

Text-Based Emotion Recognition Using Supervised Machine Learning

Shashwat Rai¹, Gaurav Singh², Mrinmoy Kayal³

^{1,2,3}*Department of SCSE Galgotias University Greater Noida, India*

Abstract—A supervised machine learning model that identifies human emotions in text, such as reviews, chat, comments, social media posts etc. Is referred to as Text Based Emotion Recognition using supervised machine learning. The system uses machine learning algorithms like Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR) to identify various emotions such as happy, sad, anger, fear, surprise and neutral expression. The proposed research was evaluated by a thorough experiment with the standard metrics, namely accuracy, precision, recall and f1-score to analyze and determine which machine learning model perform better on text emotion classification task. This proposed model was trained using a dataset with 416,809 text records in which the result showed that logistic regression outperformed the other machine learning models in terms of accuracy, thus predicting and classifying text emotionally. There are also pre-processing natural language processing techniques such as tokenization, stop-words removal, stemming, and also feature extraction techniques like TF-IDF, Bag-of-Words which were included to perform better classification. The achieved results demonstrate the effective performance of these advanced supervised machine learning models in detecting human emotions in text with minimized manual interference for emotional analysis, can revolutionize human-computer interaction and would be greatly beneficial in fields like mental health monitoring, analysis of customer feedback, chatbots, recommendation engines, learning platforms, social media monitoring etc.

Index Terms—component, formatting, style, styling, insert

I. INTRODUCTION

With the rise of the usage of digital communication such as social media, chat messenger, online forum, blogging, and reviews

website, humans are producing massive text data in every single day. Text content contains sentiment, opinion and emotions that reflect human being behavior, psychology. As more and more unstructured text-based information being produced on the web day by day, manually analyzing and understanding those emotions within the text data has become difficult. This fact has inspired researchers and developers to develop artificial system that automatically recognize and classify the emotion in text in a fast and accurate manner. The Text-Based Emotion Recognition Using Supervised Machine Learning has taken an approach by employing strong machine learning technique that will detect emotions within text and provide a meaningful analysis of emotions in real-time.

The framework for detecting emotions from different text source like, posts on social media, customer reviews, chat messages, online comments. . . are aimed at extracting valuable emotion patterns from text data where every single piece of text might reflect emotion regarding Happiness, Sadness, Anger, Fear, Surprise or Neutral. The machine learning models will look at the training dataset which are already classified and analyze which emotion are strongly relates to what words, sentence patterns. Hence the model learn relationship between text pattern and emotions. From then the classification of new unseen text to a respective emotion will be possible. And thus reducing manual working on emotion text.

Several Supervised Machine Learning Algorithms are examined including Naive Bayes, Decision Tree, Random Forest, Logistic Regression, Support Vector Machine. We have considered all models on equal ground of training and testing data with evaluation metrics including accuracy, precision, recall and F1-score. Among all those algorithms considered

advanced supervised learning models produce better prediction and classification. To improve classification accuracy and effectiveness, we propose to use Natural Language Processing techniques that include tokenization, stop-words removing, stemming and TF-IDF Vectorizer. The preprocessing step will transform the text data into suitable format that machine learning models can take in order to perform classification.

Previously it was time consuming and inefficient to perform analysis based on manual analysis or simply on the basis of keywords on text for identifying sentiments. This system have successfully reduced all the human intervention of manual analysis with an intelligent real time emotion classification within few seconds only. With the increase in size of text data, automatic emotion detection systems is important to have as it can analyze large amount of data without the need for any further manual supervision. The proposed system framework can be applied on any other domain such as medical text for monitoring patient health, for analyzing customer feedbacks to a business., educational text, review recommendation system, and chat-bot or social media analysis and etc. The framework will operate on same principle that is classifying the sentiments based on pre-trained data models that will analyze sentiments on real-time basis.

So the system, Text-Based Emotion Recognition presents itself as a data-driven, scalable, user-friendly framework that may assist humans-computer interaction and improve emotion analysis in digital world. Our main contribution is as following. (1) Design a framework of Text-Based Emotion Recognition system using Supervised Machine Learning with the incorporation of Natural Language Processing and TF-IDF Vectorizer. (2) Perform a comparative evaluation among proposed models (NB, DT, RF, LR and SVM) based on common metrics. (3) This paper proposes Logistic Regression (LR) model that the accuracy higher than Decision Tree (DT), Random Forest (RF) and Naïve Bayes (NB).

The second part of this paper describes the related literature of emotion recognition systems. Part three contains the pro-posed methodology, dataset, preprocessing methods and model development. The fourth part shows experimental results and models comparison while the last part concludes and

suggests the possible future work involving deep learning, and hybrid AI approaches.

II. LITERATURE SURVEY

In recent times, the traditionally keyword-matching based text classification has been gradually replaced by more intelligent and personalized emotion recognition systems with the use of machine learning and NLP techniques. Text-Based Emotion Recognition systems are a tool to extract the emotions such as happy, sad, angry, fear, surprise and neutral in a text content, e.g. Social media, reviews, chats, blogs, and user comments, while machine learning methods are more efficient and accurate in performing the emotion detection than manual methods, however, language ambiguities, sarcasm and lack of labeled dataset, contexts in the language, and different languages in texts still remain as a challenging task. Previous research had focus more on keyword-based and lexicon-based approach where predetermined dictionaries of emotions were used to classify text, however it fail to consider the contexts of text, for example if a user used word like” cool” as” cool is going” and other user” it’s cool to see that”, hence such system does not work well. From later stage, supervised machine learning algorithms such as Naive Bayes, Decision Tree, Logistic Regression, Random Forest and Support Vector Machine, etc. Used to perform better and predict more accurately for emotions recognition tasks than manually and lexicons-based approach, since it learns features for emotions directly from labeled data rather than relying on a pre-specified keyword set, but challenges like data collection and pre-processing remain. During research of machine learning approaches, features extraction and pre-processing methods like tokenization, stop word removal, stemming, lemmatization, Bag-of-Words and TF-IDF Vectorizer are important, and a significant boost in classification

accuracy were achieved with TF-IDF Vectorizer due to its effectiveness in capturing the important emotional words while excluding noisy text. Recently there has been a lot of research in deep learning and AI based hybrid models like Neural Networks, LSTM and transformer-based models, such as BERT, which are found to show great results in

understanding the context and semantic of the text. However, these models requires more computational resource and time and huge labeled data. Therefore for practical applications supervised machine learning models are still the preferred choice of emotion detection owing to the advantage of low computational requirements and shorter time for training and it works better for structured text data. Ensemble based algorithms and optimized models has get more and more attention for this field because it enhances the performance and accuracy in detection as it acts as a robustness parameter for classifier, regularizes model and achieves highly predictive classification model, comparing machine learning models, it's found that Random Forest, SVM achieve highest accuracy compare to other machine learning classifiers. Researchers also emphasize the need for explanation able and scalable systems which not only predict emotion accurately, but also capable of processing a stream of real-time text. Thus, proposed system” Text-Based Emotion Recognition Using Supervised Machine Learning” aim at providing scalable and intelligent system for emotions classification by adopting NLP techniques and machine learning models which aims at high accuracy in predicting emotions and reduced human effort for text analysis. Like any other recommenders and decision-support systems, the proposed model works based on previous training on text documents, predicted unseen text for emotions and generated predictions efficiently along with a scope of expansion of future systems on deep learning models, hybrid models and contextual language understanding.

III. METHODOLOGY

The proposed Text Based Emotion Recognition Using Supervised Machine Learning framework contains around 416,809 textual records from various online textual sources like social media posts, customer reviews, online comments, discussion forums, chat messages, etc. It includes unstructured text data with emotion binary labels for supervised machine learning models to learn and identify the patterns of emotions in text. Positive emotions (happiness, surprise) are labeled 1 and negative emotions (sadness, anger, fear, neutral emotion) are labeled 0 in this study. It is large-scale, which

enhances the robustness, scalability and generalization ability of the proposed framework, enabling machine learning models to learn from various linguistic structures, different vocabulary, writing styles and contextual emotional expressions. The dataset is preprocessed prior to training the models to enhance the quality of textual data and to mitigate any unwanted noise from the raw text, such as lowercasing, tokenization, removing stop words, removing punctuation, URL cleaning, stemming and lemmatisation. The preprocessing techniques aim to normalize the textual data and remove irrelevant variations in the data set. The TF-IDF (Term Frequency–Inverse Document Frequency) Vectorizer is used to transform the cleaned textual data into numerical feature vectors, where the frequency and significance of the text in the entire dataset contribute to assigning importance weights to words, allowing machine learning algorithms to better comprehend emotional information present in the text. The processed data is then split into training and testing sets for modeling and testing. The generated feature vectors are used for emotion classification using multiple supervised machine learning algorithms such as Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM). The size of the dataset is 416,809 records, which greatly helps with enhancing the ability of the models to learn, reducing overfitting and increasing the accuracy of the classifed models. The representation and distribution of emotional categories in the set is depicted in Fig.1 and the pre-processing and classification process of the proposed framework is shown in Fig.2.

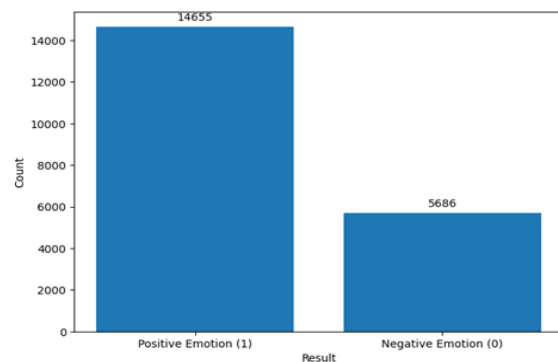


FIGURE 1. TWO labels of sentiment
 Fig. 1. Proposed Text-Based Emotion Recognition Methodology Framework

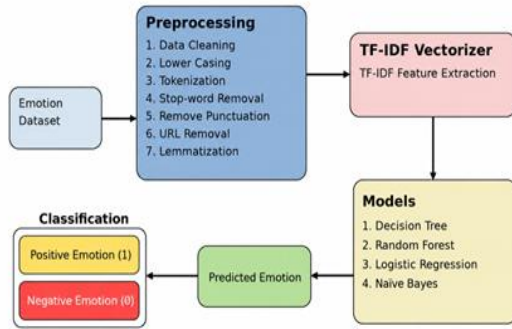


FIGURE 2. Proposed Text-Based Emotion Recognition Methodology Framework

Fig. 2. Proposed Text-Based Emotion Recognition Methodology Framework

IV. RESULTS

Multiple supervised classification algorithms namely De-cision Tree (DT), Logistic Regression (LR), Random Forest

Metric	Decision Tree	Random Forest	Logistic Regression	Naive Bayes
Accuracy	92.38	94.41	96.57	95.51
Precision(0)	94	96	97	95
Precision(1)	91	92	96	96
Recall(0)	92	93	97	97
Recall(1)	92	96	96	94
F1-Score(0)	93	95	97	96
F1-Score(1)	92	94	96	95
Sensitivity	91	96	96	94
Specificity	92	93	97	97

Fig. 3. Performance Comparison Using Different Machine Learning Classi-fiers

(RF) and Naive Bayes (NB) were utilized to evaluate the proposed Text-Based Emotion Recognition using Supervised Machine Learning framework, and were compared based on various evaluation metrics namely accuracy, precision, recall and F1-score. The experiments were carried out on the widely used large textual dataset consisting of 416,809 records after carrying out NLP preprocessing steps and using TF-IDF Vec-torizer to extract features. The results of experiments carried out highlight the effectiveness of the supervised machine learning models in correctly predicting the emotions based on the text data. Decision Tree(DT) model performed with the accuracy of 92.38 and provided good recall score for both the positive and negative classes; however, the performance in terms of precision and F1-score was relatively low compared to other classifiers because

of the inherent complexity and vagueness of the emotions expressed through text. From the analysis of the confusion matrix of DT, it is observed that most of the emotions have been classified correctly with the exception of few misclassified emotions. Logistic Regression

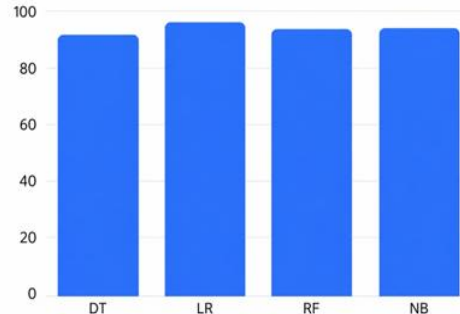


Fig. 4. (a) Performance of Accuracy

(LR) model achieved a maximum accuracy of 96.57 compared to all the models used to evaluate the proposed framework, and it provided the better prediction along with high scores for precision, recall and F1-score for both positive and negative classes. Also, the confusion matrix of LR shows the minimum prediction for the false cases, thus LR performed better than

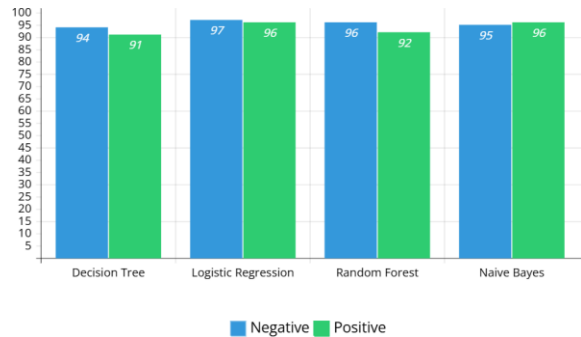


Fig. 5. (b) Performance of Precision

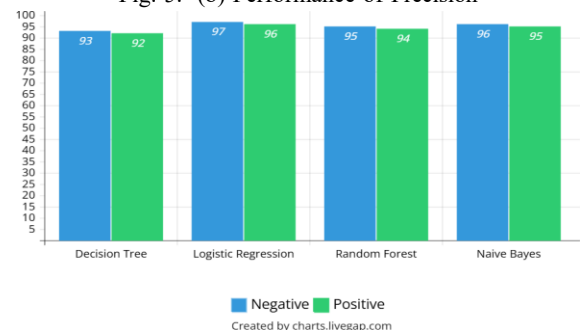


Fig. 6. (c) Performance of Recall



Fig. 7. (d) Performance of F1-Score

all other classifiers used to develop the framework. Similarly, Random Forest (RF) model achieved an accuracy of 94.41 and showed a great stability while predicting and gave good recall and F1-score, which shows that this model is very capable of extracting the patterns in the text and detecting the emotions. Also, Naive Bayes (NB) model achieved the performance with 95.51 of accuracy and performed well by effectively processing the TF-IDF vectors and requiring low computational time. Based on all the observed evaluation parameters, it is clearly shown from the comparison that the proposed LR based framework has achieved much better accuracy than the others; DT, RF and NB. From all these results, it is strongly proved that the proposed framework which is based on supervised machine learning with NLP pre-processing and TF-IDF feature extraction can achieve accurate, scalable and efficient automatic emotion detection from text. This proposed framework can be applicable in various applications like sentiment analysis, mental health monitoring, chatbot systems, customers review, and social media analytics. The comparative analysis of all the machine learning algorithms is presented in Figure 3, and the confusion matrices of each of the models used for the detection of emotions from text data are given in Figure 4.

Besides this, confusion matrix analysis show that the proposed system have strong classification ability to predict positive and negative emotions class and there were low misprediction, in the Logistic Regression it was found and this balancing classification ability is critical for actual emotion analysis system, since misclassification of emotion can result

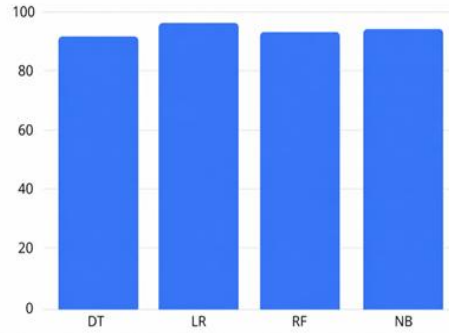


Fig. 8. (e) Performance of Sensitivity

in inappropriate and wrong analysis to system like monitoring people’s mind status, understanding consumer opinions, intelligent chat-bot. It also reinforces that integrating NLP preprocessing technique to the TF-IDF vectorization greatly increase supervised model’s ability to find latent emotion from text; consequently, the proposed framework provides robust, scalable and computationally efficient way to automated emotion recognition, and provide solid baseline to future work in deep learning and transformer-based models.

V. CONCLUSION

This model or system for Text Based Emotion Recognition using supervised machine learning is a smart, fast and automatic process of classifying emotions in the texts. In this framework of Text Based Emotion Recognition using Super-vised Machine Learning provides a smart and fast system that can automatically classify and distinguish among emotions from text data. Several supervised machine learning algorithms that are Decision Tree (DT), Random Forest (RF), Logistic Regression (LR) and Nave Bayes (NB) are applied and compare the models so to predict the appropriate model that correctly predict emotions. From the experimental results it has shown that Logistic Regression has achieve an accuracy of 96.57 which performs very well on complex text emotional patterns. The usage of NLP and TF-IDF vectorization allows to process the emotional patterns in text and thus provides very good results for text classification. Preprocessing techniques that are tokenization, stop word removal, stemming, and feature

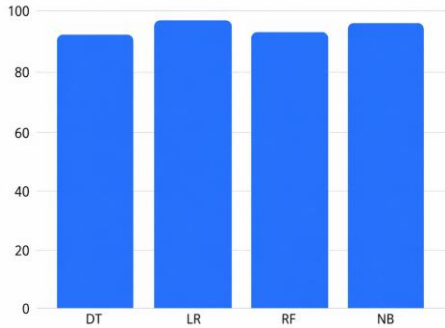


Fig. 9. (f) Performance of Specificity

extraction convert unstructured data into suitable numbers that are used in classification. This framework not only re-duces the manual effort but also provides a system which is more scalable, reliable and efficient. This system can be implemented for many applications such as sentiment analysis, mental health monitoring, customer feedback analysis, chatbot, and in an educational system as well as in social media analysis etc. It also indicates that machine learning algorithms should be compared to discover the best one for classification of emotions. Although the presented work has obtained good performance it could further be improved with the integration of deep learning, transformer-based models such as BERT, multilingual emotion detection, sarcasm in text and adaptive learning real time which will increase its accuracy. Further development of this work can also involve incorporating deep learning algorithms, transformers such as BERT, multilingual emotion classification, detection of sarcasm and understanding context, real-time adaptation, hybrid AI techniques and many more, so as to improve accuracy in predicting emotions.

ACKNOWLEDGEMENT

I would like to be thankful to you on behalf of my project guide Dr. Mrinmoy Kayal (Asistent professor) who helped in leading us and giving us. value information, and affirmative response throughout the research process of undertaking this research. Their technical aptitude as well as have prudent thoughts in the systematizing, performing, and examination of the appreciate system of recipes. has been presented in this paper. I also need to say that it was facilitated by the application

of open-source tools, by the datasets to be used, and online. a research center was put in place and this helped in designing, testing and evaluation of the system. And finally, I would like to express my gratitude to my coworkers and people who were interested in me and encouraged me. than which led me to the present study, with a high degree of commitment and perseverance.

REFERENCES

- [1] Bondevik, Jon Nicolas, Kwabena Ebo Bennin, Önder Babur, and Carsten Ersch. "A systematic review on food recommender systems." *Expert Systems with Applications* 238 (2024): 122166.
- [2] Rodrigues, Miguel Simoes, Filipe Fidalgo, and Angela Oliveira. "RecipeIS—Recipe recommendation system based on recognition of food ingredients." *Applied Sciences* 13, no. 13 (2023): 7880.
- [3] Chhipa, Shubham, Vishal Berwal, Tushar Hirapure, and Soumi Banerjee. "Recipe recommendation system using TF-IDF." In *ITM Web of Conferences*, vol. 44, p. 02006. EDP Sciences, 2022.
- [4] Teng, Chun-Yuen, Yu-Ru Lin, and Lada A. Adamic. "Recipe recommendation using ingredient networks." In *Pro-ceedings of the 4th Annual ACM Web Science Conference*, pp. 298-307. 2012.
- [5] Kayal, Mrinmoy, Mohinikanta Sahoo, and Jayadeep Pati. "Review on Mental Healthcare System Using Data Analytics and IoT." In *International Conference on Global Mental Health and Public Health Challenges and Innovation*, pp. 72-78. Cham: Springer Nature Switzerland, 2022.
- [6] Kayal, Mrinmoy, Mohinikanta Sahoo, Jayadeep Pati, and Ranjan Kumar Behera. "Quantum-Inspired Aspect-Based Sentiment Analysis Using Natural Language Processing." In *Advances in Quantum Inspired Artificial Intelligence*, pp. 151-169. Springer, Cham, 2025.
- [7] Kayal, Mrinmoy, Mohinikanta Sahoo, Jayadeep Pati, Law Kumar Singh, Rajni Mohana, and Rekha Singh. "Comparative Evaluating of Machine Learning Classifiers through Sentiment Analysis to Detect Emotions Using TF-IDF." In

- 2025 7th International Conference on Signal Processing, Computing and Control (ISPC), pp. 619-623. IEEE, 2025.
- [8] Kayal, M., Singh, L.K., and Pati, J.” Sentiment Analysis of User Comments Using TF-IDF and Machine Learning Clas-sifiers: A Comparative Study.” Recent Trends in Intelligent Computing and Communication: Volume 2, p. 449, 2025.
- [9] Kayal, M., Patra, J., Kumar, S., Pati, J., and Sawan,”Enhancing Mental Health Treatment Prediction Using BXGBoost: A Hybrid Machine Learning Approach.” In 2025 International Conference on Computing, Intelligence, and Application (CIACON), pp. 1-6. IEEE, 2025.
- [10] Kayal, M., Patra, J., Kumar, S., Pati, J., Sah, D.K., and Sawan, V. ”Co-DLGAX: A Hybrid Ensemble Framework for Emotion Detection via NLP-Driven Sentiment Analysis.” In 2025 International Conference on Computing, Intelligence, and Application (CIACON), pp. 1-6. IEEE, 2025.
- [11] Kayal, M., Pati, J., Kumar, S., and Singh, L.K. ”Para-metrically Modified BERT for Emotion Prediction through Sentiment Analysis.” Discover Computing 29, no. 1 (2026):69.
- [12] Liu, Huiting, Kun Cai, Peipei Li, Cheng Qian, Peng Zhao, and Xindong Wu. ”REDRL: A Review-Enhanced Deep Reinforcement Learning Model for Interactive Recommenda-tion.” Expert Systems with Applications 213 (2023): 118926.
- [13] Pecune, Florian, Lucile Callebert, and Stacy Marsella. ”A Recommender System for Healthy and Personalized Recipes Recommendations.” HealthRecSys@RecSys, 2020.
- [14] Lambay, Muhib Anwar, and S. Pakkir Mohideen. ”A Hybrid Approach Based Diet Recommendation System Using ML and Big Data Analytics.” Journal of Mobile Multimedia 18, no. 6 (2022): 1541-1560.
- [15] Durairaj, Sureshkumar. ”Enhanced Online Product Simi-larity Classification Using Description and Images.” PhD diss., National College of Ireland, Dublin, 2024.