshmanRangisetti@my.unt.edu) |
| 4 | Satvik Reddy Chittela | 11657165 | [Satvikreddychittela@my.unt.edu](http://Satvikreddychittela@my.unt.edu) |

**MOTIVATION**:

With the popularity of online marketplaces over the last few decades, online sellers and merchants now regularly seek customer feedback. As a result, millions of reviews are created every day, making it difficult for potential customers to decide whether to purchase a product. It takes time. This project considers the problem of classifying reviews according to their overall semantics (positive or negative). Two separate supervised machine learning techniques will be used to conduct the study. Different category products from Amazon and Walmart will be experimented with SVM, logistic regression, multinomial naive Bayes, decision trees, and ensemble classifiers.

**SIGNIFICANCE**:

People are purchasing goods through various e-commerce websites as the global commercial landscape almost entirely transitions to the online platform. And as a result, it's also a usual practice to read product reviews before making a purchase. So, in order to make the data more dynamic, it is now crucial to analyze the data from those customer evaluations. It takes a lot of time in this day and age of increasingly sophisticated machine learning-based algorithms to comprehend a product by reading thousands of reviews, though we can focus a review on a specific category to determine how well-liked it is among consumers everywhere.

The goal of this study is to classify customer reviews of various goods into positive and negative feedback, and to develop a supervised learning model to polarize a large number of reviews. Over 88% of online consumers trust reviews as much as personal suggestions, according to a study conducted on Amazon last year. Any online product with a significant number of favorable reviews offers a persuasive remark on the item's legitimacy. In contrast, a lack of reviews on books or any other online purchase makes prospective customers suspicious. Simply put, evaluations with more numbers appear more credible. People respect other people's opinions and experiences, and the only way to learn what other people think of a product is to read reviews on it. User experiences with particular goods or topics, as well as opinions gathered from those experiences, directly affect future consumer purchasing decisions. Similar to this, poor evaluations frequently result in decreased sales. The objective is to gather customer input and polarize a lot of data in accordance with that. There have been some comparable studies conducted using Walmart and Amazon datasets. To comprehend the polarized attitudes toward the products, we will conduct sentiment analysis on a small collection of datasets of Amazon and Walmart product reviews. We will contrast the performance of the two groups in a number of areas, including product rating and price.

**OBJECTIVE:**

■ The project uses sentiment analysis to analyze and compare customer reviews from Amazon and Walmart and prepare a word cloud to identify crucial keywords from the data in different categories.

■ The objective is to develop a prediction system or a recommendation system that can be used to identify which category of products has better rating and pricing among Walmart and Amazon and also it recommends the best possible outcome from the user request by suggesting the best product.

**Feature of Proposed System:**

In our proposed approach we will use several Natural Language Processing Techniques like Pre-Processing and Sentiment Analysis and then we will use Machine learning algorithms like support vector machine (SVM), Random Forest (RF), Logistic Regression (LR), TF - IDF matrix.

1. First, we will take the Amazon product review dataset and other different e-commerce websites dataset containing different classes or products.

2. We will filter the dataset according to requirements and create a new dataset that has attributes according to analysis to be done by combining all the different datasets alongside ratings and reviews.

3. We will perform Pre-Processing on the dataset.

4. We will Split the data into training and test it.

5. We will train the model with training data and then analyze the testing dataset over the classification algorithm.

6. Then, we will create a recommendation system from the positive and negative outcomes of the sentiment analysis.

7. It will recommend products of a different kind from the user input by using considering different factors like rating and prices and review by the customers simultaneously at priority.

8. Also, It will classify the products based on the ratings of a product on different websites in different categories and recommends which website would be preferred.

**Data Set Description:**

The dataset for the Amazon and Walmart reviews for different category products is obtained from the Kaggle Website. The link to access the dataset is given below:

<https://www.kaggle.com/datasets/sreevanidanthojue/walmart-product-review-dataset>

<https://www.kaggle.com/datasets/sreevanidanthojue/amazon-category-wise-product-reviews>

**FlowChart:**

Dataset

Data Analysis

Pre-Processing

Prediction Result

Graph Analysis

Pre-Processed Dataset

RF/SVM/LR

Classifier

Pre-Processed Dataset

**REFERENCES:**

Samha,Xu,Xia, Wong & Li “Opinion Annotation in Online Chinese Product Reviews.” In Proceedings of LREC conference, 2008.

<https://ieeexplore.ieee.org/abstract/document/8376299>

Nina Isabel Holleschovsky, “The social influence factor: Impact of online product review characteristics on consumer purchasing decisions”, 5 th IBA Bachelor Thesis Conference, Enschede, The Netherlands 2015

<https://essay.utwente.nl/67351/1/Holleschovsky_BA_MB.pdf>

Elli, Maria Soledad, and Yi-Fan Wang. "Amazon Reviews, business analytics with sentiment analysis." 2016

<https://docplayer.net/151407565-Amazon-reviews-business-analytics-with-sentiment-analysis.html>

Xu, Yun, Xinhui Wu, and Qinxia Wang. "Sentiment Analysis of Yelp‘s Ratings Based on Text Reviews." (2015).

[https://cs229.stanford.edu/proj2014/Yun Xu, Xinhui Wu, Qinxia Wang, Sentiment Analysis of Yelp%27s Ratings Based on Text Reviews.pdf](https://cs229.stanford.edu/proj2014/Yun%20Xu,%20Xinhui%20Wu,%20Qinxia%20Wang,%20Sentiment%20Analysis%20of%20Yelp%27s%20Ratings%20Based%20on%20Text%20Reviews.pdf)

Rain, Callen. "Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning."Swarthmore College (2013).

<https://www.sccs.swarthmore.edu/users/15/crain1/files/NLP_Final_Project.pdf>

D.M. Hussein, A Survey on Sentiment Analysis Challenges, Journal of King Saud University-Engineering Science, April 2016

<https://www.sciencedirect.com/science/article/pii/S1018363916300071>

A. Singh and R. Sathyaraj, A Comparison between Classification Algorithms on Different Datasets Methodologies Using Rapidminer, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, May 2016

[https://www.ijarcce.com/upload/2016/may-16/IJARCCE 140.pdf](https://www.ijarcce.com/upload/2016/may-16/IJARCCE%20140.pdf)