PROJECT TITLE

EMOTION CLASSIFICATION OF TWEETS USING NATURAL LANGUAGE PROCESSING

STATEMENT OF PROJECT OBJECTIVES

◦ Develop Natural Language Processing Models that classifies the Text sentences based on emotion it displays.

◦ The NLP models will be able to predict the way a sentence, a tweet in this case, has been expressed.

◦ Models will be built using Vanilla Recurrent Neural Network, Long-Short Term Memory, Gated Recurrent Unit, Bidirectional RNN.

◦ Another set of NLP models will be built by using the same dataset this time by feeding the input sentences in reverse order. These models will be evaluated by feeding the input in the actual order. This is done to reduce the effect of Vanishing Gradient.

◦ Another Important Objective of this project is the check if building model by feeding input in reverse order improves the performance of the model or not.

WHAT'S NEW IN THIS PROJECT ?

Building NLP Models by Feeding Input Sentences in Reverse Order.

◦ After developing NLP Models by using the dataset with sentences in the actual order, new NLP models are built by using the same dataset, this time with the sentences in reverse order.

  ◦ Actual Sentence: [My, Father, Felt, So, Happy, For, My, Success]

  ◦ Reverse order: [Success, My, For, Happy, So, Felt, Father, My]

◦ These new models, trained with the 'Reverse' dataset, are evaluated and performance is compared with the previously built models.

◦ 'Reverse Order' model building is performed to check if the primary task of this project can be performed in a better way or not.

STATEMENT OF VALUE

◦ Building NLP models like these presents a significant advancement in understanding and leveraging human sentiments in various fields.

◦ These models also helps in understanding the reaction, opinion or feeling of the users or society regarding a specific incident or topic or product.

◦ These models Enables the identification of distress signals in social media posts, potentially providing early support for individuals in need. Also helps in identifying threats and criminal intentions in the posts which can be tracked and stopped.

◦ These models not only opens avenues for enhanced communication and understanding between businesses and consumers but also plays a critical role in societal well-being and mental health monitoring.

REVIEW OF STATE OF THE ART AND RELEVANT WORKS

◦ <u>For NLP Model Building and Evaluation:</u>

S. M. . et. al., "Detecting Emotion from Natural Language Text Using Hybrid and NLP Pre-trained Models", TURCOMAT, vol. 12, no. 10, pp. 4095–4103, Apr. 2021.

L. Mathew and V. R. Bindu, "A Review of Natural Language Processing Techniques for Sentiment Analysis using Pre-trained Models," 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2020, pp. 340-345, doi: 10.1109/ICCMC48092.2020.ICCMC-00064.

APPROACH

- Dataset: 5 V's of Data; Volume, Variety, Velocity, Veracity, Value are checked to ensure right set of data is collected for the problem statement.

- https://www.kaggle.com/datasets/pashupatigupta/emotion-detection-from-text

- The dataset has 40,000 samples and 13 emotions with tweet_id, sentiment, content as features. It is a labeled and structured dataset.

- Natural Language Processing models: Vanilla Recurrent Neural Network, Long Short Term Memory RNN, Gated Recurrent Unit RNN, Bidirectional RNN.

- Tools: Pandas, Seaborn, Tensorflow, Keras, Numpy

- Techniques: Label Encoding, One hot encoding, Vectorization, feature extraction, tf-idf vectorization.

DELIVERABLES

The deliverables for this project include:

◦ A fully trained Natural Language Processing Model capable of Detecting emotions from the given input text.

◦ A comprehensive report detailing the development process, methodologies used and performance metrics.

◦ A Dataset compilation used for training and testing.

◦ For future access and replicability, all coding scripts will be provided along with documentation.

EVALUATION METHODOLOGY

The NLP Model's performance can be evaluated by measuring the below metrics:

- ◦ <u>Accuracy:</u> The percentage of total correct predictions out of all predictions made.
- ◦ <u>Precision:</u> The ratio of true positive predictions to the total number of positive predictions.
- ◦ <u>Recall:</u> The ratio of true positive predictions to the total number of actual positives.
- ◦ <u>F1 Score:</u> The harmonic mean of precision and recall for each emotion category.
- ◦ <u>Confusion Matrix:</u> A table that allows visualization of the performance of the algorithm, including true positives, false positives, true negatives, and false negatives.
- ◦ <u>Classification Report:</u> The Classification report provides a detailed overview of the model's performance on a per-class basis.

THANK YOU!