NLP Assessment

Assessment Description:

This assessment is about tweet categorization using a machine learning model. You should use different algorithms to train the model and evaluate the performance of the model. Based on the performance, suggest the best model for the dataset.

Dataset Description:
This is an entity-level sentiment analysis dataset of twitter. Given a message and an entity, the task is to judge the sentiment of the message about the entity. There are three classes in this dataset: Positive, Negative and Neutral. We regard messages that are not relevant to the entity (i.e. Irrelevant) as Neutral.

Usage
Please use **twitter_training.csv** as the training set and **twitter_validation.csv** as the validation set.

Step-by-step guide:
1) Download the dataset and upload the dataset in the colab.
2) Clean the dataset with different data pre-processing.
3) Tokenize the dataset tweet with appropriate nltk function.
4) Use tf-idf to vectorize the tokens of each tweet.
5) Apply the following machine learning model to train the model with the **twitter_training.csv.**
    i) Support Vector Machine
    ii) Decision Tree
    iii) Random Forest
    iv) Multinomial Naive Bayes
6) Do hyperparameter tuning to find the best hyperparameter to tune the model.
7) Use twitter_validation.csv to test the model.
8) Use appropriate evaluation methods to evaluate the model accuracy and other performance.
9) Visualize the evaluation with appropriate visualization functions.
10) Suggest the BEST algorithm.