CSCE 5290: Natural Language Processing Project Proposal

Project Title: Amazon product review analysis using Sentiment Analysis and Topic Modeling.

Team Members: S M Mazharul Hoque Chowdhury and Kiran Karidibusigari.

1. Project Description

In this project, we are planning to collect user reviews of a particular product from Amazon and apply sentiment analysis to understand users' emotions towards it [1]. Here, from the data collection to visualization, everything will be automated and user just need to provide the link of the product that they want to analyze. As the raw data may have anomalies that are not relevant to the goal. Therefore, those will be cleaned. In the next step sentiment analysis will be applied and initially all the data will be divided into three classes – Positive, Negative and Neutral. But we will try to extend into more specific emotion classes, such as – Happy, Sad, Angry, Excited and so on. Later, based on that, Topic Modeling will be applied to the data to extract features that are affecting the user [2][3]. In this part, collected data will be divided based on their sentiment class. For each sentiment class all the text will be gathered together and topic modeling will be applied to extract the features. At the end of the analysis, it will be possible to determine the sentimental state for the post and their relevant features. In Figure 1, the data flow of the model has been presented.

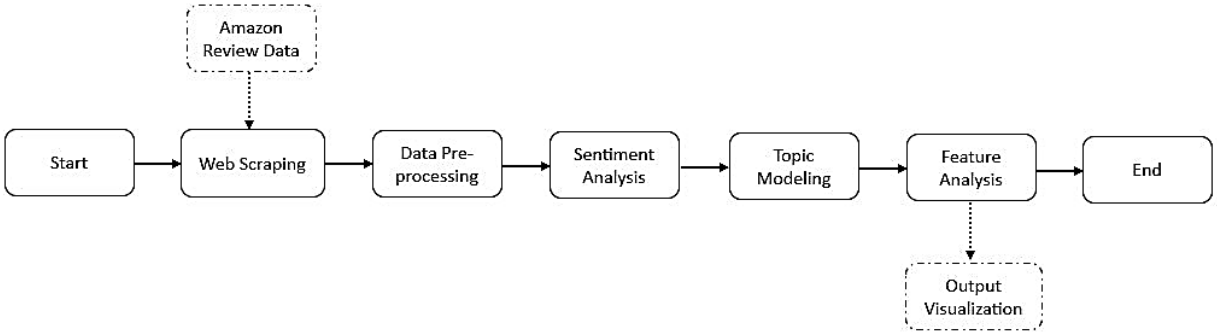

Figure 1. Data Flow Diagram of the model

2. Motivation

Day by day, we are becoming more dependent on technology, and it is not limited to using computers for fast calculation. It has become part of our lives, such as communicating with others or daily shopping. Online shopping has become very common; therefore, companies like Amazon, eBay, and AliExpress are growing quickly. Nevertheless, with this growth, one big concern is "Customer Satisfaction". Because if the customers are not happy, they won't trust the company. As a result, their sales will go down.

Considering all those, it can be said that to build a good business, it is important to focus on customers' needs and satisfaction. However, with the traditional data analysis methods, it is impossible to handle the rapid growth, and here, Natural Language Processing can help us solve the problem. Therefore, we will be working on a model that will be able to collect data from Amazon and will be able to identify users need based on the sentiment analysis and topic modeling technique. This project will be very helpful to anyone

who is involved to market research or business decision making or businesses who are trying to improve their product.

3. Significance

This project has several significant implications and potential benefits, such as –

- Customer Insights – By analyzing user reviews and sentiment, it is possible to understand customers' feelings about the product, which will give insights about the product to the companies.
- Product Improvement – Companies will be able to understand the positive and negative sides of the product and improve accordingly.
- Competitive Analysis – It will be possible to compare the product with the competing products and refine their marketing and product strategies.
- Market Research – In market research and product development, topic modeling can be extremely valuable in identifying what customers care about most.
- Customer Engagement – Addressing customer concerns and acknowledging positive feedback can help companies build a more loyal customer base.
- Business decision making – With the outcome of this project, it will be easier for the companies to make better business decisions based on user requirements.

Here are some benefits of the project, but not limited to them. There are other benefits like better marketing, business trend detection, predictive analysis, and so on.

4. Objectives and milestones

Objectives of the project are given below –

- ✓ Automating the data collection from Amazon.
- ✓ Preprocessing the raw data and tokenizing and structuring the data for processing.
- ✓ Applying sentiment analysis techniques to identify the emotion of the text.
- ✓ Classify text to emotion classes and assign sentiment scores.
- ✓ Applying topic modeling to identify key topics.
- ✓ Extract and interpret the most significant words or phrases associated with each topic.
- ✓ Analyzing the topics to understand the features that are being discussed.
- ✓ Identify the pros and cons of a product.
- ✓ Creating a visual representation of the outputs (sentiment analysis and topic modeling).

5. Features

The key features of the project are given below:

- Data Collection: The process of collecting Amazon product reviews.

- Data Preprocessing: The systematic cleaning and organization of review data.
- Sentiment Analysis: The methodical determination of user sentiment, categorized as positive, negative, or neutral.
- Topic Modeling: The systematic identification of important discussion topics.
- Feature Analysis: The comprehensive understanding of product features that are either liked or disliked.
- Visualization: The creation of charts and graphs to facilitate easy comprehension.
- User Interface: User-friendly interface for input and viewing.

6. Output Visualization

In the output visualization part, there will be a sentiment classification chart (Bar chart/pie chart) that will represent the sentiment score of the product. Later, for each sentiment class there will be a Wordcloud of the features extracted through topic modeling. We will try to create a user interface to provide input and present the output.

7. Uniqueness of the project

The uniqueness of this project lies in its holistic approach to automatically extracting user review from online stores and fusion of sentiment analysis and topic modeling in order to present valuable insights of the customers. This model will be able to address real-world business needs through data driven decisions.

8. Technical features

In this project we will use web scrapping techniques (like- BeautifulSoup) to collect the data from amazon and clean it, so that it can be processed. We may need to take help of API to do that. For sentiment analysis we can use Text Blob, which can return polarity and subjectivity as output [4]. Later spaCy or NLTK can be used for tokenization. In the end, output will be visualized as pie chart or bar chart for the sentiment analysis and Wordcloud for topic modeling and to do that we will be using matplotlib.

9. References

1. Tufchi, S., Yadav, A., Rai, V.K., Banerjee, A. (2023). Sentiment Analysis on Amazon Product Review: A Comparative Study. In: Khanna, A., Polkowski, Z., Castillo, O. (eds) Proceedings of Data Analytics and Management. Lecture Notes in Networks and Systems, vol 572. Springer, Singapore. https://doi.org/10.1007/978-981-19-7615-5_13
2. H. Nainwal, A. Garg, A. Chakraborty and D. Bathla, "Text Summarization of Amazon Customer Reviews using NLP," 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2023, pp. 43-47.
3. Abdulrahman Helan, Zainab Namh Sultani; Topic modeling methods for text data analysis: A review. AIP Conf. Proc. 2 February 2023; 2457 (1): 040006. https://doi.org/10.1063/5.0118679
4. https://www.analyticsvidhya.com/blog/2022/07/sentiment-analysis-using-python/ [Last visited: 09-14-2023]

GitHub Repository Link: https://github.com/mazharul2213/NLP-Project-Fall23