**PHASE-3**

**STOCK PRICE PREDICTION**

**DATA PREPROCESSING:**

Data preprocessing in stock price prediction is a critical step because financial data can be noisy and complex.The quality of your data and how well you preprocess it can significantly impact the results of your analysis or machine learning models. Well-preprocessed data often leads to more stable and interpretable models. This is important when you need to understand the reasons behind a model's predictions in the context of stock price movements. It significantly impacts the accuracy, stability, and interpretability of the predictive models, making it an essential aspect of building successful stock price prediction systems.

Dataset:

<https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset>

**Importing the libraries:**

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

**Importing the dataset:**

dataset=pd.read\_csv('MSFT.csv')

data=dataset.copy()

data.isnull().sum()

data=pd.read\_csv('MSFT.csv',na\_values=['?'])

data.isnull().sum()

X=data.iloc[:,~dataset.columns.isin(['Date'])].values

y=data.iloc[:,dataset.columns.isin(['Date'])].values

**Taking Care of missing data:**

from sklearn.impute import SimpleImputer

imputer=SimpleImputer(missing\_values=np.nan,strategy='mean')

imputer.fit(X)

X=imputer.transform(X)

print(X)

**Encoding the Categorical Data:**

from sklearn.compose import ColumnTransformer

from sklearn.preprocessing import OneHotEncoder

ct = ColumnTransformer(transformers=[('encoder', OneHotEncoder(),

[‘Date’])], remainder='passthrough')

y = np.array(ct.fit\_transform(X))

print(y)