Nguyen Ngoc Thai An - 220090

Final Project Report - CS313

TELECOMMUNICATIONS COMPANIES DURING 2019 - 2022

1. Nasdaq Stock Price Prediction (Nasdaq dataset)

1.1. Choosing companies and creating samples

According to Afzal (2024), 11 out of 15 largest telecommunications companies in the US are on the Nasdaq stock market.

TMUS: T-Mobile US, Inc.
CMCSA: Comcast Corporation
CHTR: Charter Communications, Inc.
LBRDA: Liberty Broadband Corporation (Series A)
FYBR: Fibertel SA
SATS: Singapore Airport Terminal Services Ltd
IRDM: Iridium Communications Inc.
CCOI: Cogent Communications Holdings, Inc.
USM: United States Cellular Corporation
GSAT: Globalstar, Inc.
CABO: Cable One, Inc.

The majority of companies' historical data was taken from the provided Nasdaq dataset except for 4 companies TMUS, USM, GSAT, and CABO taken from MarketWatch. Aiming to create a stock open price prediction model between 2019 and 2022, the historical data of all companies were selected from 02/01/2019 to 12/12/2022. This ending date was defined based on the last collected day in the provided Nasdaq dataset. Due to the fact that nearly or more than half of the historical data of FYBR and CABO was missing or mismatched, the above list was reduced to 9 remaining companies.

With FYBR: Its historical data on MakerWatch wasn't recorded from 02/01/2019 to 30/4/2021.
With CABO: Its original data type of all features is object leading to most features losing 2000 data points when changing the type into integer or float.

The modified dataset was created by stacking unshuffled 30-consecutive-day data samples of each company. For 3 models, each data sample will have the same number of features (Low, Open, High, and Close), and 1 label depending on the model's purpose (the next-day open price, the third-day ahead open price, or the next 3-day open price).

1.2. Stock Price Prediction Model

The overall architecture of the stock price prediction model consists of 2 LSTM layers followed by 3 Dense layers to generate the final prediction output(s). The total trainable parameters of those models (presented in Figures 1, 2, and 3) are around 42,000 to 135,000 parameters, which can be considered lightweight and simple models to predict stock price.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 30, 32)	4,736
lstm_1 (LSTM)	(None, 64)	24,832
dense (Dense)	(None, 100)	6,500
dense_1 (Dense)	(None, 50)	5,050
dense_2 (Dense)	(None, 1)	51

Total params: 41,169 (160.82 KB)
Trainable params: 41,169 (160.82 KB)
Non-trainable params: 0 (0.00 B)

Figure 1. Next-Day Open Price Prediction Model

Model: "sequential_2"

Layer (type)	Output Shape	Param #
lstm_4 (LSTM)	(None, 30, 64)	17,664
lstm_5 (LSTM)	(None, 128)	98,816
dense_6 (Dense)	(None, 128)	16,512
dense_7 (Dense)	(None, 64)	8,256
dense_8 (Dense)	(None, 1)	65

Total params: 141,313 (552.00 KB)
Trainable params: 141,313 (552.00 KB)
Non-trainable params: 0 (0.00 B)

Figure 2. Third-Day Ahead Open Price Prediction Model

Model: "sequential_4"

Layer (type)	Output Shape	Param #
lstm_8 (LSTM)	(None, 30, 128)	68,096
lstm_9 (LSTM)	(None, 64)	49,408
dense_12 (Dense)	(None, 128)	8,320
dense_13 (Dense)	(None, 64)	8,256
dense_14 (Dense)	(None, 3)	195

Total params: 134,275 (524.51 KB)
Trainable params: 134,275 (524.51 KB)
Non-trainable params: 0 (0.00 B)

Figure 3. Three-Day Open Price Prediction Model

Since the Nasdaq dataset is from a mature market with a long history of trading and less noise, the pattern is more predictable and less volatile. Moreover, being the largest telecommunications companies in the US, these 9 companies are likely to have similar patterns. Therefore, building a model to predict the open price for this sub-dataset or the entire Nasdaq dataset can be simple and less-resourced but comes with great performance. This can be inferred by looking at the MSE loss and the total parameters of these 3 models in Table 1.

	Next-Day Open Price Prediction Model	Third-Day Ahead Open Price Prediction Model	Three-Day Open Price Prediction Model
Total parameters	41,169	141,313	134,275
Performance (MSE loss)	0.03699983	0.06567975	0.036901742

Table 1. Performance and total parameters of 3 models

Despite being basic and less-resourced models, the functions inside them are complex enough to give more accurate outputs despite having the same MSE loss.

2. Vietnam Stock Price Prediction (Vietnam dataset)

2.1. Choosing companies and creating samples

Based on statistics on SSI iBoard, FOX (FPT Telecom Joint Stock Company) and VGI (Viettel Global Investment JSC) are 2 companies having the highest market capitalization and outstanding shares in the telecommunications industry. With the idea of choosing the largest companies in an industry like with the Nasdaq dataset, this model is designed to look at the overall trend and make open price predictions of those companies.

With the purpose of designing the model in the range of 2019 to 2022, the historical data of VGI and FOX were selected from 16/04/2019 to 30/12/2022. Starting date 16/04/2019 was chosen due to many Vietnamese companies holding annual general meetings in April, followed by dividend payments in the subsequent quarter. These dividend distributions can cause a temporary downward adjustment in share price. Therefore, choosing the starting time in the middle of April helped minimize the potential impact of dividend distributions on stock prices in 2019.

The modified dataset was created by stacking unshuffled 30-consecutive-day data samples of 2 companies. For 3 models, each data sample will have the same number of features (Low, Open, High, and Close), and 1 label depending on the model's purpose (the next-day open price, the seventh-day ahead open price, or the next 7-day open price).

2.2. Stock Price Prediction Model

The overall architecture of the stock price prediction model consists of BiLSTM + Dropout and LSTM + Dropout layers followed by 3 Dense layers to generate the final prediction output(s). The total trainable parameters of those models (presented in Figures 4, 5, and 6) are in the range of around 2 million to 8 million parameters. The performance of these 3 models are shown in Table 2.

Model: "sequential_26"

Layer (type)	Output Shape	Param #
bidirectional_51 (Bidirectional)	(None, 30, 512)	534,528
dropout_82 (Dropout)	(None, 30, 512)	0
bidirectional_52 (Bidirectional)	(None, 30, 256)	656,384
dropout_83 (Dropout)	(None, 30, 256)	0
lstm_155 (LSTM)	(None, 256)	525,312
dropout_84 (Dropout)	(None, 256)	0
dense_78 (Dense)	(None, 64)	16,448
dense_79 (Dense)	(None, 32)	2,080
dense_80 (Dense)	(None, 1)	33

Total params: 1,734,785 (6.62 MB)
Trainable params: 1,734,785 (6.62 MB)
Non-trainable params: 0 (0.00 B)

Figure 4. Next-Day Open Price Prediction Model

Model: "sequential_2"

Layer (type)	Output Shape	Param #
bidirectional_6 (Bidirectional)	(None, 30, 1024)	2,117,632
dropout_10 (Dropout)	(None, 30, 1024)	0
bidirectional_7 (Bidirectional)	(None, 30, 256)	1,180,672
dropout_11 (Dropout)	(None, 30, 256)	0
bidirectional_8 (Bidirectional)	(None, 30, 256)	394,240
dropout_12 (Dropout)	(None, 30, 256)	0
lstm_22 (LSTM)	(None, 30, 256)	525,312
dropout_13 (Dropout)	(None, 30, 256)	0
lstm_23 (LSTM)	(None, 256)	525,312
dropout_14 (Dropout)	(None, 256)	0
dense_6 (Dense)	(None, 128)	32,896
dense_7 (Dense)	(None, 64)	8,256
dense_8 (Dense)	(None, 1)	65

Total params: 4,784,385 (18.25 MB)
Trainable params: 4,784,385 (18.25 MB)
Non-trainable params: 0 (0.00 B)

Figure 5. Seventh-Day Open Price Prediction Model

Model: "sequential_24"

Layer (type)	Output Shape	Param #
bidirectional_73 (Bidirectional)	(None, 30, 1024)	2,117,632
dropout_114 (Dropout)	(None, 30, 1024)	0
bidirectional_74 (Bidirectional)	(None, 30, 512)	2,623,488
dropout_115 (Dropout)	(None, 30, 512)	0
bidirectional_75 (Bidirectional)	(None, 30, 512)	1,574,912
dropout_116 (Dropout)	(None, 30, 512)	0
bidirectional_76 (Bidirectional)	(None, 30, 256)	656,384
lstm_202 (LSTM)	(None, 30, 256)	525,312
dropout_117 (Dropout)	(None, 30, 256)	0
lstm_203 (LSTM)	(None, 256)	525,312
dense_72 (Dense)	(None, 128)	32,896
dense_73 (Dense)	(None, 64)	8,256
dense_74 (Dense)	(None, 7)	455

Total params: 8,064,647 (30.76 MB)
Trainable params: 8,064,647 (30.76 MB)
Non-trainable params: 0 (0.00 B)

Figure 6. Seven-Day Open Price Prediction Model

	Next-Day Open Price Prediction Model	**Seventh-Day Ahead Open Price Prediction Model**	**Seven-Day Open Price Prediction Model**
Total parameters	1,734,785	4,784,385	8,064,647
Performance (MSE loss)	0.037258197	0.1380079	0.06730984

Table 2. Performance and total parameters of 3 models

Compared to 3 Nasdaq models, the Vietnam stock open price prediction model is more complex and time-consuming training time. Since the Vietnam stock market has a significant amount of short-term trading, it increases the volatility and unpredictability of stock prices and the complexity of Vietnam's prediction models. In addition, the prediction models can be proven to be more complex from observations while training models.

	Overall architecture	**Observation**	**Solution**
1	Multiple LSTM layers (unit = 64, 128, or 256)	The validation loss was always greater than the training loss → Underfit	Add more LSTM layers with different units (128, 256, or 512) and some Dropout layers and adjust Dense layers.
2	Multiple LSTM layers (unit = 128, 256, or 512)	MSE loss was in an acceptable range (0.06 -	Change LSTM layers into multiple BiLSTM layers

	followed by a Dropout layer	0.15). However, the model struggled to capture the real price fluctuations. The predicted price trend aligns with the overall market direction, but fails to predict short-term price movements.	with different units (64, 128, or 256)
3	Multiple BiLSTM layers with different units (64, 128, or 256) followed by a Dropout layer	The model's performance was acceptable, but the predicted price line could more accurately follow the actual price trend.	Using the combination of multiple BiLSTM and LSTM layers with different units (128, 256, or 512) to increase complexity while trying to balance the run time.
4	Combination of multiple BiLSTM and LSTM layers with different units (128, 256, or 512) followed by a Dropout layer	The model's performance is acceptable, and the next-day price prediction model works well. The predicted line has a similar shape to the actual price line.	

Table 3. Training observations

In the training process, there are cases where two models with the same mean squared error can produce vastly different prediction plots. Basic models (Table 3, rows 2-3) are limited to classifying price movements (up or down), whereas intricate models (row 4) offer more accurate price predictions.

3. Trading signal identification for Vietnam market

The CNN-based trading signal identification model (Figure 7), whose other name is the Buy, Sell, and Hold classification model, is trained on the 30-day samples of 2 telecommunications companies such as VGI and FOX. With 4 features (Open, High, Low, Close) and a classification label, the model is designed to classify the current price into should buy (0), sell (1), or hold (2).

Ways to create the label for each current price:

1. *Define a threshold: This threshold will act as a hyperparameter to change the range of values (prices) that are considered significant enough to trigger a buy or sell signal.*
2. *At each current open price, calculate the gap between the next-day and the current price and assign a label based on the magnitude and direction of the gap relative to a predefined threshold.*
   - *If the next-day price is significantly higher than the current price (exceeding the upper threshold), a "Buy" label is assigned.*

- *If the next-day price is significantly lower (below the lower threshold), a "Sell" label is assigned.*
- *If the price movement is within the threshold range, a "Hold" label is assigned.*

Instead of relying on the closing price, as commonly used in many investment strategies, this classification model utilizes the opening price to align with its short-term investment focus.

Model: "sequential_10"

Layer (type)	Output Shape	Param #
conv1d_21 (Conv1D)	(None, 30, 128)	1,664
conv1d_22 (Conv1D)	(None, 30, 128)	49,280
max_pooling1d_9 (MaxPooling1D)	(None, 15, 128)	0
dropout_34 (Dropout)	(None, 15, 128)	0
conv1d_23 (Conv1D)	(None, 15, 128)	49,280
conv1d_24 (Conv1D)	(None, 15, 128)	49,280
batch_normalization_6 (BatchNormalization)	(None, 15, 128)	512
conv1d_25 (Conv1D)	(None, 15, 32)	12,320
max_pooling1d_10 (MaxPooling1D)	(None, 7, 32)	0
dropout_35 (Dropout)	(None, 7, 32)	0
conv1d_26 (Conv1D)	(None, 7, 64)	6,208
conv1d_27 (Conv1D)	(None, 7, 64)	12,352
batch_normalization_7 (BatchNormalization)	(None, 7, 64)	256
max_pooling1d_11 (MaxPooling1D)	(None, 3, 64)	0
dropout_36 (Dropout)	(None, 3, 64)	0
dense_30 (Dense)	(None, 3, 1000)	65,000
dense_31 (Dense)	(None, 3, 500)	500,500
dense_32 (Dense)	(None, 3, 3)	1,503

Total params: 748,155 (2.85 MB)
Trainable params: 747,771 (2.85 MB)
Non-trainable params: 384 (1.50 KB)

Figure 7. Buy, Sell, and Hold classification model

Changing the hyperparameter threshold will directly affect the model's performance since it is one of the factors deciding the multiple classification.

4. Portfolio composition, risk management and portfolio optimization for Vietnam market

According to the SSI iBoard, there are 8 companies in the telecommunications industry list.

ABC: VMG Media JSC
FOX: FPT Telecom Joint Stock Company

MFS: Mobifone Service JSC
PAI: Petroleum Information Technology Telecom and Automation JSC

PIA: Petrolimex Information Technology And Telecommunication JSC
PTP: Post Joint Stock Company
TTN: Viet Nam Technology & Telecommunication JSC
VGI: Viettel Global Investment JSC

To ensure a fair comparison, data for all companies was restricted to the period of 16/04/2019 to 30/12/2022, as MFS's historical data begins on 16/04/2019. The 8 charts in Figure 8 display the historical price and volume of 8 telecommunications companies during this period. Each chart includes price trends and trading volumes, providing insights into the performance and volatility of each company's stock. This information is relevant for making investment decisions based on historical performance and risk assessment.

Figure 8. Historical Price and Volume Chart of each company during 2019 - 2022
(Full-size version: Appendix 1)

From the charts, several inferences can be drawn based on three factors: profit, risk, and the maximization of profit while minimizing risk.

The risk scoring methodology will assess companies based on 3 factors: volatility, volume, and historical performance. Volatility measures the frequency and magnitude of price changes. High trading volume can indicate higher liquidity but may also signal *increased risk if accompanied by price volatility. Consistent performance with fewer fluctuations is considered less risky.*

Given the investment objective of maximizing returns, companies such as FOX, TTN, and ABC present compelling opportunities. VGI, characterized by its high volatility, offers potential rewards but also carries elevated risk. To mitigate investment risk, it is advisable to avoid companies such as VGI, PAI, and PIA, which exhibit higher volatility and, consequently, elevated risk profiles. To achieve a balanced portfolio that aims to maximize returns while minimizing risk, investing in companies like FOX, ABC, and MFS could be a strategic move. Their historical performance suggests a more stable trajectory compared to the highly fluctuating VGI, making them potentially attractive options for risk-conscious investors.

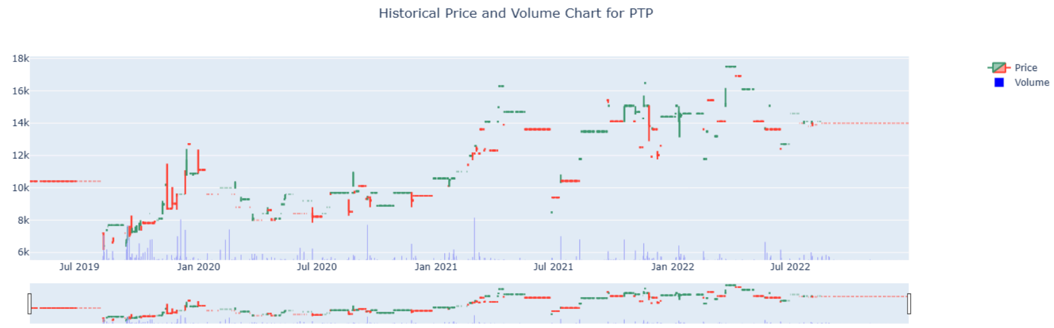
References

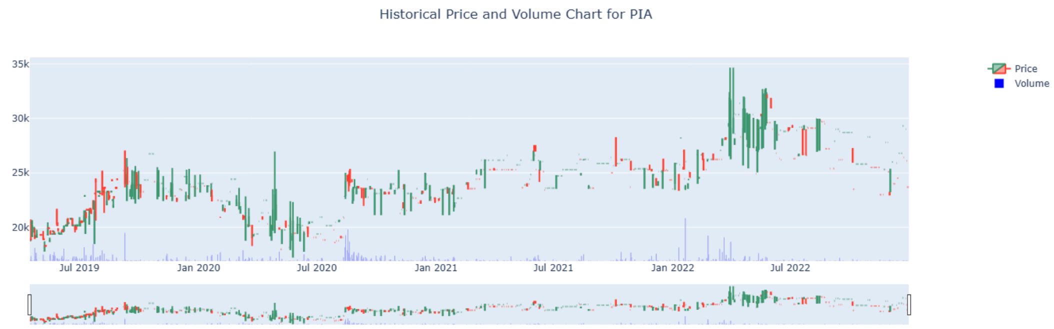
Afzal, M. (2024, March 13). *15 Biggest Telecom Companies in the US*. Yahoo!Finance. https://finance.yahoo.com/news/15-biggest-telecom-companies-us-224945747.html

Bảng giá - SSI iBoard. (n.d.). SSI iBoard. https://iboard.ssi.com.vn/

Appendix

1. Historical Price and Volume Chart of each company during 2019 - 2022





Historical Price and Volume Chart for PAI

Historical Price and Volume Chart for VGI

Historical Price and Volume Chart for ABC

Historical Price and Volume Chart for FOX