VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY
INFORMATION SYSTEM FACULTY

FINAL PROJECT

DATA ANALYTICS IN BUSINESS

TOPIC

ANALYSIS OF FPT CORPORATION'S BUSINESS PERFORMANCE AND STOCK PRICE FORECAST

Lecturer: **Dr. Trần Văn Hải Triều**

Class: **IS403.O12.TMCL**

Student group:

Nguyễn Minh Hà	21520216
Thiều Vĩnh Tiến	21521533
Trần Anh Thy	21521517
Trương Phương Thy	21520477
Nguyễn Lê Won Bin	21521870
Vũ Đình Tuấn Kiệt	21521039

Ho Chi Minh City, 14th November, 2023

ACKNOWLEDGEMENT

Firstly, we would like to express our sincerest gratitude to Professor Tran Van Hai Trieu, lecturer of the Information Systems Faculty at the University of Information Technology. The professor has been wholeheartedly supportive, providing direct guidance and instructions throughout our research and study process. During our time studying under him, we not only acquired valuable knowledge but also developed a serious and effective work ethic and research attitude. These qualities are essential for our future academic and professional endeavors.

However, we acknowledge that our technical knowledge is still limited, and each team member lacks practical experience. Therefore, the content of this report may have certain shortcomings. We sincerely hope to receive your feedback and additional guidance, Professor, to further improve our expertise. This will enable us to enhance the quality of this report and support the execution of future research projects.

Ho Chi Minh City, November, 2023

Team 3

LECTURER'S FEEDBACK

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

……

TABLE OF CONTENT

I. INTRODUCTION

1.1 Reason for choosing the topic:

Currently, the strong development of the Industrial Revolution 5.0 has been comprehensively changing all aspects of the world's economies in general and Vietnam in particular. International economic integration creates conditions for the country to develop, however, this also makes competition in economic fields become more fierce. FPT Corporation is a leading corporation in Vietnam and is a typical witness of efforts to overcome all difficulties to rise to the ranks of leading corporations in Vietnam.

In addition, highlighting the achievements during FPT Corporation's operational journey, such as accomplishments attained in a particular year and the challenges faced during the COVID-19 period, can bring significant benefits to stakeholders interested in investment and business development. These stakeholders may include the government, financial institutions, investors, and companies seeking financial expertise. It enables them to gain valuable insights and make more informed and comprehensive decisions regarding business activities, trading, and investment in the economic and financial markets.

1.2 Research objectives:

The objective of this study is to apply the knowledge from the “Data Analytics in business” course to analyze, evaluate, and provide recommendations for improving the operational efficiency of the corporation, as well as identifying the highlights in its business activities. The research will focus on collecting and analyzing financial and business data, as well as assessing both internal and external impacts on the corporation. Furthermore, it will evaluate the operations and provide recommendations for enhancing operational efficiency during the period from 2004 to 2023. Additionally, the study will forecast the operational efficiency of the corporation for the next three years.

1.3 Subject and scope of research:

The research article will focus on analyzing and evaluating the performance of FPT Corporation in Vietnam. The variables that will be used in the study include:

- Scope: FPT Corporation is listed on stock exchanges, financial reports are collected from the website https://vietstock.vn/.
- Time: From 2004 to 2023.

1.4 Research methods:

To achieve the research objective of "Evaluating the business efficiency of FPT Corporation from 2004 to 2022," team will utilize methods in time series analysis, specifically qualitative and quantitative methods, combined with prominent models such as ARIMA and linear regression.

The qualitative methods will be used to access and analyze theoretical bases and empirical evidence, design a research model, discuss research results and make related suggestions and recommendations.

The quantitative method will be used to determine the research outcomes through methods including descriptive statistics, correlation analysis, and regression analysis. For the quantitative method, team will utilize financial and business indicators such as Return on

Average Assets (ROAA), GDP, attrition rate, financial leverage, current ratio, and debt ratio. These indicators will be determined based on the results of the models run to evaluate the operational efficiency of FPT Corporation in the mentioned years.

1.5 Novelty of the topic:

Currently, there have been quite a few researches interested in the issue of business performance in both qualitative and quantitative directions, but the use of running ARIMA and LSTM models is not common.

Under the impact of Covid-19 on the economy, some industries are almost completely paralyzed. Business activities are interconnected with all industries in the economy, so the impact of the pandemic will be more difficult to predict. Team has researched and used qualitative and quantitative methods to analyze and evaluate the current state of business operations of enterprises.

The team conducted an analysis of FPT's basic business activities through business performance results and indicators such as FPT's operational efficiency. From there, empirical research evidence can be added, a reference information channel for performance assessment, helping analysts consider stock investment. Find effective ways to manage FPT's business activities to learn and implement for your business.

II. OVERVIEW OF RESEARCH OBJECTS

2.1 Overview:

Figure 1. FPT logo

FPT Group is one of the largest information technology service companies in Vietnam. FPT is always constantly innovating and creating with the desire to bring customers the most optimal technology products, solutions and services.

FPT has nearly 29,000 employees, including more than 17,000 information technology engineers, programmers, technology experts, and office systems in 22 countries and territories outside Vietnam. FPT is also a leading enterprise in the fields of Software export, System integration; Software Development; IT services. Review important milestones in the history of the formation and development of FPT Joint Stock Company.

2.2 History of formation:

- September 13, 1988: established with the first name Food Processing Joint Stock Company.
- October 27, 1990: Changed name to Technology Investment and Development Company.
- April 2002: Technology Investment and Development Company became a joint stock company.
- January 1, 2007: Established FPT Retail Company Limited
- March 13, 2007: Established FPT Advertising Joint Stock Company and Asia Pacific Software Company located in Singapore.
- 2014: FPT acquired the IT company RWE IT Slovakia (a member of the European Energy Group, RWE).
- September 12, 2017: Cooperated with investor Synnex Technology International Corporation.
- 2019: Achieved total revenue of 27,717 billion VND, growing by 19.8%.

Over the past 32 years, FPT has pioneered the development of software that brings technology to life, strongly promoting education and modernizing the nation's key economic sectors. FPT Group is spread across countries with a system of 46 offices located in 22 countries around the world and telecommunications infrastructure covering 59/63 provinces and cities in Vietnam. The distribution of network coverage everywhere helps FPT provide services/solutions to customers globally in the most effective way.

Figure 2. Overview FPT

2.3 FPT Culture:

"Tôn Đổi Đồng – Chí Gương Sáng"

- "Tôn Đổi Đồng": "Respect for individuals – Spirit of innovation – Team spirit": towards sharing among FPT members.
- "Chí Gương Sáng": "Dignity – Exemplary – Insight": refers to the necessary qualities of FPT leaders.

Orientation 2019 - 2021

In the context of digital technology 4.0, FPT Corporation aims to bring customers the best experiences by being a pioneer in the digital trend of improving itself through building, applying and converting. This is the basic goal towards a multi-sector business model. FPT will focus on two main goals:

- Comprehensive digital transformation;
- Investing in building solid foundations.

In 2018, FPT had a turning point in comprehensive digital transformation and outstanding development. This is the premise for future development goals as follows:

- Become the world's leading digital enterprise in digital transformation services;
- Providing solutions and services to customers in Vietnam and around the world to help them transform into digital businesses;
- Participate in building and providing smart technology services for Government, Health, Transportation, Education, Energy, Telecommunications, and Manufacturing;
- Meet customer needs with just one touch.

Figure 3. FPT's fields

2.4 Model of a member company of FPT group:

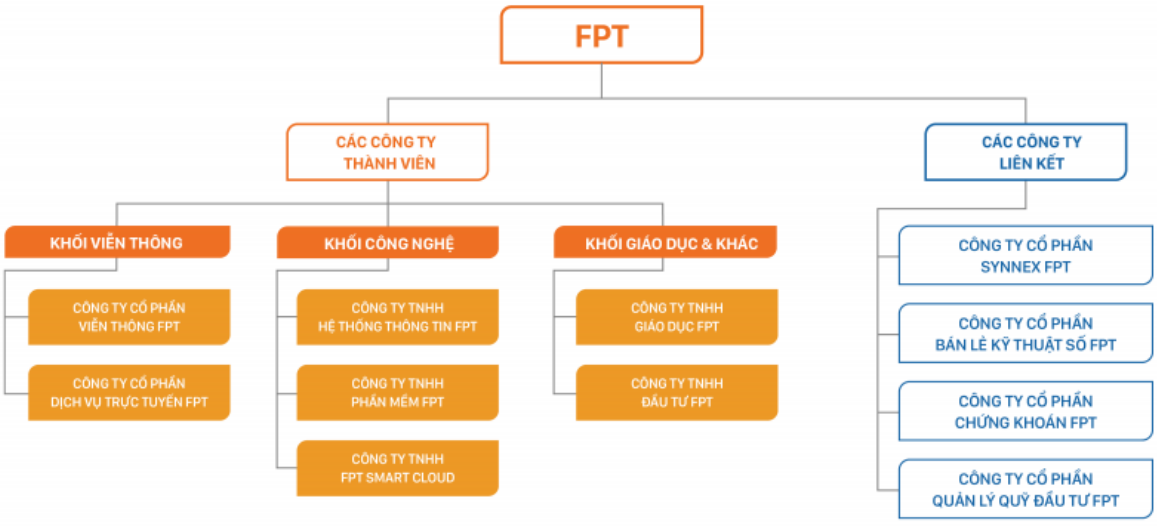


Figure 4. Model of a member company of FPT group

III. Theoretical basis

3.1 Definition of variables:

3.1.1 ROAA:

ROAA - Return on Average Assets represents the rate of return on average assets of a business.

This index is often used when evaluating the profit potential from corporate assets. In other words, ROAA will reflect the efficiency of using a business's assets, it shows everyone how profitable the business is when relying on assets.

ROAA is preferred by many experts because normally, the assets that a business owns will fluctuate within a period. Instead of using assets at the beginning of the period or assets at the end of the period to calculate, the ROAA index is calculated based on average assets, so it gives results with higher accuracy, closest to reality.

3.1.2 GDP:

GDP - Gross Domestic Product. This is a consumption index that measures the total market value of all goods and services produced in a country during a certain period.

GDP is the indicator that measures total market value. By using market prices, many types of products are combined into a single indicator of value. Market prices represent the amount consumers are willing to pay for different goods, so they accurately reflect the value of these goods.

GDP only represents goods that are legally produced and sold on the market. GDP does not include products produced and sold illegally in the underground economy. GDP reflects the value of production during a specific period of time, usually a year or a quarter.

3.1.3 Exchange rate:

An exchange rate (also known as foreign currency rate) is the rate of exchange between two different currencies. This exchange rate indicates the value of one currency compared to another. Typically, exchange rates are expressed by comparing the value of one currency in another, for example: USD/JPY (US dollar to Japanese yen exchange rate), EUR/USD (US dollar to Japanese yen exchange rate). euro price versus US dollar), and many other currency pairs.

Exchange rates play an important role in international trade, investment, and global finance. It affects the value of goods and services traded in international markets, and it can also affect the profits of internationally active companies. This makes the exchange rate an important indicator to watch and monitor in the global economy.

Exchange rates can fluctuate daily under the influence of many factors, including fluctuations in financial markets, politics, economies of countries, and even world situations. Institutions such as central banks can intervene to stabilize or influence exchange rates through monetary policy and other measures.

3.1.5 Debt ratio:

Is a financial ratio used to measure capacity and manage debt based on the total assets that a business owns. Also known as Debt to Total Assets Ratio.

Debt ratio indicates the percentage of a company's total assets that are financed by debt. A low debt ratio can indicate ineffective use of debt, while a high debt ratio indicates a large debt burden, which can lead to insolvency. Specific: A debt to total assets ratio greater than 1 indicates that a significant portion of assets is financed by debt. In other words, the company has more debt than assets, indicating that the business's Equity is negative. This is a situation that no business wants to encounter because accumulated losses over many years have exceeded the owner's contributed capital, also known as "negative equity".

3.2 Theoretical basis:

GDP positively affects the business activities of the manufacturing and service industries.

Lee (2014) investigated the relationship between macroeconomic factors such as (Economic Growth Rate, Inflation Rate) and profitability in Taiwan's Property Liability Insurance Industry. The results show that the economic growth rate has a significant influence on profitability in the operating ratio model but has an insignificant influence on profitability in the ROA model. However, an insurance company can still be considered a service business company. This is a reliable facility.

Nguyen, T.T.C., Le, A.T.H. and Nguyen, C.V. (2023), "Internal factors affecting the financial performance of an organization's business processes" investigated, researched and proved that financial leverage has a negative relationship with financial business activities.

Siminica et al. (2011) conducted a study to determine the factors that affect the return on total assets of companies by collecting data from 40 companies listed on the Bucharest Stock Exchange in Romania for four years from 2007 to 2010. The study covers two years of economic growth (2007 and 2008) and two years of economic growth and recession (2009 and 2010) in Romania. The dependent variable is a return on assets (ROA) and the independent variables include fixed assets ratio, financial stability ratio, debt ratio, financial leverage, employment ratio, current ratio, quick ratio, working capital, fixed asset financing ratio, investment scope, operating capital requirements, net working capital turnover, collection period Average, effective use of assets. After 4 years of analysis, the author has proposed a statistical model linking the dependent variable and independent variable. The results show that the financial performance of Romanian companies is going down due to the impact of the economic crisis. However, before the crisis, financial performance was significantly affected by financial structure. This is similar to the results of previous researchers such as Master & Yong (2014), Zeitun and Tian (2007), Nagy (2009).

IV. RESEARCH METHODS

4.1 Hypotheses:

Based on the theoretical basis of the previous section, the group has the following hypotheses:

- H1: GDP index has a positive impact on FPT's business performance.

- H2: Attrition rate has a negative impact on FPT's business performance.
- H3: Current ratio has a positive impact on FPT's performance.
- H4: Debt ratio has a positive impact on FPT's performance.

4.2 Research data:

- Research data is collected from FPT Corporation's financial statements by year from 2004 to 2022.
- FPT's operating situation in the period 2004-2022.
- The data on GDP and attrition ratio are taken from the WorldBank website.

4.3 Research models and methods:

4.3.1 ARIMA:

The ARIMA (Autoregressive Integrated Moving Average) model is a time series analysis method. The ARIMA model combines the autoregressive model (AR - Autoregressive), the moving average model (MA - Moving Average) and the integration (I - Integrated) to handle nonlinear time series.

In the ARIMA model, "AR" refers to an autoregressive model, meaning that the current value of the time series depends on the previous values of the time series. "MA" refers to the moving average model, meaning that the current value of a time series depends on the difference between the current value and the average value of the time series. "I" refers to integration, that is, the use of differences between the values of a time series to create a stationary series, where the values of the time series are independent of time. The ARIMA model is used to forecast the future value of a time series based on observed past values of the time series.

The parameters in the ARIMA model include: p (lag of the autoregressive model), d (level of integration) and q (lag of the moving average model). These values are determined based on considering the lag and nonlinear nature of the time series.

4.3.2 Linear Regression:

Linear Regression is one of the most basic and popular algorithms of Supervised Learning, where the prediction output is continuous. This algorithm is suitable for predicting output values that are continuous quantities such as sales or prices instead of trying to classify them into discrete quantities such as color and material of clothes, or determining whether the object in a photo is a cat or a dog, etc.

In Linear Regression, we will encounter two types of problems: Univariate Regression and Multivariate Regression. To simplify the algorithm, we will carefully learn and analyze the mathematics of univariate regression. So what is univariate linear regression?

Univariate Linear Regression is the relationship between two continuous variables on the x-axis and on the y-axis. The univariate linear regression equation has the form of a straight line equation $y = ax + b$ with x being the independent variable and y being the variable that depends on x. For Multivariate Linear Regression, you can simply understand that there will be many independent variables $x_1, x_2, \ldots, x_n$ and many coefficients $a_1, a_2, \ldots, a_n$ instead of just a single variable x.

4.3.3 LSTM:

LSTM - Long Short-Term Memory, is a special type of neural network architecture designed to solve the problem of vanishing gradients during the training of recurrent neural networks

(RNN). RNN is a type of neural network commonly used in processing sequence data, but it often has problems when processing long sequences or when there are distant dependencies between elements of the sequence.

LSTM is introduced to overcome this drawback by making it possible for information to be maintained over time without the problem of vanishing gradients. This makes LSTM more efficient at handling long strings and handling more complex situations. The basic structure of an LSTM unit consists of three main gates:

- Forget Gate: Determines which information should be discarded from the previous state.
- Input Gate: Decide what information should be added to the previous state.
- Output Gate: Determines the new state of the LSTM unit and its output.

These gates help the LSTM control how information is passed and retained across time steps, increasing the model's learning ability and performance in tasks involving sequence data. LSTMs have become an important part of many machine learning applications, especially in the fields of natural language processing and time series prediction.

4.3.4 Descriptive statistics:

Descriptive statistics is the process of using statistical methods and techniques to summarize and represent data in an overview and easy-to-understand manner. The main goal of descriptive statistics is to provide information and understanding of data without making complex conclusions or predictions. Descriptive statistics help us understand key characteristics of data, such as its mean, variance, distribution, and underlying trend.

V. ANALYSIS OF FPT'S PERFORMANCE FROM 2004 - 2022

- Selected variable: ROAA.
- Reasons to choose ROAA: This index is often used when evaluating the profit potential from business assets. In other words, ROAA will reflect the efficiency of using a business's assets, it shows everyone how profitable the business is when relying on assets.

```
df.set_index('Giai đoạn', inplace = True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 19 entries, 2004-01-01 to 2022-01-01
Data columns (total 10 columns):
 #   Column                                          Non-Null Count  Dtype
---  ------                                          --------------  -----
 0   Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)  19 non-null     float64
```

Figure 5. Prepare data

```python
# Tính giá trị trung bình ROAA
mean_roaa = np.mean(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)'])
print("Giá trị trung bình ROAA:", mean_roaa)

# Tính trung vị ROAA
median_roaa = np.median(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)'])
print("Trung vị ROAA:", median_roaa)

# Tính phạm vi ROAA
range_roaa = np.max(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)']) - np.min(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)'])
print("Phạm vi ROAA:", range_roaa)

# Tính độ lệch chuẩn ROAA
std_deviation_roaa = np.std(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)'])
print("Độ lệch chuẩn ROAA:", std_deviation_roaa)
# Tính min, max của ROAA
min_roaa = np.min(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)'])
print("Giá trị nhỏ nhất của ROAA: ", min_roaa)
max_roaa = np.max(df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)'])
print("Giá trị lớn nhất của ROAA: ",max_roaa)
data = df['Tỷ suất sinh lợi trên tổng tài sản bình quân (ROAA)']
# Tính skewness
print("Giá trị skewness: ", data.skew())
# Tính Standard Error of Mean
print("Giá trị standard Error of Mean: ", data.sem())
# Tính Kurtosis
```

Figure 6. Calculate statistical values

```
Mean of ROAA: 11.747894736842106
Median of ROAA: 10.57
Range of ROAA: 14.959999999999997
standard Deviation of ROAA: 3.60784755621959
Min of ROAA:  7.12
Max of ROAA:  22.08
skewness:  1.3406416101369873
standard Error of Mean:  0.8503778241633951
Kurtosis:  1.9823813626035212
```

Figure 7. Results

- **General identify:**

Average ROAA value is 11.75%: This shows that the average business from 2004 to 2022 is achieving a relatively high return on their average assets. This shows that they use their assets to make a profit in an effective way.

Variance: 13.74 which is relatively large, this shows that FPT's ROAA has experienced a relative level of fluctuation during the above period. This can indicate a change in the company's financial performance, which may be due to factors such as market fluctuations, changes in business strategy, or the influence of economic and political factors.

Skewness: 1.34 > 0 (skewed right) has many years with ROAA value greater than the average value compared to the normal distribution. Reflects success in some special projects or business activities in the past. At the same time, we see that FPT focuses on goals and creates long-term values, in addition, business operations are managed effectively and the ability to manage risks in the best way.

Kurtosis: 1.98 < 3 we see not many outliers. FPT's ROAA distribution has average concentration and does not have large fluctuations in profits. This can be considered as the stability of the business in generating profits.

Based on the following chart, we divide into 3 stages:

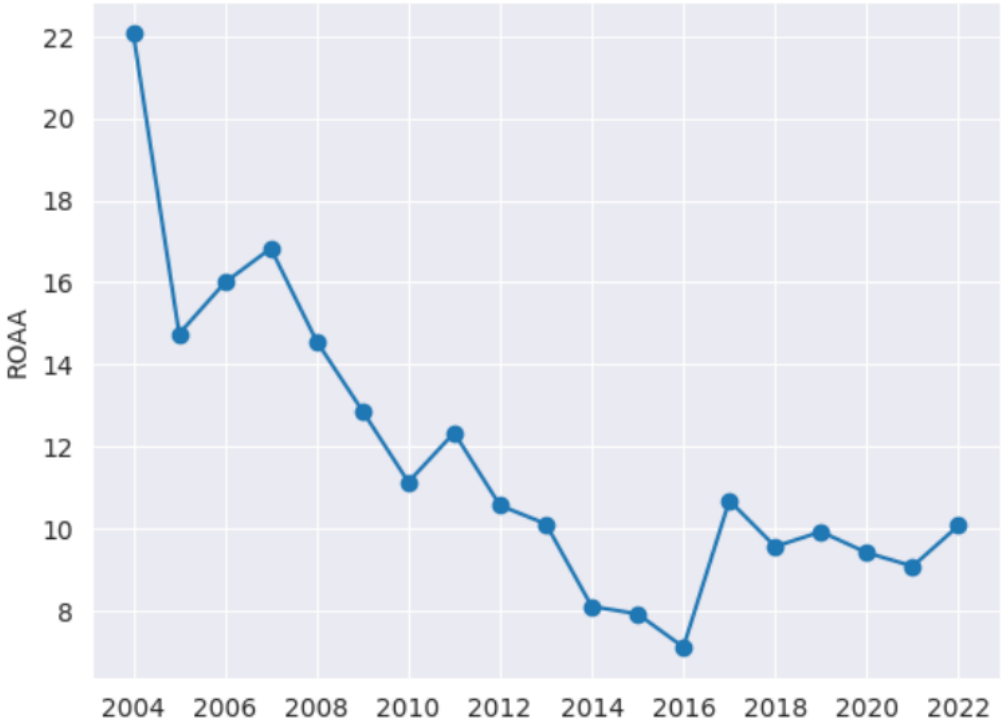


Figure 8. Chart of ROAA changes over time

5.1 Decreased sharply from the beginning of the period:

There was a sharp decline in the first years, from a high in 2004, falling significantly thereafter.

The year 2004 (The year with the highest ROAA value): looking at the financial statements, we see that only revenue and profits from other sources did not change much compared to the following years, but in the basic profit per share. incredibly high (92,130 billion VND), far exceeding the values in the following years. This is a typical reason why FPT business efficiency has increased so much. FPT and Microsoft have officially signed a memorandum of understanding on developing information technology applications in the Asia-Pacific region. Immediately after these events, FPT stock immediately had a terrible fever in the OTC market, FPT's stock price exceeded 14 times its original par value (100,000 VND/stock).

In 2006, FPT issued additional stocks to two strategic investors, receiving an investment of 36.5 million USD from Texas Pacific Group and Intel Capital. This increase in capital may have increased the company's total assets, which is good for long-term growth, but has reduced its return on assets (ROAA) in the short term. In addition, in November 2006, FPT also signed a strategic alliance with Microsoft. These strategic moves, along with internal restructuring and investments in new areas, may have impacted short-term financial performance, leading to a decline in ROAA.

Kết quả kinh doanh - FPT	2004	2005	2006	2007	2008	2009
Giai đoạn	01/01-31/12	01/01-31/12	01/01-31/12	01/01-31/12	01/01-31/12	01/01-31/12
Hợp nhất	Đơn lẻ	Đơn lẻ	Hợp nhất	Hợp nhất	Hợp nhất	Hợp nhất
Kiểm toán	Kiểm toán	Kiểm toán	Kiểm toán	Kiểm toán	Kiểm toán	Kiểm toán
Công ty kiểm toán		AASC	AASC	KPMG	KPMG	Deloitte
Ý kiến kiểm toán			Chấp nhận từng phần - Ngoại trừ	Chấp nhận từng phần - Ngoại trừ	Chấp nhận toàn phần	Chấp nhận toàn phần
Doanh thu thuần về bán hàng và cung cấp dịch vụ	**8,735.00**	**14,101.00**	**21,400.00**	**13,499.00**	**16,382.00**	**18,404.00**
Giá vốn hàng bán	8,197.00	13,180.00	20,049.00	11,537.00	13,403.00	14,719.00
Lợi nhuận gộp về bán hàng và cung cấp dịch vụ	**538.00**	**921.00**	**1,351.00**	**1,961.00**	**2,978.00**	**3,685.00**
Doanh thu hoạt động tài chính	4.00	7.00	11.00	49.00	197.00	188.00
Chi phí tài chính	28.00	40.00	74.00	72.00	495.00	445.00
Chi phí bán hàng	158.00	284.00	358.00	385.00	527.00	527.00
Chi phí quản lý doanh nghiệp	164.00	269.00	436.00	600.00	963.00	1,306.00
Lợi nhuận thuần từ hoạt động kinh doanh	**192.00**	**335.00**	**495.00**	**953.00**	**1,191.00**	**1,594.00**
Lợi nhuận khác	1.00	9.00	114.00	72.00	89.00	33.00
Phần lợi nhuận/lỗ từ công ty liên kết liên doanh	0.00	0.00	0.00	4.00	(40.00)	70.00
Tổng lợi nhuận kế toán trước thuế	**193.00**	**344.00**	**609.00**	**1,029.00**	**1,240.00**	**1,698.00**
Lợi nhuận sau thuế thu nhập doanh nghiệp	**175.00**	**301.00**	**536.00**	**880.00**	**1,051.00**	**1,406.00**
Lợi nhuận sau thuế của cổ đông Công ty mẹ	**175.00**	**280.00**	**450.00**	**737.00**	**836.00**	**1,063.00**
Lãi cơ bản trên cổ phiếu (VNĐ)	92,130.00	5,125.00	8,008.00	0.00	5,959.00	7,498.00

Figure 9. FPT's financial reports from 2004 - 2009

FPT invests heavily in the fields of finance - banking and real estate through capital contributions to establish FPT Securities Joint Stock Company (charter capital of 440 billion VND, FPT contributes 110 billion VND), Commercial Bank Tien Phong Joint Stock Company (Charter capital of 1,100 billion VND, FPT contributes 150 billion VND) and FPT real estate company (Charter capital of 30 billion VND). And after the sudden increase in 2004, FPT's ROAA tends to decrease.

5.2 Decreasing trend:

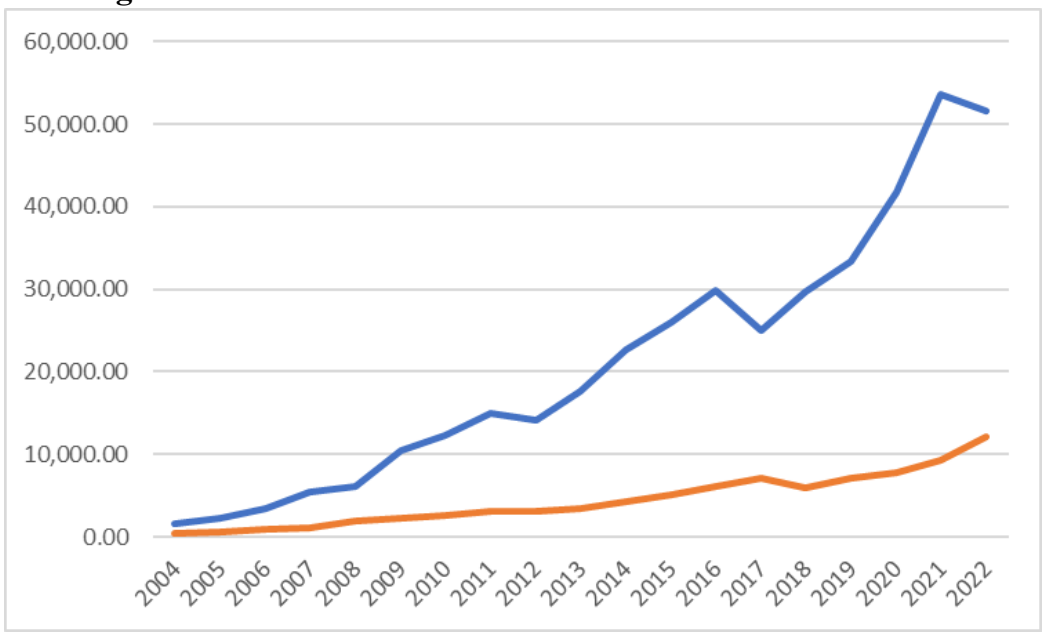


Figure 10. Chart of fluctuations of total assets and total expenses

Between 2007 and about 2014, ROAA continued to decline, possibly reflecting that the company was going through a period of restructuring or major investment that had yet to see results. During this period, if FPT has carried out restructuring or invested heavily in new

projects, the related costs may not yield immediate profits and thus negatively affect ROAA. These investments may include expansion into international markets, research and development of new products, or investments in new technology.

From 2007 to 2010, the profit ratio tended to go down, partly reflecting ineffectiveness when FPT invested relatively widely in areas outside its main business. At this time, the new strategy focusing on key business segments has been initially implemented, but FPT believes that it still needs time for the restructuring process to be as effective as expected.

The macro-economy is slow to develop and the consumption of high-tech products in the market has decreased, so FPT's distribution of mobile phones, laptops, and information technology products has reduced sales. The financial crisis in 2007 also reduced FPT's stock profit to 0. Vietnam's competitiveness index according to the World Economic Forum (WEF) decreased in the period 2006 - 2009, from 64th to 75th in 2009

In the field of telecommunications, the investment figure is estimated at about 870 billion VND, specifically 220 billion VND invested in the Asia Pacific Gateway (APG) international undersea fiber optic cable backbone, expected to be put into operation in 2014, reached shore in Da Nang; 350 billion VND invested in backbone telecommunications network infrastructure; 100 billion VND invested in the field of providing telecommunications services to 8 provinces and cities along the North-South axis. At the same time, FPT will also look outward with an investment of about 200 billion VND to provide telecommunications services in 4 major cities in Cambodia.

In the field of mobile telecommunications, after withdrawing from the EVN deal, Mr. Truong Dinh Anh said that FPT will continue to look for other opportunities to participate in the mobile telecommunications market in both aspects of mergers and acquisitions. or apply for a license to build the next-generation mobile network. In addition, the retail network will be focused on a more systematic scale with an investment of VND 250 billion for the FPT digital retail chain with a scale of 150 stores nationwide.

5.3 Stabilize and recover:

From around 2014 onwards, ROAA appeared to be more stable and showed slight signs of recovery. This could indicate that the measures the company has taken are starting to work, or that the company may have adjusted its business model to better suit market conditions.

2016: Total assets increased sharply. After the construction phase, as well as investment and development, the period from 2014 to now is a period that can be called enjoying sweet fruits. FPT Corporation has determined its development direction towards stable development. Promoting its development achievements, FPT has focused strongly on the needs and trends of the world. What is clearly proven is that in the complicated developments of the COVID-19 epidemic. In foreign markets, FPT won three contracts worth over 100 million USD in 2020, something that had never happened before the pandemic.

They developed a solution to accomplish three goals:

- Use technology to perform and monitor work;
- Control activities by project - take the leader as the center;
- Implement internal communication activities to maintain employee enthusiasm.

Not only for internal use, they sell it and have 1,500 customers.

In 2021, there were 43 internal digital transformation projects related to the automation and digitization of business processes; customer care management; HRM; Management targets, business plans... Deployed throughout the Group, saving 98 billion VND in costs and contributing 141 billion VND in revenue. For example, the Customer Data Platform analyzes data of more than 48 million customers using FPT's service platforms. From there, FPT can better understand customers' needs and consumption behaviors and provide appropriate and timely solutions in the future.

FPT's online segment earned VND 4,610 billion, growing 10% compared to the first 9 months of 2020 and accounting for 33% of the Company's total consolidated revenue. In addition, FPT Shop continues to maintain its position as the No. 1 Laptop retailer in the market with revenue in the first 9 months of 2021 reaching VND 3,350 billion, an impressive growth of 80% compared to the same period in 2020.

According to the Company's representative, this result is thanks to the Company's reasonable strategic planning, ensuring a rich source of goods and many attractive incentive programs that resonate with people's need to own a Laptop. The number of people increased during the epidemic season, especially the back-to-school season.

In addition, a representative of FPT Retail (FRT) said that the positive side of the social isolation period is that the number of online orders increased by 14-20%. Due to prior planning, this system proactively allocated reasonable resources to meet this shift.

- **Summarize:**

Rates of return over the past 18 years have shown constant trends and fluctuations. From there, we see that FPT Corporation is a very active corporation, always looking for new things and growing day by day. During 18 years, there were times of difficulty but FPT still affirmed its strength. This is a trusted corporation, a giant of the technology sector in Vietnam. At the same time, based on this analysis, we can make the following hypothesis:

H5: Stock price affects ROAA

VI. ANALYSIS OF THE IMPACT ON THE RATE OF PROFIT ON AVERAGE TOTAL ASSETS

6.1 Explain the variables in the model:

	Variable name	Notation
Dependent variable	ROAA	ROAA
Independent variable	GDP	X1
	Exchange rate	X2
	Current ratio	X3
	Debt ratio	X4
	Stock price	X5

6.2 Descriptive statistics:

	ROAA	GDP Growth	exchange rate	Current ratio	Debt ratio
ROAA	1.000000	0.285354	-0.837405	0.425994	0.506241
GDP Growth	0.285354	1.000000	-0.261565	0.177339	0.082710
exchange rate	-0.837405	-0.261565	1.000000	-0.648484	-0.502023
Current ratio	0.425994	0.177339	-0.648484	1.000000	-0.092473
Debt ratio	0.506241	0.082710	-0.502023	-0.092473	1.000000

Figure 11. Correlation statistics table

According to the correlation statistics table, we see that the variables Exchange rate and Debt ratio have a strong impact on ROAA while GDP and Current ratio have a rather weak correlation. At the same time, according to the analysis in Chapter 4, we can see that the interaction of stock returns affects ROAA. To see the details, let's go into analysis based on models.

6.3 Linear regression:

Variable	
Log-Likelihood	An index that measures the fit of a model to the data. The higher the value, the better the model fits.
AIC and BIC	Are all measures of fit, with lower values being better.
R-squared (R^2) Adj.R-squared	A measure that shows the percentage of variation in the dependent variable explained by the regression model.
F-statistic and Prob (F-statistic)	F-statistics are used to test the general properties of the model. Prob (F-statistic) indicates whether the model is statistically significant or not
Coef	Coefficient is an estimate of the average change in the dependent variable for each unit change in the corresponding independent variable.
Std Err (Standard Error)	This is a measure of the coefficient's volatility. It is commonly used to calculate confidence intervals.
t-statistic	This is the statistical value used to test the hypothesis about the value of the coefficient. If the t value is large, the hypothesis can be rejected.
P > |t| (P-value)	Purpose: Probability of the null hypothesis (coefficient equal to 0). If this value is less than a threshold (usually 0.05), we have enough evidence to reject the hypothesis. For example: If P>|t| for the variable > 0.05, there is not enough evidence to reject the hypothesis that it has no impact.

[0.025 0.975]	The 95% confidence interval for the coefficient, tells us the range within which the coefficient is likely to lie at the 95% confidence level. If the confidence interval for the GDP Growth coefficient is from -0.571 to 0.930, then there is a 95% chance that the true value of the coefficient lies within this interval.
Omnibus	Purpose: Test the hypothesis about the normal distribution of residuals. This hypothesis is important to ensure the accuracy of other statistical tests. Explanation: The Omnibus value can be used to test the hypothesis of a normal distribution of residuals. If this value is large enough, the hypothesis of normal distribution can be rejected.
Durbin-Watson	Purpose: Measure the autocorrelation between residuals. The value is usually between 0 and 4. Explanation: If the Durbin-Watson value is close to 2, it is a positive sign of relative independence between the residuals.
Skew	Purpose: Measure the bias of the residuals distribution. Explanation: The Skew value measures the bias of the residuals distribution. If this value is large, the residuals tend to be skewed.
Kurtosis	Purpose: Measure the kurtosis of the residuals distribution. Explanation: The Kurtosis value measures the kurtosis of the residuals distribution. High values may imply a thicker tail of the distribution and more concentrated values.
Cond. No.	Purpose: Measure linear independence between independent variables. Explanation: This value is often used to test the linear independence of independent variables. If this value is high, there may be a linear independence problem.

Linear regression model between the dependent variable Return on Average Assets (ROAA) and two independent variables: GDP Growth and Exchange rate

OLS Regression Results

Dep. Variable:	Return on Average Assets (ROAA)	R-squared:	0.706
Model:	OLS	Adj. R-squared:	0.669
Method:	Least Squares	F-statistic:	19.21
Date:	Thu, 16 Nov 2023	Prob (F-statistic):	5.59e-05
Time:	23:03:15	Log-Likelihood:	-39.710
No. Observations:	19	AIC:	85.42
Df Residuals:	16	BIC:	88.25
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	31.4550	4.704	6.687	0.000	21.483	41.427
GDP Growth	0.1794	0.354	0.507	0.619	-0.571	0.930
Exchange rate	-0.0010	0.000	-5.830	0.000	-0.001	-0.001

Omnibus:	6.643	Durbin-Watson:	1.278
Prob(Omnibus):	0.036	Jarque-Bera (JB):	3.983
Skew:	0.914	Prob(JB):	0.137
Kurtosis:	4.301	Cond. No.	1.95e+05

Figure 12. Linear regression model between the dependent variable Return on Average Assets (ROAA) and two independent variables: GDP Growth and Exchange rate

- R-squared: 0,706 => The model explains approximately 70.6% of the variation in ROAA.
- Adj. R-squared: 0,669
- Prob (F-statistic): 5,59e-05 => very small, indicating that the model is statistically significant.
- GDP: The Coef value is 0,1794, but with a value of P > |t| is large (0.619), so there is no statistical evidence to confirm that GDP has a significant effect on the dependent variable.
- Exchange rate: P > |t| very small (0,000), showing that Exchange rate has a significant and statistically significant influence on the dependent variable ROAA.

Conclusion: in the model, Exchange rate appears to have a significant impact on ROAA, while GDP does not appear to have a statistically significant impact.

Linear regression model between the dependent variable Return on Average Assets (ROAA) and two independent variables Current ratio (short term) and Debt ratio

OLS Regression Results

Dep. Variable:	Return on Average Assets (ROAA)	R-squared:	0.193
Model:	OLS	Adj. R-squared:	0.168
Method:	Least Squares	F-statistic:	7.788
Date:	Thu, 16 Nov 2023	Prob (F-statistic):	0.000929
Time:	23:07:50	Log-Likelihood:	-121.81
No. Observations:	68	AIC:	249.6
Df Residuals:	65	BIC:	256.3
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.1150	2.575	0.045	0.965	-5.027	5.257
Current ratio (short term)	2.7781	0.794	3.498	0.001	1.192	4.364
Finace leverage (Debt ratio)	-0.0233	0.037	-0.633	0.529	-0.097	0.050

Omnibus:	86.297	Durbin-Watson:	1.595
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1363.481
Skew:	3.634	Prob(JB):	8.39e-297
Kurtosis:	23.698	Cond. No.	786.

Figure 13. Linear regression model between the dependent variable Return on Average Assets (ROAA) and two independent variables Current ratio (short term) and Debt ratio

- R-squared is 0,193, meaning the model can only explain about 19.3% of the variation in the dependent variable ROAA.
- Adj. R-squared: 0,168
- Prob (F-statistic) (0,000929) is small, indicating that the model is statistically significant. At least one of the independent variables is important to ROAA.
- Current ratio: P > |t| very small (0,001), showing that Current ratio (short term) has a significant and statistically significant effect on the dependent variable ROAA.
- Debt raio: P > |t| large (0,529), there is no statistical evidence to confirm that Debt ratio has a significant effect on the dependent variable.

Conclusion: Current ratio (short term) has a significant influence on ROAA, while Debt ratio has no significant influence.

Linear regression model between the dependent variable Return on Average Assets (ROAA) and the independent variable Stock price

OLS Regression Results

Dep. Variable:	Return on Average Assets (ROAA)	R-squared:	0.016
Model:	OLS	Adj. R-squared:	0.000
Method:	Least Squares	F-statistic:	1.028
Date:	Fri, 17 Nov 2023	Prob (F-statistic):	0.314
Time:	15:01:51	Log-Likelihood:	-93.695
No. Observations:	67	AIC:	191.4
Df Residuals:	65	BIC:	195.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	2.8538	0.172	16.546	0.000	2.509	3.198
Price	-4.944e-06	4.88e-06	-1.014	0.314	-1.47e-05	4.79e-06

Omnibus:	23.938	Durbin-Watson:	1.444
Prob(Omnibus):	0.000	Jarque-Bera (JB):	38.332
Skew:	1.325	Prob(JB):	4.74e-09
Kurtosis:	5.589	Cond. No.	5.02e+04

Figure 14. Linear regression model between the dependent variable Return on Average Assets (ROAA) and the independent variable Stock price

- R-squared: In this case, only about 1.6% of the variation in ROAA is explained by the model.
- Prob (F-statistic) is large (0.314), the model is not statistically significant.
- Price: P > |t| is 0,314

Conclusion: The model seems inadequate to explain the dependent variable Stock price

6.4 ANOVA:

When considering, the independent variables we choose are continuous variables but usually ANOVA will use categorical variables so the ANOVA model is not suitable for our data.

6.5 Choose the appropriate model:

Through the above sections, the Linear Regression model is suitable to consider the impact of dependent variables on ROAA

Conclude:

- Accept the hypotheses: H2, H3
- Reject hypothesis H1
- There is not enough evidence to verify hypotheses H4, H5

VII. FORECASTING STOCK PRICE SITUATION

Based on the knowledge we have learned, we use the ARIMA and LSTM models to predict stock prices

7.1 LSTM model:

Step 1: Prepare data

```python
# Read the CSV file
df = pd.read_csv('D:/Desktop/FPT Historical Data.csv')
```

Python

```python
# some format data
df.set_index('Date',drop=True,inplace=True)
```

Python

+ Code + Markdown

```python
df.head(20)
```

Python

```python
# Plot using hvplot
plot = df.hvplot.line(title='Time Series Plot', xlabel='Date', ylabel='Price')

# Show the plot
plot
```

Python

```python
# Extract the 'Close' prices for modeling
prices = df['Price'].values.reshape(-1, 1)
```

```python
# Normalize the data
scaler = MinMaxScaler(feature_range=(0, 1))
prices_scaled = scaler.fit_transform(prices)
```

```python
prices_scaled
```

```
array([[0.14617319],
       [0.15503653],
       [0.16432273],
       ...,
       [0.92188598],
       [0.92084446],
       [0.93334271]])
```

Figure 15. Prepare data

Step 2: Sequence creation

```python
# Function to create sequences for LSTM
def create_sequences(data, seq_length):
    X, y = [], []
    for i in range(len(data) - seq_length):
        X.append(data[i:i + seq_length, 0])
        y.append(data[i + seq_length, 0])
    return np.array(X), np.array(y)
```

[9] Python

```python
# Set the sequence length (number of time steps to look back)
sequence_length = 10
```

[10] Python

```python
# Create sequences for training
X, y = create_sequences(prices_scaled, sequence_length)
```

[11] Python

+ Code + Markdown

Figure 16. Sequence creation

Step 3: Preprocessing

```python
# function to plit the data into train, dev, and test sets
def train_dev_test_split(data, seq_length, train_ratio=0.75, dev_ratio=0.15):
    total_size = len(data)
    train_size = int(total_size * train_ratio)
    dev_size = int(total_size * dev_ratio)

    X, y = create_sequences(data, seq_length)

    X_train, y_train = X[:train_size], y[:train_size]
    X_dev, y_dev = X[train_size:train_size + dev_size], y[train_size:train_size + dev_size]
    X_test, y_test = X[train_size + dev_size:], y[train_size + dev_size:]

    return (X_train, y_train), (X_dev, y_dev), (X_test, y_test)
```

[12] Python

```python
# X_train, y_train = X[:train_size], y[:train_size]
# X_dev, y_dev = X[train_size:train_size + dev_size], y[train_size:train_size + dev_size]
# X_test, y_test = X[train_size + dev_size:], y[train_size + dev_size:]

(X_train, y_train), (X_dev, y_dev), (X_test, y_test) = train_dev_test_split(prices_scaled, SEQUENCE_LENGTH)
```

[13] Python

Figure 17. Preprocessing

Step 4: Reshape data for LSTM

```python
# Reshape input data to be 3D for LSTM input (samples, time steps, features)
X_train = np.reshape(X_train, (X_train.shape[0], X_train.shape[1], 1))
X_dev = np.reshape(X_dev, (X_dev.shape[0], X_dev.shape[1], 1))
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1], 1))
```

```python
X_train
```

```
array([[[0.14617319],
        [0.15503653],
        [0.16432273],
        ...,
        [0.15377004],
        [0.16263442],
        [0.17149776]],

       [[0.15503653],
        [0.16432273],
        [0.17403074],
        ...,
        [0.16263442],
        [0.17149776],
        [0.17149776]],

       [[0.16432273],
        [0.17403074],
        [0.18416056],
        ...,
        [0.17149776],
        [0.17149776],
        [0.16305624]],

       ...,
...
        [0.30187213],
        ...,
        [0.30373229],
        [0.30435199],
        [0.31923323]]])
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

Figure 18. Reshape data for LSTM

Step 5: LSTM model definition and Training

```python
model = Sequential([
    LSTM(50, return_sequences=True, input_shape=(SEQUENCE_LENGTH, 1)),
    LSTM(50, return_sequences=True),
    LSTM(50),
    Dense(1)
])

model.compile(optimizer='adam', loss='mean_squared_error')
```

Python

```
WARNING:tensorflow:From E:\anaconda3\Lib\site-packages\keras\src\layers\rnn\lstm.py:148: The name tf.executing_eagerly_outside_functions is deprecated. Please use tf.compat.v1.executing_

WARNING:tensorflow:From E:\anaconda3\Lib\site-packages\keras\src\optimizers\__init__.py:309: The name tf.train.Optimizer is deprecated. Please use tf.compat.v1.train.Optimizer instead.
```

```python
# Train the model
history = model.fit(X_train, y_train, epochs=100, batch_size=32, validation_data=(X_dev, y_dev))
```

Python

```
Epoch 77/100
99/99 [==============================] - 1s 11ms/step - loss: 1.1200e-05 - val_loss: 9.7285e-05
Epoch 78/100
99/99 [==============================] - 1s 12ms/step - loss: 1.1241e-05 - val_loss: 1.2983e-04
Epoch 79/100
99/99 [==============================] - 1s 10ms/step - loss: 1.1641e-05 - val_loss: 2.0910e-04
Epoch 80/100
99/99 [==============================] - 1s 12ms/step - loss: 9.8998e-06 - val_loss: 1.0814e-04
Epoch 81/100
99/99 [==============================] - 1s 11ms/step - loss: 1.0621e-05 - val_loss: 1.1567e-04
Epoch 82/100
99/99 [==============================] - 1s 11ms/step - loss: 9.6715e-06 - val_loss: 1.0763e-04
Epoch 83/100
99/99 [==============================] - 1s 12ms/step - loss: 1.0937e-05 - val_loss: 9.6378e-05
Epoch 84/100
99/99 [==============================] - 1s 11ms/step - loss: 9.4299e-06 - val_loss: 9.5888e-05
Epoch 85/100
99/99 [==============================] - 1s 12ms/step - loss: 9.2853e-06 - val_loss: 1.0034e-04
Epoch 86/100
99/99 [==============================] - 1s 11ms/step - loss: 1.0348e-05 - val_loss: 1.0532e-04
...
Epoch 99/100
99/99 [==============================] - 1s 13ms/step - loss: 9.2419e-06 - val_loss: 1.6366e-04
Epoch 100/100
99/99 [==============================] - 1s 11ms/step - loss: 1.0562e-05 - val_loss: 1.0573e-04
```

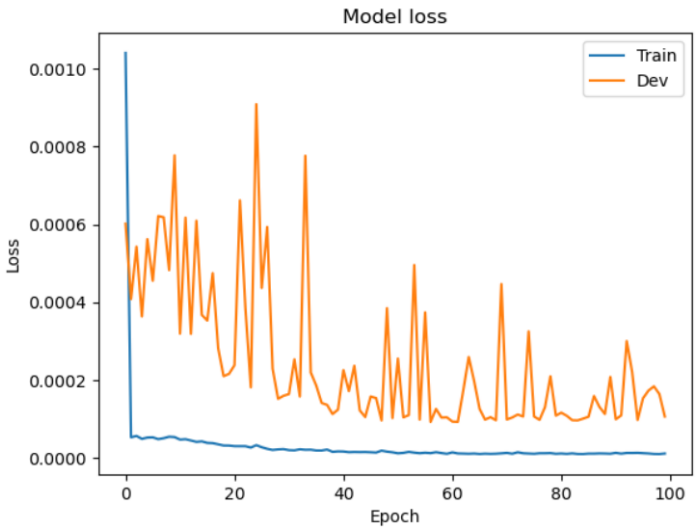
Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

Figure 19. LSTM model definition and Training

Step 6: Model Evaluation and Visualization

```python
# Plot training & validation loss values
plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Dev'], loc='upper right')
plt.show()
```

[18]



```
[19]
# Make predictions on the test set
predictions = model.predict(X_test)
```

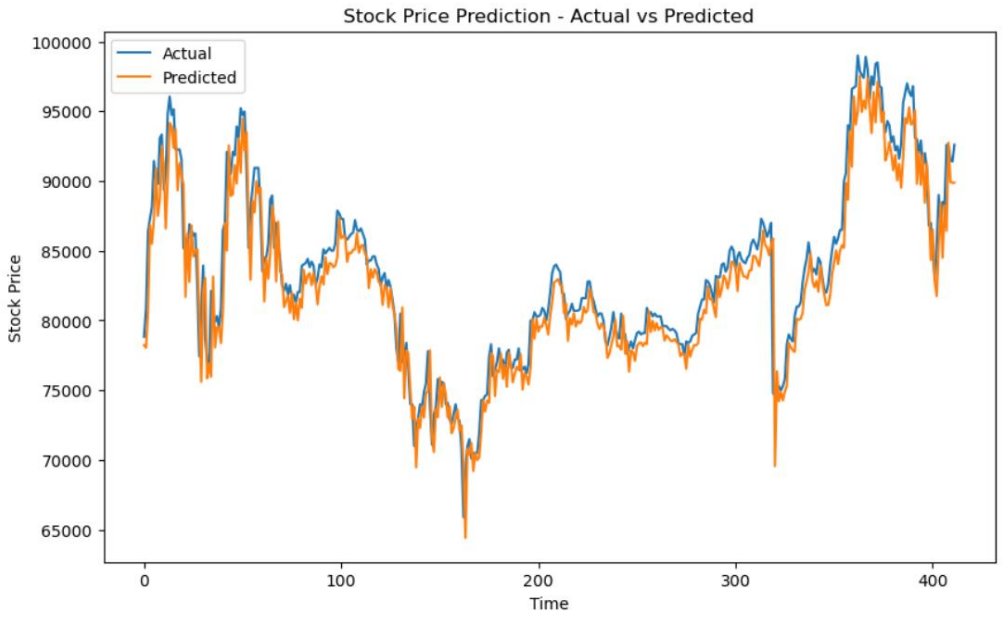
```
13/13 [==============================] - 1s 4ms/step
```

```
[20]
# Inverse transform the predictions and actual values to the original scale
predictions_inv = scaler.inverse_transform(predictions)
y_test_inv = scaler.inverse_transform(y_test.reshape(-1, 1))
```

```
[21]
# Compare predictions with actual values
comparison_df = pd.DataFrame({'Actual': y_test_inv.flatten(), 'Predicted': predictions_inv.flatten()})
print(comparison_df)
```

```
       Actual     Predicted
0     78833.0  78229.328125
1     80810.0  78054.054688
2     86412.0  80674.695312
3     87318.0  86834.726562
4     88142.0  85498.054688
..        ...           ...
407   92600.0  86426.726562
408   92500.0  92745.398438
409   91500.0  89931.062500
410   91400.0  89872.906250
411   92600.0  89879.257812

[412 rows x 2 columns]
```



```
# Calculate evaluation metrics
mse = mean_squared_error(y_test_inv, predictions_inv)
mae = mean_absolute_error(y_test_inv, predictions_inv)
r2 = r2_score(y_test_inv, predictions_inv)

# Print the evaluation metrics
print(f'Mean Squared Error (MSE): {mse:.2f}')
print(f'Mean Absolute Error (MAE): {mae:.2f}')
print(f'R-squared (R2): {r2:.2f}')
```

```
Mean Squared Error (MSE): 4149204.31
Mean Absolute Error (MAE): 1529.53
R-squared (R2): 0.90
```

Figure 20. Model Evaluation and Visualization

Step 7: Predict future price

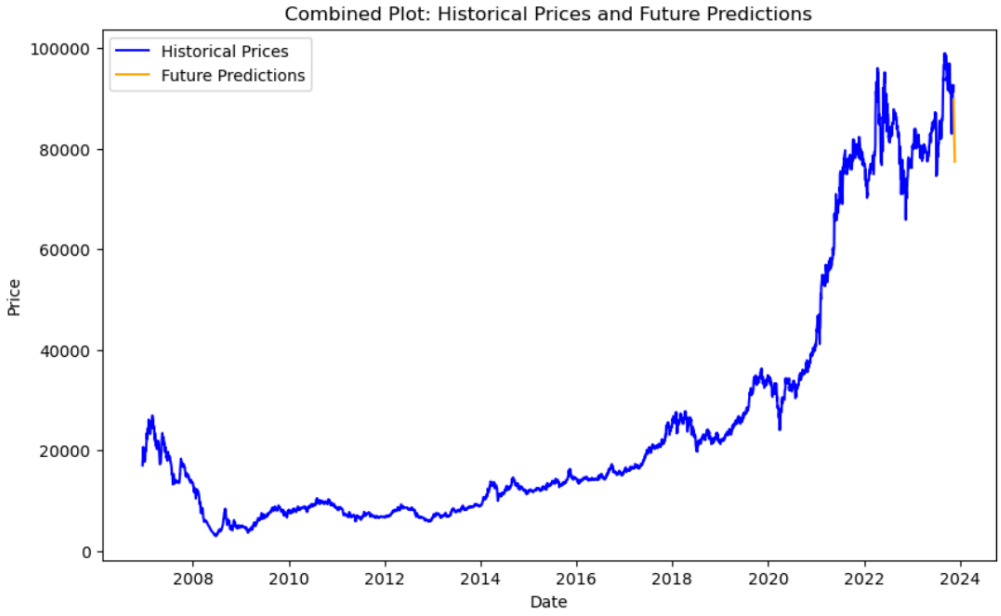


Figure 21. Predict future price

7.2 ARIMA model:

Step 1: Collect data

```
In [2]: # Read the CSV file
        df = pd.read_csv('D:/Desktop/FPT Historical Data Monthly.csv')

In [3]: # transform column 'Date' to  datetime
        df['Date'] = pd.to_datetime(df['Date'], errors='coerce')

In [4]: # some format data
        df.set_index('Date',drop=True,inplace=True)
        df.head()

        df
```

Out[4]:

Date	Price
2007-01-01	26139.5
2007-02-01	25612.7
2007-03-01	21945.5
2007-04-01	17240.0
2007-05-01	21235.6
...	...
2023-07-01	85600.0
2023-08-01	96700.0
2023-09-01	92800.0
2023-10-01	83000.0
2023-11-01	92400.0

203 rows × 1 columns

Figure 22. Collect data

Step 2: Plot and consider

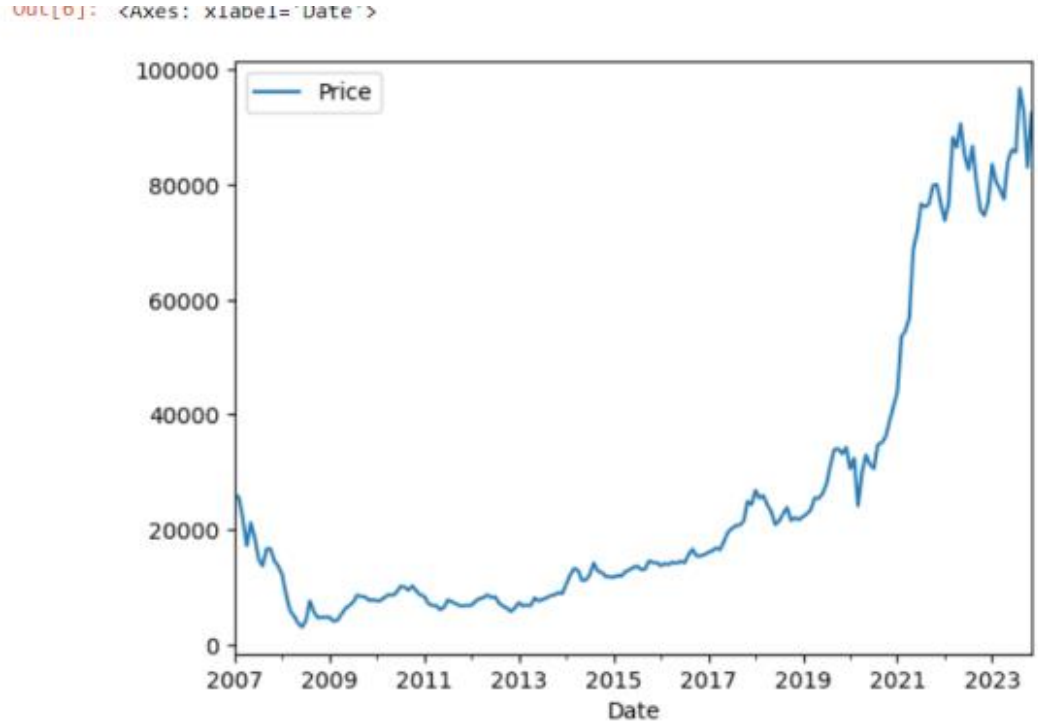


Figure 23. Plot

We can see that the series increases over time (trend series). So, we guess the series is non-stationary. The time series increases and decreases without fixed rules and does not comply with seasons. Independent growth and decline between months of the year often do not follow a fixed principle.

Step 3: Estimate whether this is non-stationary data or stationary data by using ADF

```
In [10]: # adfuller for data train

         adf_0 = adfuller(df)
         print('ADF stationary is:', adf_0[0])
         print('p-value of ADF test is:', adf_0[1])
         print('Critical values:')
         for key, value in adf_0[4].items():
             print('\t%s: .%.3f'%(key,value))

         ADF stationary is: 0.1614096592298381
         p-value of ADF test is: 0.970012201805755
         Critical values:
                 1%: .-3.466
                 5%: .-2.877
                 10%: .-2.575
```

Figure 24. ADF test

P-value is 0.97 > 0.05 => the series is non-stationary. Because it is a non-stationary data series, so that we use differencing method for that data.

```
# Function to perform differencing
def difference(data, d=1):
    """
    Perform differencing on a time series data for 'd' orders.
    data: Pandas Series or DataFrame
    d: Order of differencing (default = 1 for first difference)
    """
    if isinstance(data, pd.Series):
        # If data is a Pandas Series
        diff_series = data.diff(periods=d).dropna()
        return pd.DataFrame({f'Difference_{d}': diff_series})
    elif isinstance(data, pd.DataFrame):
        # If data is a Pandas DataFrame, operate on the specified column
        diff_series = data.iloc[:, 0].diff(periods=d).dropna()
        return pd.DataFrame({f'Difference_{d}': diff_series})
    else:
        raise ValueError("Input should be a Pandas Series or DataFrame")
```

Figure 25. Function to perform differencing

This is a function and runs this differencing function. Next, we get a data 1st difference. We ran this by adfuller again for 1st difference.

```
In [15]: adf_1 = adfuller(first_difference)
         print('ADF 1diff stationary is:', adf_1[0])
         print('p-value of ADF 1diff test is:', adf_1[1])
         print('Critical values 1diff:')
         for key, value in adf_1[4].items():
             print('\t%s: .%.3f'%(key,value))

ADF 1diff stationary is: -3.788044302133533
p-value of ADF 1diff test is: 0.0030320335820000458
Critical values 1diff:
        1%: .-3.466
        5%: .-2.877
        10%: .-2.575
```

Figure 26. ADF test for 1st difference

Now p-value < 0.05 => d = 1.

Step 4: Identify the order of the AR term (p) and MA term (q).

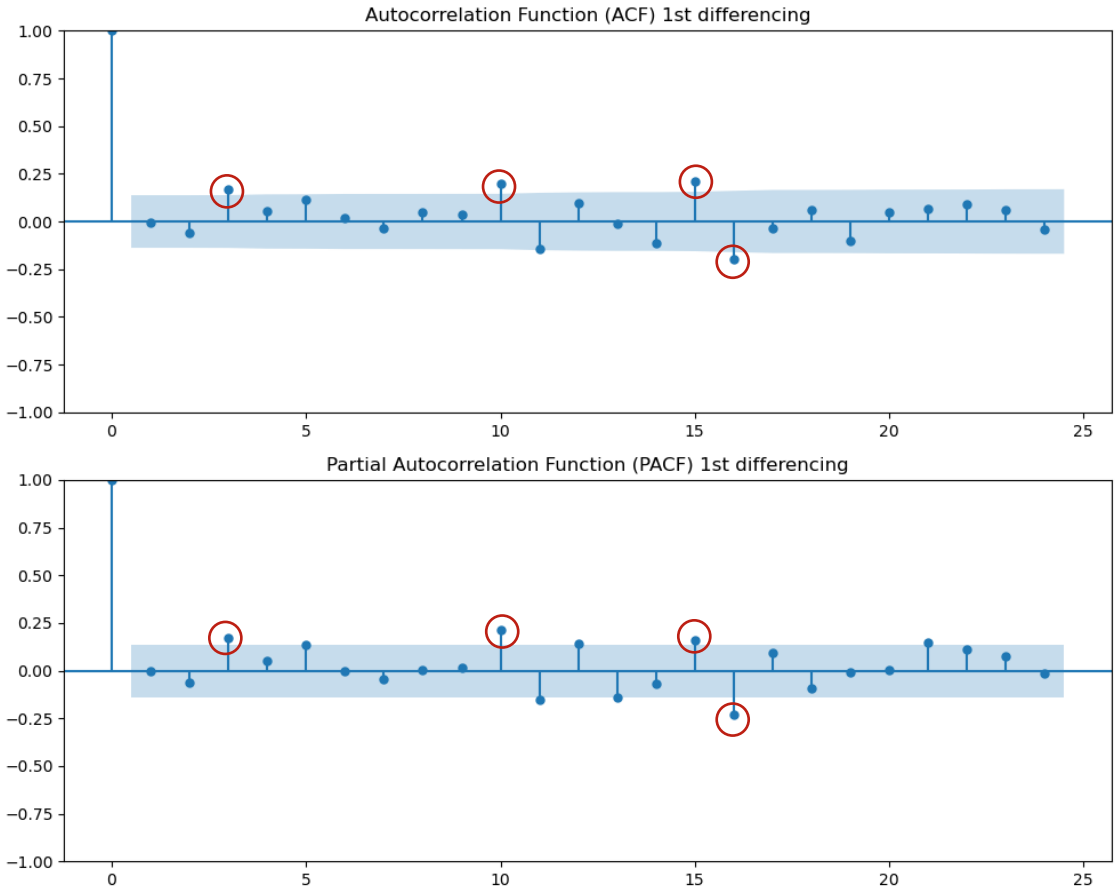


Figure 27. ACF and PACF result

```
# Define the p, d, q values
p_values = [3, 10, 15, 16]
d_value = 1
q_values = [3, 10, 15, 16]
```

Figure 28. Identify p, q, d value

Use a loop to run combinations of (p, d, q). Test different models (p, d, q) using Log likelihood, AIC, BIC, HQIC, MAE, RMSE. It is observed that the ARIMA (3, 1, 3) model has the most accurate statistics:

```
MAE: 13044.432971915538
RMSE: 14165.659699555
------------------------------------------------------------
ARIMA(10, 1, 3) model
Log likelihood: -1616.4888186699739
AIC: 3260.9776373399477
BIC: 3305.756595777669
HQIC: 3279.1319847780437
MAE: 9691.878929143617
RMSE: 11256.540774993191
------------------------------------------------------------
ARIMA(3, 1, 3) model
Log likelihood: -1619.524612227854
AIC: 3253.049224455708
BIC: 3275.438703674569
HQIC: 3262.126398174756
MAE: 9029.761385807647
RMSE: 10656.67611423437
------------------------------------------------------------
```

Figure 29. Test model

⇨ **Choose ARIMA (3, 1, 3) model.**

SARIMAX Results

Dep. Variable:	Price	No. Observations:	182
Model:	ARIMA(3, 1, 3)	Log Likelihood	-1619.525
Date:	Fri, 17 Nov 2023	AIC	3253.049
Time:	14:03:11	BIC	3275.439
Sample:	01-01-2007 - 02-01-2022	HQIC	3262.126
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	1.0269	0.164	6.259	0.000	0.705	1.348
ar.L2	-1.0448	0.122	-8.593	0.000	-1.283	-0.806
ar.L3	0.4739	0.146	3.246	0.001	0.188	0.760
ma.L1	-1.0310	0.185	-5.586	0.000	-1.393	-0.669
ma.L2	1.0985	0.150	7.343	0.000	0.805	1.392
ma.L3	-0.2901	0.186	-1.561	0.119	-0.654	0.074
sigma2	3.179e+06	1.83e+05	17.325	0.000	2.82e+06	3.54e+06

Ljung-Box (L1) (Q):	1.23	Jarque-Bera (JB):	493.06
Prob(Q):	0.27	Prob(JB):	0.00
Heteroskedasticity (H):	3.20	Skew:	0.54
Prob(H) (two-sided):	0.00	Kurtosis:	11.01

Figure 30. ARIMA model result

Explanation of parameters:

- **coef:** These are the estimated coefficients of the AR and MA components of the model. They show the relationship between the current value and the lagged values or lagged errors. For example, ar . L1 = 1.0269 means that the current value increases by 1.0269

units for each unit increase in the lag value of 1. These coefficients are both statistically significant because the p-value is less than 0.05 (P > |z|).

- **std err:** This is the standard deviation of the estimated coefficients. They indicate the accuracy of the coefficients. The smaller these values are, the better.
- **z:** This is the z-statistic value that indicates the difference between the estimated coefficient and the null hypothesis. The larger this value is, the more statistically significant the coefficient is.
- **P > |z|:** This is the probability of observing the same or higher z value if the hypothesis is not true. The smaller this value is, the more likely it is to reject the null hypothesis and accept a statistically significant coefficient. Typically, the acceptance threshold is 0.05 or 0.01.
- **[0.025 0.975]:** This is the 95% confidence interval of the estimated coefficients. It indicates the range that is likely to contain the true value of the coefficient with a 95% confidence level. If this interval does not contain the value 0, then the coefficient is statistically significant.
- **sigma2:** This is the variance of the forecast error. It indicates the magnitude of the average error when predicting future values. The smaller this value is, the more accurate the model is.
- **Ljung-Box (L1) (Q):** This is the value of the Ljung-Box test statistic for lag error 1. It tests whether a linear correlation exists between lag errors. If so, then the model is not capturing enough information in the data. The smaller this value is, the more it proves the existence of correlation.
- **Prob (Q):** This is the probability of observing the same or higher Q value if there is no correlation between the lag errors. The larger this value is, the more it proves that the model has captured enough information in the data. Typically, the acceptance threshold is 0.05 or 0.01.
- **Jarque-Bera (JB):** This is the value of the Jarque-Bera test statistic for the normal distribution of errors. It checks whether the errors follow a normal distribution or not. If not, then the model may be biased or noisy. The larger this value, the greater the difference between the distribution of errors and the normal distribution.
- **Prob (JB):** This is the probability of observing the same or higher JB value if the errors follow a normal distribution. The smaller this value is, the more we reject the hypothesis of a normal distribution of errors. Typically, the acceptance threshold is 0.05 or 0.01.
- **Heteroskedasticity (H):** This is the Breusch-Pagan test statistic value for covariance of errors. It checks whether the variance of the errors depends on the forecast value. If so, the model may have missing or residual variables. The larger this value is, the more it proves the existence of covariance.
- **Prob(H) (two-sided):** This is the probability of observing the same H value or higher if the errors have homogeneous variance. The smaller this value is, the more we reject the hypothesis of co-variance of the errors. Typically, the acceptance threshold is 0.05 or 0.01.
- **Skew:** This is the skewness of the distribution of errors. It shows the shape of the distribution. If this value is 0, then the distribution is symmetric. If this value is positive, then the distribution is skewed to the right. If this value is negative, the distribution is skewed to the left.
- **Kurtosis:** This is the kurtosis of the distribution of errors. It shows the concentration of values around the mean. If this value is 3, then the distribution is normal. If this value

is greater than 3, the distribution is more kurtotic and has more outliers. If this value is less than 3, the distribution is smoother and has fewer outliers.

From the above data, we can conclude that the ARIMA (3, 1, 3) model can be a suitable model for stock price data.

- The coefficients AR and MA are both statistically significant with a high level of confidence.
- The variance of the forecast error is quite small compared to the range of the data.
- The errors are not linearly correlated with each other, showing that the model has captured enough information in the data.
- The errors have uniform variance, indicating that the model does not have missing or residual variables.
- The errors have a nearly symmetrical distribution and are not too sharp, indicating that the model is not biased or noisy.

However, the model also has some limitations:

- The coefficient ma.L3 is not statistically significant, indicating that the model may have an excess MA component.
- The errors have higher kurtosis than a normal distribution, indicating that the model may be affected by outliers or unusual fluctuations.

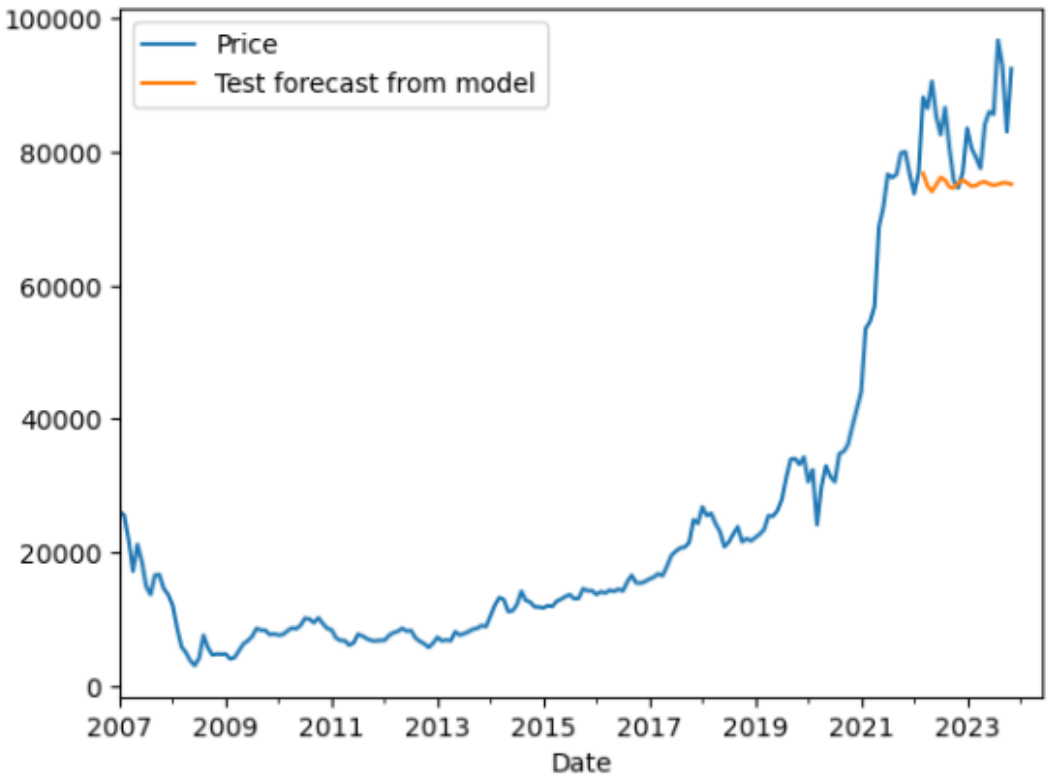


Figure 31. Prediction chart

Use ARIMA (3,1,3) to predict the next 3 months

```
# predict for6 month in future
forecast__313 = model_fit_313.forecast(len(test)+3)
print("\nForecast:")
print(forecast__313)
```

```
Forecast:
2022-03-01    76743.630573
2022-04-01    74873.368964
2022-05-01    74041.129553
2022-06-01    75087.070675
2022-07-01    76144.246725
2022-08-01    75742.677361
2022-09-01    74721.525206
2022-10-01    74593.481478
2022-11-01    75338.529426
2022-12-01    75753.431133
2023-01-01    75340.415497
2023-02-01    74835.924879
2023-03-01    74946.001609
2023-04-01    75390.365413
```

Figure 32. Prediction algorithm

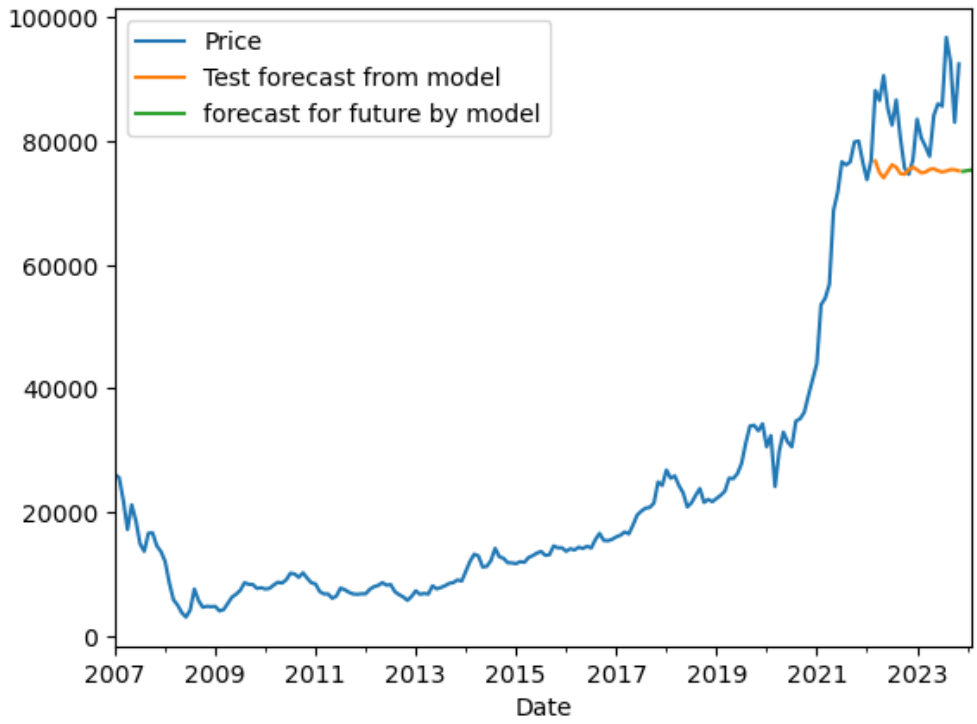


Figure 33. 3-month stock price prediction chart

7.3 Compare 2 Models:

	LSTM	ARIMA
RMSE	2036,9	10656,68
MAE	1529,53	9020,76

According to theory, the less deviation of RMSE and MAE, the more effective it is. So we see that LSTM is more suitable for prediction than the ARIMA model.

VIII. CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions:

Linear Regression model is suitable for analyzing this data

Independent variables do not always affect the dependent variable, but in some cases the independent variable is still affected by the dependent variable and the interaction of the variables should be considered at the same time.

From the results, it can be concluded that the independent variables X1, X2, X3, X4, H5 have an interaction with the ROAA variable. The obtained results satisfy the hypothesis H1. H2, H3, H4, H5.

8.2 Recommendations:

As we can see, X2 and X3 have an impact on ROAA and according to the evidence of analysis from information from chapter 4, we have the following recommendations:

- Businesses should follow FPT's example in effectively managing risks and financial reserves. There must be long-term plans to develop businesses in a sustainable way. Always update world trends to chart appropriate development directions. Understanding the impact of financial variables on business performance is important for making smart decisions.
- Use financial tools such as exchange rate risk management (hedging) to minimize the negative impact of exchange rate fluctuations on business finances.
- Diversify your business and consumer markets to minimize dependence on a specific market or currency.
- Building good relationships with suppliers and customers can help improve payment conditions and reduce the negative impact of debt on business operations.

REFERENCES

[1] Vietstock webite. Link: https://vietstock.vn/

[2] Effect of the Micro and Macro Factors on the Performance of the Listed Jordanian Insurance Companies - Mohammad W. Alomari (2019). Link: researchgate.net

[3] FPT Corporation

[4] Financial report – FPT Online. Link: fptonline.net

[5] Annual Report & Company Administration – FPT Online. Link: fptonline.net

WORK ASSIGNMENT SHEET

No.	Task	Person
1	Research and choose a topic	All
2	Develop analysis direction and understand development direction	Tiến
3	Learn the theoretical basis for variable selection	Tiến, Phương Thy, Anh Thy
4	Build the ARIMA model	Anh Thy, Won Bin, Tiến
5	Build LSTM model	Kiệt, Hà, Anh Thy
6	Analyze business efficiency	Tiến, Won Bin, Hà
7	Prepare reports	Tiến, Hà, Phương Thy
8	PowerPoint	All
9	Present the project	Tiến, Anh Thy
10	Learn the meaning of ARIMA theory	Bin, Anh Thy, Phương Thy
11	Learn the meaning of Linear Regression theory	Anh Thy, Bin
12	Learn the meaning of LSTM theory	Kiệt, Hà