03/15/2024

Project Progress Summary:

Name: Muthu Selvam

Project Topic (Refined): Sentiment Analysis on Large Scale Amazon Product Reviews

GitHub Repo: <https://github.com/muthucse7/ml_itcs_5156-project/>

Self-Assessment: Reading summary completed for final presentation slides.

**Abstract**

* E-commerce is booming, emphasizing the importance of online reviews.
* Sentiment analysis helps understand customer sentiments towards products.
* Supervised learning on large-scale Amazon datasets achieved satisfactory accuracy.

**Introduction**

* Online commerce relies heavily on customer reviews for product selection.
* Importance of sentiment analysis for understanding product popularity.
* Objective: Categorize positive/negative feedback using supervised learning.
* Trust in reviews: Over 88% of online shoppers trust reviews as much as personal recommendations.

**Related Works**

* Various approaches in sentiment analysis and opinion mining.
* Examples: Business analytics, probabilistic machine learning, feature selection methods.
* Active learning, feature extraction techniques, and different classifiers used in previous studies.

**Methodology**

* Utilized Amazon product data for analysis.
* Data Acquisition: Three categories selected – Electronics, Cell Phones, Musical Instruments.
* Active learning approach for data labelling.
* Data Pre-Processing: Tokenization, stop words removal, POS tagging.
* Feature Extraction: Bag of Words, TF-IDF, Chi-Square.
* Evaluation Measures: Accuracy, Precision, Recall, F1 Score.

**Results**

* Experimented with various machine learning algorithms (e.g., SVM, Naïve Bayes, Random Forest).
* Achieved high accuracies ranging from 88% to 94% across different datasets.
* SVM consistently provided the highest accuracy.

**Comparative Analysis**

* Compared project results with previous related works.
* Achieved significantly higher accuracies compared to earlier studies.
* Improved methodologies in pre-processing, feature extraction, and classification.

**Conclusion and Future Works**

* Proposed supervised learning model demonstrated effectiveness.
* Achieved over 90% accuracy in sentiment analysis.
* Future works include automation of data labelling, integration with customer interaction systems, and generalization to diverse text-based reviews.