GOCE DELCEV UNIVERSITY - STIP
FACULTY OF COMPUTER SCIENCE

ISSN 2545-4803 on line
DOI: 10.46763/BJAMI

BALKAN JOURNAL OF APPLIED MATHEMATICS AND INFORMATICS (BJAMI)

YEAR 2025

VOLUME 8, Number 2

AIMS AND SCOPE:
BJAMI publishes original research articles in the areas of applied mathematics and informatics.

Topics:
1. Computer science;
2. Computer and software engineering;
3. Information technology;
4. Computer security;
5. Electrical engineering;
6. Telecommunication;
7. Mathematics and its applications;
8. Articles of interdisciplinary of computer and information sciences with education, economics, environmental, health, and engineering.

BALKAN JOURNAL
OF APPLIED MATHEMATICS AND INFORMATICS
(BJAMI), Vol 8

ISSN 2545-4803 on line
Vol. 8, No. 2, Year 2025

EDITORIAL BOARD

TABLE OF C O N T E N TS

Aleksandra Risteska-Kamcheski
SOLUTION OF DIDO'S PROBLEM USING VARIATIONS ..………………………………….. 7

Mirjana Kocaleva Vitanova, Elena Karamazova Gelova, Zoran Zlatev, Aleksandar Krstev
ENHANCING GEOGRAPHIC INFORMATION SYSTEMS WITH SPATIAL DATA MINING …………………………………………………………..……………………. 19

Violeta Krcheva, Misa Tomic
ADVANCED TOOLPATH VERIFICATION IN CNC DRILLING: APPLYING NEWTON'S INTERPOLATION THROUGH MATLAB …………………………………………………... 31

Martin Tanchev, Saso Koceski
WEB-BASED EDUCATIONAL GAME FOR EARLY SCREENING AND SUPPORT OF DYSCALCULIA ………………………………………………………………………………43

Maja Kukuseva Paneva, Elena Zafirova, Sara Stefanova, Goce Stefanov
MONITORING AND TRANSMISSION OF THE PROGRESS PARAMETERS ON AGRO INDUSTRIAL FACULITY IN A GSM NETWORK …….……………...…………………... 55

Qazim Tahiri, Natasa Koceska
METHODS OF EXTRACTION AND ANALYSIS OF PEOPLE'S SENTIMENTS FROM SOCIAL MEDIA ……………………....…………………………….…………………………………... 69

Ana Eftimova, Saso Gelev
DESIGN AND SIMULATION OF A SCADA – CONTROLLED GREENHOUSE FOR OPTIMIZED ROSE CULTIVATION ……………………………………………………………………….….. 81

Milka Anceva, Saso Koceski
A FHIR – CENTRIC APPROACH FOR INTEROPERABLE REMOTE PATIENT MONITORING ……………………………………………………………………………..….. 93

Jordan Pop-Kartov, Aleksandra Mileva, Cveta Martinovska Bande
COMPARATIVE EVALUTION AND ANALYSIS OF DIFFERENT DEEPFAKE DETECTORS ………………………………………………………………………………………………….….. 103

Vesna Hristovska, Aleksandar Velinov, Natasa Koceska
SECURITY CHALLENGES AND SOLUTIONS IN ROBOTIC AND INTERNET OF ROBOTIC THINGS (IoRT) SYSTEMS: A SCOPING REVIEW 115

Violeta Krcheva, Misa Tomic
CNC LATHE PROGRAMMING: DESIGN AND DEVELOPMENT OF A PROGRAM CODE FOR SIMULATING LINEAR INTERPOLATION MOTION ..……………………………………... 127

Jawad Ettayb
NEW RESULTS ON FIXED POINT THEOREMS IN 2-BANACH SPACES ……….....……... 139

Balkan Journal of Applied Mathematics and Informatics Online ISSN 2545-4803
Volume 8 Number 2 Year 2025 UDC: 004.421-045.38:[077.5:159.942.52]

METHODS OF EXTRACTION AND ANALYSIS OF PEOPLE'S SENTIMENTS FROM SOCIAL MEDIA

Qazim Tahiri, Natasa Koceska

Abstract. This study explores the methods for extracting and analyzing people's sentiments from social media, utilizing advanced natural language processing and machine learning techniques. The goal is to recognize sentiments and based on them to categorize posts and comments as positive, negative, or neutral in order to understand the users' attitudes and emotions. The algorithms used for sentiment classification in this research include Naive Bayes, SVM, Logistic Regression, and Random Forest. Additionally, the study analyzes and compares the performance of these algorithms in terms of accuracy, recall, and F1 score, providing a comprehensive overview of their effectiveness. It also emphasizes the importance of hyper-parameter tuning to improve the accuracy of classification algorithms. The results of this study can be used to assist social media platforms, researchers, and policymakers in developing strategies to manage and improve user experiences, as well as to make informed decisions based on user reactions.

1. Introduction

Social media platforms like Facebook, Twitter, Instagram, X (formerlyTwitter) and others have created a vast space where millions of users post their thoughts, experiences and emotions daily on various topics. These posts represent a valuable source of data which can be analyzed to understand public opinions and sentiments regarding current events, services and many other aspects of life. Companies use this data to improve the quality of products and services, identify pain points and predict customer behavior, which would lead to increased customer satisfaction. The most effective method, nowadays, for determining people's opinions, attitudes or emotions expressed in a given text is sentiment analysis. The goal of sentiment analysis is to classify data according to their polarity as positive, negative or neutral or even to make a more complex emotion classification such as happiness, sadness, anger, etc. [1]. It uses natural language processing (NLP) and machine learning (ML) technologies to train computer software to analyze and interpret text in a way similar to humans. These technologies are a key part of many existing applications such as AI and chatbots, which are used by researchers to solve various tasks [2]. Classification models are a type of ML model that divides data into predefined groups called classes. The model can make a prediction between two classes (binary classification) or between more than two classes, as in our case (multi-class).

There are many different types of algorithms that can be used for classification. In this paper, we will use some well-known classification algorithms, such as Naive Bayes, Support Vector Machine (SVM), Logistic Regression, and Random Forest. We will compare them in terms of their ability to capture human emotions, taking into account metrics such as accuracy, recall, precision and F1 score. This approach seeks to explore the effectiveness of various classification algorithms in detecting sentiments on social media and to identify the strengths and weaknesses of each algorithm. We will also demonstrate which of the algorithms is most suitable for the topic we are exploring and use hyperparameter tuning for the lowest-ranked algorithm, in order to determine whether there will be an improvement in prediction capabilities or not.

2. Related work

One of the first research studies on polarity classification in tweets was conducted by Go et al. [3]. They used emoticons (eg. positive emoticons ":)", negative emoticons ":(", etc.) that often appear in tweets, to classify positive or negative tweets. The validity of this method was previously demonstrated by Read [4]. The authors used several classifiers for their data and obtained the best results using the Naive Bayes classifier. However, their approach showed poor performance when used with three classes ("positive", "negative" and "neutral"). In the same line of thinking was the research by Pak and Paroubek [5] who generated a corpus of positive, negative and neutral tweets, using emoticons. They used three classification algorithms: SVM, Naive Bayes and Conditional Random Fields (CRF), and also obtained the best results with the Naive Bayes classifier, but when used n-grams and post-tags as characteristics of the tweets. Another approach for analyzing sentiments in tweets is the one proposed by Zhang et al. [6]. The authors used the so called hybrid strategy, combining supervised learning with understanding of words that convey feelings, taken from the DAL (Dictionary of Affect in Language) sentiment dictionary. They employ various supervised learning algorithms to categorize tweets into positive and negative. Thelwall et al. [7] analyze the intensity of opinions expressed in social networks. They developed an algorithm (SentiStrenght) that assigned a positivity or negativity score of each message, on a scale of 1 (negative) to 5 (very positive). The algorithm, among other features, takes into account the emoticons used in the published messages. The conclusion of the research was that the intensity of opinions depends on the importance of the events commented on Twitter. Several other researchers have worked on sentiment analyses focusing on the data related to product reviews [8, 9], or even on real-time political tweets [10]

This research differentiates itself from previous studies in several ways. The dataset used in this research is the Twitter US Airline Sentiment dataset, initially published by Crowdflower [11]. This dataset differs from others used in similar studies as it focuses specifically on airline-related conversations rather than general tweets. The data includes sentiment confidence scores, potential reasons for negative feedback, and tweet metadata such as location and creation time, providing a rich source for sentiment analysis. Also, none of the above-mentioned approaches demonstrated the importance of hyperparameter tuning while training a classifier.

3. Methodology

This study followed a structured methodology consisting of several key phases.

- **Data Collection:** The data used in this research was from Kaggle's Twitter US Airline Sentiment dataset. The dataset contains a total of 14,640 tweets categorized into three sentiment classes: positive (2,363 tweets), negative (9,178) and neutral (3,099). In the case of sentiment analysis for social media, the negative, neutral, and positive classes represent the different sentiments in the comments. The negative class includes comments with negative sentiment, the neutral class includes comments with neutral sentiment, and the positive class includes comments with positive sentiment. Figure 1 shows the distribution of sentiments across social media posts of this dataset. From the graph it is evident that the largest number of comments in the dataset is negative, which means that the dataset is imbalanced. The reason for this may be that people generally use social media platforms mostly to convey their unsatisfactory remarks.
  The dataset contains no personal or sensitive user data such as names, usernames, or profile photos, ensuring user privacy.

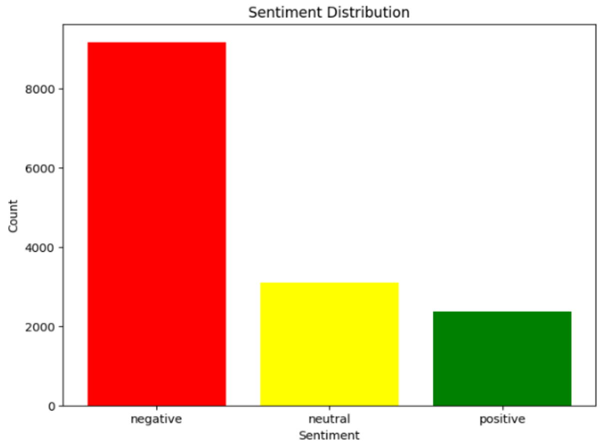

Fig. 1. *Polarity frequency in the dataset.*

- **Data Preprocessing**: To prepare the data for analysis, the following steps were performed:
  - Removal of Irrelevant Features: Features such as user IDs and timestamps were excluded as they do not influence sentiment.
  - Text Cleaning: Tweets were converted to lowercase. URLs, emojis, punctuation, numbers, hashtags, and user mentions were removed using regex techniques.
  - Tokenization and Stopword Removal: Tokenization was performed using Natural Language Toolkit (NLTK), followed by removing stopwords based on the NLTK corpus.
  - Lemmatization: Each word was lemmatized to its base form using the WordNetLemmatizer from NLTK.
  - Normalization: TF-IDF vectorization was used to normalize the textual data and reduce bias from frequent but less informative words.

- **Feature Selection:** Chi-square feature selection was used to identify the most influential words that contributed to sentiment prediction.
- **Classification:** The dataset was split into training and testing subsets with an 80:20 ratio. This approach ensured that the model was trained on the majority of the data and then evaluated on a separate portion to assess its generalization performance. Various classification algorithms including Naive Bayes, SVM, Logistic Regression and Random Forest were used for training and testing the subsets.
  - Naive Bayes: Naive Bayes is a probabilistic classifier based on Bayes' Theorem. It assumes feature independence and calculates the likelihood of a class based on input features. Despite its simplicity, it performs well in text classification tasks such as sentiment analysis.
  - Support Vector Machine (SVM): SVM is a powerful supervised learning model that identifies the hyper plane which best separates data into classes. It is effective in high-dimensional spaces and is widely used for binary and multi-class text classification.
  - Logistic Regression: Logistic Regression estimates probabilities using the logistic function and is suitable for binary and multi-class classification. It is computationally efficient, interpretable, and often performs competitively with more complex models.
  - Random Forest: Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs. It is robust against overfitting and effective for both classification and regression tasks.
- **Evaluation:** Since the dataset is imbalanced, the following metrics were used for the models' evaluation: accuracy, precision, recall and F1-score, as well as the weighted average of precision, recall, and F1-Score. A confusion matrix was also presented for each of the classification algorithms. The matrix gives an overview of the model's performance for each class and is useful for evaluating the interactions between classes.
  - Accuracy is the proportion of all classifications that were correct, whether positive or negative. It is defined as:

$$\text{Accuracy} = \frac{\mathrm{Tp+Tn}}{\mathrm{Tp+Tn+Fp+Fn}} \tag{3.1}$$

    where Tp is True positive (comments correctly predicted as positive), Tn is True negative (comments accurately predicted as negative), Fp is False positive (comments wrongly predicted as positive when they were negative), and Fn is False negative (comments wrongly predicted as negative when they were positive). Accuracy alone may not provide a complete picture of model performance, especially when working with imbalance datasets. Therefore, the other metrics like precision, recall and F1-score are used.
  - Precision is the proportion of all the model's positive classifications that are ctually positive. It is defined as:

$$\text{Precision} = \frac{\mathrm{Tp}}{\mathrm{Tp+Fp}} \tag{3.2}$$

Precision can be thought of as a quality metric; higher precision indicates that an algorithm provides more relevant results than irrelevant ones.

- Recall is the proportion of all actual positives that were classified correctly as positives. It is defined as:

$$\text{Recall} = \frac{\text{Tp}}{\text{Tp}+Fn} \tag{3.3}$$

Recall is also called sensitivity. A high recall score indicates that the model predicts the majority of the comments correctly. However, although recall and precision provide important information, they have limitations when viewed separately. Therefore, it is necessary to calculate an F1-score.

- F1 Score is a measure that combines recall and precision. It is actually a trade-off between precision and recall and can therefore be used to measure how effectively our models make that trade-off.

$$\text{F1} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision}+\text{Recall}} \tag{3.4}$$

A well-balanced performance is often indicated by a high F1-score, which shows that the model can simultaneously achieve high accuracy and recall. On the other hand, a low F1-score indicates that the model struggles to achieve this equilibrium.

Based on the analysis of these results, the best and worst performing classification algorithms for the given dataset were selected. For the worst performing classification algorithm (Random Forest), additional evaluation was performed after hyperparameter tuning.

- **Improvement:** Hyper-parameter tuning was performed on the algorithm with the worst accuracy on the given dataset, according to the previously performed analysis.

4. Research Framework: Questions and Hypotheses

This study explores the application and comparison of different ML algorithms for sentiment classification of social media data, with a focus on Twitter posts. To guide the research, the following questions and hypotheses were established.

Research Questions:

- *RQ1:* What are the main classification algorithms used for sentiment prediction on social media?
- *RQ2:* What performance do these classification algorithms demonstrate in terms of accuracy, recall, precision, and F1-score?
- *RQ3:* What is the impact of hyperparameter tuning on the performance of classification algorithms?

Research Hypotheses:

- *H0:* There is no significant difference in the performance of traditional machine learning models for sentiment classification.

- *H1:* One of the traditional machine learning models, used in this study, performs significantly better than others.
- *H2:* Hyperparameter tuning significantly improves the performance of traditional machine learning models.

To test these hypotheses, various classification algorithms were implemented, and their performance was evaluated using standard metrics. Statistical tests and comparative analysis were applied to validate or refute the hypotheses.

5. Results and discussion

5.1 Performance of the Naive Bayes Algorithm

The Naive Bayes algorithm was implemented as a baseline classifier to assess its ability to categorize tweets into sentiment classes: positive, neutral, and negative. Evaluation metrics such as precision, recall, and F1-score were used to measure the model's performance on each class. The macro average presenting the average precision, recall, and F1-score for all classes was also calculated, as well as the weighted average - which shows the average based on the number of cases for each class.

Using the Naive Bayes algorithm, the following results were achieved (Table 1):

Table 1.*Classification Report according to Naive Baye*

	precision	recall	F1-score
negative	0.79	0.97	0.87
neutral	0.71	0.40	0.51
positive	0.88	0.54	0.67
accuracy	/	/	0.79
macro avg	0.79	0.64	0.68
weighted avg	0.79	0.79	0.77

Negative class - using the Naive Bayes algorithm, the negative class has a precision of 0.79, a recall of 0.97, and an F1-score of 0.87. This means that predictions for the negative class are correct in 79% of cases, 97% of true negative cases are correctly identified, and the F1-score of 0.87 indicates a good performance for prediction.

Neutral class – the precision for the neutral class is 0.71, indicating that 71% of the comments predicted as neutral are actually correct. However, the recall and F1-score were quite lower than in negative class, showing the values of 0.40, and 0.51 respectively, implying that the model has trouble striking a balance between precision and recall.

Positive class – this class achieved a very good precision score 0.88, which means that 88% of the comments are correctly predicted as positive. However, the results for the recall are not so good, showing that only 54% of actual positive comments are classified as positive. The F1-score was also low with a value of 0.54.

Overall accuracy – The overall accuracy for Naive Bayes algorithm is 79%.

Based on the results, the Naive Bayes algorithm demonstrated a good performance in the negative class but a weaker performance in the neutral and positive classes.

The confusion matrix, which shows the number of correctly predicted cases by the algorithm and the number of actual cases that were correctly classified, is presented in Fig. 2. The main diagonal of the matrix shows the total number of correctly classified tweets. This confusion matrix helps to identify how well the algorithm predicted sentiment in all given categories.

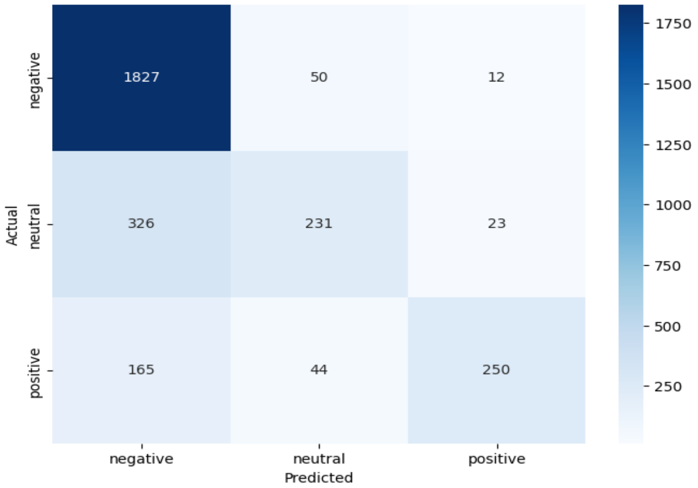

Fig. 1. *Confusion Matrix for the Naive Bayes Algorithm.*

5.2 Performance of the SVM Algorithm

The Support Vector Machine (SVM) is a supervised machine learning algorithm commonly used in sentiment analyses. The results obtained using this algorithm are not significantly different from those of the Naive Bayes algorithm. Table 2 presents the classification results obtained for the SVM algorithm:

Table 2. *Classification Report by the SVM Model*

	precision	recall	F1-score
negative	0.86	0.87	0.86
neutral	0.57	0.56	0.56
positive	0.71	0.70	0.71
accuracy	/	/	0.78
macro avg	0.71	0.71	0.71
weighted avg	0.78	0.78	0.78

Negative Class: The SVM model achieved a precision of 0.86, indicating that 86% of the comments predicted as negative are actually correct. It also has a recall score of 0.87, meaning that 87% of the actual negative comments were correctly identified. The F1-score for this class was 86%, showing a good balance between precision and recall. This suggests that the model is quite effective at accurately identifying negative comments, with high accuracy in this classification.

Neutral Class: For the neutral class, the precision, recall and F1-score are very similar. The precision score is only 0.57, meaning that only 57% of the comments predicted as neutral are correct. The recall has a value of 0.56, meaning that only 56% of the actual neutral comments were correctly identified. The F1-score is the same as recall, 0.56, suggesting a moderate performance of the model in identifying neutral comments.

Positive Class: For the positive class, the precision, recall and F1-score are also similar, but much better than for the neutral class. The precision for the positive class is 0.71, indicating that 71% of the predictions for this class are correct. The recall is 0.70, meaning that 70% of the actual positive comments were correctly identified. The F1-score for this class is 0.71, indicating a good balance between precision and recall for positive comments.

Overall Accuracy: The overall accuracy of the model in classifying social media comments is 78%, indicating a satisfactory overall performance.

The confusion matrix for the SVM algorithm is presented in Fig. 3.

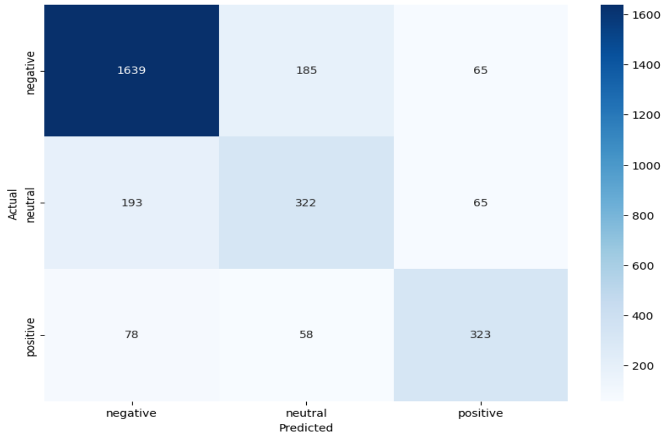

Fig. 2. *Confusion Matrix for the SVM Algorithm.*

5.3 Performance of Logistic Regression

The classification report presented in Table 3 shows the results obtained from the Logistic regression classification model applied to the social media dataset.

Table 3.*Classification Report According to Logistic Regression*

	precision	recall	F1-score
negative	0.87	0.88	0.87
neutral	0.60	0.59	0.60
positive	0.74	0.70	0.72
accuracy	/	/	0.80
macro avg	0.74	0.72	0.73
weighted avg	0.79	0.80	0.79

Negative Class: For this class the Logistic regression model achieved quite similar results: 0.87 for precision and F1-score, and 0.88 for recall. This means that 87% of the comments were correctly predicted as negative, and 88% of the actual negative comments were correctly identified. The F1-score of 87% shows a good balance between precision and recall.

Neutral Class: The achieved precision and F1-score for this class are also the same, but lower than for the negative class. The 0.6 score for precision means that only 60% of the comments predicted as neutral are correct. The recall is 0.59, indicating that 59% of the actual neutral comments were correctly identified. The obtained 60% of F1-score suggests a moderate performance of the Logistic regression model in identifying neutral comments.

Positive Class: Using the Logistic regression model for the positive class, the following results were obtained: 74% for precision, 70% for recall, and 72% for the F1-score, which shows a moderate performance of the algorithm for this class.

Overall Accuracy: The overall accuracy of the model in classifying social media comments is 80%, indicating a satisfactory overall performance.

The confusion matrix for the Logistic regression model is presented in Fig. 4.

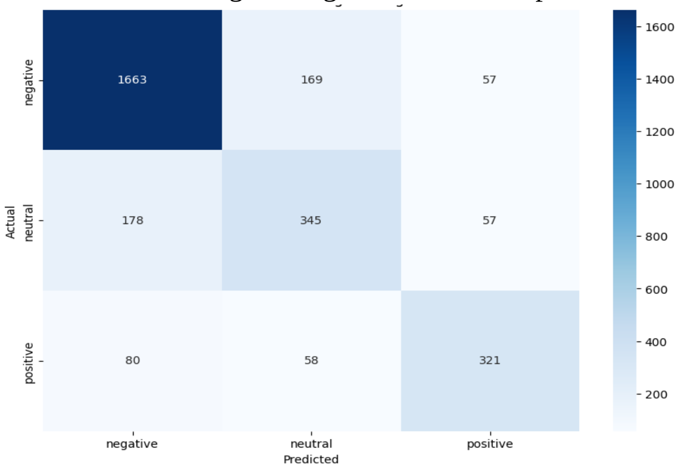

Fig. 3. *Confusion Matrix for the Logistic Regression Algorithm.*

5.4 Performance of the Random Forest Algorithm

Another algorithm that was used in our study was Random Forest. Using this model and the data set provided, the following results were achieved (Table 4):

Table 4.*Classification report according to Random Forest*

	precision	recall	F1-score
negative	0.78	0.96	0.86
neutral	0.67	0.39	0.49
positive	0.82	0.49	0.61
accuracy	/	/	0.77
macro avg	0.76	0.61	0.66
weighted avg	0.77	0.77	0.75

Negative Class: The results for precision indicate that 78% of the comments predicted as negative were actually correct. For recall the score of 0.96 was obtained, which means that as many as 96% of the actual negative comments were correctly identified. The F1-score of 0.86 shows a good balance between precision and recall.

Neutral Class: The achieved precision for the neutral class is 0.67, which means that 67% of the comments are correctly predicted as neutral. The recall score for this class is only 0.39, which is the worst of all used models in the research. This indicates that only 39% of the actual neutral comments were correctly identified. The F1-score for the neutral class is 49%, which is also the lowest score of all models.

Positive Class: Despite the poor results for the neutral class, the precision for the positive class was quite good (0.82). This means that 82% of the predictions for this class were correctly predicted as positive. The result for the recall shows that 49% of the actual positive comments were correctly identified. The F1-score of 61% for this class indicates a moderate balance between precision and recall for positive comments.

Overall Accuracy: Based on the results, the overall accuracy of the Random Forest model is 77%, which is the lowest of all models used in this research.

The confusion matrix for the Random Forest model is presented in Fig. 5.

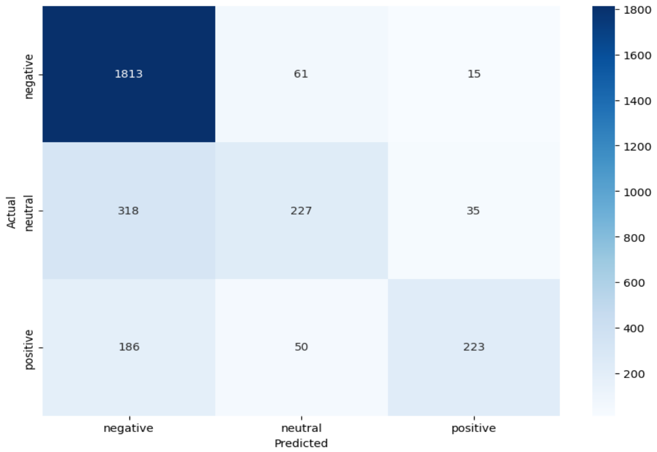

Fig. 4.*Confusion Matrix for the Random Forest algorithm.*

5.5 Hyperparameter Tuning and Comparison

To test H2, hyperparameter tuning was conducted on the Random Forest model using GridSearchCV. The following parameters were considered:

Python:

```
param_grid = {
    'n_estimators': [50, 100, 150],
    'max_depth': [None, 10, 20],
    'min_samples_split': [2, 5] }
```

After tuning, the model achieved an accuracy of 81%, which is slightly better compared to 77% before tuning. This improvement confirms that hyperparameter tuning enhances classification performance. The best parameters found were: n_estimators = 150; max_depth = 20; min_samples_split = 2.

The tuning process demonstrated that fine-tuning model parameters have a notable impact on improving prediction capabilities, especially in imbalanced datasets like the one used in this study. This supports the acceptance of H2.

5.6 Data comparison

Figure 6 shows the graphical representation of performance scores of the four machine learning models. The weighted average for precision, recall and F1-score, as well as the accuracy of each of the algorithms, are presented in order to make a better understanding of which algorithm gives the best/worst results.

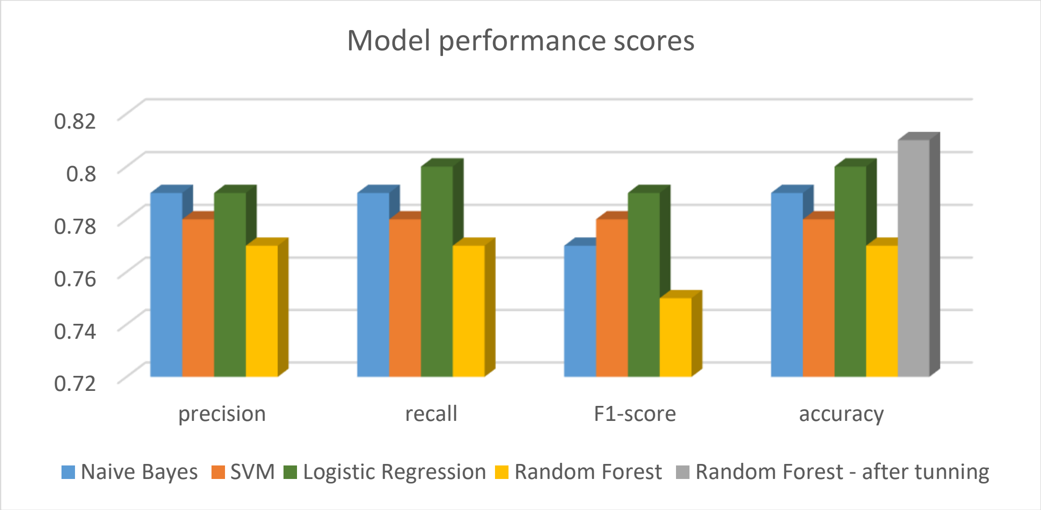

Fig. 5. *Performance measure of different ML models*

6. Conclusion

This paper gives an empirical contribution to the field of data science and sentiment analysis. In this study, we applied and compared multiple machine learning algorithms to classify sentiments from social media posts. Based on the experimental results, the following conclusions are drawn:

- **H0:** Rejected. The performance differences between models were statistically and practically significant.
- **H1:** Accepted. Logistic Regression outperformed other models with an accuracy of 80%, particularly in handling neutral and positive sentiment classes.
- **H2:** Accepted. Hyperparameter tuning improved the performance of Random Forest from 77% to 81%, demonstrating the effectiveness of model optimization.

These results suggest that Logistic Regression and Random Forest, especially when tuned, are effective for sentiment classification tasks. Future work may involve extending the models to multilingual datasets, integrating deep learning approaches, and applying real-time sentiment tracking for social media monitoring.

References

[1] *Kocaleva Vitanova M., Stojanova Ilievska A., Koceska N. and Koceski S.* (2025), "Emotion Detection from Physiological Markers Using Machine Learning". Informatica.

[2] *Koceski, S., Koceska, N., Lazarova, L. K., & Zlatanovska, B.* (2025). Exploring the performance of ChatGPT for numerical solution of ordinary differential equations. JOTSE, 15(1), 18-34.

[3] *Go, A., Bhayani, R., & Huang, L.* (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.

[4] *Read, J.* (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In Proceedings of the ACL student research workshop (pp. 43-48).

[5] *Pak, A., & Paroubek, P.* (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010, pp. 1320-1326).

[6] *Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B.* (2011). Combining lexicon-based and learning-based methods for Twitter sentiment analysis. HP Laboratories, Technical Report HPL-2011, 89, 1-8.

[7] *Thelwall, M., Buckley, K., & Paltoglou, G.* (2011). Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406-418.

[8] *Yadav, N., Kudale, O., Rao, A., Gupta, S., & Shitole, A.* (2021). Twitter sentiment analysis using supervised machine learning. In Intelligent data communication technologies and internet of things: Proceedings of ICICI 2020 (pp. 631-642). Springer Singapore.

[9] *Nafees, M., Dar, H., Lali, I. U., & Tiwana, S.* (2018). Sentiment analysis of polarity in product reviews in social media. In 2018 14th International conference on emerging technologies (ICET) (pp. 1-6). IEEE.

[10] *Sheikh, H. A., & Jaiswal, J.* (2020). Implementing sentiment analysis on real-time twitter data. J ETIR, 7(9).

[11]*Crowdflower*, “Twitter US Airline Sentiment Dataset,” [Online]. Available: https://data.world/crowdflower/twitter-airline-sentiment.

Qazim Tahiri
GoceDelcev University
Faculty of Computer Science
Stip, R.N. Macedonia
E-mail address: kazim.31028@student.ugd.edu.mk

Natasa Koceska
GoceDelcev University
Faculty of Computer Science
Stip, R.N. Macedonia
E-mail address: natasa.koceska@ugd.edu.mk