Predicting Emotions from Twitter Posts: A Comparative Study of Machine Learning Methods

Peihang Li[1,*]

[1] University of Washington, Seattle 98195, USA
*pli8@uw@uw.edu

Abstract. With the increasing importance of social media platforms such as Twitter, understanding the emotions expressed in text data has become crucial for various applications. Manual analysis of the vast amount of user-generated content is impractical, highlighting the need for automated classification techniques. This study focuses on evaluating different machine learning methods for predicting emotions from Twitter posts, specifically examining Multinomial Naive Bayes (MultinomiaNB), Support Vector Machines (SVM), and the Random Forest. A dataset containing over 4000 labeled tweets, categorized as positive, neutral, or negative, is used for evaluation purposes. The challenges associated with predicting emotions from Twitter text, including natural language ambiguity and noise, are carefully considered. The results demonstrate that all models perform well, with SVM exhibiting a slight advantage. This study contributes to a deeper understanding of user emotions and public opinion in social media contexts. Future research directions include refining preprocessing techniques, exploring advanced methods like deep learning, incorporating additional features, and leveraging ensemble learning approaches in order for higher accuracy.

Keywords: emotion prediction, sentiment analysis, random forest, Multinomial Naive Bayes, SVM.

1 Introduction

Social media platforms like Twitter have become increasingly significant in people's daily lives. With millions of tweets posted worldwide each day, these platforms facilitate information exchange and emotional expression [1]. The vast amount of textual data encompasses a range of opinions on various topics, accompanied by emotions. Analyzing these tweets provides insights into public sentiment toward specific subjects. Consequently, understanding the underlying emotions in the text is crucial for marketing, political analysis, and mental health surveillance. However, the sheer volume of user-generated content makes manual analysis impractical, necessitating the application of machine learning techniques [2, 3]. Leveraging these techniques can automate the classification of emotions in large datasets, saving time and effort. This work intends to contribute to the literature by investigating the application of machine learning approaches for identifying emotions in Twitter data

P. Kar et al. (eds.), *Proceedings of the 2023 International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2023)*, Advances in Computer Science Research 108,
https://doi.org/10.2991/978-94-6463-300-9_13

given the potential advantages of emotion prediction, enabling more comprehensive and precise sentiment analysis in the realm of social media [4, 5].

The objective of this study is to implement and compare Random Forest, Multinomial Naive Bayes, and SVM, for emotion classification in Twitter data. By evaluating their relative performance, the study aims to identify the most effective method. The dataset utilized in this research comprises over 4000 labeled tweets, where emotions are categorized as positive, neutral, or negative [6, 7]. However, the study also acknowledges the challenges associated with predicting emotions from Twitter text, such as the ambiguity of natural language, noise present in social media data (e.g., irrelevant posts, spam), and the usage of informal language elements like slang, emoticons, and abbreviations. Ultimately, this study endeavors to enhance the accuracy of emotion classification in Twitter data, enabling more precise sentiment analysis and yielding practical insights for marketers and policymakers. Moreover, it contributes to a deeper understanding of user emotions and public opinion in the context of social media.

2 Literature review

Tang et al. (2014) introduced a system called Coooolll for message-level emotion classification on Twitter using deep learning. The system achieved remarkable results, ranking 2nd among over 40 systems in the SemEval 2014 Task 9, demonstrating its effectiveness in the sentiment classification of tweets [8, 9]. Severyn and Moschitti (2015) developed a system for analyzing emotion in tweets that utilized unsupervised neural language models for word embedding initialization. Their system achieved top-ranking performance in both phrase-level and message-level sentiment analysis subtasks [10]. Jain and Jain (2019) conducted sentiment analysis on tweets data related to renewable energy, by existing machine learning algorithms. They showed that utilizing selecting feature techniques improved classification accuracy, with the SVM and CfsSubsetEval methods achieving the highest accuracy of 93% [11].

In a comparative study by Poornima and Priya (2020), using data from Twitter, the effectiveness of the MultinomiaNB, SVM, and Logistic Regression for the categorization of words in line with sentiment analysis was assessed. The research discovered that using a Bigram model, logistic regression had the best accuracy, at around 86% [12]. Another example study by Zahoor and Rohilla (2020) used machine learning algorithms to analyze sentiment on Twitter. The research found that the Naive Bayes, SVM, Random Forest Classifier, and LSTM algorithms are capable of precise sentiment prediction. These algorithms have demonstrated effectiveness in natural language processing and sentiment analysis, although their accuracy may vary [13]. Aslan et al. utilized a CNN-based method, TSA-CNN-AOA, for sentiment analysis of COVID-19-Related Twitter data. Their approach achieved high accuracy rates, demonstrating its effectiveness in understanding public sentiments towards the pandemic [14].

Despite the progress made, sentiment classification from Twitter data still faces several challenges. The ambiguity of natural language in tweets, characterized by informal language, slang, abbreviations, and emoticons, poses a significant challenge. The presence of irrelevant posts and spam adds noise to the data, further complicating the classification task. Additionally, the imbalanced distribution of positive, neutral, and negative sentiments introduces biases in the models and reduces the performance of minority classes. The current study aims to address these challenges while concentrating on the use of machine learning methods for the categorization of emotions in Twitter data.

3 Methodology

3.1 Data Collection and Preprocessing

The enriched multi-view sentiment analysis dataset (MVSA) was used, it consists of more than 4000 tweets, each labeled as positive, neutral, or negative sentiment. The dataset has been enriched with caption-generated labels using sentiment analysis, providing a reliable ground truth for emotion classification.

The first step in preprocessing was data cleaning. Unwanted elements, such as mentions, stock market tickers, retweet text ("RT"), hyperlinks, hash signs, commas, numbers, colons, and white spaces, were removed from the tweets. Regular expressions were utilized to identify and eliminate these unwanted elements. Additionally, all tweets were converted to lowercase to ensure uniformity. After the initial cleaning, the dataset was checked for duplicate entries and removed to ensure the uniqueness of each instance. Stop words are commonly occurring words, such as "is," "and," and "the," that often have nothing to do with the meaning of a line of words. Removing stop words helped reduce the feature space and emphasized meaningful words in the tweets. A predefined list of stop words from the Natural Language Toolkit (NLTK) was employed for this purpose.

Words are boiled down to their fundamental forms through the process of steaming. As an example, the words "jumping" and "jumps" would be stemmed from "jump." Combining different spellings of the same word into a single phrase streamlines the dataset. The Porter stemming algorithm, implemented in NLTK, was used for stemming. lemmatization, which is analogous to stemming, breaks down the words into their lexical or root forms, or lemma. But lemmatization, which takes into account the sentence's context and the word's part of speech, is a more advanced procedure. The spaCy library was employed for lemmatization.

The final preprocessing step was vectorization, which transformed the text data into a numerical representation for the algorithms to understand. The Term Frequency-Inverse Document Frequency (TF-IDF) vectorizer was applied in this investigation. The statistical tool TF-IDF assesses the significance of a word inside a corpus or document. By emphasizing qualities that only occasionally appear in the training corpus, it lessens the influence of often recurring words that are less informative. The cleaned and lemmatized tweets were then passed through the TF-IDF vectorizer, resulting in a matrix of TF-IDF features. These features were used as input for the machine learning models. After completing these preprocessing steps, the dataset transformed from raw and noisy tweet text into a cleaned and structured form suitable for input into the machine learning algorithms.

3.2 Model Selection and Training

For model selection, three models have been chosen: MultinomialNB, Random Forest, and SVM. Each model offers distinct advantages and is well-suited for specific tasks.

Multinomial Naive Bayes is a probabilistic classifier that excels in multiclass classification tasks with discrete feature values. It is well-suited for analyzing text data, particularly word counts. By assuming feature independence, the algorithm simplifies computation and efficiently learns from the data. Utilizing term frequency vectors, the Multinomial Naive Bayes classifier can categorize tweets into positive or negative sentiments, allowing for the evaluation of sentiment analysis precision. Formula 1 represents the computation of the feature probability P(c|y) in the Multinomial Naive Bayes methodology. Here, 'c' represents potential results or classes, and 'y' corresponds to the specific instance that requires classification, indicating specific characteristics.

$$P(c|y) = \frac{P(y|c)*P(c)}{P(y)} \tag{1}$$

The Random Forest ensemble learning method builds numerous decision trees and outputs the class that shows up most frequently among the various trees. It is widely used in emotion classification tasks due to its robustness, ease of implementation, and ability to handle high-dimensional data. Random Forests are resilient to the inclusion of irrelevant features and provide reliable baseline results, making them a suitable choice for this study.

SVMs are supervised learning models employed for classification and regression analysis. SVMs are particularly adept at handling high-dimensional data, which is a common characteristic of text data. By constructing hyperplanes in a high-dimensional space, SVMs aim to achieve maximum margin separation of tweets in the feature space. The usefulness and efficiency of SVMs in text categorization tasks have been shown in several research. The discriminative function of SVM is defined as formula 2. There are vectors X, w, and b, which stand for the feature, weights, and bias. $\phi()$ maps the input space to the feature space.

$$g(X) = w^T \phi(X) + b \tag{2}$$

4 Results

The performance of the three machine learning models (MultinomialNB, Random Forest, and SVM) was evaluated using three metrics: precision, recall, and F1-score. This paper first used precision to evaluate the performance. Precision is defined as formula 3. It indicates the classifier's ability to avoid mislabeling negative samples as positive. Then the performance was evaluated using recall. Formula 4 gives how it is calculated. The result gives the performance of classifying positive samples. This paper also used the F1 score to evaluate the models. The F1 score is given by formula 5. It shows the accuracy of the models.

$$\text{Precision} = \frac{True\ positive}{True\ positive + False\ positive} \tag{3}$$

$$recall = \frac{True\ positive}{True\ positive + False\ negative} \tag{4}$$

$$F1\ score = \frac{2 \times precision \times recall}{precision + recall} \tag{5}$$

Fig. 1 illustrates a comparison of the models using precision, recall, and F1-score scores obtained from running the Twitter dataset. Among the three models, SVM achieved identical precision, recall, and F1-score, performing slightly better than the other two models. Random Forest also exhibited identical statistics, while MultinomialNB showed slightly lower recall and F1-score compared to SVM and Random Forest.

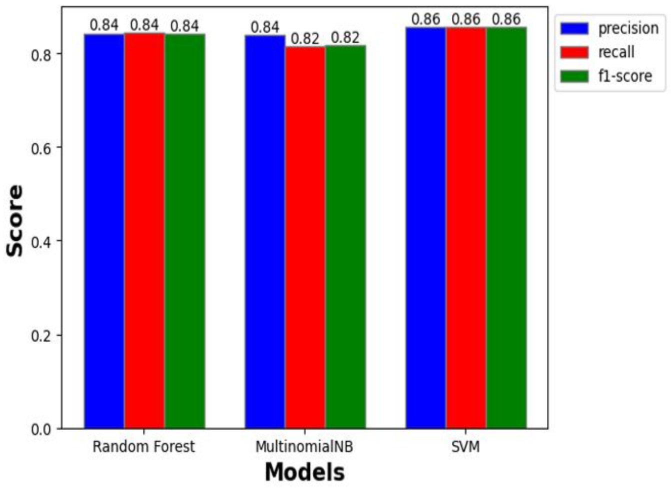

Fig. 1. The measurement results of the three models (Photo/Picture credit: Original)

The performance of a classification model can be assessed using a confusion matrix. It lists the model's predictions in comparison to the actual data labels. The number of true and false positives, true and false negatives is shown in the confusion matrix.

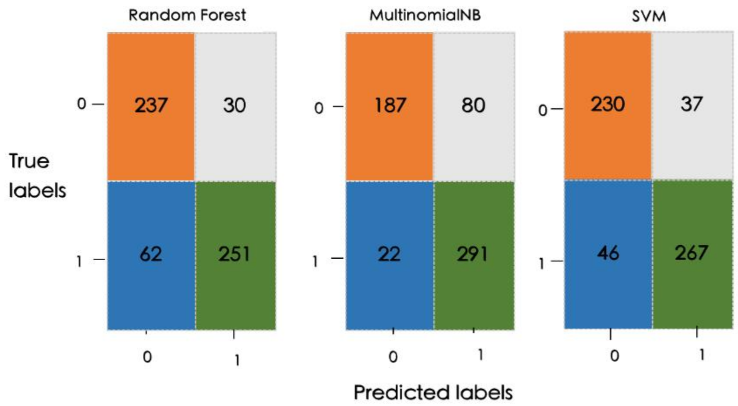

Fig. 2. The results of the three models (Photo/Picture credit: Original)

Fig. 2 illustrates the relationship between the true labels and the predicted labels among the three models. Analyzing these results provides valuable insights into the performance of each model. Random Forest predicted the highest number of true positives and demonstrated moderate performance in classifying negatives. MultinomialNB had the highest number of true negatives but performed relatively weaker in classifying true positives compared to the other models. SVM showed overall good performance, accurately classifying both positive and negative emotions.

In conclusion, the task of classifying emotions in Twitter data was successfully completed by the SVM, Random Forest, and models, with SVM showing a little edge among the three. But each model has advantages and disadvantages of its own. MultinomialNB excels at identifying negative sentiments but struggles with positive ones. Random Forest performs well in recognizing positive emotions but falls short in identifying negative sentiments. SVM demonstrates balanced performance in both positive and negative classification tasks.

5 Conclusion

This study examined the application of three machine learning models, namely Multinomial NB, Random Forest, and SVM, in the task of emotion classification using Twitter data. However, it is important to point out that each model has its own strengths and limitations, highlighting the need to choose the appropriate model based on particular classification requirements. The results showed that while all models performed well, SVM showed a slight lead in terms of the tests. Despite the promising results, several challenges were identified, including the inherent ambiguity of natural language, the presence of noise in social media data, and the usage of informal language, slang, and abbreviations. These challenges emphasize the complexity of the task and underscore the necessity for ongoing refinement of classification models and preprocessing techniques.

Future research should focus on addressing these identified challenges. This could involve refining preprocessing techniques to better handle data noise and the complexities of natural language. Additionally, exploring advanced methods such as deep learning and transformer models, which have shown promising results in natural language processing, could further enhance the accuracy of emotion classification. Furthermore, the incorporation of additional features, such as metadata from tweets (e.g., timestamps, locations, and user information), could be explored to enrich the feature set and improve model performance. An alternative approach worth considering is the integration of multiple models to leverage their complementary strengths. Techniques such as ensemble learning and stacking provide opportunities to harness the collective power of multiple models. By building upon the insights gained from this study and continuously refining the methods, it is hoped that more accurate and robust emotion classification models can be developed to better understand and utilize the abundant emotional information embedded within social media data.

References

1. B. Gupta, et al. Study of Twitter sentiment analysis using machine learning algorithms on Python. International Journal of Computer Applications 165.9, 29-34 (2017).
2. M. S. Neethu and R. Rajasree, Sentiment analysis in twitter using machine learning techniques. 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp. 1-5. Tiruchengode, India (2013).
3. R. Wang and D. Ji. Twitter Sentiment Classification Method Based on Convolutional Neural Network and Multi-Feature Fusion. Computer Engineering 44.2, 210-219 (2018).
4. W. Zhu. Research on sentiment classification and visualization of Twitter. Wuhan University of Technology (2023).
5. S. Zhang, et al. Emotional Contagion Phenomenon in Twitter. Journal of Shandong University 51.1, 71-76 (2016).
6. D. J. Mohammed and H. J. Aleqabie. The Enrichment Of MVSA Twitter Data Via Caption-Generated Label Using Sentiment Analysis. 2022 Iraqi International Conference on Communication and Information Technologies (IICCIT), pp. 322-327. Basrah, Iraq (2022).
7. Niu, Teng, et al. Sentiment analysis on multi-view social data. MultiMedia Modeling: 22nd International Conference, pp. 15-27. Miami, USA (2016).
8. Duyu Tang, et al. Coooolll: A deep learning system for twitter sentiment classification. Proceedings of the 8th international workshop on semantic evaluation. pp. 208-212. Dublin, Ireland (2014).
9. Al-Smadi, Mohammad, et al. Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. Journal of computational science 27, 386-393 (2018).
10. A. Severyn, A. Moschitti. Unitn: Training deep convolutional neural network for twitter sentiment classification. Proceedings of the 9th international workshop on semantic evaluation. pp. 464-469. Denver, USA (2015).
11. A. Jain, and Jain Vanita. Sentiment classification of twitter data belonging to renewable energy using machine learning. Journal of information and optimization sciences 40.2, 521-533 (2019).

12. Poornima and K. S. Priya. A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques. 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). pp. 493-496. Coimbatore, India (2020).
13. S. Zahoor and R. Rohilla. Twitter Sentiment Analysis Using Machine Learning Algorithms: A Case Study. 2020 International Conference on Advances in Computing, Communication & Materials (ICACCM). pp. 194-199. Dehradun, India (2020).
14. Aslan, Serpil, K. Soner, and S. Eser. TSA-CNN-AOA: Twitter sentiment analysis using CNN optimized via arithmetic optimization algorithm. Neural Computing and Applications 35, 1-18 (2023).