Introduction

Amazon Stock Price dataset was used for time series forecasting. In addition to statistical operations and calculations, stock prediction systems were created by implementing SARIMA, Prophet, LSTM, CNN and LSTM models. The systems were evaluated with metrics such as MAPE, MAE, R^2.

Exploratory Data Analysis (EDA)

The dataset contains 5842 rows of information. As can be seen in Table 1, the data contains information between 1997 and 2020.

Columns	Description	Data Type
Date	Prices on the date in yy-mm-dd format.	datetime64
Open	The price of the stock at market open	float64
High	Highest price reached in the day	float64
Low	Lowest price reached in the day	float64
Close	The stock closing at the end of the Market hours	float64
Adj Close	The closing price after adjustments for all applicable splits and dividend distributions	float64
Volume	Number of shares traded	float64

Table 1: Amazon Stock Price dataset

Some of the data in the dataset are as shown in Table 2.

Date	Open	High	Low	Close	Volume	Adj Close
1997-05-15	2.437	2.5	1.927	1.958	72156000	1.958
1997-05-16	1.968	1.979	1.708	1.729	14700000	1.729
1997-05-17	1.760	1.770	1.625	1.708	6106800	1.708
1997-05-18	1.726	1.75	1.635	1.635	5467200	1.635

Table 2: Examples in the Amazon Stock Price dataset

The data has been analyzed since 1997 and it has been observed that the close prices of Amazon shares have followed a similar line until 2020, and the open, high and low prices have followed a similar line. Image 1 shows the line of the close price between 1997 and 2020.

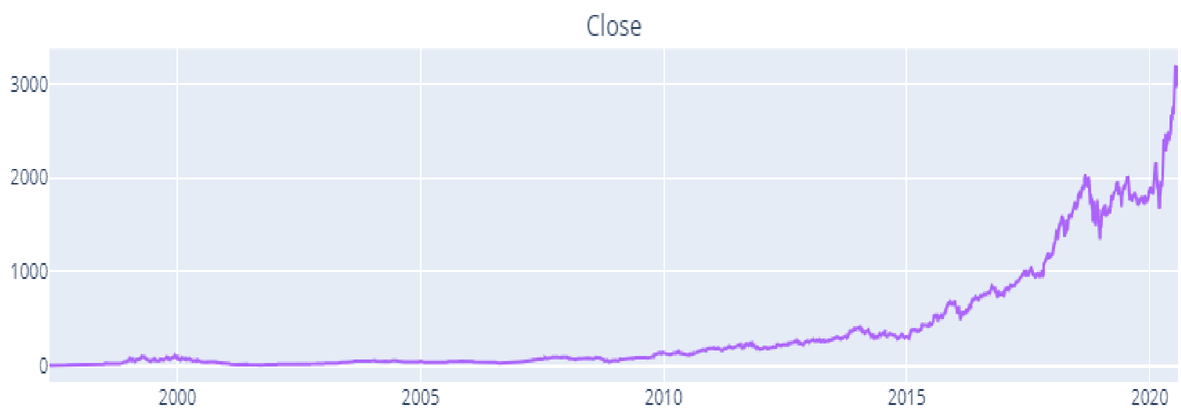

Image 1: Close price trend line between 1997 and 2020

Image 2 shows the histogram graph of the transaction volume using the Volume information.

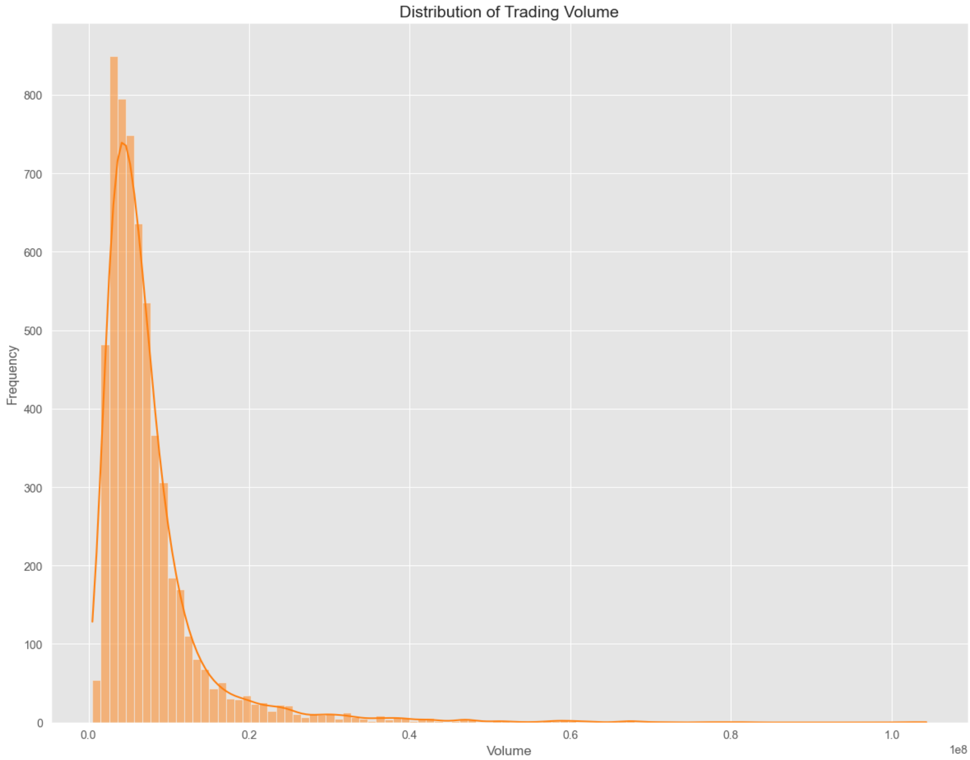

Image 2: Distribution of trading volume

The distribution of trading volume ranges widely, with most of it concentrated in low volume areas. Low volume usually means that there is less liquidity in the market and prices can be more volatile [1]. The chart in Image 2 shows that on most trading days, Amazon shares traded on relatively low volumes. This can also be attributed to the increase in Amazon share prices in recent years. It is known to be riskier to enter the market on low volume days, while on high volume days the market can be more predictable [2].

Using Close information in the dataset, a feature was derived to examine medium and long-term price movements. MA_50 for medium-term price movements and MA_250 for long-term price movements were extracted, resulting in the graph in Image 3.

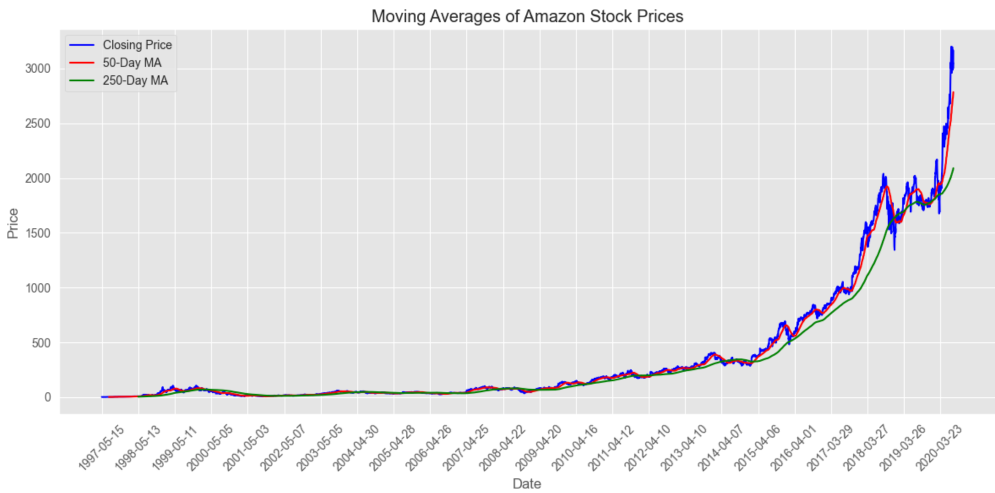

Image 3: Moving averages of close prices

When the chart in Image 3 is analyzed, the observation window is set at 50 and 250 and trend components are calculated. When the trend of medium and long-term price movements cross each other upwards, it can generally be considered a buy signal, indicating that the stock has entered an uptrend. This indicates a potential buying opportunity for investors [3]. The trend components prepared with this intention are shown in Image 4.

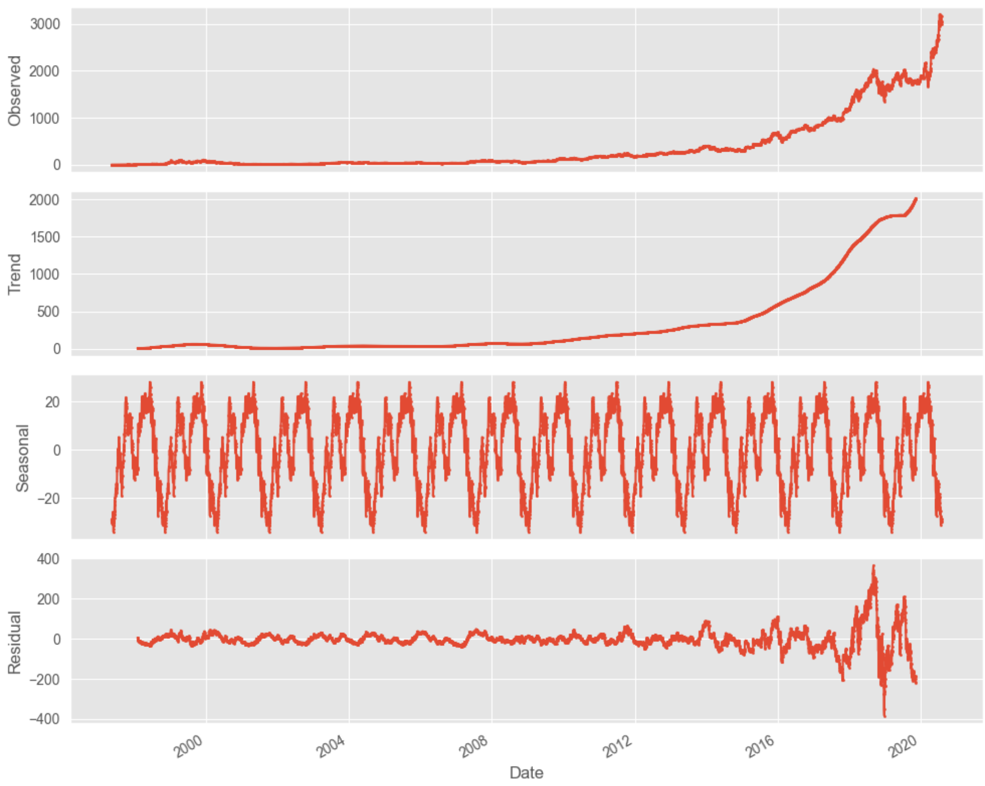

Image 4: Seasonal decompose

It has been observed that the fluctuation of Amazon share prices and their repetition in certain periods depend on parameters such as year-end closes, company balance sheet announcements [4]. In order to analyze the daily volatility of the stock, Daily Return feature was extracted from the close information. Image 5 shows the volatility of stock price.

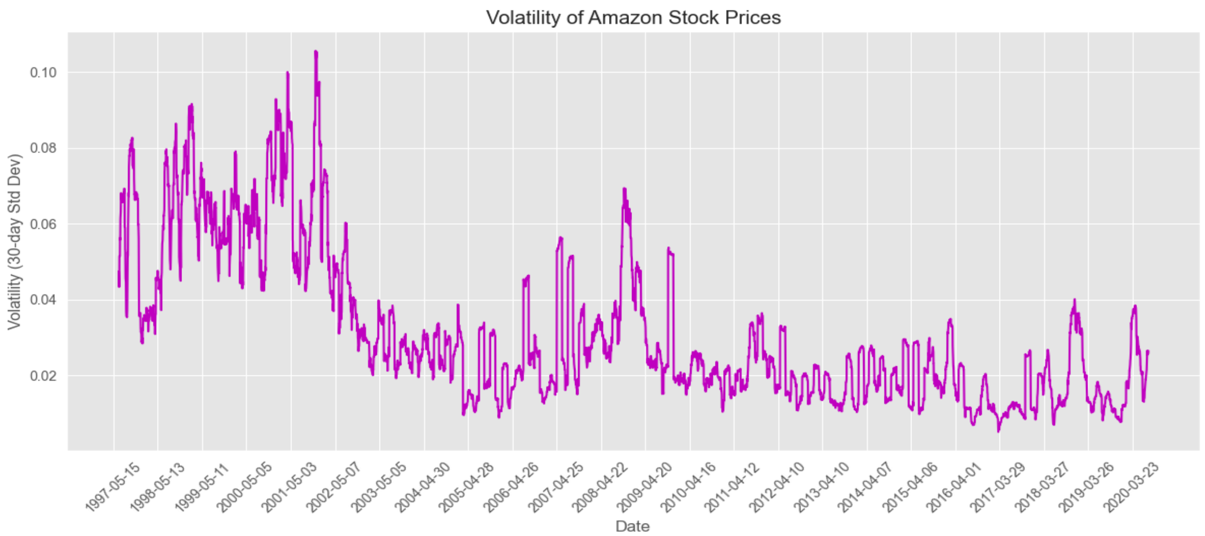

Image 5: Volatility of stock prices

Low volatility is seen as a stable investment instrument, while high volatility means that the risk is high. In addition, a fluctuating volatility also indicates that the stock may be risky [5]. It can be interpreted that the risk is high in the first few years of the data and that the risk decreases towards 2020 and that Amazon stock can be preferred as an investment instrument. In addition, in order to make comments with the same logic, the Annual Return feature is generated from the Adj Close feature and the Annual Volume feature is generated from the Volume feature and graphed in Image 6.

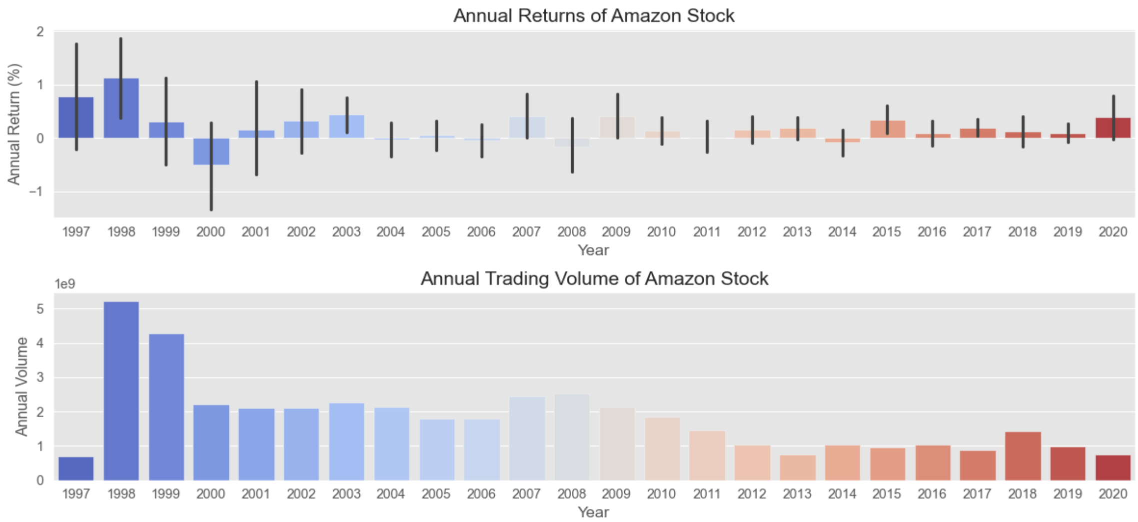

Image 6: Annual volume and returns of Amazon stock

Annual Return is mostly positive, which means that in the long run investors are making gains. The trend in the annual trading volume is upward, indicating that investors are trying to dispose of stocks. Thus, the characteristics appear to support the comments made for Volatility. This gives the impression that investors are making gains. It can be shown that in the early years Amazon stock was a risky but potentially high-yielding investment.

Models

In addition to statistical operations and calculations, stock prediction systems were created by implementing SARIMA, Prophet, LSTM, CNN and LSTM models. Short evaluations were made respectively with appropriate metrics. The overall evaluation is done in the last section by comparing the results between the models.

SARIMA

In addition to Exploratory Data Analysis (EDA), the Augmented Dickey-Fuller (ADF) test revealed that the time series was non-stationary, and trend correction was performed. This provided a stationary series with a constant mean, variance and covariance over time for the use of SARIMA. The ADF test was applied to evaluate the SARIMA model where only the Close column from the dataset was given as input and the resulting outputs are shown in Table 3.

	Before Detrending	**After Detrending**
ADF Test Statistic	5.773923749756864	-6.1979022347352135
p-value	1.0	0.000000059037296
Critic Value 1%	-3.431124850687746	-3.431476604615945
Critical Value 5%	-2.8618824309525883	-2.8620378526446375
Critical Value 10%	-2.5669522711136055	-2.567035004946881
Hypothesis	Residuals are not stationary.	Residuals are stationary.

Table 3: ADF test results

The correlation map of the features added to the dataset and the final version of the data is shown in Image 7.

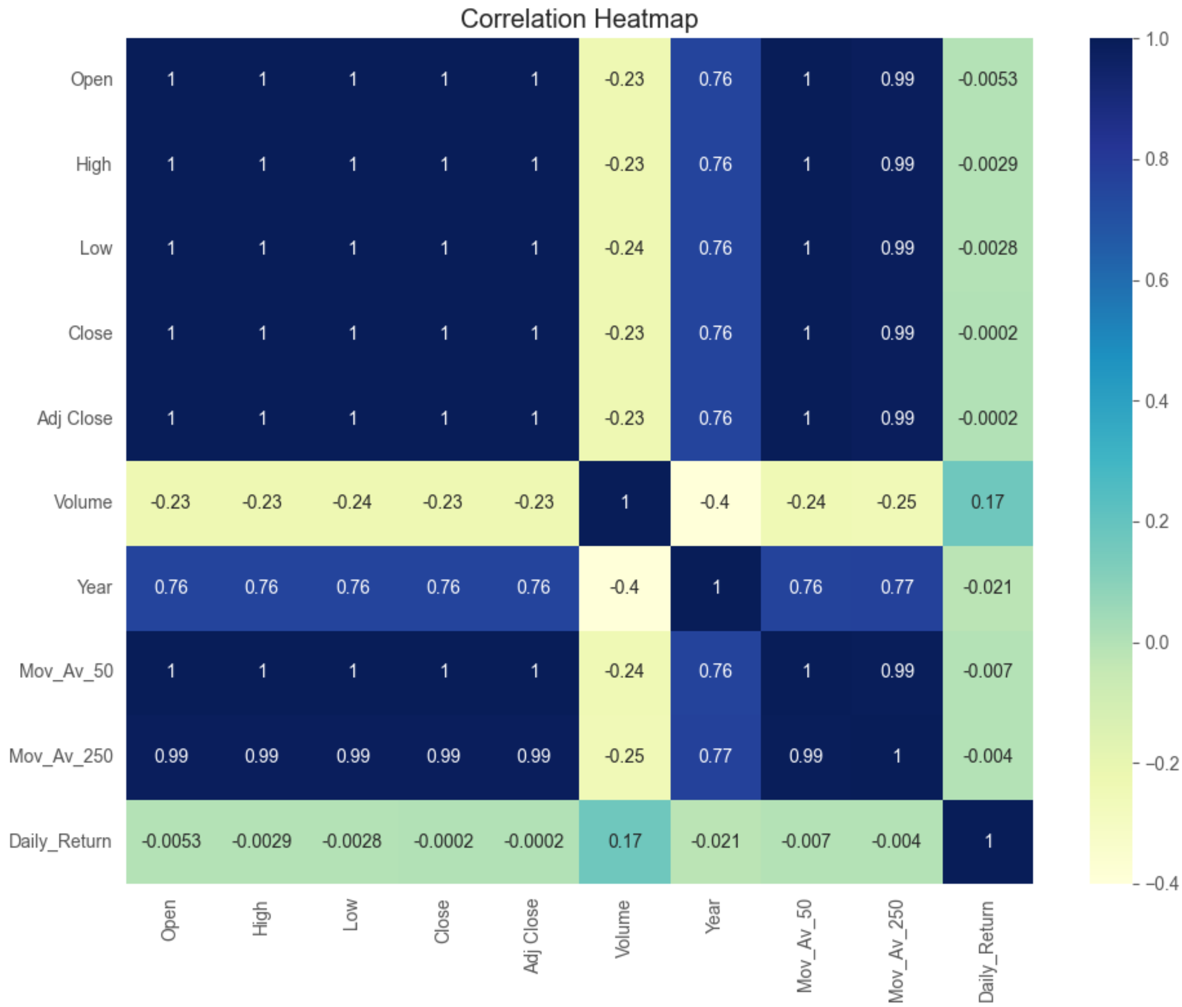

Image 7: Correlation matrix

Hyperparameter optimization was performed for the SARIMA model by finding the most optimal autoregressive (AR), differencing (I) and moving average (MA) values with the auto_arima function in the pmdarima library. The seasonality (S) value is chosen as 7 since there are 5 trading days and a weekend. The best model was found to be ARIMA(2,1,2)(0,0,0)[7].

Evaluation

The SARIMA model was measured with MAE, MAPE and R^2 metrics in common with the other models. The measured metrics are shown in Table 4.

Model	MAE	MAPE	R^2
SARIMA	363.234236	17.413345	-0.971901

Table 4: Evaluation metrics

The MAE shows that the value of the Amazon stock price is in the range of approximately positive to negative 363 differences, while the MAPE shows this in percentage terms. These values are quite large and while it can be seen that it does not predict the data well, the fact that it does not explain the data well is evidenced by the negative R^2 metric. The SARIMA prediction graph is shown in Image 8.

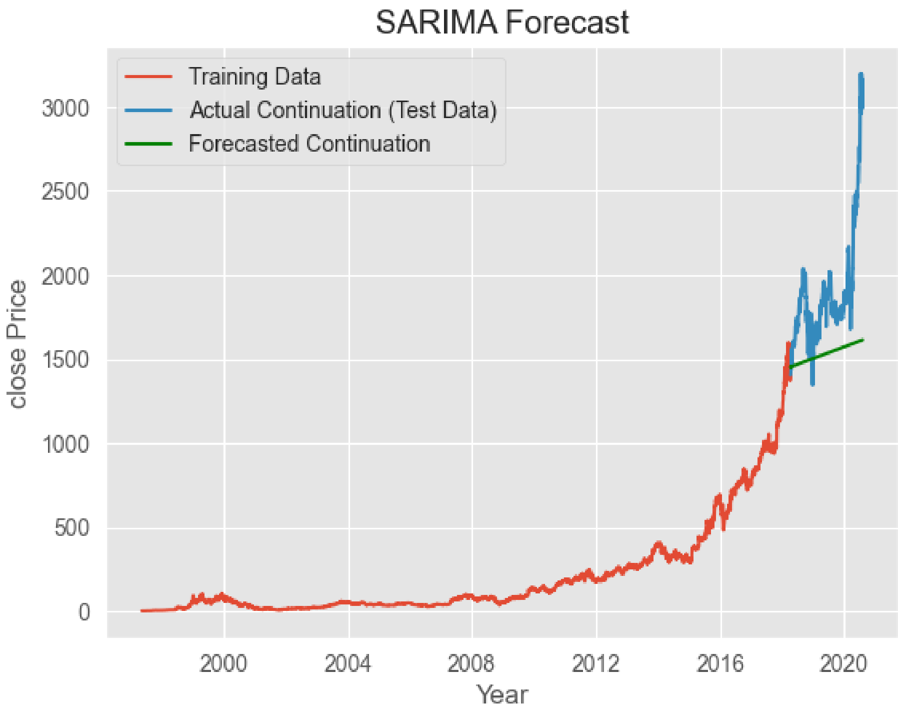

Image 8: SARIMA forecast graph

It has also been experienced that the model shows the trend trend to be partially accurate at an acceptable level.

Random Forest

The model was given the dataset in Table 2. Date and Close columns were not used in the model training and were dropped. To improve the performance of the model, hyperparameter optimization was performed with GridSearchCV and the hyperparameters in Table 5 were adjusted.

Hyperparameter	Description	Values			Best Values
n_estimators	number of trees used	100	200	500	500
max_depth	maximum depth of each tree	10	20	30	20
in_samples_split	minimum number of samples required to split a node	2	5	10	2
min_samples_leaf	minimum number of instances that must be at a leaf node	1	2	4	1

Table 5: Hyperparameters of Random Forest

For each hyperparameter combination, the training dataset was divided into three parts (cv=3) and the performance of the model was tested by Cross-Validation. Thus, the best parameters are displayed in the Best Values column in Table 5.

Evaluation and Conclusion

The best hyperparameters with the Random Forest model were applied to the dataset and Table 6 shows the performance metric results.

Model	MAE	MAPE	R^2
Random Forest	0.546649	0.154870	0.999984

Table 6: Evaluation metrics

The metrics show the fit of the model to the data and the likelihood of overfitting.

Prophet

Developed for time series prediction, Prophet uses only the date and the target variable to be predicted. Thus, this simple model was used to predict future values. The variables in Table 2 are given one by one in order with the Date column.

Evaluation

In the implemented code, Close was found to be the best performing prediction variable among all features. Image 9 shows the forecast with changepoints for the Close feature.

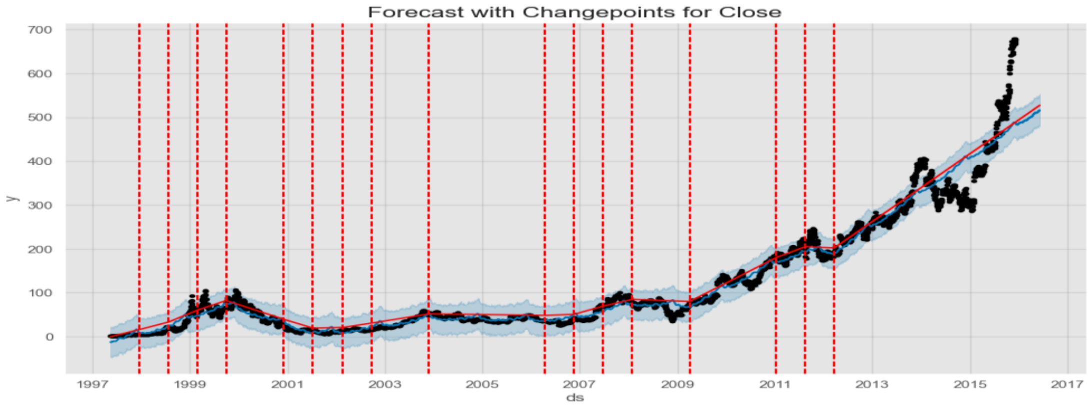

Image 9: Forecast with changepoints for close

This graph, which is seen in the data given for Training, was able to show a trend that can be considered consistent. In addition, the fact that Close was the most performant variable proved that the purpose of the study was appropriate in this respect. Table 7 shows the evaluation metric results.

Model	MAE	MAPE	R^2
Prophet	737.702311	46.619738	-1.304842

Table 7: Evaluation metrics

The model's predicted values are far from the positive and negative range with a fairly high MAE and a large deviation from the MAPE. This therefore indicates that the model fails to predict Close. Likewise, R^2 shows that the model predicts very poorly, worse than predicting with the mean value and even randomly due to its negativity.

LSTM

Additionally to the columns in the Amazon stock price dataset in Table 2, a Daily Return variable was generated, which was thought to provide meaningful information as described in EDA. In addition to this variable, Moving Average variables were created to carry short (10), medium (50) and long (100) term information. The dataset to be given to the model is shown in Table 8.

Date	Open	High	Low	Volume	MA for 10 days	MA for 50 days	MA for 100 days	Daily Return
1997-05-15	2.437	2.5	1.927	72156000	370.959051	363.507729	355.892522	0.001944
1997-05-16	1.968	1.979	1.708	14700000	370.959051	363.507729	355.892522	-0.11702
1997-05-17	1.760	1.770	1.625	6106800	370.959051	363.507729	355.892522	-0.01204
1997-05-18	1.726	1.75	1.635	5467200	370.959051	363.507729	355.892522	-0.04268

Table 8: Processing dataset given model

The data without Close, Adj Close and Date was split into 60-day sequences and fed into the LSTM model in Figure 1.

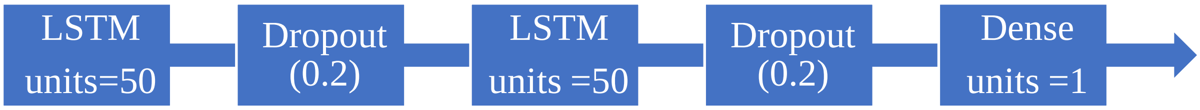

Figure 1: LSTM model architecture

The number of model epochs is 20 and the data batch size is 32.

Evaluation

The LSTM model is evaluated with MAE, MAPE and R^2 metrics. Image 10 shows the LSTM Amazon close price forecast graph for the first 6 years.

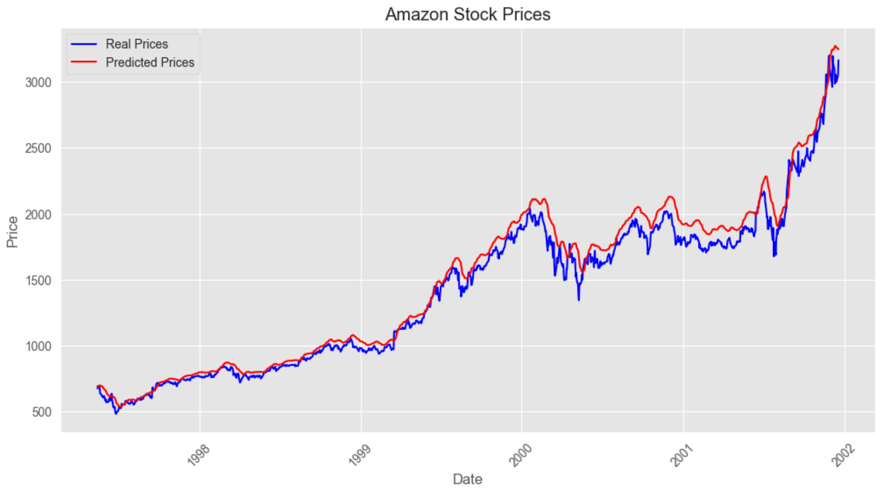

Image 10: LSTM close price prediction graph

The first year's forecast graph, which was preferred to be shown, successfully predicted the close price. This is also evidenced by the metrics in Table 9.

Model	MAE	MAPE	R^2
LSTM	76.244591	5.162967	0.970561

Table 9: Evaluation metrics

The LSTM model shows that it has a good understanding of the data, as evidenced by its R^2 score, as well as its low percentage deviation in prediction. It is also seen that the MAE value and the range of positive and negative differences are low.

LSTM Model with CNN for Raw Data

Considering the previous LSTM model, the model layers were organized and the CNN layer was added. In this model, it was preferred to train with the columns of the dataset in Table 2, i.e. in its raw form. Figure 2 shows this model.

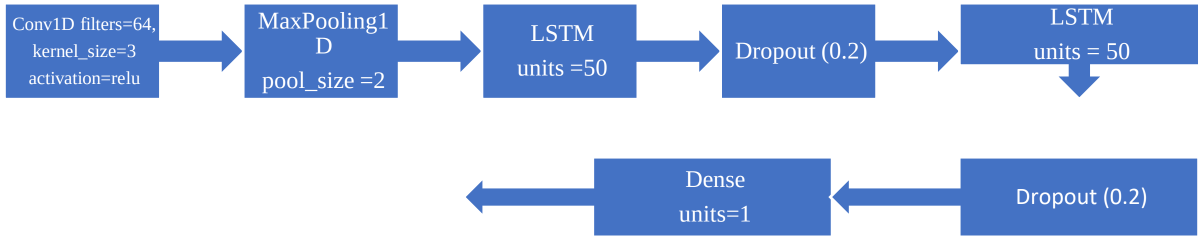

Figure 2: LSTM with CNN model architecture for Raw Data

As in the previous LSTM model, the number of epochs is 20 and the batch size is 32.

Evaluation

The designed model architecture has changed compared to the previous one. This model was developed for feature extraction with a CNN layer and the metrics improved. Image 11 shows the prediction graph of the model for close prices for the first 6 years.

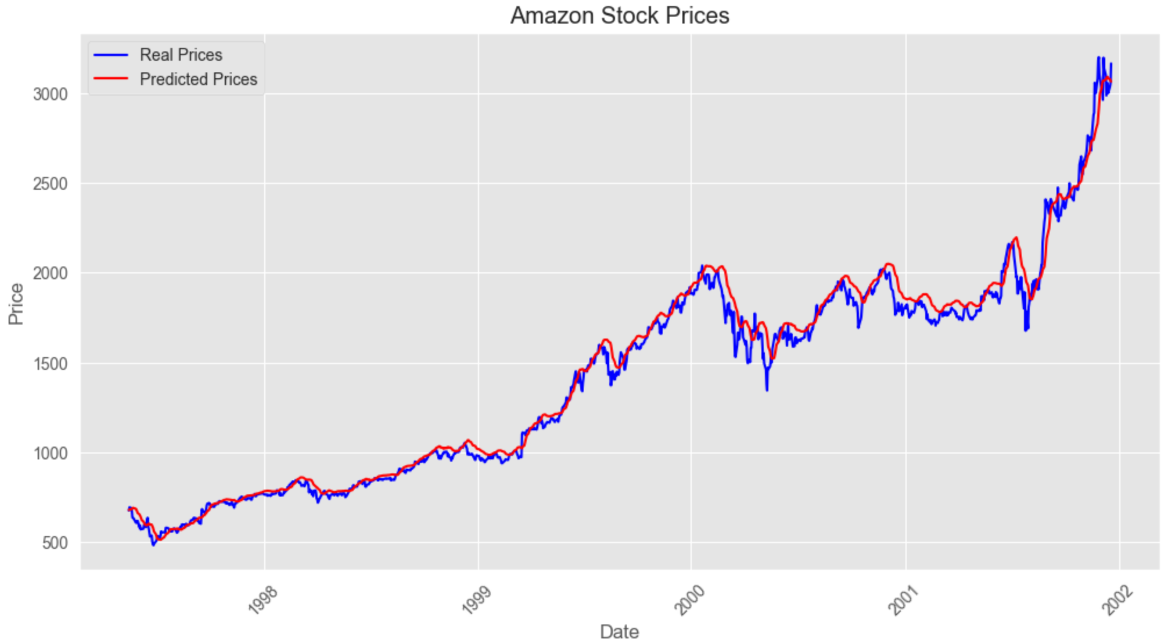

Image 11: LSTM with CNN close price prediction graph

In addition to predicting a closer value than the previous LSTM model, the evaluation metric results also provide support in this direction. Table 10 shows the evaluation metrics.

Model	MAE	MAPE	R^2
LSTM with CNN for Raw Data	47.107138	3.326102	0.985757

Table 10: Evaluation metrics

LSTM Model with CNN for Processing Data

The LSTM model with CNN processing data was established by considering the previous LSTM model with raw data and the LSTM model with CNN model and the data in Table 8 together and by developing some layers. Figure 3 shows the model architecture.

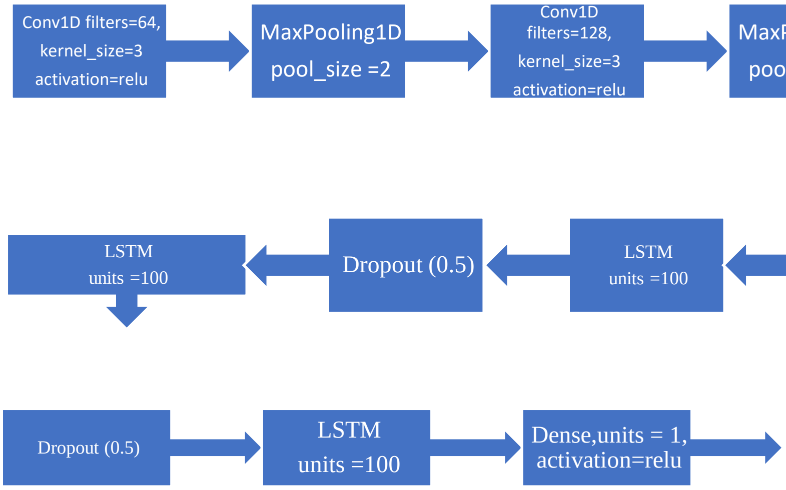

Figure 3: LSTM with CNN model architecture for Processing Data

As in the previous LSTM model, the number of epochs was set to 20 while the batch size was chosen to be 40.

Evaluation

The LSTM model was given the Processing data in Table 8 and the model with the architecture in Figure 3 was given the same dataset and the results were seen with the evaluation metrics. Image 12 shows a graph of the model's predictions for the close data.

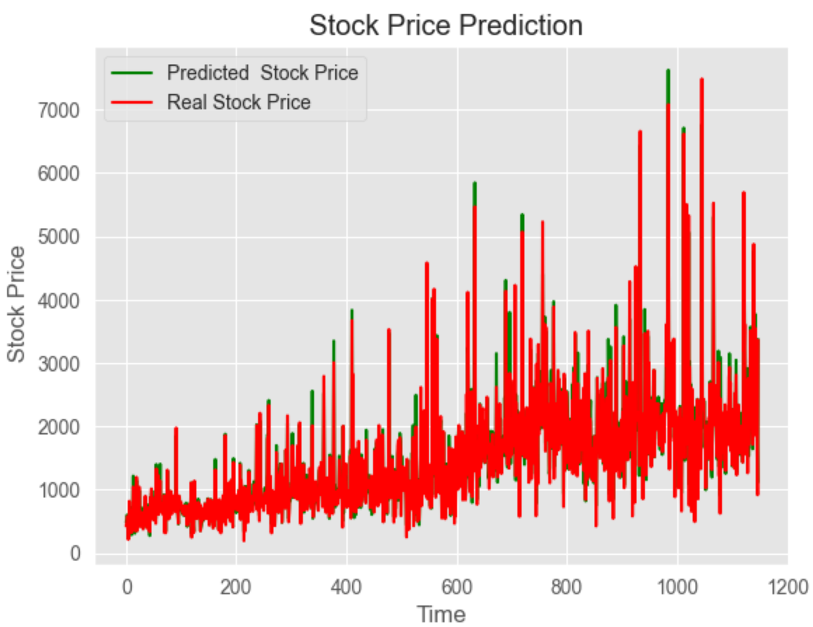

Image 12: Stock price prediction for LSTM with CNN for processing data

In this graph, which cannot be considered as successful as other LSTM models, the close data predictions are close enough to be almost spot-on and are considered to be overfitting. Table 11 shows the evaluation metrics calculated.

Model	MAE	MAPE	R^2
LSTM with CNN for Processed Data	0.074119	71.746713	0.945001

Table 11: Evaluation metrics

As mentioned earlier, while the R^2 value seems to indicate that the model has a good understanding of the data, it is evident from the MAE metric that it predicts a close price that is almost spot on. Not to be outdone, the MAPE value reveals a deviation of up to 71%. The results of the metrics show that the model can make large errors in forecasting in some cases.

Evaluation and Conclusion

The SARIMA model was implemented, statistical operations and calculations were performed and hybrid models such as Prophet, LSTM and LSTM with CNN were tested on the designed dataset. Brief evaluations were made by interpreting the metrics and prediction graphs. Table 12 shows the metric results of the evaluation of the models applied to the Amazon Stock Price 1997 to 2020 dataset.

Model	**MAE**	**MAPE**	**R^2**
SARIMA	363.234236	17.413345	-0.971901
Random Forest	0.546649	0.154870	0.999984
Prophet	737.702311	46.619738	-1.304842
LSTM	76.244591	5.162967	0.970561
LSTM with CNN for Raw Data	47.107138	3.326102	0.985757
LSTM with CNN for Processed Data	0.074119	71.746713	0.945001

Table 12: Results of the evaluation metrics of the implemented models

Table 12 shows that the implemented Random Forest model has the lowest MAE and MAPE values with its calculated metrics, while it has the highest R^2 score. Given the dataset in Table 2 without feature extraction, these results are obtained. Therefore, it is considered that it cannot be considered a generalizable model. In the SARIMA model, only the Close variable was given as input and tested. The detrending process was intended to eliminate trend structures and considering the seasonality of the data, the SARIMA model was found to be an unsuccessful forecasting tool in the direction supported by the calculated metrics.

Prophet model Date and Close column were given as input and it was seen that a worse model was created than SARIMA. When the metrics evaluated together with its predictions were analyzed, it was the system that gave the worst results among the models tested. On the contrary, when deep learning-based models are involved, it has been experienced that when trained with more than one column, which can be considered as complex as it is here, considerably successful models emerge.

When the LSTM model was trained with the data in Table 8, it was seen that the system was able to make significantly better predictions than SARIMA, Prophet and Random Forest when the calculated metrics were analyzed. When the CNN layers are added and the model architecture is slightly improved, the MAE value is quite low, but the MAPE value is quite high, indicating that this process is more unsuccessful than the model without CNN. Finally, when the model is trained on raw data by adding CNN layers to the LSTM model, it is seen that the results are more generalizable, give better results than the others and better interpret the relationships in the data. Therefore, the best model is LSTM with CNN for Raw Data and this model is saved in .keras format.

References

[1] Web link: https://www.unlumenkul.com/blog/hacim-volume-nedir/

[2] Web link: https://finans.mynet.com/haber/detay/borsa/borsada-islem-hacmi-nedir-islem-hacmi-yuksek-olursa-veya-duserse-ne-olur/467087/

[3] Web link: https://tcmb.gov.tr/wps/wcm/connect/5fb0919f-5abd-4175-89ea-55129f551e1e/3b22_i.pdf?MOD=AJPERES

[4] Web link: https://kobikasa.com/bilanco-nedir-bilanco-tarihleri-8099

[5] İŞLETME ARAŞTIRMALARI DERGİSİ JOURNAL OF BUSINESS RESEARCH-TURK 2021, 13(1), 904-911 https://doi.org/10.20491/isarder.2021.1173