Edge Hill University | Department of Computer Science

Evaluation of Sentiment Analysis Algorithms for Android Application Reviews

Practical Data Analysis (CIS4515)

Level 7

Coursework 2
2023/2024

Student Name: Sunday Idika
ID:****

Email: **@edgehill.ac.uk**

Table of Contents

Cover page.

1. Introduction

- Background Information

The background information provides essential context for sentiment analysis of Android applications. It outlines the significance of customer reviews in influencing consumer behaviour and the importance of sentiment analysis in understanding these reviews.

According to W. Medhat et al. (2014), sentiment analysis is a process of determining people's points of view towards an entity. It can be described as a classification process that targets an individual's opinion, identifies the sentiment, and then classifies their polarity (i.e. either positive, negative, or neutral opinion).

An overview of the sentiment analysis process for Android application reviews is shown in Figure 1. Starting from 1 (collection of user's reviews) to 9 (evaluation of the model performance).

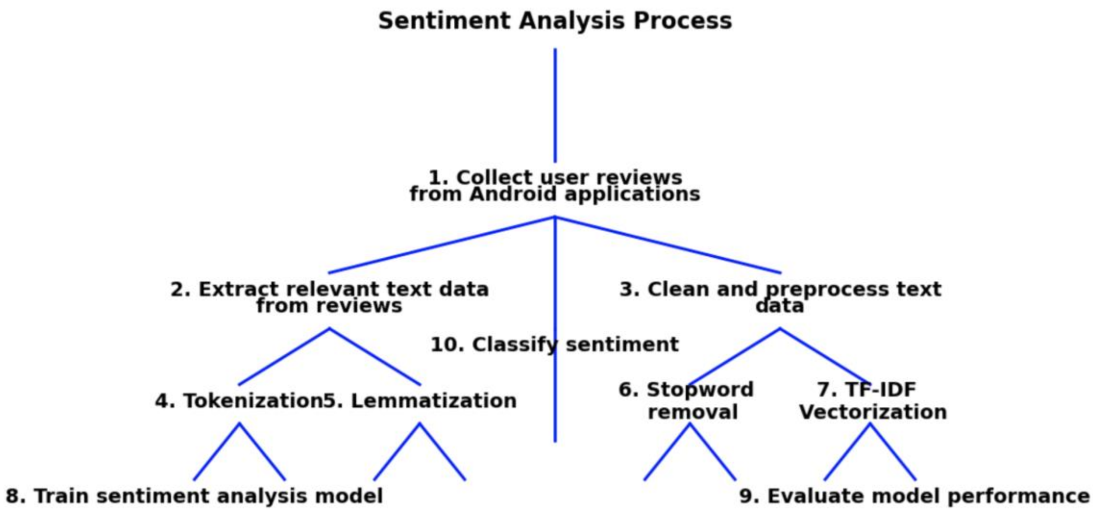

Figure 1.

- Objectives and Aim

The objective is to evaluate sentiment analysis algorithms applied to reviews of Android applications (See Figure 1). By analyzing the reviews, gain insights into the strengths and weaknesses of each of the AAD company's applications and make informed decisions regarding our investment. The aim is to assess the performance of these algorithms in accurately classifying the sentiment expressed in the reviews and provide insights into the effectiveness of different approaches to sentiment analysis.

2. Literature Review

- Existing Systems for Sentiment Analysis

Sentiment analysis has recently attracted considerable attention from both academic researchers and industry practitioners due to its versatile applications and significant implications. Existing systems for sentiment analysis encompass various approaches, machine learning models (e.g., SVM, Naive Bayes), deep learning techniques (e.g., RNNs, CNNs), and lexicon-based approaches, (Rodríguez-Ibánez et al. 2023).

- Advantages and Limitations

Machine learning algorithms like Support Vector Machines (SVM) and Multinomial Naive Bayes (MNB) are used for sentiment analysis tasks. SVM, a supervised learning algorithm, aims to find an optimal hyperplane to separate data points into different classes based on their features. Its popularity in sentiment analysis is attributed to its ability to handle high-dimensional datasets and achieve high accuracy levels (Saravanan and Sujatha 2018).

MNB is a probabilistic classifier based on Bayes' theorem with an assumption of conditional independence between features given the class label. It calculates the probability of a class label given the features and selects the label with the highest probability (Surya et al. 2019).

Lexicon-based methods rely on predefined sentiment lexicons or dictionaries containing words annotated with sentiment polarities. These approaches assign sentiment scores to text based on the presence of positive and negative words. Although lexicon-based methods are computationally efficient, they may struggle with context-specific sentiments and nuanced expressions (Basiri and Kabiri 2017).

Deep learning models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), have shown promise in sentiment analysis tasks. These models excel in capturing complex patterns and relationships in textual data, leading to improved sentiment classification performance (Bhatt et al. 2021) (Justus et al. 2018).

Despite advancements in sentiment analysis methodologies, challenges persist, including the accurate interpretation of sarcasm, irony, and context-dependent sentiments (Hussein

201). Additionally, domain-specific nuances and cultural differences pose further challenges in accurately capturing sentiment from the text (Bhonde and Prasad 2015). Continued research efforts are essential to address these challenges and advance the field of sentiment analysis.

3. Methodology

SVM and MVB algorithms were used to perform sentiment analysis on the provided dataset containing reviews of Android applications. The dataset comprised features such as review texts and their associated sentiment labels (positive, neutral, or negative). The training set is for algorithm training and the test set is for performance evaluation.

- **Description of Algorithms and Learning Mechanisms**

SVM discerns between instances of different classes by pinpointing the hyperplane that maximizes the margin between classes in feature space.

SVM Algorithm:

I. SVM was trained using the training dataset, where the text of the reviews was pre-processed to remove stopwords, and punctuation, and perform lemmatization.
II. The pre-processed text data was then transformed into numerical features using TF-IDF vectorization.
III. SVM was trained to classify the reviews into one of the three sentiment classes (positive, neutral, or negative) based on the Term Frequency-Inverse Document Frequency (TF-IDF) features.
IV. The performance of the SVM algorithm was evaluated using metrics such as precision, recall, F1-score, and accuracy.

- Strengths and Weaknesses

SVM excels with high-dimensional data and can handle non-linear decision boundaries using kernel tricks. However, its performance may decline with large datasets and noisy data.

Multinomial Naive Bayes (MNB): MNB is a probabilistic classifier based on Bayes' theorem with an assumption of conditional independence between features given the class label. It calculates the probability of a class label given the features and selects the label with the highest probability.

Strengths: Simple and efficient, works well with high-dimensional sparse data, computationally inexpensive.

Weaknesses: Assumes independence between features and may not capture complex relationships in the data.

- **Dataset Statistics Exercise**

Overview:

The dataset consists of reviews for Android applications extracted from Amazon.

Class Distribution:

Class labels represent the sentiment expressed in the reviews:

Class 1: Negative sentiment

Class 2: Neutral sentiment

Class 3: Positive sentiment

Distribution of classes in the training dataset:

Negative reviews: [2,983]

Neutral reviews: [2,206]

Positive reviews: [14,812]

Total = 20,001 reviews

In Figure 2, relationships between different variables in the review (train) dataset were explored using a bar chart.

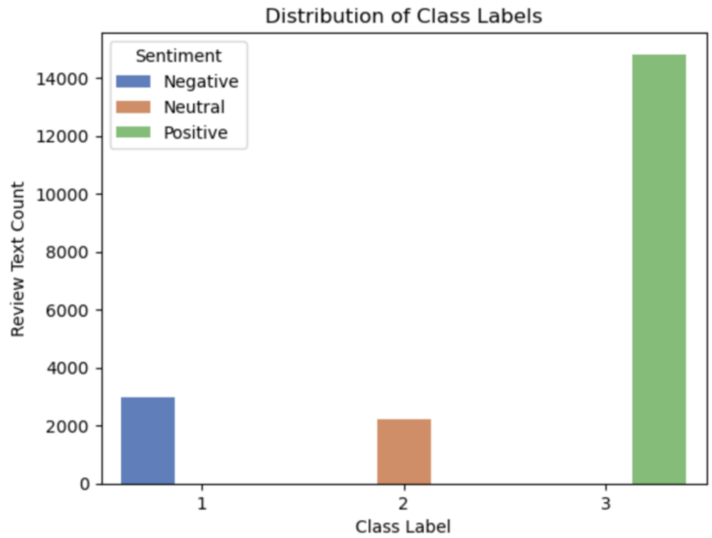

Figure 2.

Distribution of review lengths: The length of each review (number of words) was plotted in a histogram to visualize the distribution of review lengths (See Figure 3).

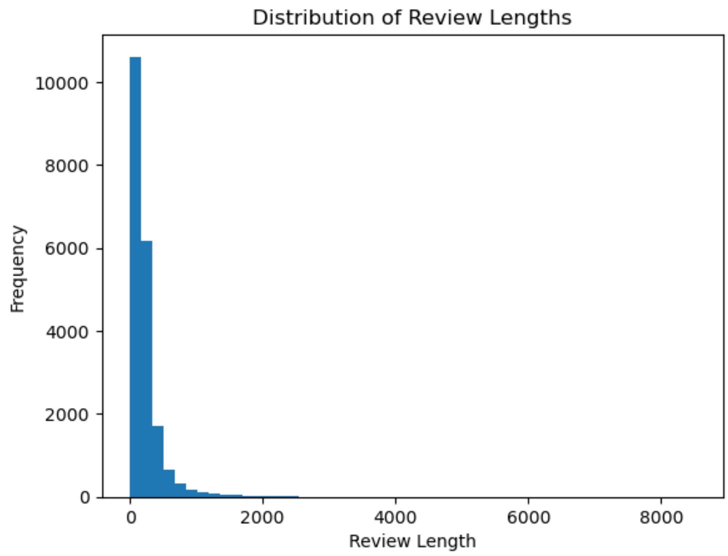

Figure 3. (Train dataset)

Distribution of classes in the test dataset:

Negative reviews: [3,469]

Neutral reviews: [2,087]

Positive reviews: [14,443]

Total = 19,999 reviews

In Figure 4, relationships between different variables in the review (test) dataset were also explored using a bar chart.

Figure 4

The length of each review (number of words) was calculated using a histogram to visualize the distribution of review lengths for the test dataset (See Figure 5).

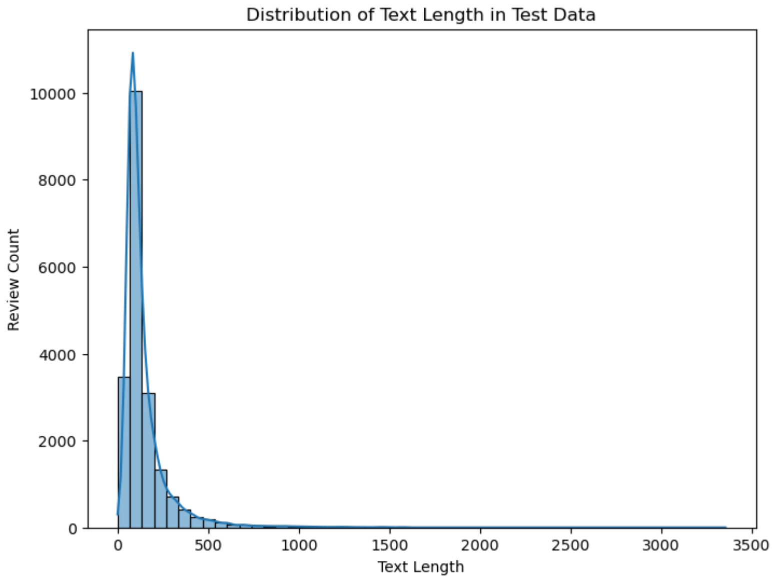

Figure 5

In Figure 6, the first row of the training dataset was printed to better understand the structure of the data.

```
First row of the training dataset:
```

	class label	ID code	review text
0	2	B004A9SDD8	Loves the song, so he really couldn't wait to ...
1	3	B004A9SDD8	Oh, how my little grandson loves this app. He'...
2	3	B004A9SDD8	I found this at a perfect time since my daught...
3	3	B004A9SDD8	My 1 year old goes back to this game over and ...
4	3	B004A9SDD8	There are three different versions of the song...

Figure 6.

The first row of the test dataset was printed as well. (See Figure 7)

```
First few rows of the test data:
```

	class label	ID code	review text	clean_text	text_length
0	3	B004K4RY9M	I am a person who has always enjoyed word game...	person always enjoyed word game thiis one exce...	67
1	3	B004K4RY9M	Love this. I try to beat my own time to see h...	love try beat time see fast complete keep mind...	54
2	3	B004K4RY9M	This game is fun and it can also be alearning ...	game fun also alearning game recomend age good...	61
3	3	B004K4RY9M	I enjoy these puzzles have books of them keep ...	enjoy puzzle book keep entertained hour great ...	60
4	3	B004K4RY9M	Have spent many enjoyable hours playing this g...	spent many enjoyable hour playing game would r...	87

Figure 7

A word cloud (See Figure 8) was generated to visualize the most frequent words in the reviews (train) data. This shows the common themes or topics discussed in the reviews.

Figure 8

A word cloud was also generated for the test dataset (See Figure 9).

Figure 9

4. Experiments

The experiment was done using SVM and MNB algorithms for sentiment analysis.

Performance Metrics

Test Accuracy: Proportion of correctly classified instances.

Precision: Proportion of true positive instances among all instances predicted as positive.

Recall: Proportion of true positive instances correctly identified.

F1-score: Harmonic mean of precision and recall.

5. Analysis Results

- Performance Comparison

SVM Model Evaluation table:

Metric	Class 1	Class 2	Class 3	Overall
Precision	0.67	0.52	0.82	
Recall	0.51	0.11	0.96	
F1-score	0.58	0.18	0.88	
Accuracy				0.79

Table 1.

Table 1. presents precision, recall, and F1-score metrics for three distinct classes (Class 1, Class 2, and Class 3), alongside an aggregate accuracy score. Precision values, ranging from 0.67 to 0.82, denote the accuracy of positive predictions within each class. Recall values, varying between 0.11 and 0.96, signify the completeness of the classifier's predictions for each class. F1-score, spanning from 0.18 to 0.88, provides a harmonic mean that balances precision and recall. The overall accuracy, recorded as 0.79, reflects the proportion of accurately classified instances relative to the entire dataset.

MNB Model Evaluation table:

Metric	Class 1	Class 2	Class 3	Overall
Precision	0.85	0.00	0.73	
Recall	0.09	0.00	1.00	
F1-score	0.16	0.00	0.85	
Accuracy				0.74

Table 2.

Similarly, Table 2. offers comparable metrics for the same classes (Class 1, Class 2, and Class 3), alongside an overarching accuracy metric. Precision values, spanning from 0.00 to 0.85, indicate the precision of positive predictions within classes 2 and 3 but nothing in class 2. Recall values, ranging between 0.00 and 1.00, depict the classifier's completeness in capturing instances belonging to each class. F1-scores, varying from 0.16 to 0.85, represent the harmonic mean of precision and recall for each class. The overall accuracy, quantified at 0.74, highlights the proportion of accurately classified instances relative to the entire dataset.

The summary of Tables 1 and 2 was represented in a graph for better visualization as shown in. Figure 2.

Figure 10.

In Figure 10, we observe two sets of bars representing the performance metrics of SVM and MNB across three classes (Class 1, Class 2, and Class 3).

- Interpretation of Results

The SVM bars consistently stand taller, indicating higher precision in Class 2 and Class 3, SVM also maintains higher performance in Recall within Class 1 and Class 2, and F1-score across all classes compared to the MNB bars. Notably, Class 2 of SVM demonstrates the highest values in all the performance metrics, suggesting superior classification accuracy for this class. Conversely, the MNB generally has zero value in all metrics across Class 2.
SVM demonstrates better capability in handling class imbalances and capturing complex relationships in the data compared to MNB.

- AAD Company Investment

After applying the SVM algorithm to the test data, I was able to draw insight from each AAD company. Below are results gotten from each company:

AAD_1: Positive - 105 occurrences, Neutral - 0 occurrences, Negative - 2 occurrences

AAD_2: Positive - 203 occurrences, Neutral - 13 occurrences, Negative - 80 occurrences

AAD_3: Positive - 100 occurrences, Neutral - 2 occurrences, Negative - 66 occurrences

While sentiment analysis can provide valuable insights, it's essential to consider other factors such as the overall sentiment trend over time, the nature of the negative reviews (if any), the competitiveness of the market, financial performance, and future growth potential.

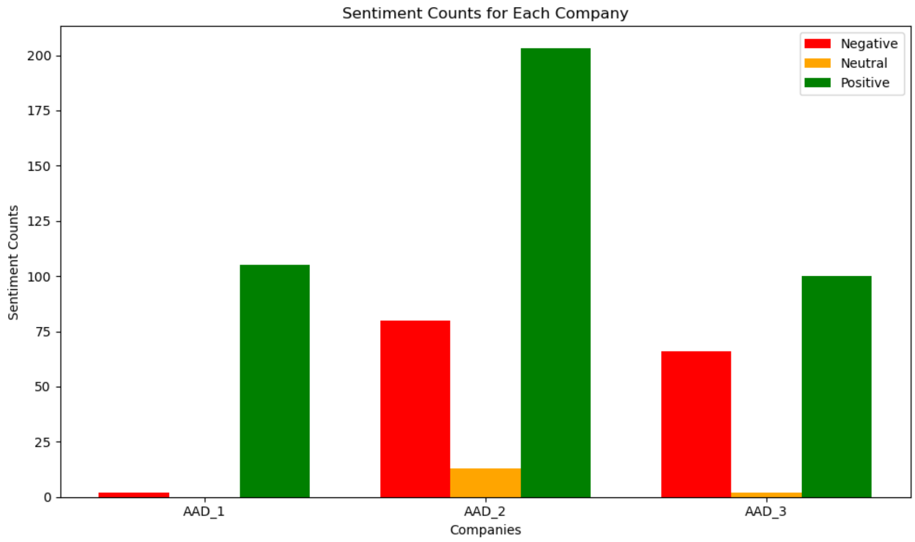

Figure 11.

In this scenario, AAD_2 appears to have the highest number of positive occurrences (203) compared to AAD_1 (105) and AAD_3 (100). However, it's important to analyze the reasons behind the negative sentiments and the overall context of the reviews.

Therefore, based solely on sentiment counts, AAD_2 might be considered more positively by customers, but a comprehensive analysis considering multiple factors would be necessary to make a well-informed investment decision.

6. Conclusion

Summary of Aims:

The study aimed to compare the performance of MNB and SVM classifiers in sentiment analysis of Amazon reviews for Android applications and was able to decide which company to invest with.

Achievement of Objectives:

Both classifiers were evaluated on the test dataset, and their performance metrics were compared. The SVM classifier achieved higher accuracy and better performance across sentiment classes compared to the MNB classifier.

Implications and Future Directions:

The findings suggest that the SVM classifier is more suitable for sentiment analysis tasks requiring accurate classification of text data with imbalanced classes.

Future research could explore ensemble methods, deep learning models, or fine-tuning existing classifiers to further improve performance and address specific challenges in sentiment analysis tasks.

This report provides insights into the effectiveness of NB and SVM classifiers in sentiment analysis and highlights the importance of choosing appropriate algorithms for different text classification tasks.

Reference

[1] W. Medhat, A. Hassan and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Engineering Journal,* vol. 5, no. 4, pp. 1093-1113, 2014.

[2] M. Rodríguez-Ibánez, A. Casánez-Ventura, F. Castejón-Mateos and P.-M. Cuenca-Jiménez, “A review on sentiment analysis from social media platforms,” *Expert Systems with Applications,* vol. 223, p. 119862, 2023.

[3] R. Saravanan and P. Sujatha, “A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification,” in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2018.

[4] A. Mucherino, P. Papajorgji, P. Pardalos, A. Mucherino, P. Papajorgji and P. Pardalos, “K-nearest neighbor classification,” *Data Mining in Agriculture. Springer Optimization and Its Applications,* vol. 34, pp. 83-106., 05 September 2009.

[5] M. E. Basiri and A. Kabiri, “Translation is not enough: Comparing Lexicon-based methods for sentiment analysis in Persian,” in *2017 International Symposium on Computer Science and Software Engineering Conference (CSSE)*, Shiraz, Iran, 2017.

[6] C. Bhatt, I. Kumar, V. Vijayakumar, K. Singh and A. Kumar, “The state of the art of deep learning models in medical science and their challenges,” *Multimedia Systems,* vol. 27, pp. 599-613, 2021.

[7] J. B. D. Justus, S. Bonner and A. S. McGough, “Predicting the Computational Cost of Deep Learning Models,” in *2018 IEEE International Conference on Big Data (Big Data), Seattle*, Seattle, WA, USA, 2018.

[8] D. M. E.-D. M. Hussein, “A survey on sentiment analysis challenges,” *Journal of King Saud University - Engineering Sciences,* vol. 30, no. 4, pp. 330-338, 2018.

[9] S. B. Bhonde and J. R. Prasad, “Sentiment Analysis-Methods, Applications \& Challenges,” *International Journal of Electronics Communication and Computer Engineering,* vol. 6, p. 634, 2015.

[10] P. P. Surya, L. V. Seetha and B. Subbulakshmi, "Analysis of user emotions and opinion using Multinomial Naive Bayes Classifier," 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2019, pp. 410-415, doi: 10.1109/ICECA.2019.8822096.