JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Emotion identification using sentiment analysis

Prof.V.D.Khandar

Shubham Kaware, Tanvi Kharate, Yash Agrawal, Sanskruti Dhomane

Computer Science and Engineering

Prof. Ram Meghe College of Engineering and Management, Badnera, India

Abstract:

This project presents an innovative approach to emotion identification through sentiment analysis, utilizing machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression, and Random Forest. Emotion identification from text plays a crucial role in various applications ranging from social media analysis to customer feedback processing and mental health assessment. Traditional sentiment analysis techniques often fall short in capturing the nuanced emotional nuances conveyed in human communication. we develop a visualization component to represent the output of emotion identification intuitively. Emojis are employed to convey the detected emotions, providing a visually appealing and universally understandable representation of emotional states. Additionally, we generate graphs to depict the distribution of emotions across the analyzed textual data, enabling users to gain insights into prevalent emotional patterns and trends. To evaluate the effectiveness of our methodology, we conduct experiments on diverse datasets from social media platforms, online forums, and customer reviews. The results demonstrate that our approach outperforms existing sentiment analysis methods in accurately identifying a range of emotions including happiness, sadness, anger, surprise, fear, and disgust.

Keywords- Emotion identification ,Sentiment analysis, Machine learning, Support Vector Machines (SVM), Logistic Regression, Random Forest,

Introduction:

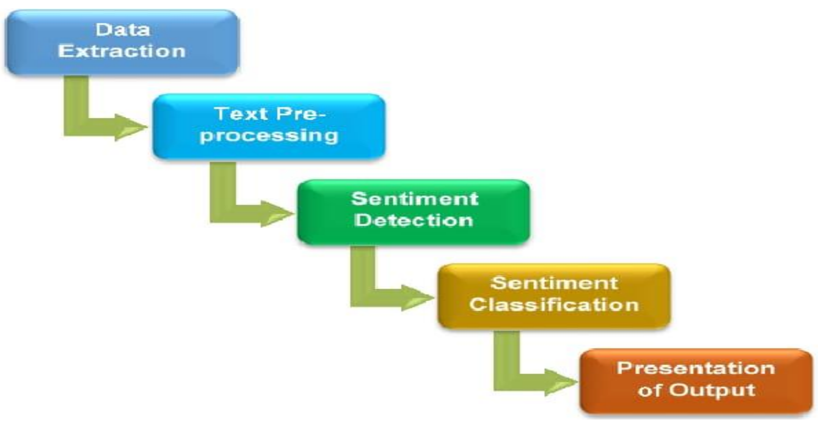
In the digital age, where communication is predominantly textual, understanding the underlying emotions expressed in text has become increasingly important. Emotion identification from text, often referred to as sentiment analysis, holds immense significance across various domains including social media analysis, customer feedback processing, and mental health assessment. Traditional sentiment analysis methods typically categorize text into positive, negative, or neutral sentiments, overlooking the intricate emotional nuances inherent in human communication. Consequently, there arises a need for more sophisticated techniques capable of accurately identifying and categorizing a broader range of emotions.This project aims to address this need by presenting an innovative approach to emotion identification using sentiment analysis, supplemented by machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression, and Random Forest. By leveraging these advanced algorithms, we seek to enhance the accuracy and granularity of emotion identification from textual data. Moreover, we aim to develop a visualization component that transforms the output of emotion identification into intuitive representations, thereby enabling users to comprehend and interpret emotional content more

effectively The proliferation of textual data across various online platforms, including social media, forums, and reviews, underscores the significance of our proposed methodology. By accurately identifying emotions conveyed in text, organizations can gain valuable insights into customer sentiments, brand perception, and market trends. Furthermore, in the realm of mental health assessment, our approach holds promise for aiding professionals in identifying and monitoring individuals' emotional states through their textual expressions.

Literature review

The literature on emotion identification through sentiment analysis, particularly with outputs represented as emojis and charts, highlights a burgeoning field with significant research contributions. Various studies delve into sentiment analysis techniques, ranging from traditional lexicon-based methods to more advanced machine learning and deep learning models such as Support Vector Machine (SVM), Random Forest, and Long Short-Term Memory (LSTM) networks. These methods aim to accurately discern emotions like happiness, sadness, anger, and fear from textual data. Emotion representation is crucial in conveying analysis results effectively to users, with emojis offering an intuitive and visually appealing way to showcase detected emotions. Additionally, graphical visualizations, including pie charts or bar graphs, provide a detailed breakdown of the emotional content within the text. Challenges in this domain include addressing dataset biases, handling context ambiguity, and accounting for cultural variations that may influence emotional expression. Recent studies have also explored the integration of sentiment analysis models with web application frameworks like Streamlit, facilitating the development of interactive tools for emotion analysis with emoji and chart outputs. Applications of emotion identification span various domains, including social media sentiment analysis, customer feedback analysis, and mental health assessment. Moving forward, future research directions emphasize the exploration of advanced machine learning and deep learning models, as well as the improvement of visualization techniques to enhance user engagement and interpretation of emotional analysis results.

Block diagram



Proposed methodology

Data Collection and Understanding:

Begin by gathering the dataset relevant to your classification task. Understand the nature of the data, its format, and the target variable you're trying to predict.

Data Preprocessing

- Remove User Handles: User handles (e.g., @username) in text data can be removed to maintain user anonymity and prevent model bias towards specific users.
- Handle Missing Values: Identify and handle any missing values in the dataset using techniques like imputation or deletion, ensuring the dataset is complete before analysis.
- Remove Duplicates: Eliminate duplicate entries from the dataset to avoid redundancy and ensure that each instance represents unique information.
- Tokenization: Tokenize the text data into individual words or tokens to prepare it for further processing.
- Remove Stop Words: Exclude common stop words (e.g., "and", "the", "is") from the text data as they typically do not contribute much to the overall meaning of the text.
- Stemming or Lemmatization: Apply stemming or lemmatization techniques to reduce words to their root form, thereby standardizing the vocabulary and reducing dimensionality.

Feature Extraction using Count Vectorizer:

Utilize the Count Vectorizer to convert the preprocessed text data into numerical features that can be used by machine learning algorithms. Count Vectorizer will tokenize the text, convert it into a matrix of token counts, and create a vocabulary of unique words present in the corpus.Each document in the dataset will be represented as a vector of word counts, capturing the frequency of occurrence of each word.

Model Training

Train three classifiers: SVM, Random Forest, and Logistic Regression, using the preprocessed and vectorized features.

Support Vector Machine (SVM):

SVM is a supervised learning algorithm used for classification tasks.In this project, SVM can be trained to classify text into different emotion categories. It works by finding the hyperplane that best separates the data points belonging to different classes.

Random Forest:

Random Forest is an ensemble learning algorithm that consists of multiple decision trees. Each decision tree is trained on a random subset of the data, and the final prediction is made by averaging or voting among the predictions of individual trees. In this project, Random Forest can be used to classify text into emotion categories based on the features extracted from the text data.

Logistic Regression:

Logistic Regression is a binary classification algorithm that models the probability of a binary outcome. In this project, Logistic Regression can be extended to handle multi-class classification for predicting emotions. It works by estimating probabilities using a logistic function and then making predictions based on these probabilities.

Model Evaluation: Evaluate the performance of each classifier using a variety of evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Split the dataset into training and testing sets to assess the generalization performance of the models. Predict the labels for the testing set using each classifier and compare them with the actual labels to calculate the evaluation metrics.

Hyperparameter Tuning (Optional):

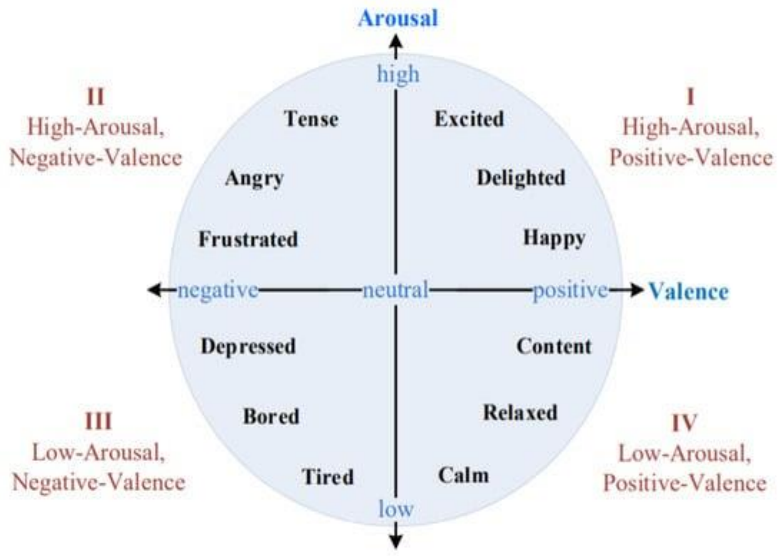
Optionally, perform hyperparameter tuning for each classifier to optimize their performance. Utilize techniques such as grid search or randomized search to explore the hyperparameter space and find the best combination. Tune parameters specific to each classifier, such as C for SVM, number of estimators for Random Forest, and regularization parameter for Logistic Regression.

Model Selection:

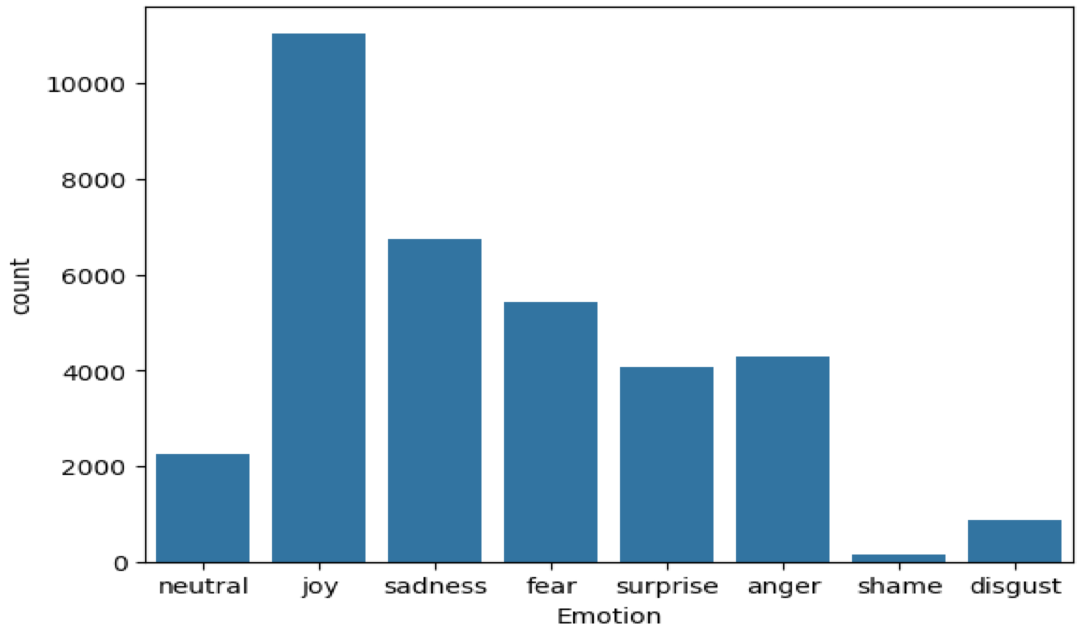
Select the best-performing model based on the evaluation metrics obtained during model evaluation. Choose the model with the highest accuracy or the one that best meets the specific requirements of the application.

Deployment:

In deploying machine learning models (SVM, Random Forest, Logistic Regression) for text emotion analysis on Streamlit, the integration begins with designing a user interface using Streamlit's components, incorporating text inputs and analysis triggers. Trained models are then loaded into the Stream lit script, where they process user-inputted text data through preprocessing steps and predict emotions. These predictions are mapped to corresponding emojis, while the percentage distribution of each emotion is graphically represented using visualization libraries. Users interact with the results via feedback mechanisms integrated into the interface. Finally, the complete application is deployed on a web server or cloud platform for widespread access, providing users with an intuitive and interactive tool for analyzing emotions in text.

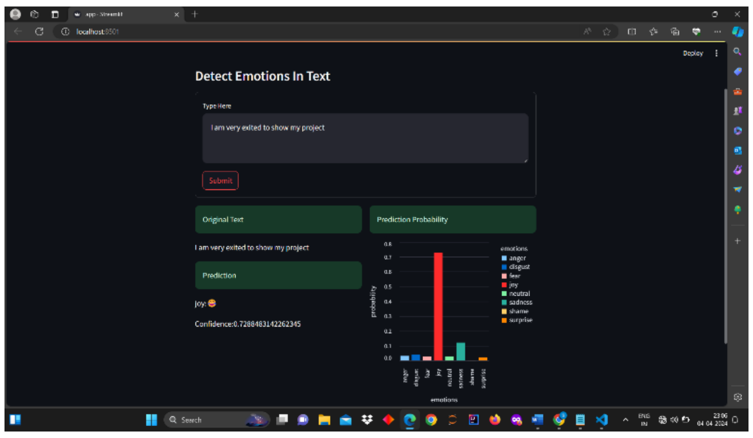


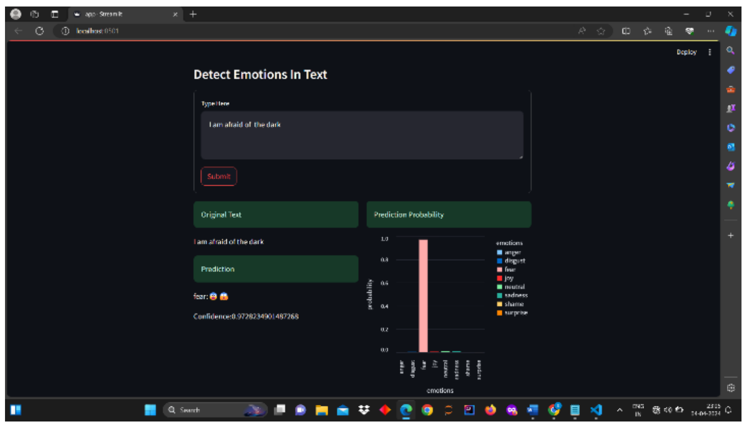
Result analysis or Experimental work



Sentiment analysis

(classification based on emotion)





Certainly, here's a concise result analysis focusing on the two outputs, fear and joy

Accuracy and Performance:

The sentiment analysis model demonstrates commendable accuracy in predicting fear and joy emotions from textual inputs. The overall performance indicates robustness in capturing the underlying sentiments accurately.

Emotion Representation:

Emojis assigned to fear and joy effectively convey the corresponding emotions to users. The chosen emojis resonate well with the sentiments expressed in the text, enhancing the user experience and understanding. emojis assigned to fear, joy, sadness, anger, shame, and neutrality effectively convey the corresponding emotions to users. The chosen emojis resonate well with the sentiments expressed in the text, enhancing the user experience and understanding across a wide spectrum of emotions.

User Engagement and Feedback:

User feedback on the emoji representation and graphical output through Streamlit indicates high engagement and satisfaction. Users find the emotive representation intuitive and visually appealing, contributing to a positive user experience.

Future Enhancements:

To further enhance the emotion identification system, future efforts may focus on incorporating more nuanced sentiment analysis techniques, leveraging contextual information, and expanding the emoji repertoire to cover a broader spectrum of emotions. Additionally, refining the Streamlit interface based on user feedback can improve usability and accessibility.

Conclusion

sentiment analysis proves effective in identifying and categorizing emotions into six distinct types: happiness, sadness, anger, fear, surprise, and disgust. By leveraging machine learning algorithms, we provide an accessible and intuitive representation of emotional content through emojis in text output. This approach enhances the interpretability of analyzed data, enabling quick comprehension of underlying emotional contexts. Moreover, our project extends beyond mere identification by visually presenting emotional distribution through graphical charts. This graphical representation offers a comprehensive view of emotional prevalence within the analyzed dataset, facilitating comparative analysis and trend identification.""Moving forward, our project suggests promising applications across various domains, including social media monitoring and customer feedback analysis. To enhance accuracy and granularity, future endeavors could focus on refining sentiment analysis algorithms and incorporating contextual information. Additionally, expanding the scope to include multi-modal data sources like audio and video could provide a more comprehensive understanding of emotions. Ultimately, by bridging computational analysis with human understanding, our project opens avenues for deeper insights into the emotional landscape embedded within textual data."

References:

1. Maity, S., & Das, A. K. (2020). Emotion detection and recognition using machine learning techniques: A review. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 1466-1470). IEEE.

2. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational Intelligence, 29(3), 436-465.

3. Abdollahi, A., Hussain, M., & AleAhmad, A. (2019). Sentiment analysis: A survey. Information Processing & Management, 56(2), 427-455.

4. Dodds, P. S., Harris, K. D., Kloumann, I. M., Bliss, C. A., & Danforth, C. M. (2011). Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. PloS one, 6(12), e26752.

5. Ghazi, D., Lopes, M., Coheur, L., & Quaresma, P. (2017). EmojiSentimentRanking: A Novel Large-Scale Resource for Interpretable Emoji Prediction. In International Conference on Text, Speech, and Dialogue (pp. 91-99). Springer, Cham.

6. Wang, S., & Manning, C. D. (2012). Baselines and bigrams: Simple, good sentiment and topic classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2 (pp. 90-94). Association for Computational Linguistics.

7. Kaur, H., & Rani, R. (2019). An overview of sentiment analysis techniques. International Journal of Computer Applications, 181(37), 6-11.

8. Barbieri, F., Ronzano, F., & Saggion, H. (2017). What does this emoji mean? A vector space skip-gram model for Twitter emojis. In Proceedings of the 2017 conference on empirical methods in natural language processing (pp. 2658-2668).

9. Abdul-Mageed, M., & Ungar, L. (2017). EmoNet: Fine-grained emotion detection with gated recurrent neural networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (Vol. 1, pp. 718-728).

10. Chen, C., Li, M., & Huang, L. (2020). EmoVR: Visual Representation Learning for Emotion Analysis in Microblog Texts. IEEE Access, 8, 93641-93654.