Humorous Emotion Detection Model for Social Media Posts

1st Faiaz Mohammad Tiham
Dept. of CSE
Brac University
Dhaka, Bangladesh
faiaz.mohammad.tiham@g.bracu.ac.bd

2nd Sajid Imam Mahir
Dept. of CSE
Brac University
Dhaka, Bangladesh
sajid.imam.mahir@g.bracu.ac.bd

3rd Fahad Al Mannan
Dept. of CSE
Brac University
Dhaka, Bangladesh
fahad.al.mannan@g.bracu.ac.bd

4th Navid Alvi Ahsan
Dept. of CSE
Brac University
Dhaka, Bangladesh
navid.alvi.ahsan@g.bracu.ac.bd

***Abstract*—This research paper proposes a text humor detection model using machine learning for social media posts. The model utilizes a combination of feature engineering with pre-processed data to identify humorous posts on social media. The proposed approach is evaluated on a dataset of social media meme images and compared to several baseline models. The experimental results demonstrate how accurate the current baseline models are in detecting humorous posts. Overall, the proposed text humor detection model can be valuable for various applications, including social media content moderation, sentiment analysis, and recommendation systems.**

***Index Terms*—Humor detection, Classification, Naive Bayes.**

I. Introduction

In natural language processing, detecting humor in texts is becoming increasingly important day by day. Natural Language Processing allows the machine, or computer, to understand human language. The one we are focusing on is the written English human language acquired from images, and to give optimal replies to those texts, a machine should be well-prepared. A usually written text has one meaning, but someone can potentially mean something completely different by writing the same thing. We see this practice in our daily lives, particularly on social media. In our daily lives, we observe that a user may occasionally be banned from using specific social media platforms, such as Twitter, Facebook, etc. These texts or meme images were mainly marked as "hate speech" by the AI model on the back end of these social media platforms. So, our main purpose is to classify texts as humorous and their overall feel without marking them directly as hate speech. Yes, obviously there are shortcomings in this process, but directly classifying non-hate-speech as hate speech and banning users is not a good end result for anyone. Our proposed model focuses on both linguistic and contextual aspects of the social media posts to actually classify them as humorous if they are. Our system was evaluated on baseline models for now which are: Decision Tree, Logistic Regression, and Naive Bayes.

II. Pre-processing

For training and testing the Humorous Emotion Detection Model for Social Media Posts, the provided data needs to be processed beforehand to ensure that the data is in a suitable format for modeling and analysis. Data preprocessing is an important phase for machine learning and analysis. For this project, the data is preprocessed by the following steps:

A. Data Loading

The dataset is first loaded from the provided source. We have mounted the dataset in our Google Drive. The dataset consists of social media posts that have been already Optical Character Recognition (OCR) labeled. The labels include categorical features such as humor, sarcasm, motivation, offensiveness, and target feature overall sentiment. The dataset we picked is the famous 'Memotion Dataset' [1] with almost seven thousand entries in a `.csv` file.

B. Handling Missing Values

Missing values can lead to inaccurate analyses resulting in incorrect results. This can be addressed by the following steps:

- **Finding**: Checking and finding the missing or null values in the dataset using the `isnull()` function.
- **Filling**: We have Imputed or filled the missing values in the 'text_corrected' column with the mode value of the same column to ensure data completeness.

C. Text Preprocessing

Text preprocessing is also an important phase where the texts from the dataset are cleaned and standardized. We have done text preprocessing by the following steps:

- **Lowercasing**: All texts in the `text_corrected` column were converted to lowercase.
- **URL Removal**: URLs in the text were replaced with spaces to eliminate unnecessary links.
- **Special Character Removal**: Non-alphanumeric characters were removed from the text, including punctuation marks and symbols.
- **Keyword Replacement**: Specific keywords, such as website domains and common web-related terms, were replaced with spaces.

D. Label Encoding

The categorical labels in the dataset were encoded into numerical values using scikit-learn's LabelEncoder. This allows machine learning algorithms to work with the categorical labels effectively.

E. Data Splitting

The dataset was split into training and testing sets using the `train_test_split` function from scikit-learn. The training set was used to train the machine learning models, while the testing set was used to evaluate their performance the ratio here is Training set 80% and Testing set 20%.

F. Feature Scaling

To ensure that features had similar scales in a fixed range, we applied feature scaling using the StandardScaler from scikit-learn. This step is essential when working with algorithms sensitive to feature scales, such as SVM.

III. METHODOLOGY

Methodology refers to a systematic and structured set of techniques and processes used to achieve the project's objectives. It is crucial as it describes how research is conducted and what steps are followed in an orderly manner. The methodology employed for our humorous emotion detection model for social media posts is as follows:

A. Data Preprocessing

Data Preprocessing is a method of processing datasets to a suitable condition before using it for training and testing to increase the efficiency of the model. We have performed data preprocessing in our project by the following:

- Data Loading: We have loaded data from the provided source by mounting Google Drive and linking the path to the dataset.
- Handling Missing Data: We have handled the missing values in the dataset by imputing with the mode value.
- Text Preprocessing: We have processed and standardized texts from the dataset.
- Data Splitting: We have split the data into training and testing datasets.
- Feature Scaling: We have scaled and encoded categorical data and labels

B. Feature Engineering

Feature engineering is crucial when working with machine learning models sensitive to features and their scales [2]. This helps to increase a model's efficiency and ensure the model's success. We have engineered the features for this project by Encoding categorical data and Feature Scaling:

- Encoding Categorical Labels: We encoded categorical labels which are humor-related such as humor, sarcasm, motivational, offensive, and overall sentiment, into numerical values using Label Encoding.
- Feature Scaling: We scaled the features using StandardScaler from scikit-learn to ensure that all the features are on a similar scale.

C. Model Selection

We have initially considered five different machine learning algorithms for our projects, namely Decision Trees, Logistic Regression, Naive Bayes, K-Nearest Neighbors (KNN), and Support Vector Machines (SVM). We then selected three models from these five models after a detailed comparison based on various criteria, such as follows:

1) Comparison of Models: After training and testing the datasets with each of these models, we measure the accuracy of these models in classifying humorous emotions. The results of this comparative analysis based on their accuracy, shown in Table I, revealed significant variations in model performance.

TABLE I
COMPARISON OF MODEL PERFORMANCE

Model	Accuracy
Decision Tree	0.4753
Logistic Regression	0.4103
Naive Bayes	0.4396
KNN	0.4210
SVM	0.4375

2) Model Selection Criteria: We have compared and selected our Machine Learning models based on certain criteria that adequately assess the efficiency of models. The criteria are as follows:

- **Accuracy**: Accuracy refers to the proportion of correctly predicted instances out of the total instances in a dataset. We have calculated the accuracy of each model, which indicates the ability to correctly classify humorous emotions. Models with higher accuracy were preferred.
- **Robustness**: Robustness refers to the degree to which the accuracy of a model changes when testing with a variety of data. We have assessed the robustness of models against noise and variations in social media text data. The models that have performed consistently across a variety of different datasets were favored.
- **Computational Efficiency**: Computational Efficiency refers to the time or memory required to efficiently

execute a machine learning algorithm. We have examined the computational efficiency and the models with less memory and other resources required were fancied.

D. Model Choice

After considering and comparing the five machine learning models based on the specific criteria, we have selected three machine learning algorithms out of these five, namely Decision Trees, Naive Bayes, and Support Vector Machines (SVM). These models have displayed distinction in terms of accuracy, robustness, and computational efficiency in the context of detecting humor from social media text data. The Models are as follows:

- **Decision Trees**: Decision trees are supervised learning algorithms that are effectively used for classification and regression tasks. Decision trees are interpretable, and robust, with their accuracy being the highest at 0.4753 amongst the five machine learning models, making it an excellent choice.
- **Naive Bayes**: Naive Bayes is a probabilistic machine learning algorithm that is used mostly for classification tasks. It is known for its efficiency in handling text data and has an accuracy of 0.4396, making it a suitable choice for this project.
- **SVM**: Support Vector Machines (SVM) is another example of a supervised machine learning algorithm that is used for classification and regression tasks. Its robustness and ability to handle high volumes of data [4] like social media texts along with its third-highest accuracy of 0.4375, makes it a preferable choice to use in our project.

E. Model Training

Model training is a process of machine learning algorithm, where the model learns different behavioral patterns from the dataset and gets ready to execute the classification and regression tasks accurately. We have trained the model utilizing the training data which was split, using the fit function from sci-kit learn models.

F. Performance Metrics

To evaluate model performance, we employed a set of performance metrics, including accuracy, precision, recall, and F1-score. These metrics were chosen due to their relevance to the problem of humorous emotion detection.

G. Results Interpretation

Finally, we have interpreted the results of each model by analyzing confusion matrices, visualization of predicted vs. true labels, and histograms of accuracy scores. The insights gained from these analyses are discussed in more detail in the Result Analysis section.

IV. Result Analysis

In this section, we present and analyze the results obtained from our experiments in developing the Humorous Emotion Detection Model for Social Media Posts. The analysis provides insights into the model's performance and its effectiveness in detecting humorous emotions in social media posts.

A. Model Performance Metrics

We assessed the performance of each model using a range of evaluation metrics, including accuracy, precision, recall, and F1-score. These metrics offer a comprehensive view of the models' abilities to correctly classify emotions in social media content.

B. Comparison of Models

To identify the most suitable model for the emotion detection task, we compared the accuracy scores of various models. The following observations were made:

TABLE II
Comparison of Accuracy, Precision, Recall, and F1 Score

Model	Accuracy	Precision	Recall	F1 Score
Decision Tree (DT)	0.4753	0.448	0.475	0.421
Naive Bayes (NB)	0.4396	0.328	0.440	0.273
Support Vector Machine (SVM)	0.4375	0.191	0.437	0.266

- The Decision Tree model exhibited 0.4753 accuracy.
- Multinomial Naive Bayes demonstrated 0.4396 accuracy.
- The Support Vector Machine (SVM) with a linear kernel attained 0.4375 accuracy.

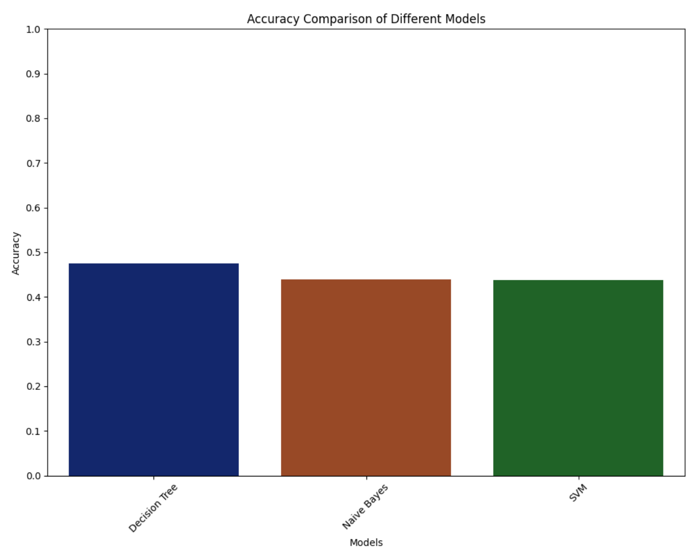

Fig. 1. Accuracy Comparison

Based on these results, the Decision Tree Model emerges as the most promising choice for humorous emotion detection.

C. Confusion Matrices and Visualizations

Confusion matrices were generated to further scrutinize model performance, allowing us to examine true positives, true negatives, false positives, and false negatives. Additionally, we visualized the confusion matrices for each model using `countplots` to gain insights into the distribution of predicted and true sentiment labels. Here is the best among them, Decision Tree:

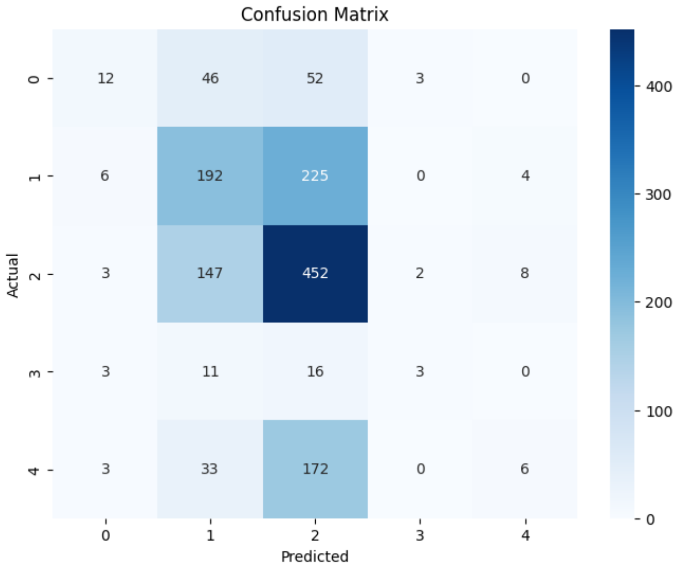

Fig. 2. Confusion Matrix of Decision Tree

D. Histograms of Accuracy Scores

Histograms were created to illustrate the distribution of accuracy scores for different models. This visualization aids in understanding the variability in model performance.

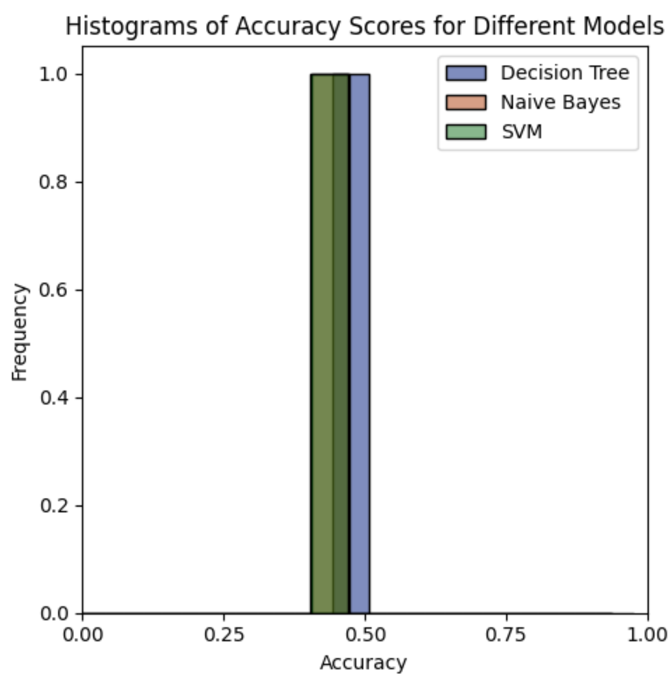

Fig. 3. Histogram comparing accuracy scores.

E. Discussion

The results analysis shows the importance of rigorous evaluation and model selection. The 'Decision Tree' demonstrates superior accuracy and holds great promise for real-world applications in humorous emotion detection on social media platforms.

Future Plans

Working on video, picture, and voice inputs to find out humor from them too if they are appropriate for social media or not is a possibility to work on our baseline model more. This will help in reviewing other media automatically too.

Humor is just the starting classifications like jokes, puns, and double-meaning can also be added in the future to correctly justify a social media post. This will allow a better understanding of any machine to understand the text.

Our proposed model still gives some false results as we are doing the task sentence-wise so, there is more scope for improving it and adding a whole social media post and the user's previous posts to get an idea of their behavior pattern to correctly justify their current post to humorous or not.

Conclusion

In natural language processing, the ability to recognize the humor in the text is becoming more and more crucial. It is critical to distinguish between amusing content and hate speech when using social media. Currently used AI models frequently mistake non-hate speech for hate speech, which results in user bans. A proposed methodology to address this accurately classifies social media posts as hilarious by taking into account their language and contextual characteristics. By expanding the baseline models to assess voice, picture, and video inputs, more advancements can be made. Future revisions might include categorizing additional kinds of humor and taking into account a user's earlier posts. Although the suggested methodology has promise, there is still an opportunity for improvement and growth in the area of text emotion recognition.

References

[1] C. Sharma, W. Scott Paka, D. Bhageria, A. Das, S. Poria, T. Chakraborty, B. Gambäck. Task Report: Memotion Analysis 1.0 @SemEval 2020: The Visuo-Lingual Metaphor! In *14th International Workshop on Semantic Evaluation (SemEval-2020)*, Barcelona, Spain, Sep 2020. Association for Computational Linguistics.

[2] C. Reid Turner, A. Fuggetta, L. Lavazza, A. L. Wolf. A conceptual basis for feature engineering, Journal of Systems and Software 49.1 (1999): 3-15.

[3] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[4] Y. Shihong, L. Ping, H. Peiyi. SVM CLASSIFICATION :ITS CONTENTS AND CHALLENGES Applied Mathematics-A Journal of Chinese Universities 18 (2003): 332-342.