Stock Price Prediction

Stock price analysis has been a critical area of research and is one of the top applications of machine learning. This tutorial will teach you how to perform stock price prediction using machine learning and deep learning techniques. Here, you will use an LSTM network to train your model with Google stocks data.

DATASET :

Dataset Link: https://www.kaggle.com/datasets/prasoonkottarathil/microsoft-lifetime-stocks-dataset

1986-03-13 0.088542 0.101563 0.088542 0.097222 0.062549 1031788800

1986-03-14 0.097222 0.102431 0.097222 0.100694 0.064783 308160000

1986-03-17 0.100694 0.103299 0.100694 0.102431 0.065899 133171200

1986-03-18 0.102431 0.103299 0.098958 0.099826 0.064224 67766400

1986-03-19 0.099826 0.100694 0.097222 0.09809 0.063107 47894400

1986-03-20 0.09809 0.09809 0.094618 0.095486 0.061432 58435200

1986-03-21 0.095486 0.097222 0.091146 0.092882 0.059756 59990400

1986-03-24 0.092882 0.092882 0.08941 0.090278 0.058081 65289600

1986-03-25 0.090278 0.092014 0.08941 0.092014 0.059198 32083200

1986-03-26 0.092014 0.095486 0.091146 0.094618 0.060873 22752000

1986-03-27 0.094618 0.096354 0.094618 0.096354 0.06199 16848000

1986-03-31 0.096354 0.096354 0.09375 0.095486 0.061432 12873600

1986-04-01 0.095486 0.095486 0.094618 0.094618 0.060873 11088000

1986-04-02 0.094618 0.097222 0.094618 0.095486 0.061432 27014400

1986-04-03 0.096354 0.098958 0.096354 0.096354 0.06199 23040000

1986-04-04 0.096354 0.097222 0.096354 0.096354 0.06199 26582400

1986-04-07 0.096354 0.097222 0.092882 0.094618 0.060873 16560000

1986-04-08 0.094618 0.097222 0.094618 0.095486 0.061432 10252800

1986-04-09 0.095486 0.09809 0.095486 0.097222 0.062549 12153600

COLUMNS USED :

Date

Open

High

Low

Close

Adj Close

Volume

LIBRARIES USED :

The Python 3 environment comes with many helpful analytics libraries installed and several

helpful packages to load.

The essential libraries used in this project are :

• Importing OS (for kaggle inputs)

• Numpy and Pandas libraries

• Matplotlib

TRAIN AND TEST

Training the dataset by describe(), isnull().sum(), drop(), show(), and by using k-means

algorithm we train the data

Testing the data by importing sklearn.cluster from k-means with ensuring the plot range and axis

labels producing the k value, scattering the data by kmeans.cluster_centers and producing 3D

plot.

REST OF THE EXPLANATIONS

Data Collection

The process involves gathering customer data, which includes information about their purchase

history, demographics, and interaction patterns.

Data Preprocessing

The task involves preparing and cleaning data, handling missing values, and converting

categorical features into numerical representations.

Feature Engineering

Data preparation and cleaning, handling missing values, and the transformation of

categorical features into numerical representations are all part of the task.
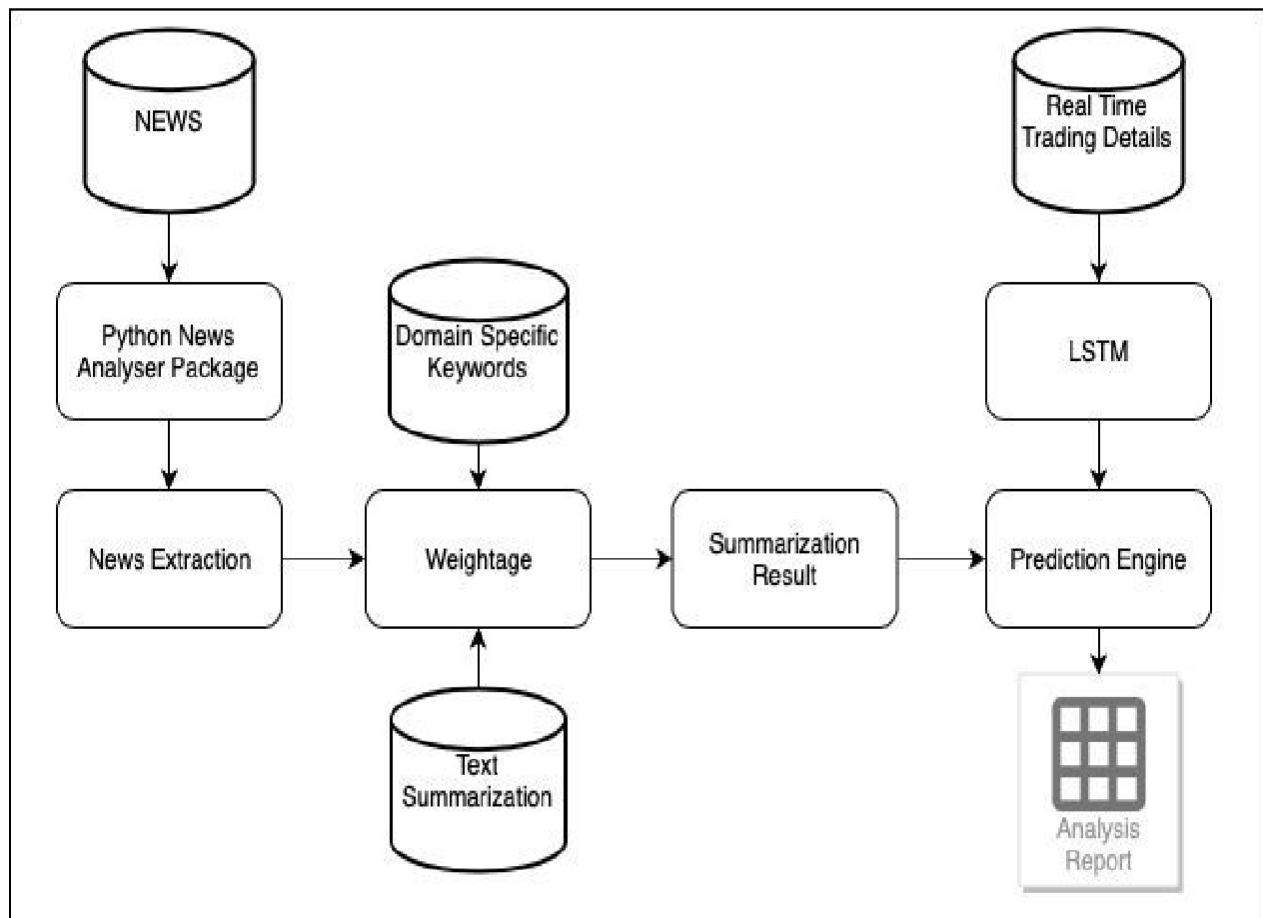
ALGORITHMS USED

Alorithms like

- regression

- classifier

- support vector machine (SVM).

- Long Short-Term Memory (LSTM), were used

Visualization: Visualize the customer segments using techniques like scatter plots, bar

charts, and heatmaps. Interpretation: Analyze and interpret the characteristics of each customer

segment to derive actionable insights for marketing strategies.

DESIGN AND DATAFLOW :

Prediction

Prediction Process

START

Historical stock price/ returns
Technical indicators
Fundamental variables
Financial news

Data collection

Data cleaning

Data pre-processing

Scaling features
Feature generation
Feature selection
Feature extraction

Optional

Data splitting

Training sample
Validation sample
Testing sample

Fitting model parameters
Fitting model structure
Model validation

Model training

Model selection

Optimized structure
Optimized parameters

Prediction of future values