Saudi stock Exchange (Tadawul)

Machine Learning Project

Instructor: Amro Handousa

Group	Trainee Names
Girls Power	Shouq bakry
	Layla Alshehri
	Ebtisam Almutairi
	Wafa Alshehri
	Jana Aldarwish

Table of Contents

About Saudi Exchange 3
Introduction 3
Machine Learning Project Report 3
1- Problem Statement: 4
2. Data Review 4
3. Exploratory Data Analysis (EDA) 4
4.The approach: 6
5.Machine Learning (ML) algorithms Implementation for Stocks Prediction: 6
6.Findngs and Result: 8
7.Conclusion 9

About Saudi Exchange

The Saudi Exchange is a fully owned subsidiary of Saudi Tadawul Group and was established in March 2021 following the transformation of the Saudi Stock Exchange (Tadawul) into a holding company, Saudi Tadawul Group.

As Saudi Arabia's dedicated stock exchange and the largest stock exchange in the Middle East, the exchange carries out listing and trading in securities for local and international investors. The official source of all market information, Saudi Exchange, is instrumental in achieving long-term growth plans for the group and providing market participants with attractive and diversified investment opportunities.

The Saudi Exchange is the 9th largest stock market among the 67 members of the World Federation of Exchanges and is the dominant market in the Gulf Cooperation Council (GCC). It is the 3rd largest stock market among its emerging market peers and is an affiliate member of the International Organization of Securities Commissions (IOSCO), the World Federation of Exchanges (WFE), and the Arab Federation of Exchanges.

Introduction

Saudi stock market is being large goal and attract many investors as one of goal of 2030 to be expanded and include more investors so, we decide to apply machine learning on this major (Tadawul) to predict stock price and can use it by many organizations.

Machine Learning Project Report

Stock price predictions are very important among many businesspeople and the public. People can make a lot of money or lose their financial income from a stock market job. Algorithm predictions and models can be used to make future predictions applied to historical data. Predicting the future has been a daunting task, one that many have found difficult to understand. This type of prediction is even more appealing when it involves money and risks such as Stock Market speculation. Researchers are conducting research on stock market forecasts from a variety of fields, including computer science and business. Researchers have tried a variety of methods to predict the market, including different strategies and algorithms and the combination of indicators. The attribute that makes a prediction model depends on factors on which market performance can depend.

We worked here with historical data about the stock prices of a publicly listed company. We implemented a mix of machine learning algorithms to predict the future stock price of this company, starting with simple algorithms like averaging and linear regression, and then move on to advanced techniques like LSTM. Long-Short-term Memory (LSTM) is one of a variety of RNNs structures. LSTM replaces traditional artificial neurons in the hidden network layer into the most useful memory cells. With these memory cells, networks can better associate memory with remote input over time, which

is why it is worthwhile to understand the formation of strong data over time with great predictive power. How to improve the accuracy of stock prices is an open question today so, we hope our model we apply it to help more for predict stock price for organizations, company, and bank. specially in 2030 the stock market expands and serves more sectors.

1- Problem Statement:

Our project is focused around helping investors to invest in Saudi market exchange Tadawal. Stock is an unpredictable curve. Prediction in stock market is covered with complexity and instability. In this project we attempt to implement machine learning approach to predict stock prices. Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices to make more informed and accurate investment decisions. We propose a stock price prediction system uses machine learning for the purpose of achieving better stock prediction accuracy and issuing profitable trades.

2. Data Review:

Saudi Stock Exchange (Tadawul) dataset which used in machine learning project taken from Kaggle.

This is the data of Saudi stock market companies since 2000-01-01 until 5-3-2020. It was collected from Saudi Stock Exchange (Tadawul). It contains 14 column and more than half million rows also, as there are many sectors we decide to choose two of them Energy & Communication service.

https://www.kaggle.com/salwaalzahrani/saudi-stock-exchange-tadawul

3. Exploratory Data Analysis (EDA)

- Retype some column's names to make them more consistent.

```
In [4]:
df.rename(columns={'no_trades ': 'no_trades',
                   'trading_name ': 'trading_name',
                   'volume_traded ': 'volume_traded',
                   }, inplace=True)


df.columns
```

```
Out[4]: Index(['symbol', 'name', 'trading_name', 'sectoer', 'date', 'open', 'high',
       'low', 'close', 'change', 'perc_Change', 'volume_traded',
       'value_traded', 'no_trades'],
      dtype='object')
```

- Change date data type.

```
In [28]:
#change the date data type
df['date']= pd.to_datetime(df['date'])
```

- Handling with nulls values:
  - Check number of zero values in close column because it is our target.

In [35]:
```
# Check number of zero values
# we observe the zeros value in the close column they are meaningless in the whole rows
df[df['close']==0]
```

Out[35]:

symbol	name	trading_name	sectoer	date	open	high	low	close	change	perc_Change	volume_traded	value_traded	no_trades

- Filter the sectors by (Energy and communication services)
  - The Energy Sector.

In [49]:
```
Energy = df[df['sectoer']=='Energy'].round(1)
Energy.sort_values('date', inplace=True)
Energy.head()
```

Out[49]:

	symbol	name	trading_name	sectoer	date	open	high	low	close	change	perc_Change	volume_traded	value_traded	trades_Number
12301	4030	National Shipping Company of Saudi Arabia	BAHRI	Energy	2001-12-31	9.2	9.2	9.2	9.2	9.2	0.0	256935.0	2367154.5	44.0
12300	4030	National Shipping Company of Saudi Arabia	BAHRI	Energy	2002-01-01	9.2	9.2	9.2	9.2	0.0	0.0	13693.0	125845.2	7.0
12299	4030	National Shipping Company of Saudi Arabia	BAHRI	Energy	2002-01-02	9.2	9.2	9.0	9.2	0.0	0.0	30793.0	281304.8	12.0
12298	4030	National Shipping Company of Saudi Arabia	BAHRI	Energy	2002-01-03	9.2	9.0	9.0	9.0	-0.2	-2.7	47315.0	430902.0	9.0
12297	4030	National Shipping Company of Saudi Arabia	BAHRI	Energy	2002-01-05	9.0	9.2	9.0	9.2	0.2	2.8	235129.0	2167297.2	27.0

- The communication services Sector.

In [66]:
```
cs = df[df['sectoer']=='Communication Services'].round(1)
cs.sort_values('date', inplace=True)
cs.head()
```

Out[66]:

	symbol	name	trading_name	sectoer	date	open	high	low	close	change	perc_Change	volume_traded	value_traded	trades_Number
263765	4070	Tihama Advertising and Public Relations Co.	TAPRCO	Communication Services	2001-12-31	21.0	21.0	21.0	21.0	-31.0	-59.6	3750.0	77933.8	3.0
263764	4070	Tihama Advertising and Public Relations Co.	TAPRCO	Communication Services	2002-01-01	21.0	20.5	20.5	20.5	-0.5	-2.4	5213.0	107898.8	6.0
263763	4070	Tihama Advertising and Public Relations Co.	TAPRCO	Communication Services	2002-01-02	20.5	20.5	20.5	20.5	0.0	0.0	6250.0	129375.0	2.0

- Filter the companies (Tow companies from each sector):
  - National Shipping Company of Saudi Arabia Co. (BAHRI).
  - Saudi Pharmaceutical Industries & Medical Appliances Corporation (SARCO).
  - Tihama Advertising and Public Relations Co. (TAPRCO).
  - Saudi Telecom Co. (STC).

In [53]:
```
grouped_BAHRI = (grouped_df.loc["BAHRI",:]).reset_index()
grouped_BAHRI.round(2).info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4755 entries, 0 to 4754
Data columns (total 8 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   date           4755 non-null   datetime64[ns]
 1   open           4755 non-null   float64
 2   high           4755 non-null   float64
 3   low            4755 non-null   float64
 4   close          4755 non-null   float64
 5   change         4755 non-null   float64
 6   value_traded   4755 non-null   float64
 7   trades_Number  4755 non-null   float64
dtypes: datetime64[ns](1), float64(7)
memory usage: 297.3 KB
```

4.The approach:

In fact, in this project we used different machine learning algorithms and compare each one to find the best one which achieves better stock prediction accuracy and issuing profitable trades.

5.Machine Learning (ML) algorithms Implementation for Stocks Prediction:

First, we did preprocess and EDA techniques on the data and chose only two sectors energy and communication service to make it ready for the machine learning algorithms. Second, we apply our first model which is Linear Regression we used some features to predict our target which is close values. The features are symbol, which is unique number for each company, value traded which is the stock company sales and buy for the day, and no trades which is the number of trades for the day. Then we split data to train and test and scaled them. After that we built the linear regression model. Then we applied the prediction with 15% error tolerance after that we get true and false predation. Also, we apply the same logic in different model decision tree and random forecast. Below you can see true predation percentage of the model for each sector.

Model Name	Energy	Communication Service
Linear	33%	23%
Decision tree	44%	-
Random Forest	55%	21%

The time constraint and our logic in data splitting that the main cause of our issues here, that is why we start searching about another approach that related with the time series.

5.1 Moving Average:

We applied the Moving Average, we made two arrays for the close price dependent on the date variable:

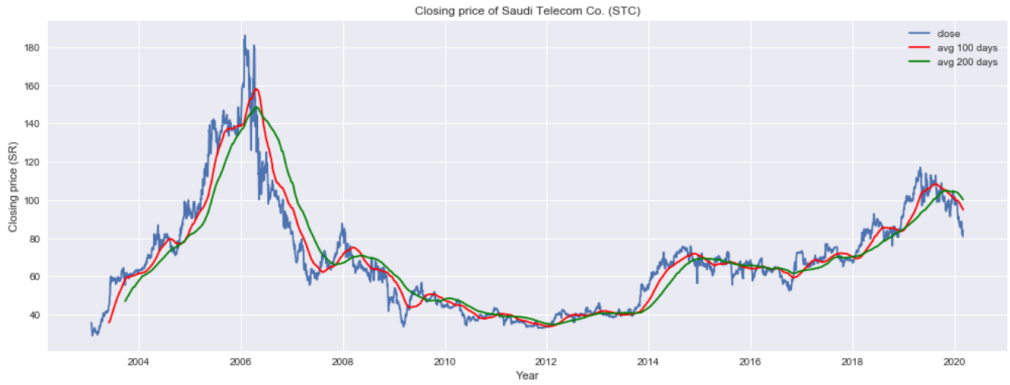
-The first moving average starting after 100 days:

```
In [23]: mean_pre_100=date_index['close'].rolling(100).mean()
         mean_pre_100
```

-The second moving average starting after 200 days:

```
In [26]: mean_pre_200=date_index['close'].rolling(200).mean()
         mean_pre_200
```

This chart shows the Moving Average with shifting 100 days " the red line" and shifting 200 days " the green line":

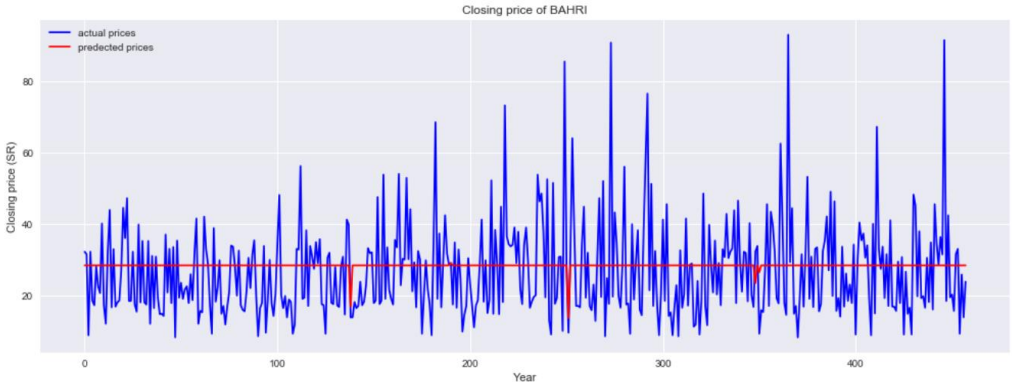


It is simple technique to predict the data with time series, that just shows the behavior of the data, we can't use this technique to predict the close prices

5.2 Support Vector Machine (SVM):

The true prediction percentage obtained by SVM model was 11%. While applying Grid search techniques the true prediction percentage increased by 10%. After improving the model, the true prediction percentage was less than 70%, as a result, we can't use this model to predict the close prices also.

This chart shows the actual and predict closing prices with SVM model:



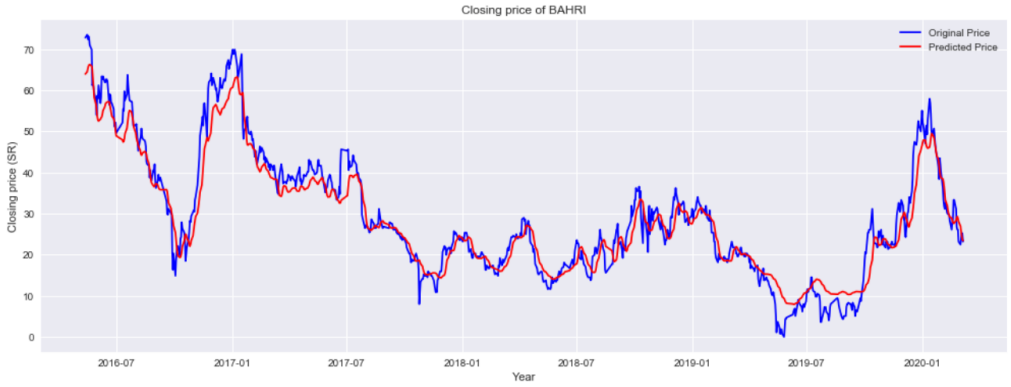
5.3 Long Short-Term memory (LSTM):

- Making the date column as index.
- Splitting the data into training with 80% and testing with 20%.
- Scaling the training array by MinMaxScaler().
- Making a loop for splitting the training data into x_train to train 180 days and y_train to predict the day 181 and so on.
- Applying Sequential model.
- Adding 4 layers for the model.
- Compiling the model.
- Scaling and splitting the testing data as what we did with training data.
- Finding the prediction price.
- Evaluating regression models by Mean Square Error and Mean Absolute Error.
- Calculate the true prediction to evaluate performance the model.
- Exporting csv file have actual price and predict price for each company to importing it in the tableau program to making a dashboard for the model result.

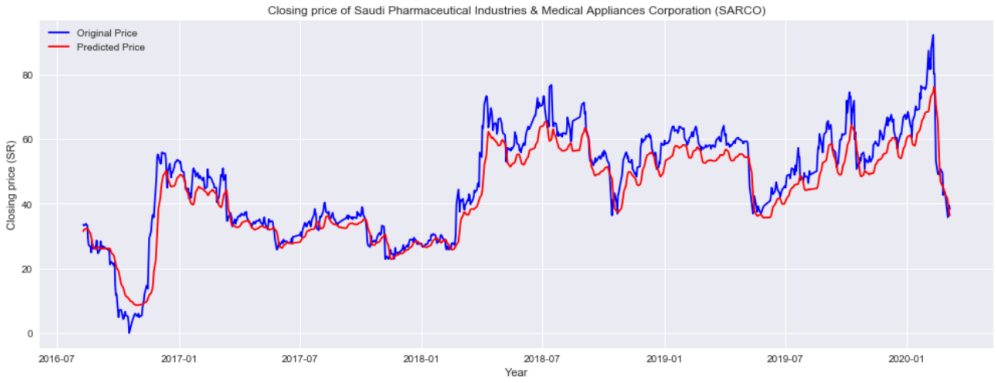
6.Findngs and Result:

As we explained above, we tried many models and we reached to one of them give us a high accuracy so, we decided to go and tried another model and depending on research that we found the best one for predict stock price is Long short-term memory (LSTM). We chose to apply LSTM model in four company, the oldest tow from each sector Energy & Communication Service with our target column close.

- First company Bahri Business Units for Energy sector:

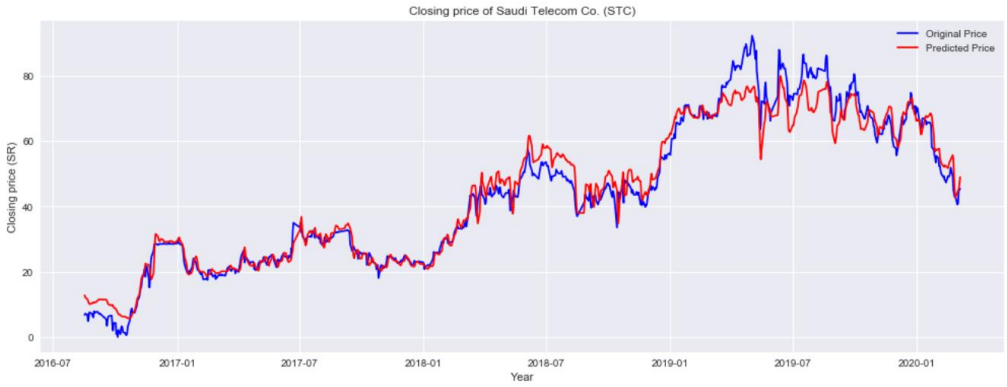


- Second Company Saudi Pharmaceutical Industries & Medical Appliances Corporation (SARCO) for Energy sector:



- Third company Tihama Advertising and Public Relations Co (TAPRCO) for Communication Service Sector:

- Forth company Saudi Telecom Co. (STC):



7.Conclusion

Finally, we learn many things when we worked in this project and faced many challenges to complete. Some of lesson we learned time management, searching better as a teamwork at the same time we faced some challenges when we apply LSTM model but at the end, we reach to satisfying result and we got a useful lesson.