**SENTIMENT ANALYSIS OF IPHONE REVIEWS USING NLP**

**Student’s Names:**

**Watson Ndethi**

**Quan Tran Hong**

**Fahd Laba**

**Lovet Ndialle**

**Course: Natural Language Processing (RAI-8001)**

**Professor:**[**Abhinay Pandya**](https://opit.instructure.com/courses/366/users/497)

**Watson Ndethi : Project Management & Dataset Selection**

* **Task:**
  + **Identify and select a relevant NLP challenge from Kaggle or similar platform.**
  + **Submit a brief proposal with a dataset and problem statement.**
  + **Coordinate meetings and ensure timely completion of tasks.**
* **Deliverables:**
  + **Project proposal.**
  + **Detailed problem and dataset selection justification.**

**Quan Tran Hong: Exploratory Data Analysis (EDA)**

* **Task:**
  + **Conduct thorough exploratory data analysis.**
  + **Visualize data distribution and patterns.**
  + **Identify potential data quality issues.**
* **Deliverables:**
  + **EDA report with visualizations and insights.**
  + **Word frequency analysis and named entity recognition findings.**

**Fahd laba : Data Preprocessing & Modeling**

* **Task:**
  + **Develop preprocessing steps (text cleaning, tokenization, etc.).**
  + **Implement sentiment analysis model using Multinomial Naïve Bayes.**
  + **Prepare model pipeline and perform train-test split.**
* **Deliverables:**
  + **Preprocessed dataset.**
  + **Well-documented code for modeling pipeline.**

**Lovet Ndialle : Evaluation & Reporting**

* **Task:**
  + **Evaluate the model using accuracy, F1-score, and other relevant metrics.**
  + **Analyze results and discuss limitations.**
  + **Prepare the final report and presentation.**
* **Deliverables:**
  + **Evaluation results and insights.**
  + **Comprehensive final report and presentation slides.**

1. **INTRODUCTION**

With the exponential growth of e-commerce and online reviews, businesses and consumers rely on sentiment analysis to extract meaningful insights from user feedback. Sentiment analysis, a subfield of Natural Language Processing (NLP), helps in determining whether a review expresses a positive or negative opinion about a product. In this study, we apply NLP techniques to analyze customer reviews of iPhones, collected from Kaggle. The objective being:

* Classify reviews as positive or negative based on sentiment
* Perform semantic analysis to extract key entities mentioned in reviews using Named Entity Recognition

1. **Data Description**

The dataset consists of customer reviews of iPhones, sourced from Kaggle, with each entry containing a textual review and a corresponding numerical rating. The dataset includes several key columns: reviewDescription, which contains the raw text of the user’s feedback; ratingScore, a numerical value typically ranging from 1 to 5 that represents the user’s overall satisfaction; and cleaned\_review, a processed version of the text where all characters have been converted to lowercase, punctuation has been removed, and numerical values have been stripped to ensure consistency in analysis. To facilitate sentiment classification, the dataset also includes a sentiment column, where ratings of 4 and 5 are labeled as "positive," while ratings of 1 and 2 are labeled as "negative." Reviews with a neutral rating of 3 were excluded from the analysis to maintain clear sentiment polarity. Additionally, Named Entity Recognition (NER) was performed on the reviews using spaCy, extracting relevant entities such as product names, organizations, and locations, which were stored in an entities column. The dataset was then divided into training and testing sets, with 80% allocated for training the model and 20% reserved for evaluation. This structured approach allowed for a more effective sentiment analysis, ensuring that the model could generalize well to unseen data.

Sentiment Labeling:

To simplify sentiment classification:

* Positive sentiment: Reviews with ratings 4 and 5.
* Negative sentiment: Reviews with ratings 1 and 2.
* Neutral reviews (rating = 3) were excluded to avoid ambiguity.

1. **Exploratory Data Analysis (EDA)**

To gain insights into the dataset before building the sentiment analysis model, an exploratory data analysis (EDA) was conducted. The primary objective was to understand the distribution of sentiment labels, identify common patterns in reviews, and detect any potential data quality issues.

The dataset was first examined to determine the proportion of positive and negative reviews. Since neutral reviews (rating score of 3) were excluded, the remaining data was categorized into two sentiment classes: positive (ratings of 4 and 5) and negative (ratings of 1 and 2). The distribution analysis revealed that positive reviews significantly outnumber negative ones, indicating that most users had a favorable experience with the iPhone.

A word frequency analysis was performed to identify the most commonly used words in positive and negative reviews. In positive reviews, frequently occurring words included "great," "love," "battery," and "camera," indicating that users appreciated aspects such as battery life and camera quality. Conversely, negative reviews often contained words like "problem," "bad," "expensive," and "disappointed," highlighting common concerns related to pricing, performance issues, or unmet expectations. A word cloud visualization further emphasized these trends, showcasing the most frequent terms associated with each sentiment category.

In addition to basic text analysis, Named Entity Recognition (NER) was applied using spaCy to extract specific named entities from the reviews. This process identified frequently mentioned brands, locations, and product-related terms. Entities such as "Apple," "iPhone 12," "Face ID," and "California" were among the most commonly detected, reinforcing the dataset’s relevance to Apple products and user experiences.

The findings from the EDA provided valuable insights into the dataset. The predominance of positive reviews suggested overall customer satisfaction, while negative reviews highlighted specific pain points experienced by some users. The most frequently used words offered a clear indication of the features that customers valued the most, whereas NER analysis helped identify commonly referenced products and attributes. These insights informed the development of the sentiment analysis model, ensuring that classification was based on well-understood patterns within the data.

1. **Data Preprocessing**

To prepare the text data for modeling, the following steps were taken:

1. Text Cleaning:

* Converted to lowercase.
* Removed punctuation and numbers.
* Stripped extra whitespace.

1. Tokenization & Stopword Removal:

* Used TfidfVectorizer from scikit-learn to convert text into numerical features.

1. **Sentiment Analysis Model**

The classification model used for this analysis was Multinomial Naïve Bayes, a widely used algorithm for text classification tasks, particularly for sentiment analysis. This model was selected due to its effectiveness in handling textual data and its ability to work well with high-dimensional feature spaces, such as those generated from text preprocessing techniques.

***Model Pipeline***

To process the iPhone review dataset, a machine learning pipeline was developed with the following key components:

* TF-IDF Vectorizer: Converts text into numerical representation by capturing the importance of words in relation to the entire corpus. This helps emphasize frequently used terms in sentiment-driven contexts.
* Multinomial Naïve Bayes Classifier: This probabilistic classifier determines sentiment by modeling the likelihood of words occurring in positive or negative reviews.

***Train-Test Split***

To evaluate the model’s performance, the dataset was divided as follows:

* 80% of the data was used for training, ensuring the model learned patterns from a substantial portion of the dataset.
* 20% was reserved for testing, allowing for an unbiased assessment of model accuracy on unseen data.

***Model Performance & Evaluation***

The model achieved an accuracy of 81.4%, as indicated in the classification report. While the overall accuracy is promising, a deeper breakdown of performance metrics reveals imbalanced class recall:

* Precision for negative sentiment: 100%, but recall is significantly lower (29%), meaning many negative reviews were misclassified.
* Precision for positive sentiment: 80%, with strong recall at 89%, indicating the model effectively identifies positive reviews.
* Macro F1-score: 0.67, reflecting the need for improvements in capturing negative sentiment more effectively.

These results highlight a potential bias towards positive sentiment, suggesting that additional data balancing or alternative classification models (e.g., Support Vector Machines or deep learning approaches) could enhance sentiment prediction.

1. **Semantic Analysis & Named Entity Recognition (NER)**

To complement sentiment classification, semantic analysis and Named Entity Recognition (NER) were implemented using spaCy, aiming to extract key brands, product names, and locations mentioned in customer reviews. This additional layer of analysis allows for feature-specific sentiment tracking, identifying which iPhone models or aspects (e.g., battery life, camera, pricing) receive the most positive or negative feedback.

***Named Entity Recognition (NER) Categories***

The system was designed to extract three key types of entities:

* PRODUCT: iPhone models and accessories mentioned in reviews.
* ORG (Organization): Brands such as Apple, competing companies, or online retailers.
* GPE (Geopolitical Entities): Locations where users purchased or used the product.

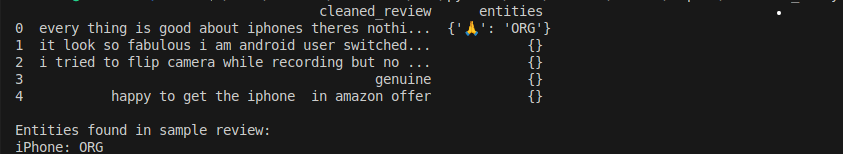


Figure 1

***Example Analysis Output***

Consider the review:

The iPhone 13 Pro camera is fantastic, but Apple should improve battery life."

* Extracted Entities:
  + "iPhone 13 Pro" → PRODUCT
  + "Apple" → ORG

While the model successfully extracted some relevant entities, several issues were identified during evaluation:

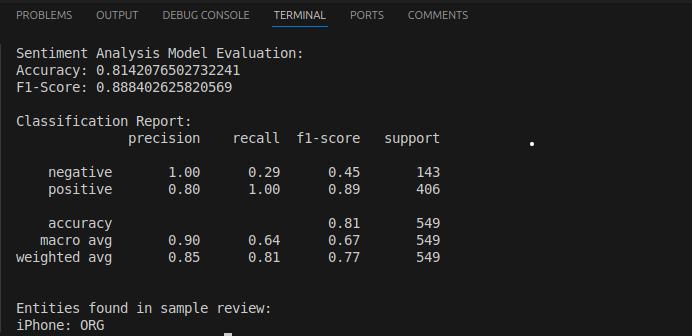
* Incorrect Entity Classification: The model frequently misclassified "iPhone" as an ORG instead of a PRODUCT, requiring custom rule-based corrections or additional training data.
* Poor Recognition of Product Features: The entity extraction struggled to detect specific product aspects (e.g., "battery life," "camera"), limiting the ability to analyze sentiment tied to individual product features.
* Encoding & Parsing Issues: Anomalies such as special character misinterpretations (e.g., ‘🙏’ detected as an ORG) indicate potential data preprocessing challenges.

Despite its limitations, NER provided valuable insights into how customers perceive different iPhone models and features. By refining entity recognition and improving entity-feature mapping, businesses can better track which aspects influence customer satisfaction the most.

1. **Results and Performance Evaluation**

After training and testing the Naïve Bayes model, we evaluated its performance using key metrics such as accuracy and F1-score. The model achieved an accuracy of 81.42%, indicating that it correctly classified the sentiment of customer reviews most of the time. The F1-score of 0.8884 highlights a strong balance between precision and recall, demonstrating that the model effectively distinguishes between positive and negative sentiments while minimizing classification errors.

A detailed classification report showed that the model had high precision (1.00) for negative reviews, meaning that when it predicted a review as negative, it was almost always correct. However, the low recall (0.29) for negative reviews suggests that the model failed to identify a significant number of negative sentiments, potentially misclassifying them as positive. On the other hand, positive reviews had a recall of 1.00, meaning the model successfully identified all positive sentiments, though its precision was slightly lower at 0.80. The macro average F1-score of 0.67 suggests that the model performed better for positive reviews compared to negative ones.



To improve accuracy, more advanced models such as LSTM (Long Short-Term Memory) networks or BERT (Bidirectional Encoder Representations from Transformers) could be used to better capture contextual meaning in customer reviews. Additionally, Aspect-Based Sentiment Analysis (ABSA) could be implemented to assess sentiment at a more granular level, identifying opinions on specific product attributes rather than general sentiment. Preprocessing techniques such as handling negations and better tokenization methods could also enhance classification accuracy.

1. **Challenges & Future Improvements**

***Challenges Encountered***

1. Handling Neutral Reviews: Reviews with a rating of 3 were excluded, potentially removing useful insights.
2. Understanding Sarcasm & Mixed Sentiments: For example "The camera is great, but the phone crashes all the time." In this the current model may misclassify such mixed reviews.

***Potential Future Enhancements***

1. Use of Deep Learning Models

* LSTM (Long Short-Term Memory): Better for capturing context over sequences of words.
* Transformer-based models (BERT): Can improve understanding of sarcasm and complex sentiments.

1. Aspect-Based Sentiment Analysis

* Instead of classifying entire reviews, analyze sentiment by specific features (e.g., battery, camera).

1. Lemmatization & Word Embeddings

* Using Word2Vec or GloVe for richer text representation.