Tadawul Stock Exchange

Stock Price Prediction with Deep Learning

Prepared By:

Amjad	**Aisha**	**Ibtisiam**
Manar	**Nawal**	**Suliman**

CONTENTS

INTRODUCTION

Problem statement:

A growing stock market named Tadawul wants to attract more investors to the Saudi market to diversify and increase the market size of Saudi Arabia. A comprehensive market study will illustrate the market opportunity and potential for investments. The study will thoroughly examine the market history and include predictions for the company's future.

The goal of the case study:

You are required to model the Stock Price predictions using Deep Learning. The results thus obtained will be public to explore market future opportunities.

DATA DEVIEW

Dataset content:

- 600k Records
- 14 Columns
- Stocks from 2001 to 2020

DATA PREPROCESSING

Preprocessing steps:

1. Drop the missing values. The missing values were 14,388 records, which is 2.4% of the whole dataset.

2. Split the **date** column, split the date column into **Year**, **Month**, and **Day** columns for more efficiency in visualization.

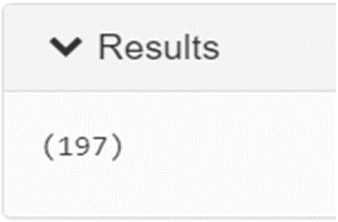
3. Delete the year **2001**. The data we have started from (**2001** to **2020**), and after investigating, there is only the date **31 Dec 2001** for the whole **year 2001**. Having one day for an entire year will not be valuable in our work, so it needs to be drooped.

DATA EXPLORATION

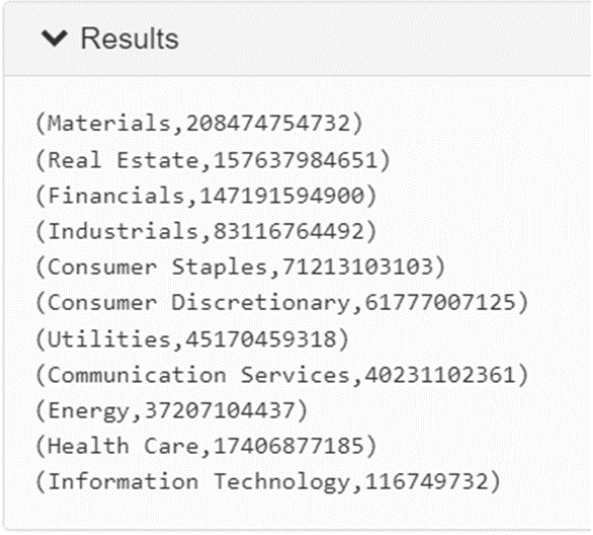
The data exploration was conducted using the **Pig Latin** script to handle the size and **Matplotlib** for visualizing the data.

Pig Latin:

- **How many trading names are in Tadawul?**



- **What is the total traded volume for each sector?**



DATA EXPLORATION

- **What is the total number of trades each year?**

Results

```
(2006,89918060)
(2007,62323043)
(2008,46226837)
(2005,44226918)
(2012,38443886)
(2009,33188081)
(2014,33089759)
(2015,28074743)
(2013,27114578)
(2019,25923806)
(2016,25287220)
(2011,23458580)
(2018,22863979)
(2017,20237598)
(2010,17827796)
(2004,12676179)
(2020,7489562)
(2003,3691935)
(2002,953780)
```

- **What are the top 10 companies with the highest volume trades?**

Results

```
(DAR ALARKAN,64002647604)
(SAUDI ELECTRICITY,41411385834)
(SAUDI KAYAN,37574860323)
(SABIC,25909558739)
(ARDCO,25284130681)
(EMAAR EC,23907597861)
(TASNEE,19558624607)
(SAPTCO,18308296195)
(BAHRI,17827286226)
(ZAIN KSA,17024847447)
```

DATA EXPLORATION

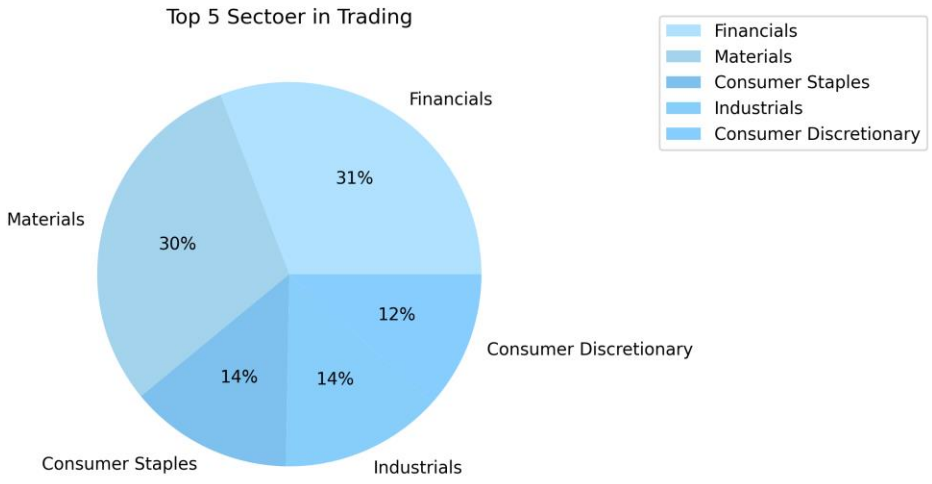
- **What are the top 10 companies with the highest value of trades?**

Results

```
(SABIC,2204591477984)
(SAUDI ELECTRICITY,748446834392)
(STC,719674093744)
(DAR ALARKAN,610659172468)
(BAHRI,556800991909)
(SAUDI KAYAN,540942918992)
(ANAAM HOLDING,441305821338)
(SAPTCO,424087176852)
(ARDCO,414404072058)
(TASNEE,407897599320)
```

Matplotlib:

- **Top 5 Sector in Trading**

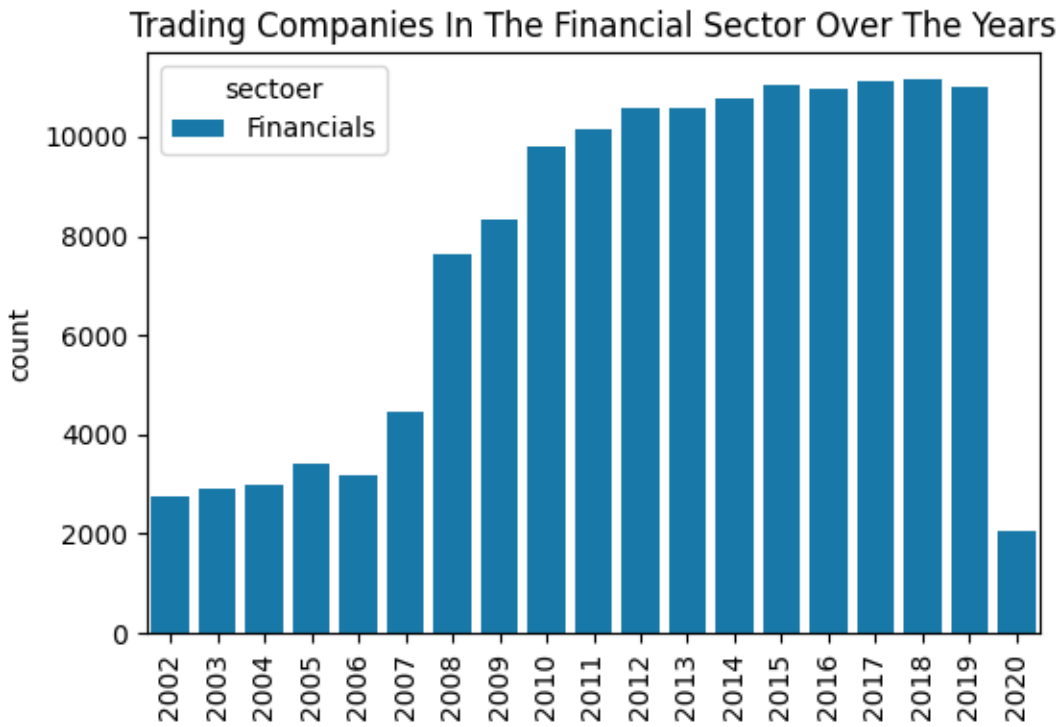


DATA EXPLORATION

The Financials sector is the largest in the Saudi exchange. Now let us explore the top three sectors.

1. Financials Sector

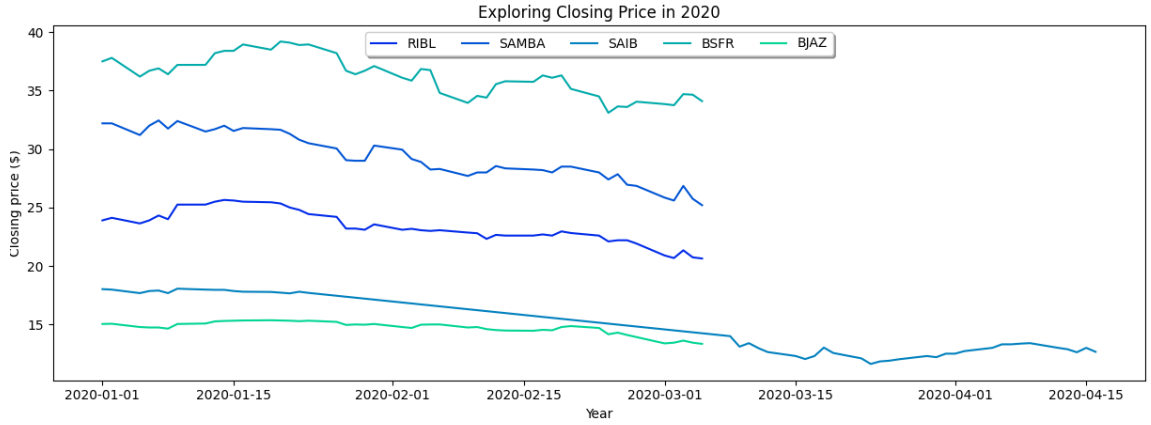
- **Trading companies in Financial Sector over the Years**



Here we observe the increase in companies whose shares are traded in the financial sector between 2002–2014 and the stability from 2014–2019. Then a significant drop happened in 2020 due to Covid-19.

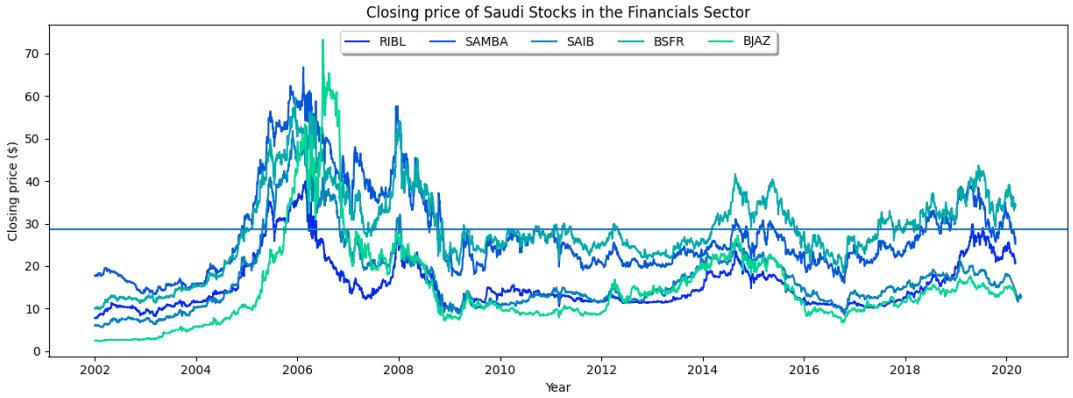
DATA EXPLORATION

- **Exploring year 2020 in Financials sector**



Here we observed the decrease in the closure index in the year 2020, precisely during March, which is the beginning of the Covid-19 virus.

- **The closing price of Saudi Stocks in the Financial Sector**

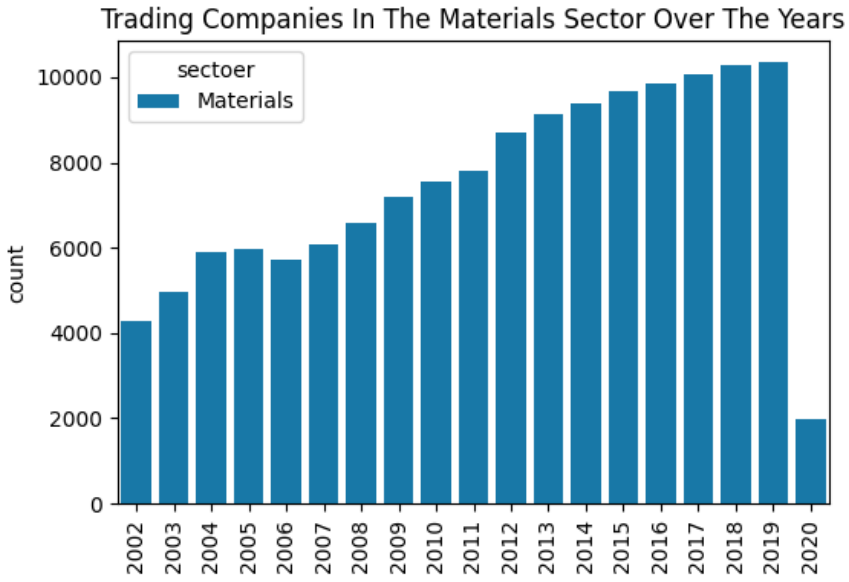


- The highest closing of Bank Aljazeera shares (BJAZ) was around 70 in 2006.
- The highest close of Banque Saudi Fransi (BSFR) shares was around 60 in 2006.
- The highest close of Saudi Investment Bank (SAIB) shares was around 50 in 2006.
- The highest close of Samba Financial Group (SAMBA) shares was around 75 in 2006.
- The highest close of Riyad Bank (RIBL) shares was around 40 in 2006.

All the companies have comparable prices above the average.

2. Materials Sector

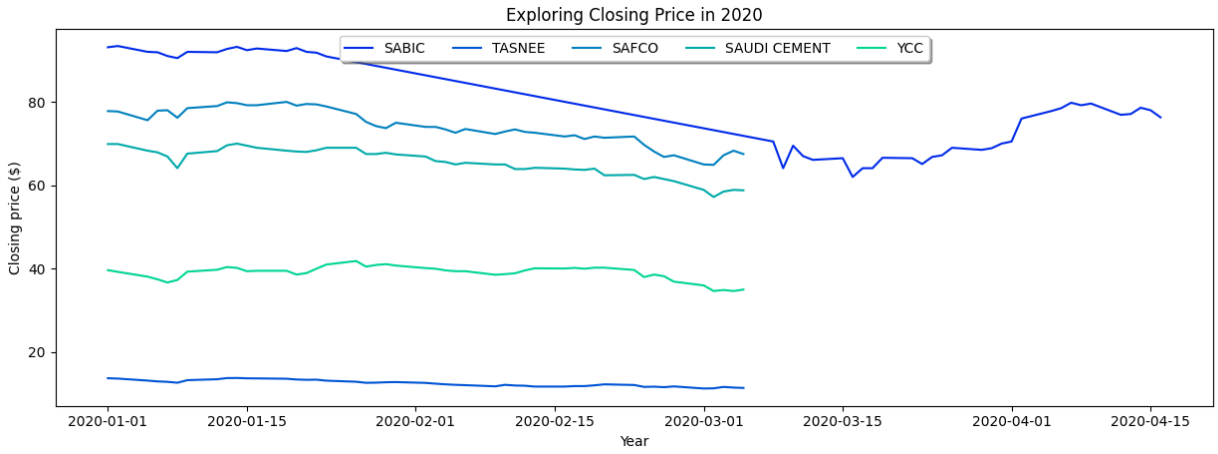
- **Trading companies in the Materials Sector over the Years**



DATA EXPLORATION

Here we observe the increase in companies whose shares are traded in the Materials sector between 2002–2014 and the stability from 2014–2019. Then a significant drop happened in 2020 due to Covid-19.

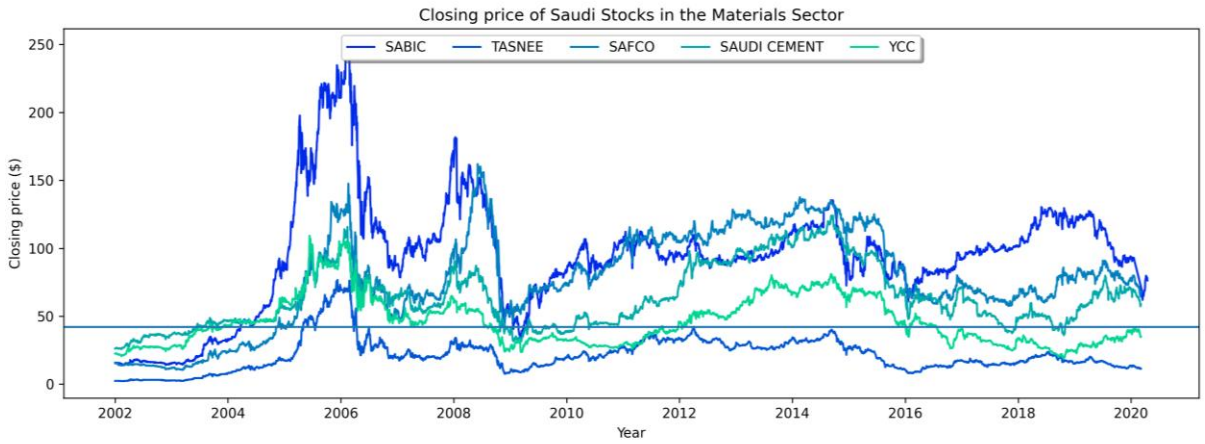
- **Exploring 2020 in Materials Sector**



Here we observed the decrease in the closure index in the year 2020, precisely during March, which is the beginning of the Covid-19 virus.

DATA EXPLORATION

- **The closing price of Saudi Stocks in the Materials Sector**

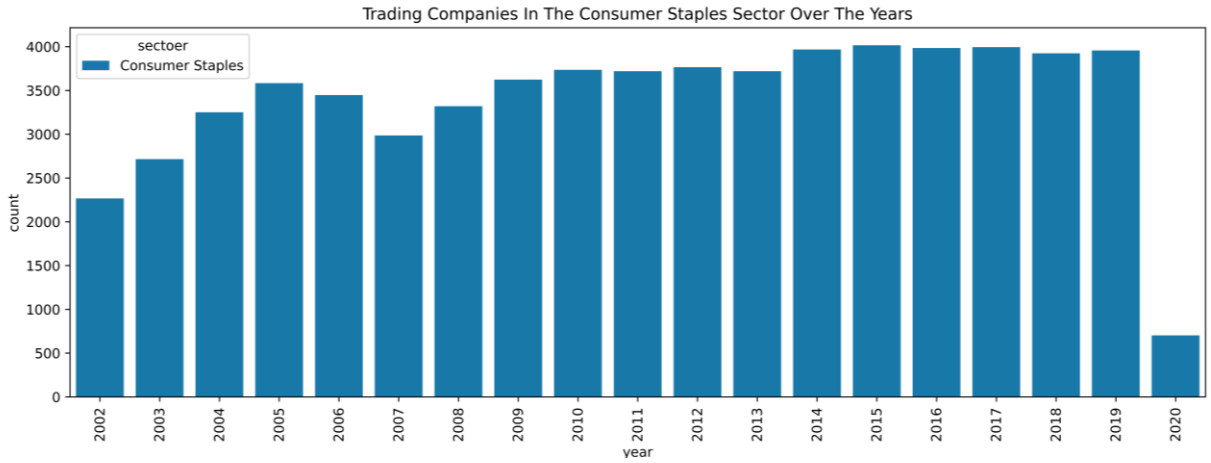


- The highest closing of Bank Aljazeera shares (YCC) was around 110 in 2005.
- The highest close of Saudi Cement Co (SAUDI CEMENT) shares was around 125 in 2015.
- The highest close of Saudi Arabian Fertilizer Co (SAFCO) shares was around 150 in 2009.
- The highest close of National Industrialization Co (TASNEE) shares was around 75 in 2006.
- The highest close of Saudi Basic Industries Corp (SABIC) shares was around 250 in 2006.

DATA EXPLORATION

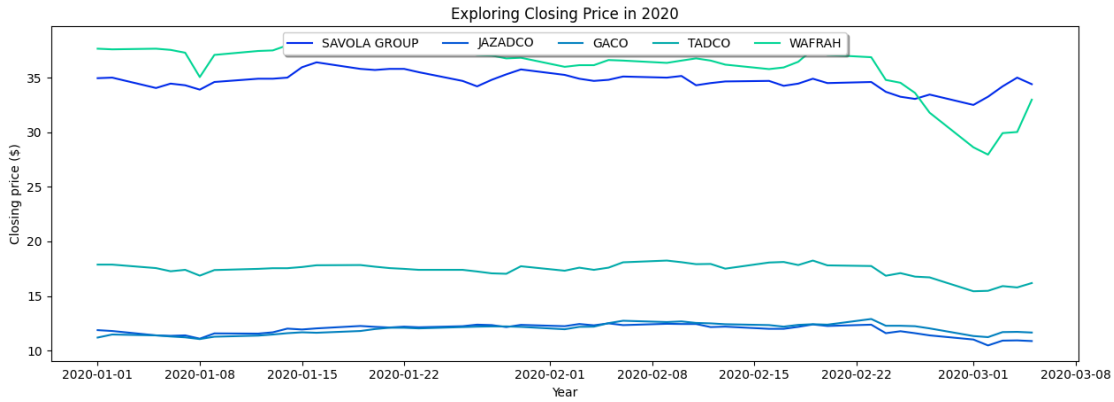
3. Consumer Staples Sector

- **Trading companies in Consumer Staples Sector over the Years**



Here we observe the increase in all years. Then a significant drop happened in 2020 due to Covid-19.

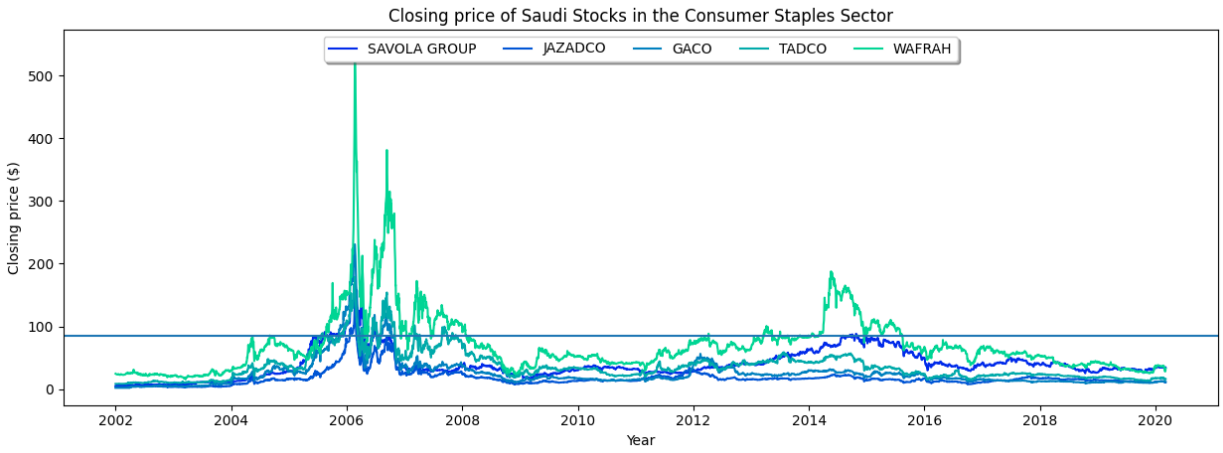
- **Exploring 2020 in Consumer Staples Sector**



DATA EXPLORATION

Here we observed a decrease in the closure in 2020, precisely during March, the beginning of Covid-19.

- **The closing price of the Consumer Staples Sector**



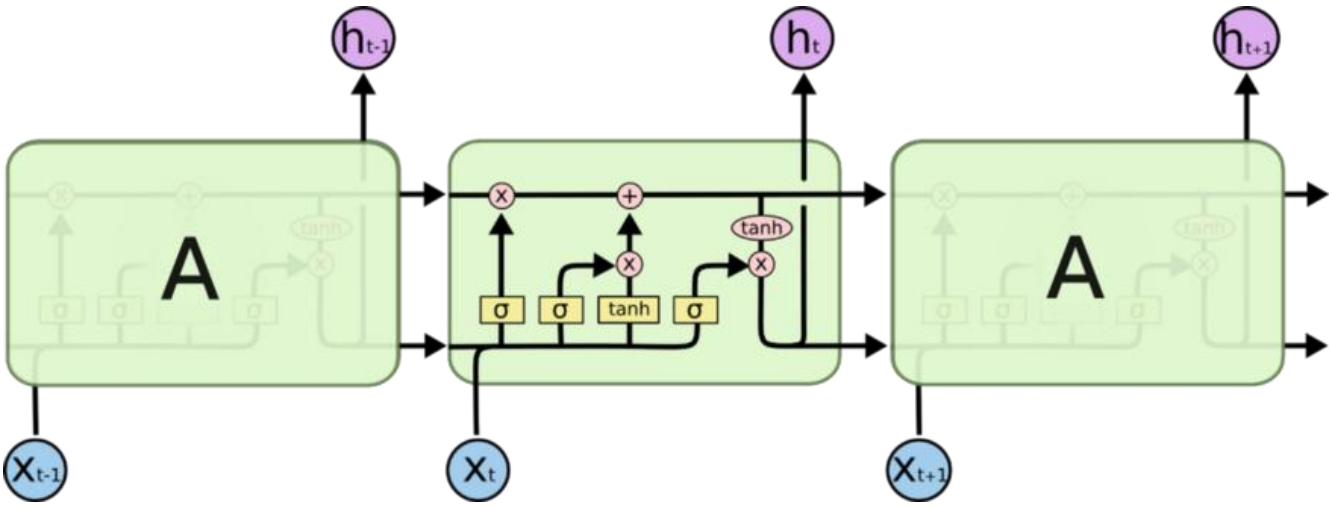
- The highest closing of Mobile Telecommunication Company Saudi Arabia(Zain KSA) shares was around 110 in 2008.
- The highest close of Saudi Research and Marketing Group (SRMG) shares was around 100 in 2018.
- The highest close of Etihad Etisalat Co (ETIHAD ETISALAT) shares was around 100 in 2006.
- The highest close of Saudi Telecom Co (STC) shares was around 180 in 2006.
- The highest close of Tehama Advertising and Public Relations Co (TAPCO) shares was around 800 in 2014.

Long short-term memory (LSTM) networks:

LSTM is a Deep Learning model. Recurrent neural networks (RNNs) can learn long-term dependencies, especially in sequence prediction problems. LSTM has feedback connections and can process the entire data sequence, apart from single data points such as images. This model finds application in speech recognition and machine translation. LSTM is a special RNN that shows outstanding performance on data sequence problems.

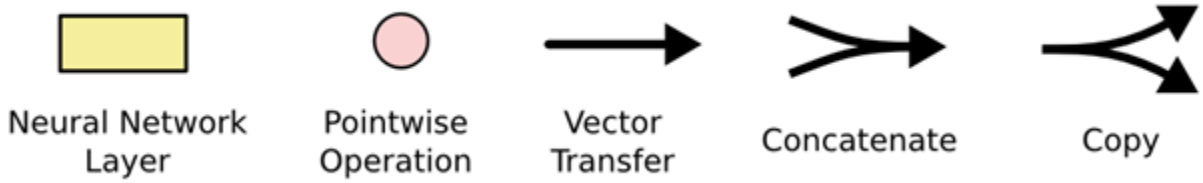
The Logic Behind LSTM:

The central role of an LSTM model is held by a memory cell known as a 'cell state' that maintains its state over time. The cell state is the horizontal line that runs through the top of the below diagram. It can be visualized as a conveyor belt through which information flows unchanged.

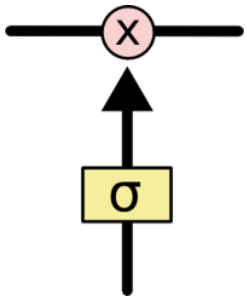


DEEP LEARNING MODEL

Information can be added to or removed from the cell state in LSTM and is regulated by gates. These gates optionally let the information flow in and out of the cell. It contains a pointwise multiplication operation and a sigmoid neural net layer that assist the mechanism.



In the above diagram, each line carries an entire vector from one node's output to others' inputs. The pink circles represent pointwise operations, like vector addition, while the yellow boxes are learned neural network layers. Lines merging denote concatenation, while a line forking denotes its content being copied and the copies going to different locations.

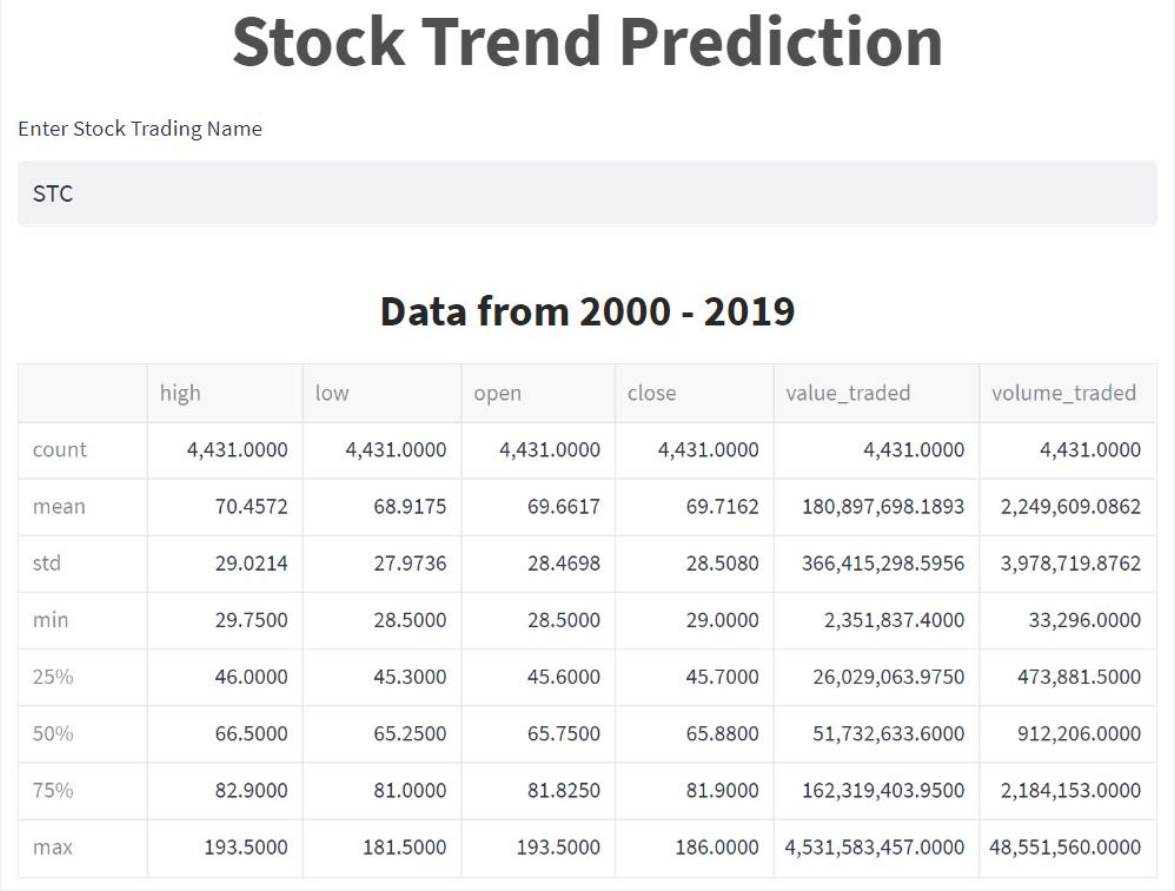


The sigmoid layer gives out numbers between zero and one, where zero means ‘nothing should be let through’ and one means ‘everything should be let through’.

DEEP LEARNING MODEL

LSTM Implementation for Stock Price Prediction:

Click HERE to visit the web app



Stock Trend Prediction

Enter Stock Trading Name

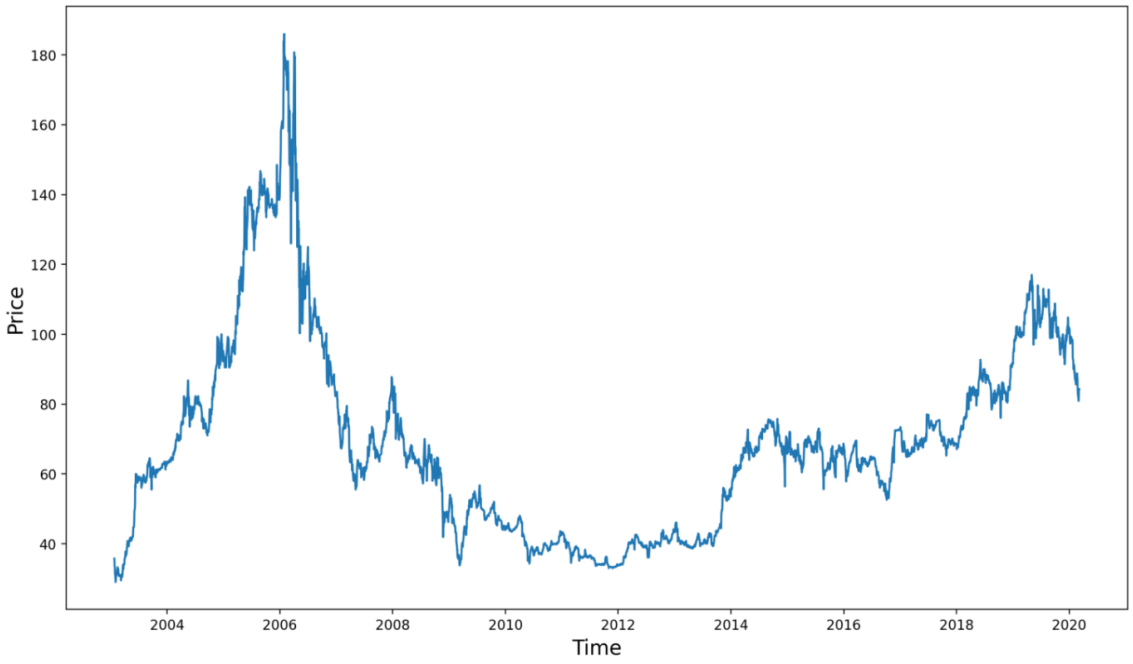
STC

Data from 2000 - 2019

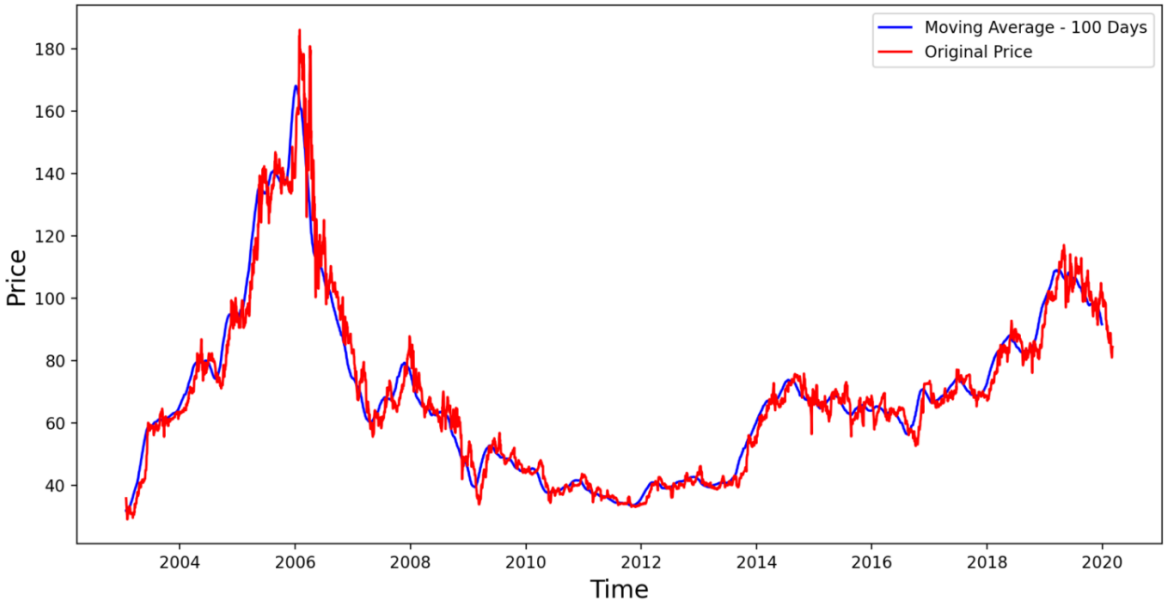
	high	low	open	close	value_traded	volume_traded
count	4,431.0000	4,431.0000	4,431.0000	4,431.0000	4,431.0000	4,431.0000
mean	70.4572	68.9175	69.6617	69.7162	180,897,698.1893	2,249,609.0862
std	29.0214	27.9736	28.4698	28.5080	366,415,298.5956	3,978,719.8762
min	29.7500	28.5000	28.5000	29.0000	2,351,837.4000	33,296.0000
25%	46.0000	45.3000	45.6000	45.7000	26,029,063.9750	473,881.5000
50%	66.5000	65.2500	65.7500	65.8800	51,732,633.6000	912,206.0000
75%	82.9000	81.0000	81.8250	81.9000	162,319,403.9500	2,184,153.0000
max	193.5000	181.5000	193.5000	186.0000	4,531,583,457.0000	48,551,560.0000

DEEP LEARNING MODEL

Closing Price vs Time Chart

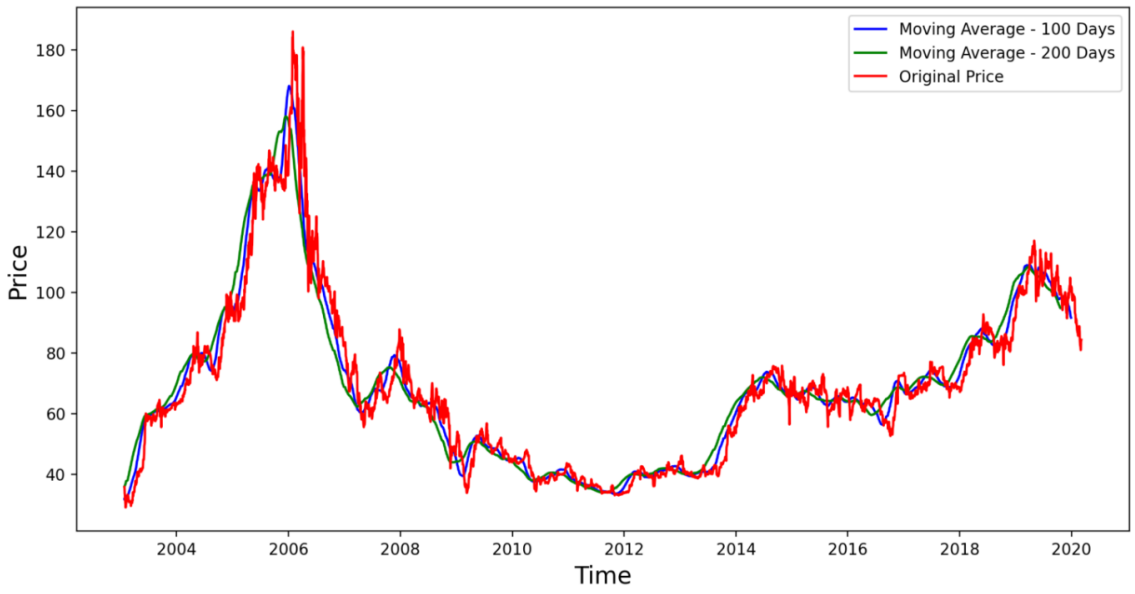


Closing Price vs Time Chart with 100 days Moving Average

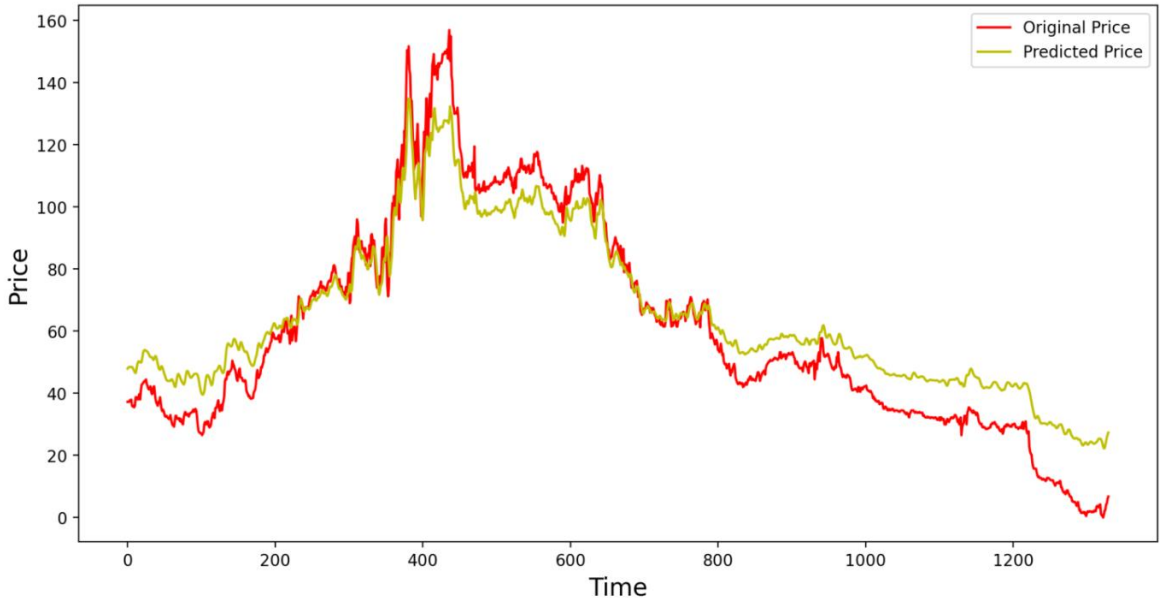


DEEP LEARNING MODEL

Closing Price vs Time Chart with 100 & 200 days Moving Averages



Predictions vs Original



RESULT

The stock market is one of the most complex systems that humankind has created. AI and machines can fail when they have to counter massive market fluctuations. Humans have an edge. We can make informed decisions by analyzing future catalysts of an asset and the reasons why the price movement goes up or down, and that's a strategy AI has yet to beat us.

Going through the (Data Exploration) in this project, we were able to analyze and investigate the Tadawul dataset and summarize the main characteristics, which helped us to understand patterns within the data better, detect outliers or abnormal events, and find exciting relations among the variables.

For the main objective of this project, long short-term memory (LSTM) is a deep learning architecture based on an artificial recurrent neural network (RNN). This architecture was applied to predict the closing price of each company in the Saudi Stock Market.

To help inform the overall picture of a stock's inherent value with its actual market fluctuation, moving average (MA) calculation was used, a stock indicator commonly used in technical analysis, to help smooth out price data by creating a constantly updated average price.

However, after building the Deep Learning model, there was apparent resistance from the model to predict low closing prices accurately due to the Saudi Stock Market regulation, where there is some limitation over how the loss and profits cannot exceed 10% percent.

FUTURE WORKS

Our future works:

1. Try different methods and techniques to predict the stock price models.
2. Add new columns from Tadawul to have more insight.
3. Enhance the performance of the applied model.
4. Build a live dashboard for Tadawul Stock Exchange.