ML PROJECT PROPOSAL

PROJECT: NETFLIX STOCK PROCE PREDICTION USING RANDOM FOREST

# PROBLEM STATEMENT

Netflix stock price prediction dataset is to predict the future stock price of Netflix based on historical data. The dataset contains daily stock prices of Netflix from 2002 to 2021, along with relevant financial indicators such as volume, market capitalization, and dividend yield.

The purpose here is to make a prediction of a continuous numerical value (the stock price) using a different set of input characteristics, making this a regression issue. For this purpose, the well-known machine learning method known as "Random Forest" might be quite useful.

The machine learning method Random Forest may be used for both regression and classification purposes. It is an ensemble learning technique that generates many decision trees during training and uses their average prediction to get a final prediction. By using Netflix's stock price history for training and making predictions for the future, Random Forest may be utilised to tackle this issue.

The result of the Random Forest method is the combination of decision trees built on independently-generated subsets of data. Since it can account for non-linear correlations between the characteristics and the objective variable, Random Forest is an effective technique for forecasting stock prices.

# METHODS USED FOR THE PROJECT

Random forest is a popular machine learning algorithm that can be used for regression tasks like Netflix stock price prediction. Here are the steps on how to use random forest in this project:

The first step is data pre-processing, which may entail, among other things, eliminating duplicates, reducing outliers, and scaling the data. Utilizing techniques such as correlation analysis, mutual information, and principal component analysis, identify the traits that are most strongly associated with the dependent variable.

Thirdly, divide the data into test and training sets. Multiple decision trees are built using distinct random samples of the training data in random forest, and their combined predictions are utilised to create a single prediction. After selecting the ideal model, the following step is to

optimise its hyperparameters for the best outcomes. Grid search and randomised search may be used to optimise the random forest method's hyperparameters, such as the number of trees, for best model performance. Next step involves fitting the random forest model to the training data. Put the model to the test by evaluating its performance based on the test data. For this aim, metrics such as R-squared, mean squared error, and mean absolute error may be used.

# EVALUATION CRITERIA

The evaluation criteria for predicting stock prices using Random Forest can include the following metrics:

* The RMSE is calculated as the square root of the mean of the squared discrepancies between actual and anticipated stock values. It is a statistical measure of the average deviation of forecasts from reality. Lower levels of root-mean-squared error (RMSE) indicate enhanced performance.
* R2 indicates the proportion of the dependent variable's variability that can be explained by the model. The optimal score would be 1, whereas a score of 0 would indicate no progress. If R2 is high, performance is outstanding.
* MAE is the average of the absolute differences between actual and predicted stock prices. This statistic measures the average deviation of expectations from reality. The closer the MAE value is to zero, the greater the performance.
* Mean Absolute Percentage Error is the average of the absolute percentage mistakes between actual and anticipated stock prices (MAPE). It measures the deviation between the average prediction and the actual outcomes. In general, a lower MAPE value reflects more performance.

These metrics may be used to determine the accuracy with which the Random Forest model anticipated Netflix stock prices. For a comprehensive understanding of the model's effectiveness, it is essential to use a number of evaluation methods.

# INFERENCE AND TAKEAWAYS

The model can be used to predict the future stock prices of Netflix, which can help investors make informed decisions about buying and selling the stock. The model can be used to identify trends and patterns in the stock prices of Netflix, which can be useful for investors and traders in making informed decisions about buying and selling the stock.

Using Random Forest to predict Netflix stock prices is a very successful strategy. Thus, it is able to identify complex data patterns and handle nonlinear relationships between attributes and the end measure of interest. The beginning price, closing price, highest and lowest prices, and volume are simply a few of the variables in the dataset that significantly impact the price of Netflix stock. The Random Forest algorithm is an effective way for predicting Netflix stock values, and it is feasible to get trustworthy findings by analysing historical patterns.