**BIG DATA ANALYTICS WITH IBM CLOUD DATABASES**

**INTRODUCTION:**

Big data is a new and emerging buss word in today’s times. Stock market is an up and ever evolving,volatile, uncertain and intriguingly potential niche, which is an important extension in finance and business growth and prediction. Stock market has to deal with a large amount of vast and distinct data to function and draw meaningful conclusions. Stock market trends depend broadly on two analyses; technical and fundamental.Technical analysis is carried out using historical trends and market values. On the other hand, fundamental analysis is done based on the sentiments, values and social media data and responses. Since large, complex and complicated and exponentially growing data is involved, we use big data analysis to help assist in the prediction and drawing accurate business decisions and profitable investments

**PROBLEM DEFINITION:**

Develop a machine learning model to predict the future closing price of a specific stock based on historical price data and relevant financial indicators using the Random Forest algorithm.Stock price prediction is a challenging task with significant financial implications. Investors and traders often seek accurate forecasts to make informed decisions. Historical stock price data will be used as the primary dataset.The Random Forest algorithm, known for its robustness and ability to handle complex data, can be a valuable tool in this context.

**Module1:** Extracting the suitable Stock Dataset for the prediction and checking for missing values.

Data collection occurs in the stock market. Two sets of data are used for this purpose: the data from the earnings calendar and the daily stock market information Various websites can be used to gather daily stock market information.

**Module2 :** Creating or designing a Model for predicting the future price of the stocks.

Many ML algorithms are used for stock markets prediction. The models have two main categories to address this issue which are regression models that seek to predict the movements of stock price such as the closing price of shares, the second one is rating models that seek to support investors while making decisions related to retain, sell, or buy stocks.

**Module 3 :** Final predicted model is compared with the real values.

Based on the final graph that is plotted we can understand how accurate our model is and this modelcan be used for predicting the value of the future stock.

**COMPONENTS REQUIRED :**

**Integrated Development Environment (IDE):** You'll need an IDE for writing, testing, and running your Python code. Eg:Jupyter Notebook,PyCharm,Visual Studio Code (VSCode).

**Python Libraries:** Install the necessary Python libraries using a package manager like pip or conda. Eg:scikit-learn,numpy,matplotlib and seaborn,ta-lib.

**Cloud Services :** If you don't have access to a powerful local machine, you can consider using cloud-based services like Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure.

**Storage:** We need sufficient storage space to store your datasets, code, and any other related files. An SSD (Solid State Drive) is recommended for faster data access and model development.

**Data Source:** We need access to historical stock price data. This data can be obtained from financial data providers, APIs (e.g., Alpha Vantage, Yahoo Finance), or by web scraping financial websites.

**Machine Learning Model Management :** If you want to manage and deploy your machine learning models efficiently, you can use platforms like TensorFlow Serving, MLflow, or Kubeflow.

Project aims to predict the best model for predicting the future value of stock with the help of historical data using a random forest algorithm.