1. Abstract:
The paper presents a comprehensive approach for emotion classification and intensity prediction in text data. Leveraging datasets containing emotions (anger, fear, joy, and sadness), we employ CountVectorizer for feature extraction in emotion classification and simple text length for intensity prediction. Our models, Multinomial Naive Bayes for emotion and Linear Regression for intensity, are trained and evaluated on the respective datasets. The results demonstrate the effectiveness of the proposed approach in capturing and predicting emotions and intensity in textual content.

2. Introduction:
Emotion analysis in text data is a critical aspect of natural language processing, with applications ranging from sentiment analysis to mental health monitoring. This paper focuses on the classification of emotions (anger, fear, joy, and sadness) and the prediction of intensity in textual content. The significance lies in understanding the emotional context of messages and the varying degrees of intensity associated with different emotions.

3. Related Work:
A review of existing literature reveals various approaches to emotion analysis, including machine learning techniques and feature extraction methods. Prior studies often employ sentiment analysis techniques and diverse datasets to address similar problems. Notably, the use of CountVectorizer for feature extraction in emotion classification and simple text length for intensity prediction is explored in this paper.

4. Data Collection:
Datasets encompassing emotions (anger, fear, joy, sadness) are collected from diverse sources. Each dataset is processed to include relevant columns such as ’id,’ ’text,’ ’Emotions,’ and ’intensity.’ The datasets are then merged to create a comprehensive dataset for training and testing.

5. Data Preprocessing:
Raw data undergoes cleaning and formatting to ensure consistency and eliminate noise. Text data is preprocessed to remove irrelevant characters, tokenize sentences, and handle stopwords. The resulting cleaned data is ready for further processing.

6. Feature Extraction:
For emotion classification, the CountVectorizer is employed to convert textual data into a numerical format. Each document is represented by a vector of term frequencies. For intensity prediction, a simple heuristic is applied, using the length of the text as a feature.

7. Model Training:
The Multinomial Naive Bayes classifier is trained for emotion classification using the transformed features, while Linear Regression is trained for intensity prediction using text length as a feature. The dataset is split into training and testing sets to assess model performance.

8. Evaluation:
The evaluation of the emotion classification model includes accuracy, precision, recall, and F1 score metrics. For intensity prediction, Mean Squared Error is utilized. These metrics provide a comprehensive analysis of the models’ performance.

9. Results:
Results indicate the successful classification of emotions and the accurate prediction of intensity levels in text data. The evaluation metrics demonstrate the models’ efficacy in capturing the nuances of emotional content.

10. Discussion:
The discussion interprets the results in the context of the initial problem statement. Comparisons are drawn between emotion classification and intensity prediction. Potential challenges and areas for improvement are highlighted, encouraging further exploration.

11. Conclusion:

In conclusion, the proposed approach demonstrates promising results in emotion classification and intensity prediction. The models exhibit effectiveness in capturing emotional nuances, paving the way for applications in sentiment analysis and mental health monitoring. The paper concludes with suggestions for future research directions.

12. References:

Competition link: https://competitions.codalab.org/competitions/16380

Additional info: http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html

13. Code Availability (Optional):

https://github.com/CHETHANBRAAJ/Emotion_ML