Major Project

- **Project Name:**

Machine Learning December Major Project

- **Project Description:**

Problem statement: Create a classification model to predict the sentiment either (1 or 0) based on Amazon Alexa reviews

Context: This dataset consists of a nearly 3000 Amazon customer reviews (input text), star ratings, date of review, variant and feedback of various amazon Alexa products like Alexa Echo, Echo dots, Alexa Firesticks etc. for learning how to train Machine for sentiment analysis.

Dataset:

https://drive.google.com/file/d/1HF5ycbFEFO6EwiCcMqCzDV3QbwBl8rVo/view?usp=share_link

Details of features:

The columns are described as follows:

1) rating: Product rating

2) Date: date on which the product was rated

3) variation: variation of the product

4) verfified_reviews: the verified reviews for the alexa

5) feedback: 1(Positive) or 0 (Negative)

Steps to consider:

1) Read the dataset

2) Remove handle null values (if any).

3) Preprocess the Amazon Alexa reviews based on the following parameter:

a) Tokenizing words

b) Convert words to lower case

c) Removing Punctuations

d) Removing Stop words

e) Stemming or lemmatizing the words

4) Transform the words into vectors using

a) Count Vectorizer OR

b) TF-IDF Vectorizer

5) Split data into training and test data.

6) Apply the following models on the training dataset and generate the predicted value for the test dataset

a) Multinomial Naïve Bayes Classification

b) Logistic Regression

c) KNN Classification

7) Predict the feedback for test data

8) Compute Confusion matrix and classification report for each of these models

9) Report the model with the best accuracy