Amazon Product Reviews Sentiment Analysis with NLP Project Proposal

Introduction

The Amazon Product Reviews Sentiment Analysis project aims to develop a system to analyze the sentiment of product reviews on Amazon using NLP techniques. This system aims to identify positive or negative sentiments regarding products. To achieve this goal, machine learning algorithms will be utilized and trained on user reviews to extract meaningful insights.

Proposal Details

1. **Project Title:** Amazon Product Reviews Sentiment Analysis with NLP
2. **Team Members:**
   - (21 07 06 039) Berke ENSEL
   - (21 07 06 041) Esma ŞEN
   - (21 07 06 015) Harun Yahya ÜNAL
3. **Problem Description:** The project aims to analyze the sentiment of product reviews on Amazon to understand customer opinions and feedback better. By classifying reviews as positive or negative, we can provide insights to businesses about their products and improve customer satisfaction.
4. **Dataset Details:** This dataset consist of nearly 3000 Amazon customer reviews (input text), star ratings, date of review, variants, and feedback for various Amazon Alexa products such as Alexa Echo, Echo Dots, Alexa Firesticks, etc. to train machine learning models for sentiment analysis.
5. **Planned Performance Evaluation Methods:** Planned performance evaluation methods for our project include accuracy, precision, recall, and F1-Score. Accuracy measures the percentage of reviews correctly classified by the model, while precision calculates the rate of true positives among reviews classified as positive. Recall indicates the rate at which truly positive reviews are correctly classified. F1-Score, on the other hand, considers both precision and recall values to comprehensively assess a model's performance. These metrics will help us accurately evaluate the success of our sentiment analysis model.
6. **Proposed Methods for Solving the Problem:** The proposed methods include first applying data preprocessing steps on the dataset of Amazon product reviews. In this step, the text data will be tokenized, unnecessary characters and tokens will be removed. Additionally, stemming or lemmatization may be applied. Next, the text data will be transformed into numerical features using techniques such as TF-IDF or word embeddings. Finally, various machine learning algorithms including Naive Bayes, Logistic Regression, Support Vector Machines (SVM), Random Forest, and Deep Neural Networks will be utilized to develop the sentiment analysis model.
7. **Rough Timeline:**

Week	**Tasks**
Week 1	Data Preparation and Cleaning
Week 2-3-4	Model Development
Week 5	Model Evaluation and Adjustments
Week 6-7-8-9–12	Progress report & Presentation Preparation& Final Report

8. **Expected Milestones:** By the deadline, a sentiment analysis accuracy of at least 85% is aimed to be achieved on the Amazon product reviews dataset.