BIG DATA AND MACHINE LEARNING APPLICATIONS

STOCK PRICE PREDICTION

2205708

ABSTRACT

- Stocks represent ownership in a company
- Investors purchase shares of stock to become partial owners of a company
- Stocks are bought and sold on stock exchanges such as NYSE or NASDAQ
- Stock prices fluctuate based on supply and demand, company performance, and market conditions

PROBLEM STATEMENT

- Growing demand for stock market analysis and prediction
- Need for accurate and reliable stock price prediction
- Potential benefits of predicting stock prices for investors
- Desire for a system capable of predicting stock prices for any company worldwide

DATASET

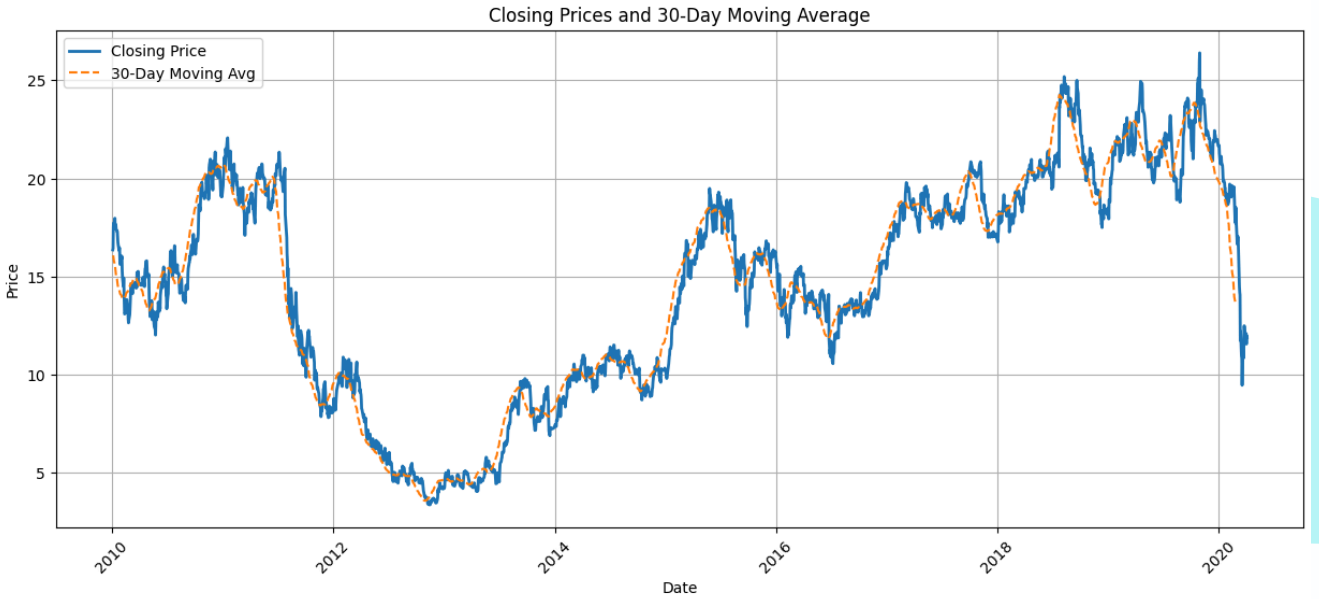
- Dataset sourced from the New York Stock Exchange (NYSE) obtained from kaggle.
- Contains historical prices and fundamental data of 500 companies (AMEX, NASDAQ, and NYSE markets)
- The dataset comprises 97,648 rows and 8 columns.
- Spans a period of 10 years from January 2010 to April 2020


```
df.head()
```

	Unnamed: 0	Name	Date	Open	Closing_Price	Daily_High	Daily_Low	Volume
0	0	Accor	2020-04-03	22.99	23.40	23.40	22.99	67
1	1	Accor	2020-04-02	23.91	22.99	23.91	22.99	250
2	2	Accor	2020-04-01	24.10	23.83	24.10	23.83	37
3	3	Accor	2020-03-31	25.04	25.00	25.24	24.99	336
4	4	Accor	2020-03-30	26.50	25.02	26.50	24.99	415

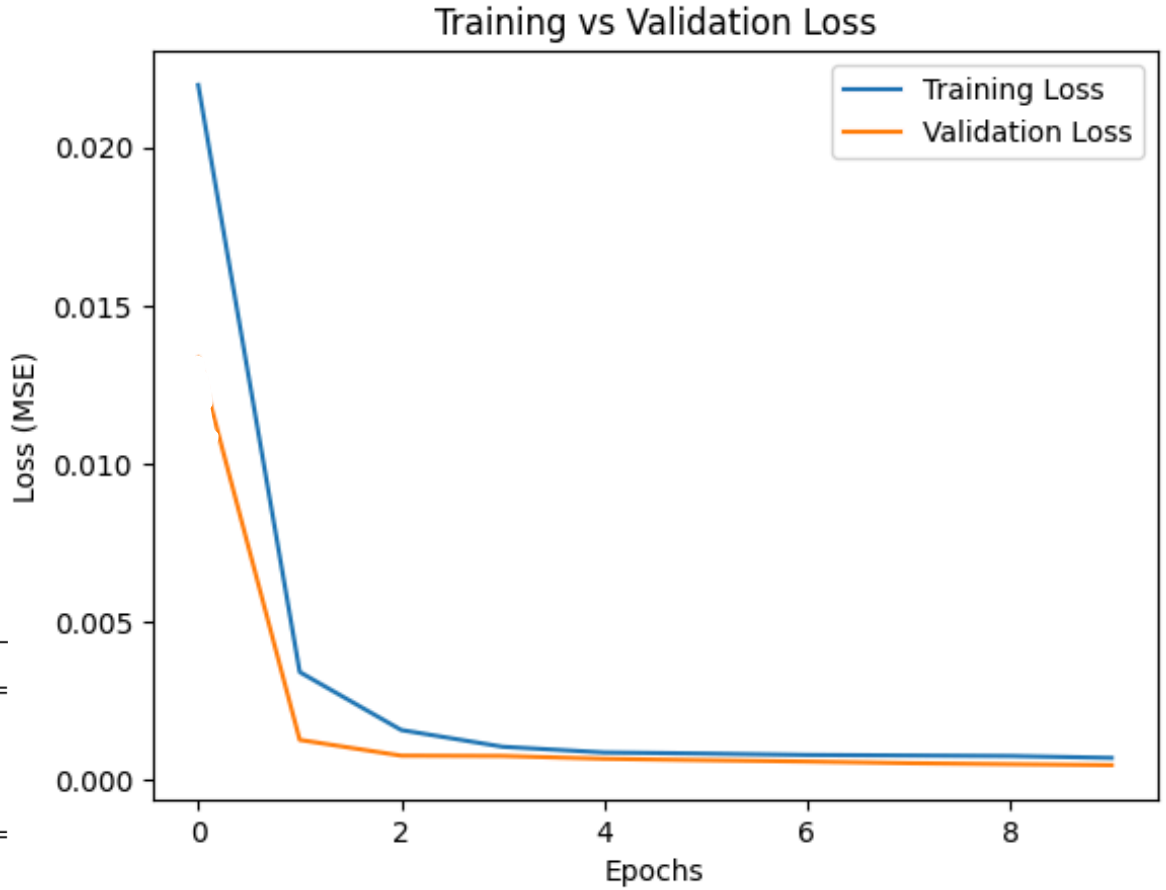
DATASET

- Dropping unnecessary columns such as "Unnamed: 0" from the dataset
- Data cleaning : No missing values
- Using the function to obtain data for a specific company “Peugeot” between 2010 and 2020
- Utilised MinMaxScaler to normalise the stock data, scaled between -1 to 1
- Defined the sequence length (number of past time steps) as 70 days
- “Closing Price” is used as target variable
- Data split into training, testing, and validation sets:
  Training data: 70%
  Testing data: 20%
  Validation data: 10%



SIMPLE RNN

- **Number of Epochs:** 10
- **Number of Neurons :** 50
- **Train MSE :** 243.54
- **Train RMSE :** 15.60
- **Model Architecture:**
  Simple RNN with 1 layer
  Dense layer with 1 output unit

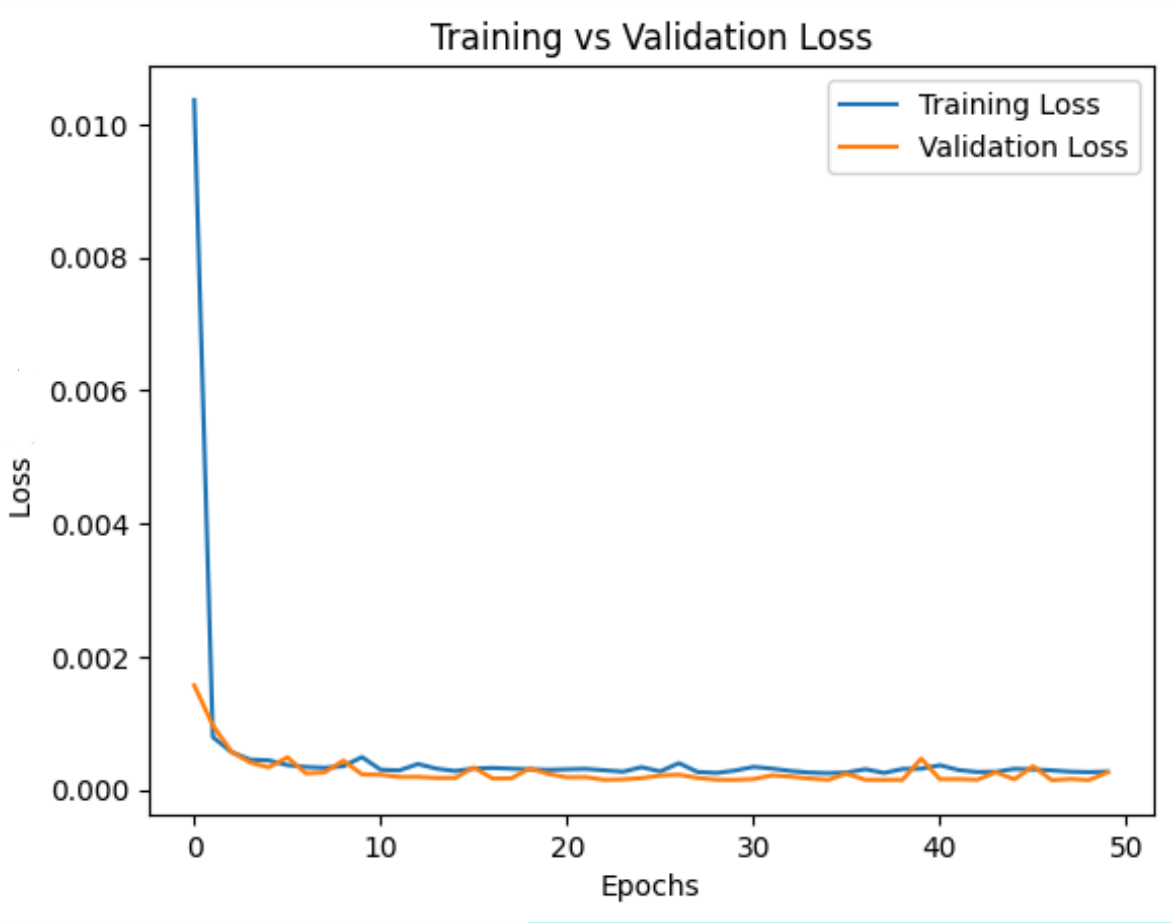


Layer (type)	Output Shape	Param #
simple_rnn_1 (SimpleRNN)	(None, 50)	2600
dense_1 (Dense)	(None, 1)	51

SIMPLE STACKED RNN

- **Number of Epochs:** 50
- **Number of Neurons :** 50
- **Train MSE :** 244.71
- **Train RMSE :** 15.64
- **Model Architecture:**
  Simple RNN with 2 layers
  Dense layer with 1 output unit

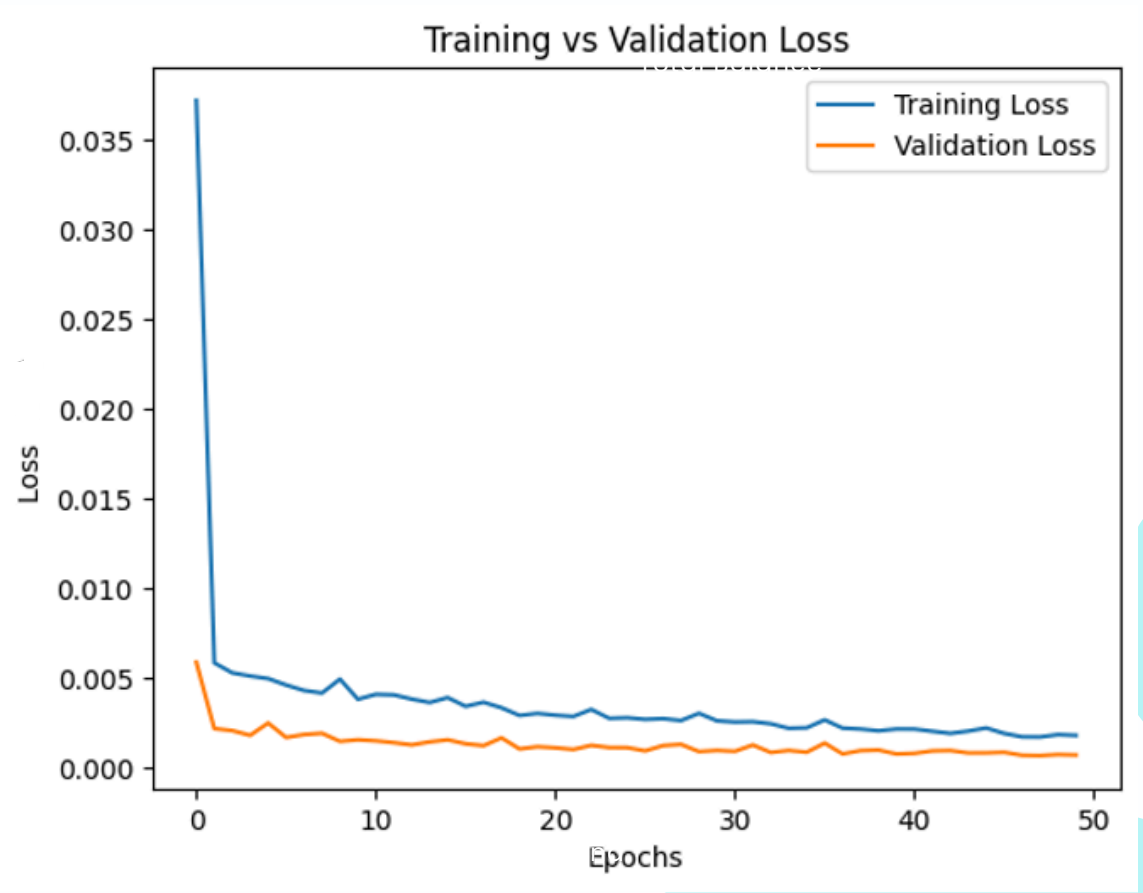
Layer (type)	Output Shape	Param #
simple_rnn_42 (SimpleRNN)	(None, 70, 50)	2600
simple_rnn_43 (SimpleRNN)	(None, 50)	5050
dense_21 (Dense)	(None, 1)	51



LSTM RNN

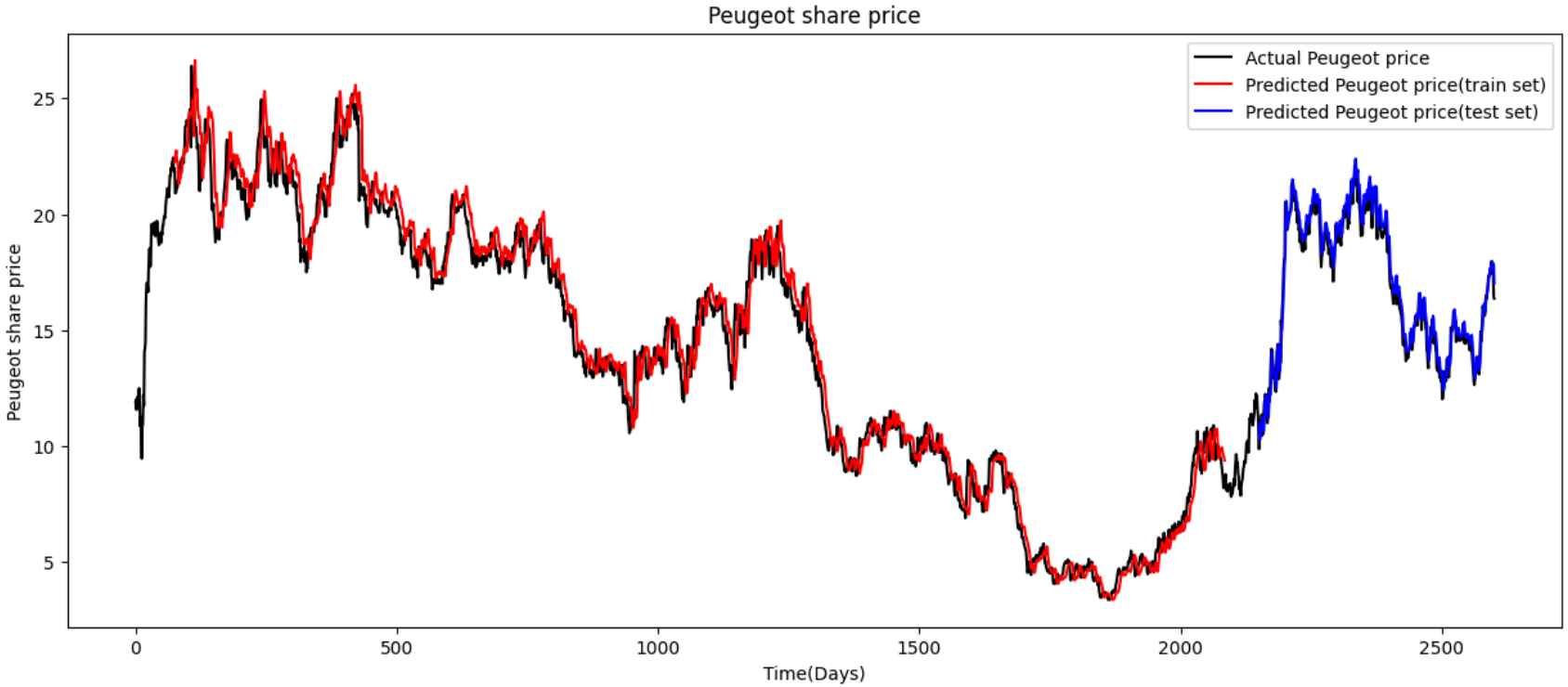
- **Number of Epochs:** 100
- **Number of Neurons :** 50
- **Train MSE :** 229.31
- **Train RMSE :** 15.14
- **Model Architecture:**
  LSTM Layers : 3
  Dropout Layers : 3
  Dense Layer : 1

Layer (type)	Output Shape	Param #
lstm_28 (LSTM)	(None, 70, 50)	10400
dropout_26 (Dropout)	(None, 70, 50)	0
lstm_29 (LSTM)	(None, 70, 50)	20200
dropout_27 (Dropout)	(None, 70, 50)	0
lstm_30 (LSTM)	(None, 50)	20200
dropout_28 (Dropout)	(None, 50)	0
dense_9 (Dense)	(None, 1)	51



LSTM RNN

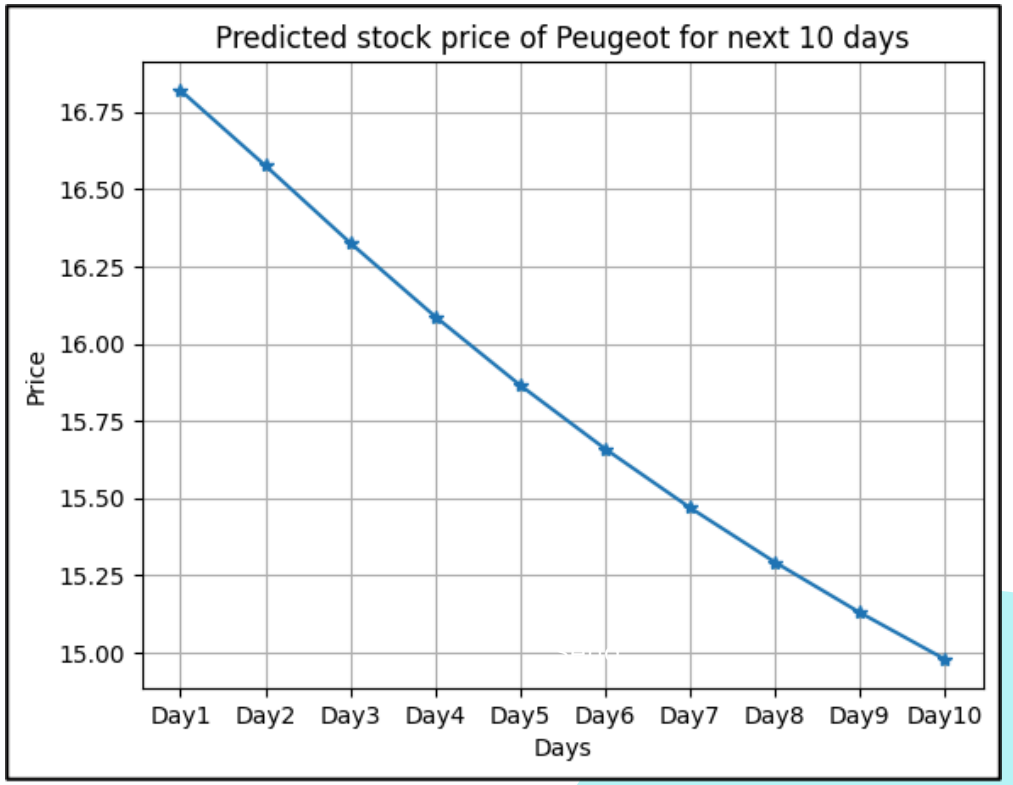
Prediction of Train/Test price



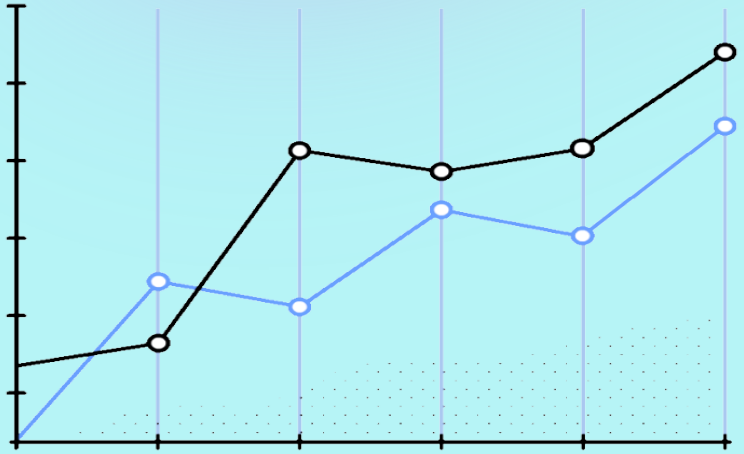
LSTM RNN

Prediction of stock price for next 10 days

- Utilized the last “sequence length” days of data to generate predictions for the next 10 days.



THANK YOU



Reference : (New York Stock Exchange)
https://www.kaggle.com/datasets/dgawlik/nyse/data