PROJECT REPORT

ON

# EMOTION INTENSITY

**by**

**K Sravan Kumar**

**Abstract:**

This paper presents an analysis of emotion intensity in textual data using machine learning techniques. The dataset used for this analysis consists of tweets labeled with emotions and corresponding intensity scores. The objective of this study is to predict the intensity of emotions expressed in text using features extracted from the textual data. We employ various machine learning models and evaluate their performance using metrics such as mean squared error, mean absolute error, and R-squared. Additionally, we discuss strategies to improve accuracy and reduce overfitting in emotion intensity prediction tasks.

**Introduction**:

The analysis of emotions in textual data has gained significant attention due to its applications in sentiment analysis, opinion mining, and understanding user behavior on social media platforms. Emotion intensity, which quantifies the strength or degree of emotions expressed in text, plays a crucial role in understanding the affective content of messages. In this paper, we focus on predicting emotion intensity in tweets, which are short textual messages often expressing various emotions such as anger, joy, sadness, and others.

**Approach:**

**Data Preprocessing:** The dataset consists of tweets labeled with emotions and intensity scores. We preprocess the textual data by tokenizing, removing stop words, and converting text into numerical features using techniques such as TF-IDF vectorization. Additionally, we perform sentiment analysis using the VADER sentiment analyzer to extract sentiment scores as features.

**Feature Engineering:** We combine the text features obtained from TF-IDF vectorization with sentiment scores to create a feature matrix for modeling. This combined feature matrix serves as input to machine learning models for predicting emotion intensity.

**Model Selection and Hyperparameter Tuning:** We experiment with various machine learning models such as linear regression, random forest regression, and support vector regression (SVR). We use grid search with cross-validation to tune hyperparameters and select the best-performing model based on mean squared error.

**Model Evaluation:** We evaluate the selected model on a holdout test set using metrics such as mean squared error, mean absolute error, R-squared, and explained variance score. These metrics provide insights into the accuracy and goodness of fit of the model.

**Analysis and Interpretation:** We analyze the performance of the model, identify factors contributing to prediction errors, and discuss strategies to improve accuracy and reduce overfitting.

**Findings**:

Our experimental results show that the 85% accuracy in Deep learning model and achieves 0.03424466773867607 the lowest mean squared error and R-squared compared to other models. However, we observe some difficult in predicting Intensity. To improve accuracy and reduce overfitting, we recommend Feature Engineering,Regularization,Cross-Validation,HyperparameterTuning,EnsembleMethods,Featureselection,Data Augmentation, Model Selection such as feature selection.

**Conclusion**:

In conclusion, this paper demonstrates the application of machine learning techniques and Deep Learning to predict emotion intensity in textual data. By combining text features with sentiment scores and using appropriate machine learning models, we can effectively model and predict emotion intensity in tweets. Further research can explore advanced techniques for feature engineering and modeling to improve the accuracy of emotion intensity prediction tasks.

* **Codalab Username:** SravanKumarReddy