

# IDENTIFY THE FACTORS INFLUENCING THE MOVEMENT OF THE VN-INDEX THROUGH MACROECONOMIC NEWS AND TECHNICAL ANALYSIS BY MACHINE LEARNING



Class: K20414C

Student: Nguyen Thanh Phuong Uyen

Student Code: K204141939

Instructor: Dr. Pham Thi Thanh Xuan

# TABLE OF CONTENT

**Introduction**

**Literature background**

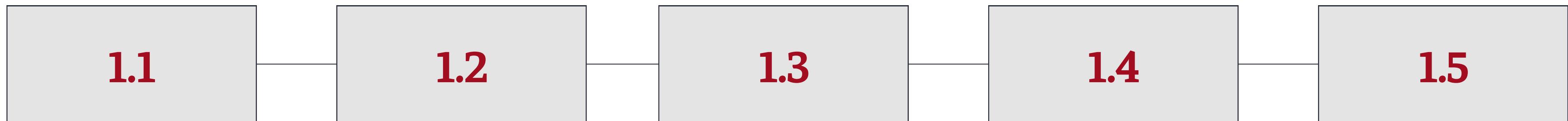
**Data and Methodology**

**Results and discussions**

**Conclusions and  
Recommendations**

**Thank You**

# 1. INTRODUCTION



**Topic's  
Reasonable  
choice**

**Research  
objectives**

**Research  
scale and  
methodology**

**Research  
scientific and  
practical  
contribution**

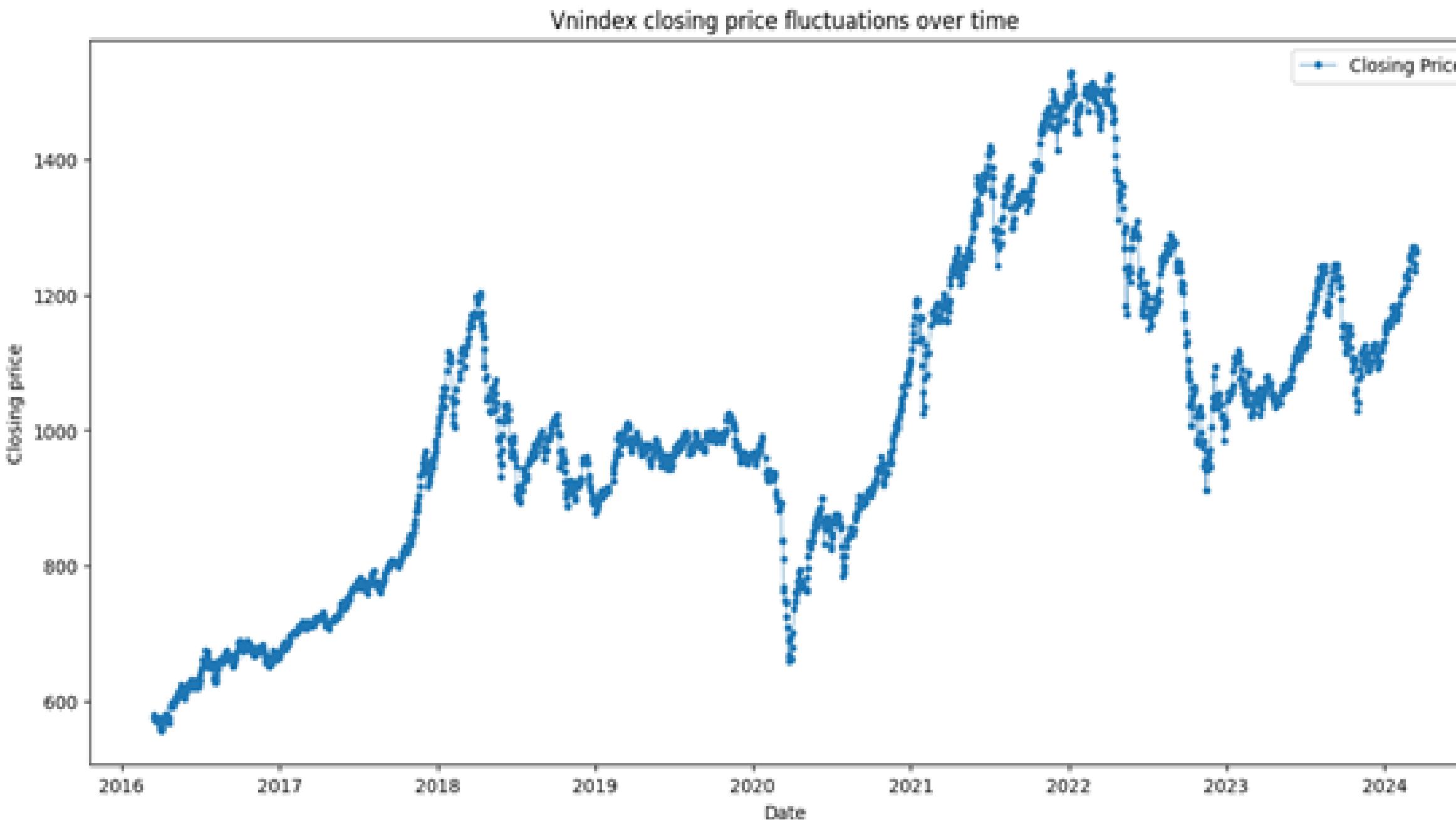
**Report  
structure**

## 1.1. Topic's Reasonable choice

- ➡ According to Mishkin (2004), the stock market in the financial system **plays an important role** in providing long-term capital for economic development
- ➡ Hafer & Hein (2007), the development of financial markets, especially the stock market, is an essential part of **a country's economic growth**.
- ➡ Dr. Tran (2018), the stock market acts as an indicator of economic growth, an upward market trend is often understood as a sign of upcoming economic expansion.
- ➡ Levine & Zervos, 1998), the role of the stock market in **minimizing economic risk** and its impact on the **banking and financial system**



# 1.1. Topic's Reasonable choice



The market has become more attractive thanks to **investment promotion policies** and **rapid economic growth** in Vietnam, especially with the **36% increase** of the VN-Index in **2021**, showing that the market condition is **too hot**.

Figure 1. VN-Index from 2016 to 2024

Source: Author's summary

# 1.1. Topic's Reasonable choice



## Problem 1

---

Vietnam's stock market is still considered an emerging market with a relatively **small scale**

## Problem 2

---

Akbar et al (2019); Singhal, Choudhary, and Biswal (2019); González and his team (2018), **empirical studies** have identified **many factors that influence the stock market**, stemming from **both macroeconomic factors and company-specific factors**.

## Problem 3

---

The application of machine learning algorithms is becoming increasingly popular, with evidence showing that **they outperform traditional statistical methods** in **stock price forecasting** (Bhattacharjee and Bhattacharja, 2019)

## Problem 4

---

**Traditional financial analysis methods** still retain their value, but they increasingly face **limitations in capturing and predicting** the **multifaceted nature of market movements**.

# 1.2. Research objectives



**Objectives 1**

**Compare and find the most effective machine learning model to predict price fluctuations in the Vietnamese stock market based on the VN-Index**



**Objectives 2**

Provide the clearest results to answer the question of which **macroeconomic factors have a strong impact on the Vietnamese stock market**.

# 1.3. Research scale and methodology

The dependent variable is **VNINDEX (VNI)**

## Independent variables

- CPI
- Exchange rate  
(VND/USD)
- Oil price
- S&P500 index
- Gold price
- Money supply M2
- Interest rate

**Time:** Uses daily statistics between **March 2016** and **March 2024**

- Decision Tree Regression,
- Random Forest
- XGBoost
- Linear Regression
- KNN
- ANN
- SVM

SHAP

Feature Important

- RMSE
- MSE
- MAPE
- MAE
- R<sup>2</sup>
- K-Fold cross-validation
- Learning curves

# 1.4. RESEARCH SCIENTIFIC AND PRACTICAL CONTRIBUTION



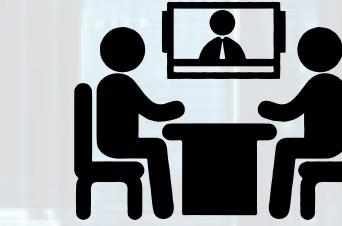
## SCIENCE

Provides new insights into how macroeconomic factors and market factors influence the VN-Index



## SCIENCE

Develop more theory about the relationship between macroeconomics and the stock market



## SCIENCE

Using machine learning techniques to analyze macroeconomic news represents a significant methodological advance



## SCIENCE

Provides a powerful and flexible method for analyzing and predicting stock market behavior

## 1.4. RESEARCH SCIENTIFIC AND PRACTICAL CONTRIBUTION



### PRACTICAL

Investors and fund managers have a deeper understanding of the factors affecting VN-Index, thereby **making more informed investment decisions**



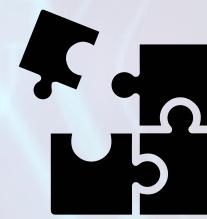
### PRACTICAL

High-performance prediction models provide **valuable information** for market planning and forecasting



### PRACTICAL

Adjust your investment strategy to **accurately reflect market risks and opportunities**



### PRACTICAL

Improve regulatory policies and **enhance** the curriculum of **finance-related academic program**

# 1.5. Report structure

## CONTENT OF THE CHAPTERS

- Abstract
- Introduction
- Literatural background
- Data và Methodology
- Results and discussions
- Conclusions and Recommendations

## OTHER SUBSECTIONS

- Title
- Comments of the instructor
- Acknowledgement
- Table of contents
- List of tables, figures
- List of abbreviations, special characters
- Appendix
- References



## 2. Literature background



2.1. Literature review  
on Macroeconomic  
Factors



2.2. Literature  
review on Machine  
Learning to identify  
Macroeconomic  
Factors

## 2.1. LITERATURE REVIEW ON MACROECONOMIC FACTORS

Kieu and Diep (2013)

specifically analyzed the Vietnamese market, revealing **long-term positive** relationships between the **VN-Index** and **domestic gold prices and money supply M2**, with a **negative correlation to inflation**.

Nhu Quynh and Huong Linh (2019)

extended this analysis, finding long-term inflation positively impacts the VN-Index, whereas **interest rates negatively affect it**

(Sujit & Kumar, 2011; Nandha & Singh, 2011; Akbar et al., 2019; González et al., 2018)

The impact of **oil prices, gold prices, interest rates, inflation, exchange rates, and money supply** on the stock market.

**CASE STUDIES AND REAL-WORLD APPLICATIONS**

## 2.1. LITERATURE REVIEW ON MACROECONOMIC FACTORS

Oil prices	<p><b>Sujit &amp; Kumar (2011) and Nandha &amp; Singh (2011)</b></p> <p>The complex relationship between oil prices and stock markets</p>
Gold Prices	<p><b>(Akbar et al., 2019)</b></p> <p>its prices to inversely affect stock markets</p>
Interest Rates	<p><b>(González et al., 2018)</b></p> <p>An inverse relationship with stock prices</p>
CPI	<p><b>Subhani, Gul, and Osman (2010); Geetha, Mohidin, Chandran, and Chong (2011); Mousa, Safi, Hasoneh, and Mohammad (2012)</b></p> <p>An inverse relationship between inflation and stock market performance</p>

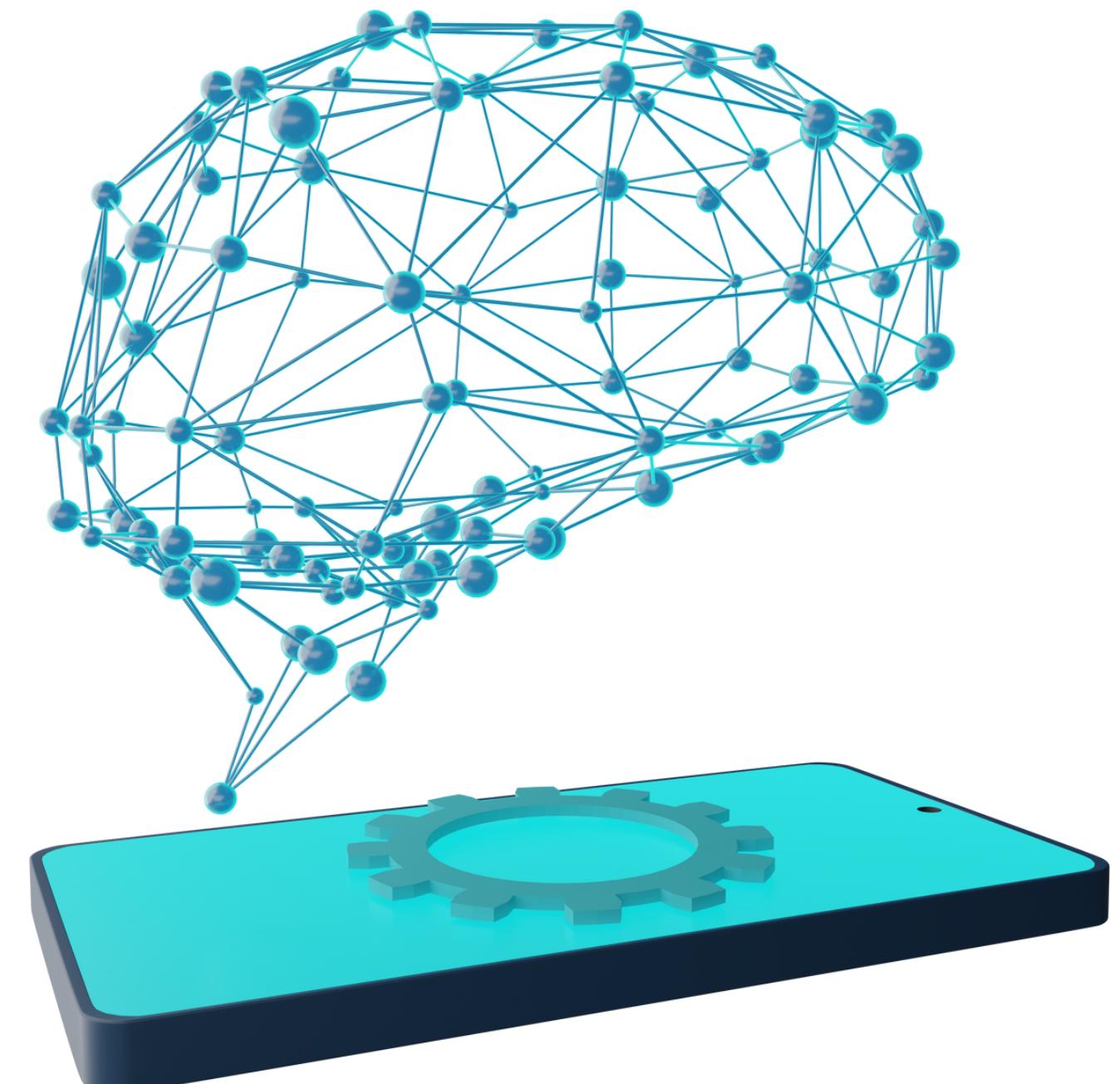
## 2.1. LITERATURE REVIEW ON MACROECONOMIC FACTORS

Exchange Rates	<p><b>Dinh và Nguyen (2008)</b> Exchange rate volatility has adverse effects on the stock market</p> <p><b>Ho và Huang (2015)</b> The impact of exchange rates on stock markets can be either positive or negative depending on the country</p>
Money Supply (M2)	<p><b>Friedman and Schwartz (1965); Nguyen and Nguyen (2013)</b> The existence of a direct bond between the money supply and stock market performance</p>
S&P 500	<p><b>Mai NC (2016); Hussainey K, Khanh Ngoc L (2009); Tien NH (2021), Kim et al., 2015; Li &amp; Giles, 2015</b> A robust positive correlation between Vietnamese stock prices and the US industrial sector and money market Significant one-way shock transmission and volatility from the U.S. market to emerging Asian markets</p>

## 2.2. LITERATURE REVIEW ON MACHINE LEARNING TO IDENTIFY MACROECONOMIC FACTORS

### MACHINE LEARNING

Early studies in the field began with the application of **simple linear models** and **time-series analysis** for **market prediction**. However, as financial markets grew in complexity, these models often **fell short in capturing non-linear relationships** and **adapting to new data**. This limitation led researchers to explore machine learning techniques that could **dynamically learn and adapt to new information**.



## 2.2. LITERATURE REVIEW ON MACHINE LEARNING TO IDENTIFY MACROECONOMIC FACTORS

### Alessi và Detken (2018)

The random forest algorithm to develop an alert system for EU banking crises, using macroeconomic indicators as predictors

Discovering that ML forecasts could, in some instances, surpass traditional regression-based strategies.

Decision trees and neural networks were identified as particularly effective, capturing complex interactions among predictors.

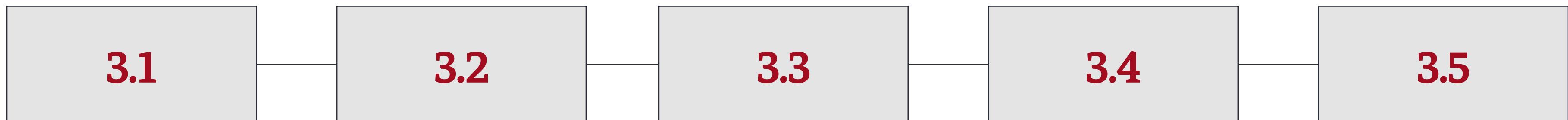
### Beutel et al. (2019)

Found that machine learning methods not only achieved high levels of in-sample accuracy but also outperformed traditional logit models in out-of-sample predictions.



### CASE STUDIES AND REAL-WORLD APPLICATIONS

# 3. DATA AND METHODOLOGY



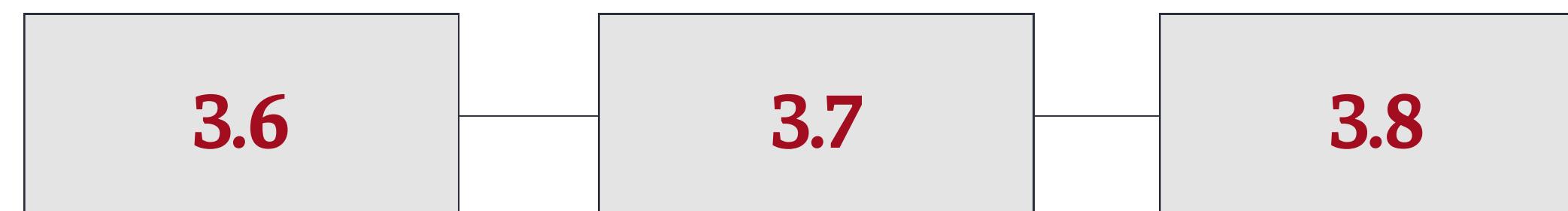
Research process

Data source and data sampling

Model specification

Data preprocessing

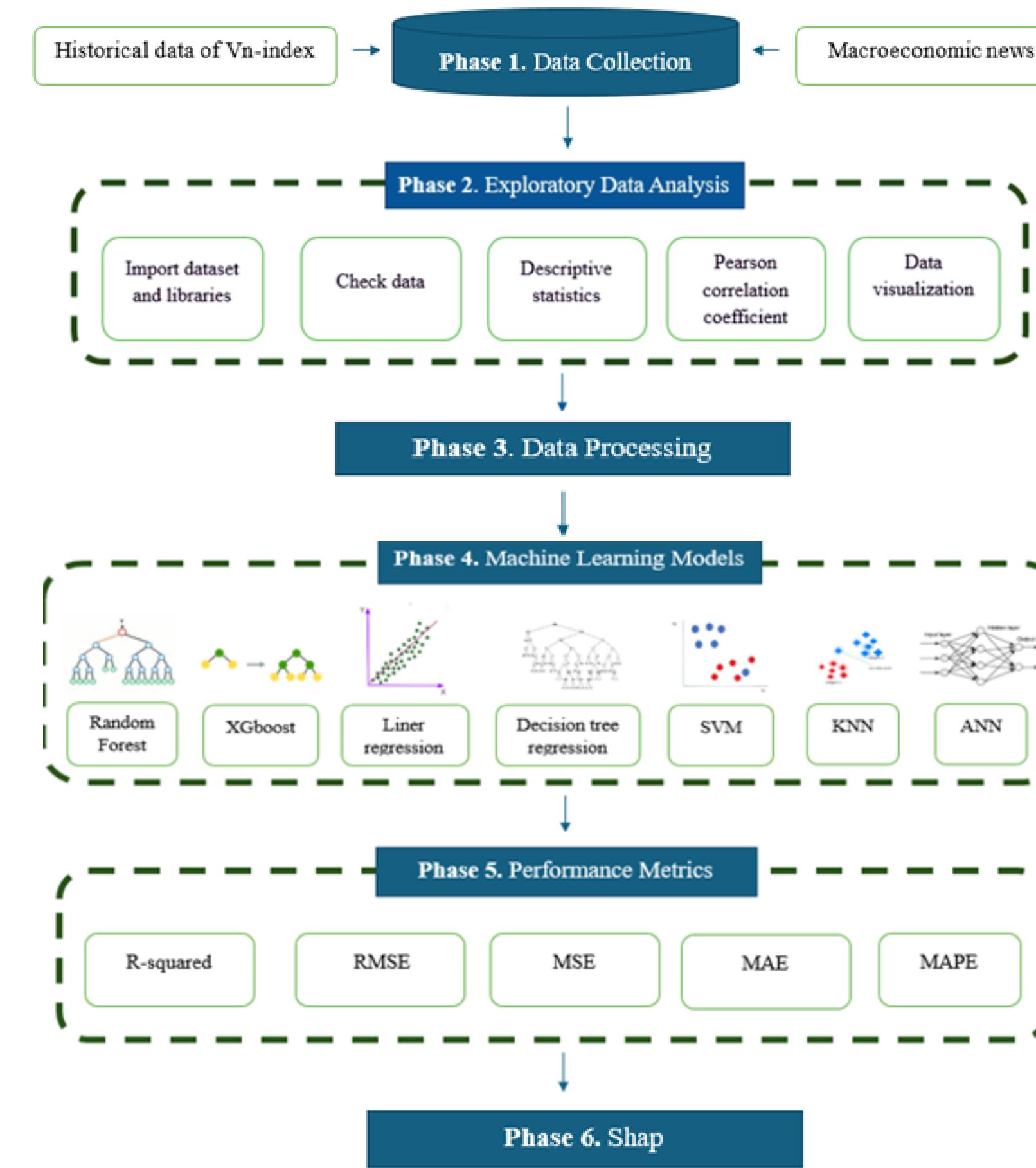
Exploratory Data Analysis



Feature Importance tools

Reasonable choices of Models

Models' performance check



### 3.1. A DIAGRAM OF THE RESEARCH PROCESS

## 3.2. DATA SOURCE AND DATA SAMPLING

### Study Scope:

- Analysis of seven key factors' impact on the VN-Index: global gold prices, global oil prices, M2 money supply volume, interest rates, exchange rates, consumer price index (CPI), and the S&P 500 index.
- Time frame: **March 2016 to March 2024**, providing a comprehensive view of economic and financial factors' influence on the stock market.

### VN-Index Data:

- **Source:** Ho Chi Minh City Stock Exchange <https://www.hsx.vn/>
- **Importance:** Primary index reflecting the stock market movement in Vietnam, with data recorded based on the last trading session's closing prices of each month.

## 3.2. DATA SOURCE AND DATA SAMPLING

### Oil Prices Data:

- Source: <https://www.investing.com/>
- Detail: Monthly average WTI crude oil price listed on the New York Stock Exchange (NYSE), considering crude oil's physical characteristics and sulfur content.

### Money Supply M2 Data:

- Source: Vietnam Securities <https://finance.vietstock.vn/>
- Insight: Represents the total circulating money, crucial for understanding liquidity and economic stability.

### Interest and Exchange Rates Data:

- Sources: State Bank of Vietnam <https://www.sbv.gov.vn/>
- Application: Assessing the economic policy impact through the refinancing rate and the USD/VND exchange rate's influence on the stock market.

### Consumer Price Index (CPI) :

- CPI Source: Vietnam Securities <https://finance.vietstock.vn/>

### Gold Prices Data:

- Source: <https://finance.yahoo.com/>
- Context: CPI measures economy-wide price increases, while gold prices are tracked globally, reflecting investment sentiment.

### S&P 500 Index Data:

- Source: S&P Dow Jones Indices <https://tradingeconomics.com/>
- Relevance: Acts as a gauge for U.S. stock market performance, influencing global market sentiment and the VN-Index.

## 3.3. MODEL SPECIFICATION

### Target Variable

The target variable of the machine learning model in this study is '**Closing\_price**', which represents the **closing price of the security**. The main aim is to **predict the exact closing price based on the defined input variables**.

### FEATURE VARIABLES

Features	Description
Oil price	International crude oil price.
CPI	Consumer price index (which measures inflation) of Vietnam.
Exchange rate	Exchange rate of Vietnamese Dongs to US Dollars.
S&P 500	S&P 500 index, representing the performance of the US stock market..
XAU/USD	World gold price
Money supply M2	The money supply in the Vietnamese economy is categorized into the M2 classification
Interest rate	Base interest rate of the State Bank of Vietnam

## 3.4. DATA PREPROCESSING

### Data Preprocessing:

- Excluded 'Date' for its non-contributory nature in price predictions and 'Closing\_price' as the target variable.
- Utilized `train_test_split` with a `test_size=0.2` and `random_state=42` to ensure a reproducible split, allocating **80% of data for training and 20% for testing**

### Machine Learning Models:

- Applied Models: `RandomForestRegressor`, `LinearRegression`, `DecisionTreeRegressor`, `KneighborsRegressor`, `MLPRegressor`, `SVR`, and `XGBRegressor`.
- Objective: To **learn from training data and predict the closing price accurately**.

### Model Evaluation:

- Metrics Used: `MSE`, `MAE`, `MAPE`, `r2_score` for assessing model performance.
- Comparison and Selection: Models are evaluated based on their predictive accuracy on the test set, guiding the choice of the most effective model for stock price prediction.

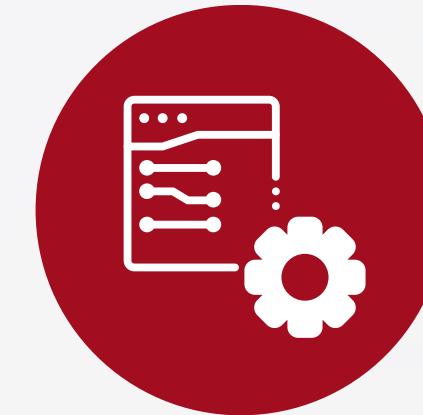
## 3.5. EXPLORATORY DATA ANALYSIS



Descriptive  
statistics



Pearson  
correlation  
coefficient



Data  
visualization

## 3.6. FEATURE IMPORTANCE TOOLS

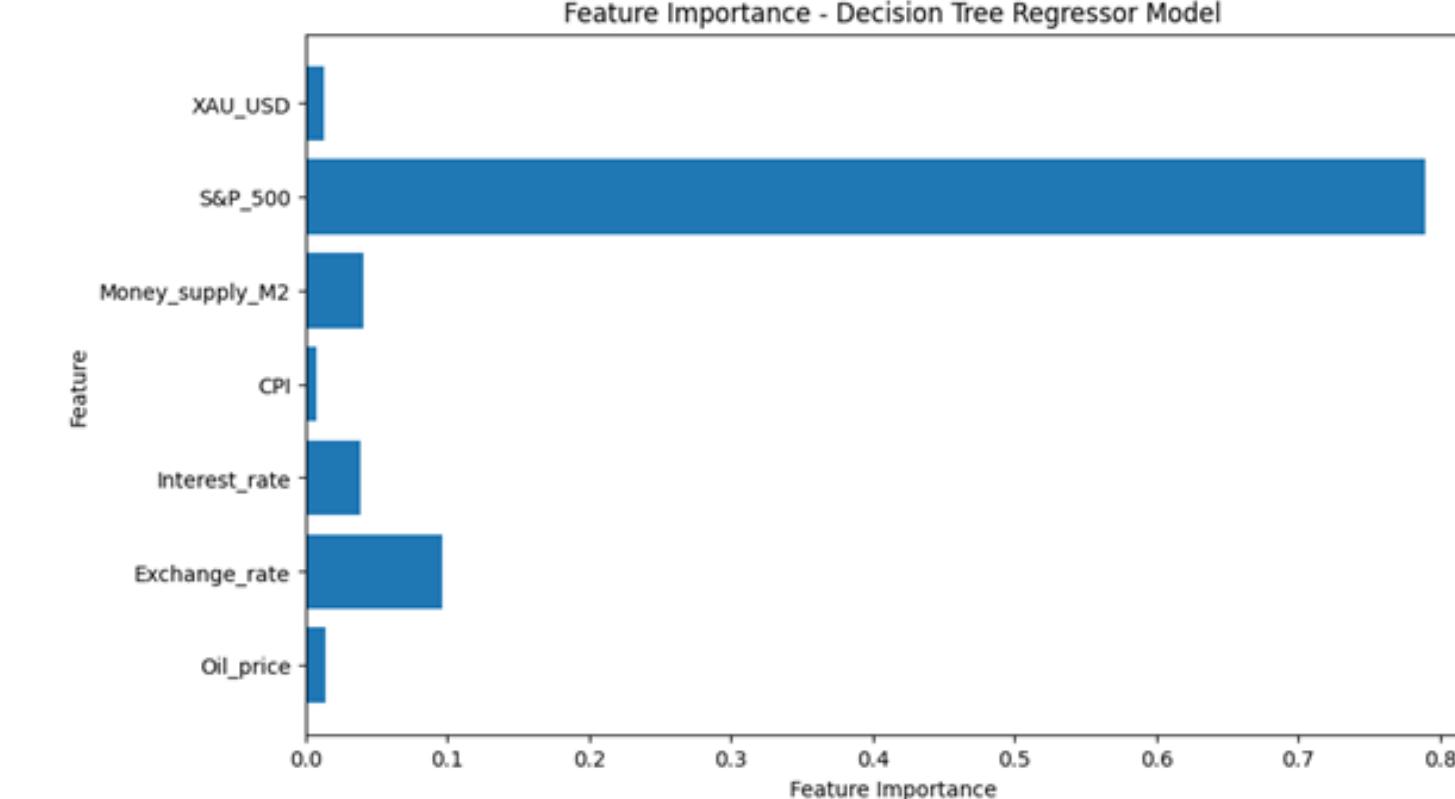
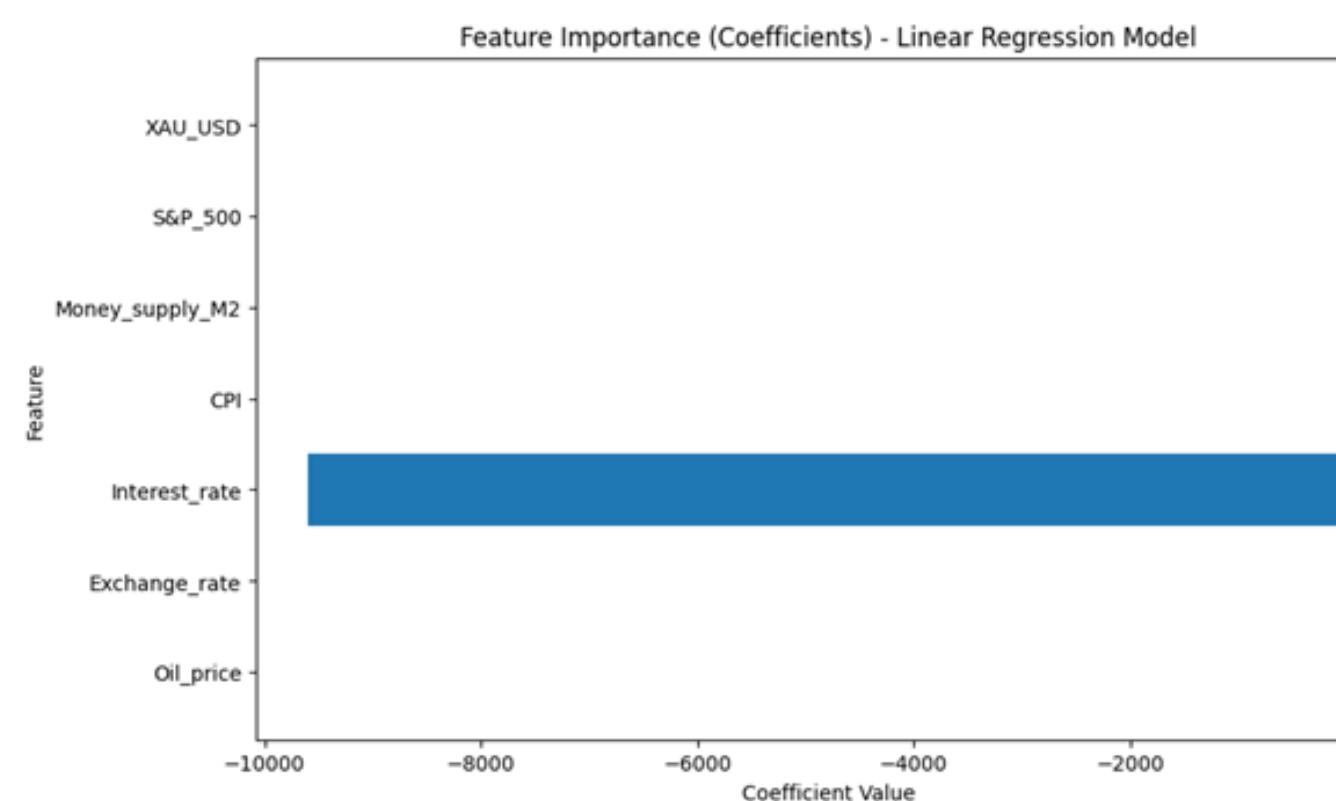
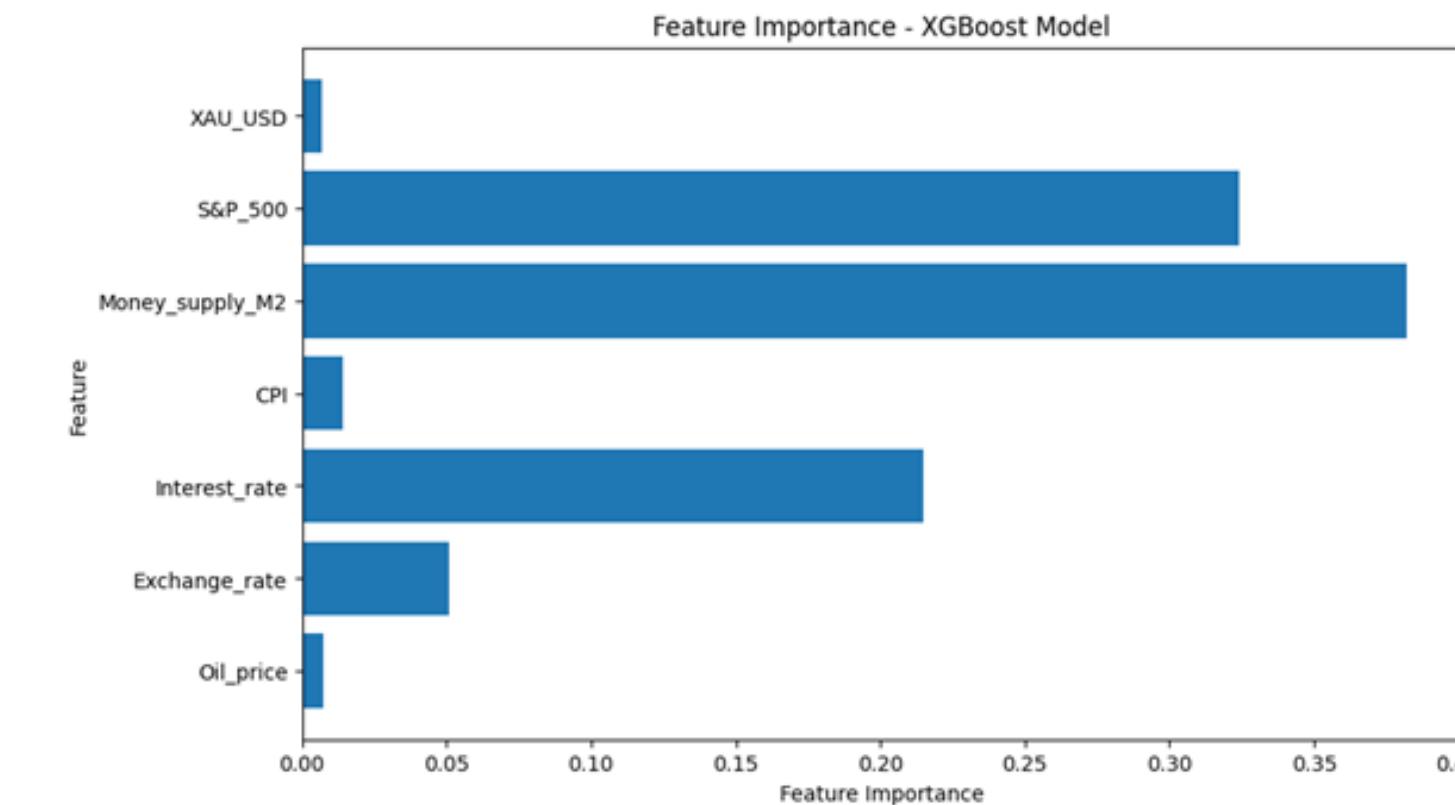
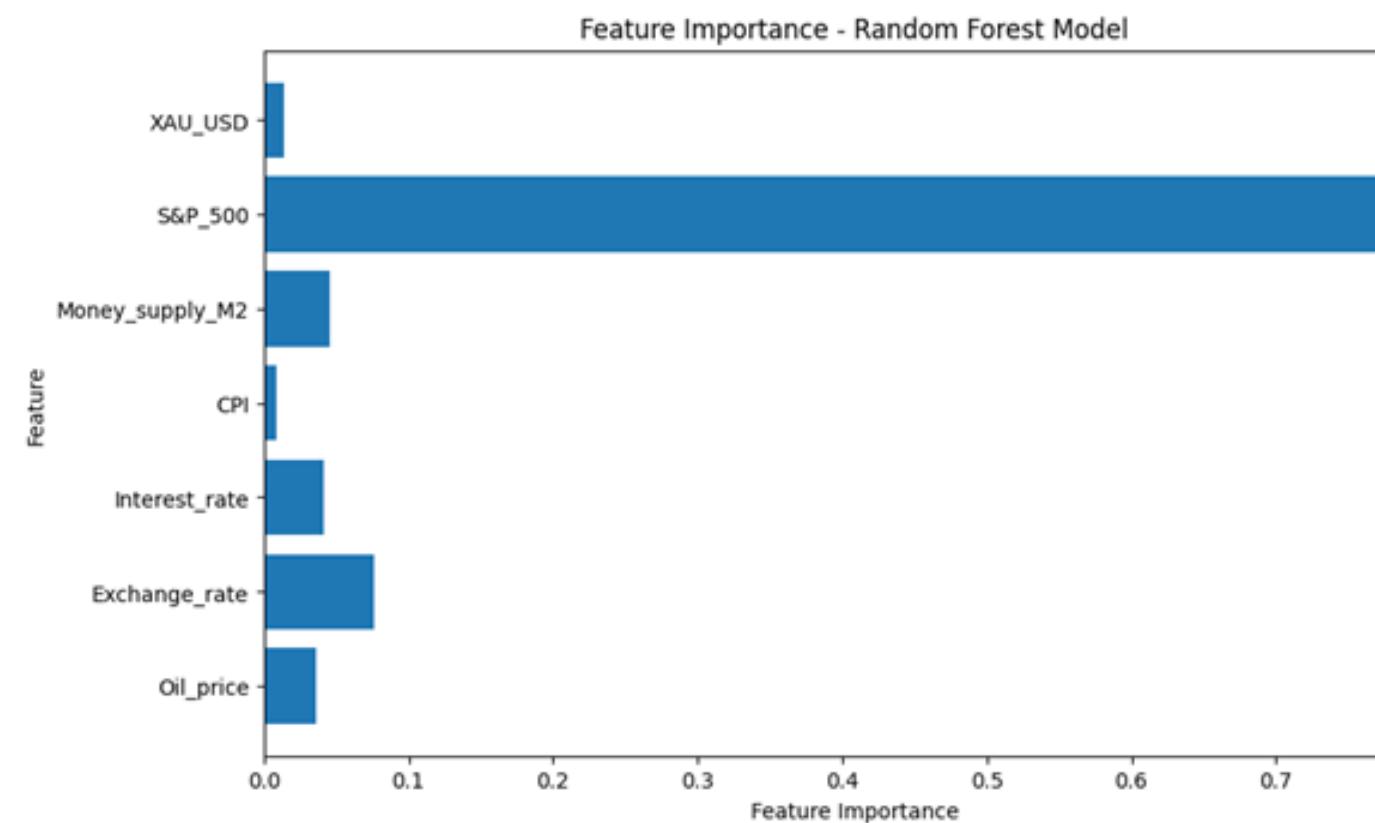
3.6.1

Feature  
Importance

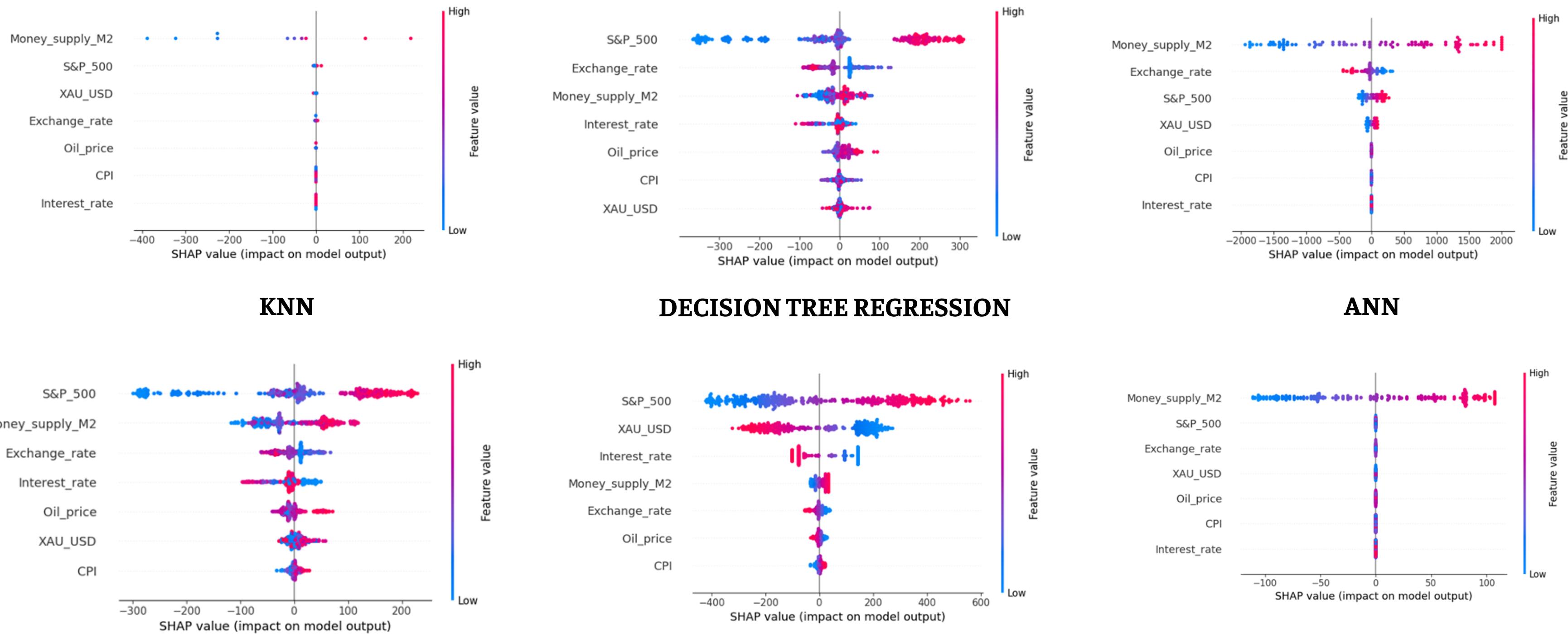
3.6.2

SHAP

## 3.6.1. FEATURE IMPORTANCE



## 3.6.2. SHAP



**XG BOOST**

**LINER REGRESSION**

**SVM**

## 3.6. FEATURE IMPORTANCE TOOLS

**S&P 500** with the strongest influence, reflecting the trend of the global stock market; money supply (**M2**), which shows the influence of economic liquidity; **exchange rate** fluctuations, impacting market confidence and investment decisions; **interest rate**, with an inverse relationship with the market and predicting the movement of the VN-index; Oil prices and gold prices, both have a significant influence with oil due to its economic importance and gold as a haven asset; and finally the Consumer Price Index (CPI) with the least impact, suggesting inflation is not an immediate concern

## 3.7. Reasonable choices of Models

### Versatility

Selected models like Random Forest, XGBoost, and Decision Tree Regression for their ability to **handle complex, nonlinear data relationships** prevalent in financial datasets.

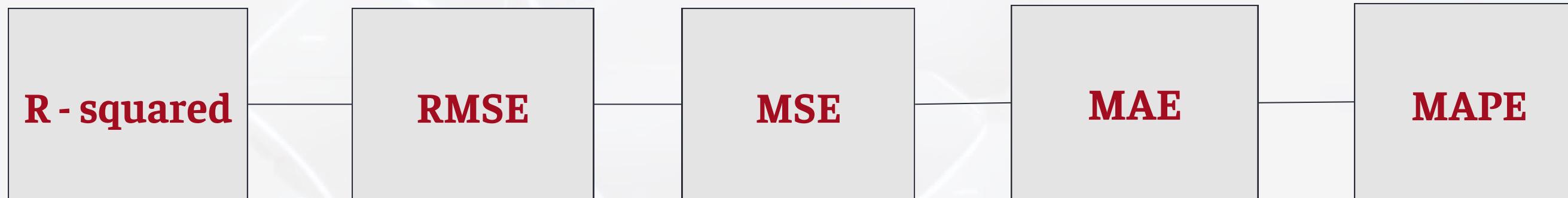
### Overfitting Prevention

Emphasis on models capable of **avoiding overfitting** is crucial for **maintaining model reliability** over time and across varying market conditions.

### Feature Importance Analysis

Chose models that offer insights into the **significance** of different predictors affecting the VN-Index, enabling a more informed interpretation of model outcomes.

## 3.8. MODELS' PERFORMANCE CHECK



## 4. RESULTS AND DISCUSSIONS

4.1

Results of the  
**Exploratory  
Data Analysis**

4.2

Results of the  
**Performance  
Metrics**

4.3

Results of  
**Model  
Comparison**

## 4.1. Results of the Exploratory Data Analysis

	Closing_price	Oil_price	Exchange_rate	Interest_rate	CPI	Money_supply_M2	S&P_500	XAU_USD
count	2002.000000	2002.000000	2002.000000	2002.000000	2002.000000	2.002000e+03	2002.000000	2002.000000
mean	1008.562133	65.705604	23095.115897	0.053988	0.264266	1.089839e+07	3323.271973	1564.271224
std	228.499609	21.026712	612.182465	0.010220	0.440917	2.782728e+06	838.329327	280.207245
min	555.820000	7.790000	21779.000000	0.040000	-1.540000	6.267958e+06	2000.540000	1128.430000
25%	856.285000	55.670000	22727.500000	0.044000	0.010000	8.521098e+06	2642.247500	1284.822500
50%	988.315000	66.820000	23171.000000	0.060000	0.250000	1.075802e+07	3099.015000	1552.695000
75%	1166.967500	77.655000	23321.574707	0.062500	0.510000	1.376102e+07	4129.542500	1824.365000
max	1528.570000	128.260000	24871.000000	0.065000	1.520000	1.535412e+07	5088.800000	2063.810000

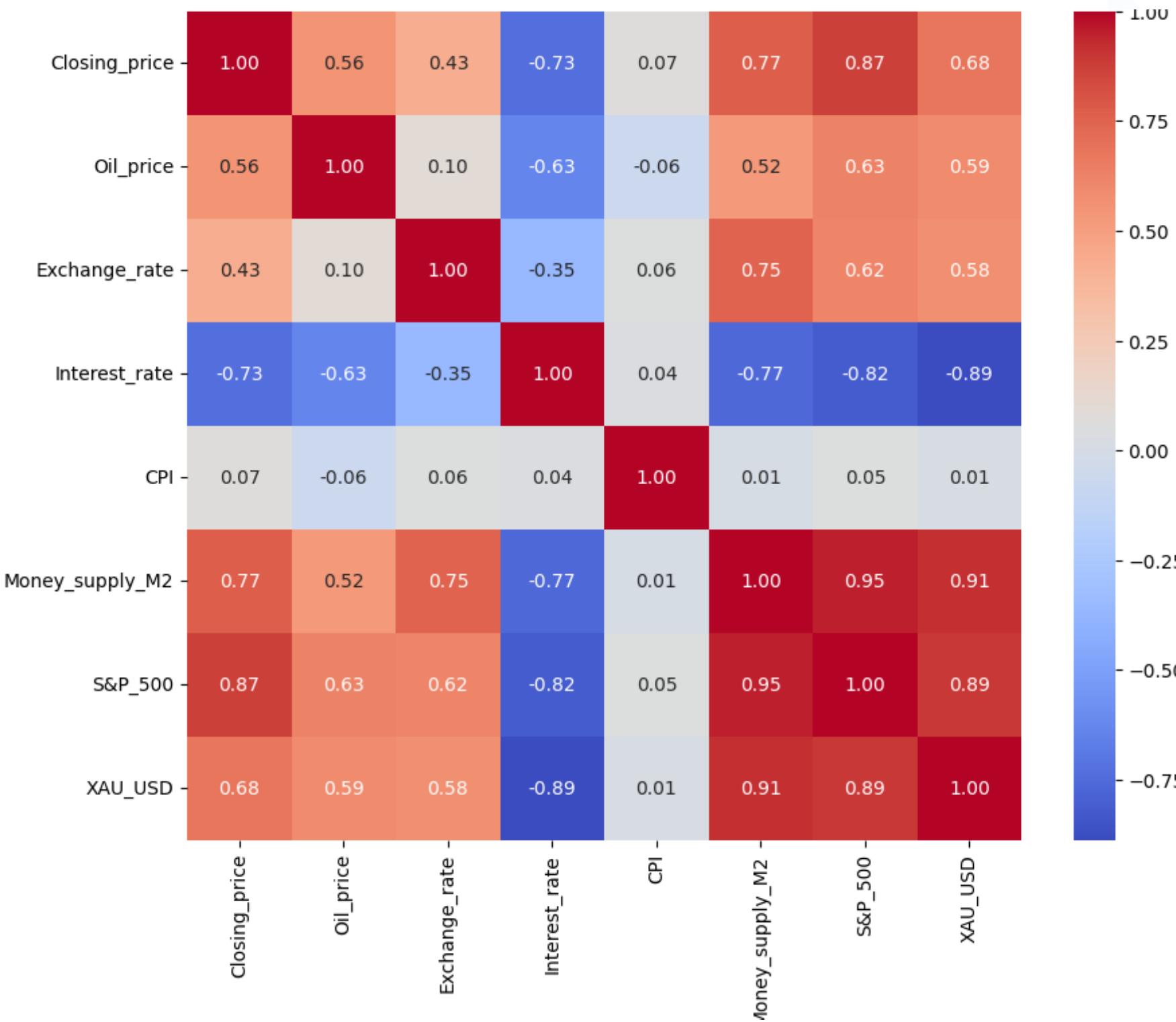
STATISTICAL DESCRIPTION OF NUMERIC VARIABLE

## 4.1. Results of the Exploratory Data Analysis

Oil prices, with an average of 65.71 units, and the S&P 500 index demonstrated considerable volatility, whereas interest rates showed relative stability with a mean of 5.39%



# 4.1. Results of the Exploratory Data Analysis

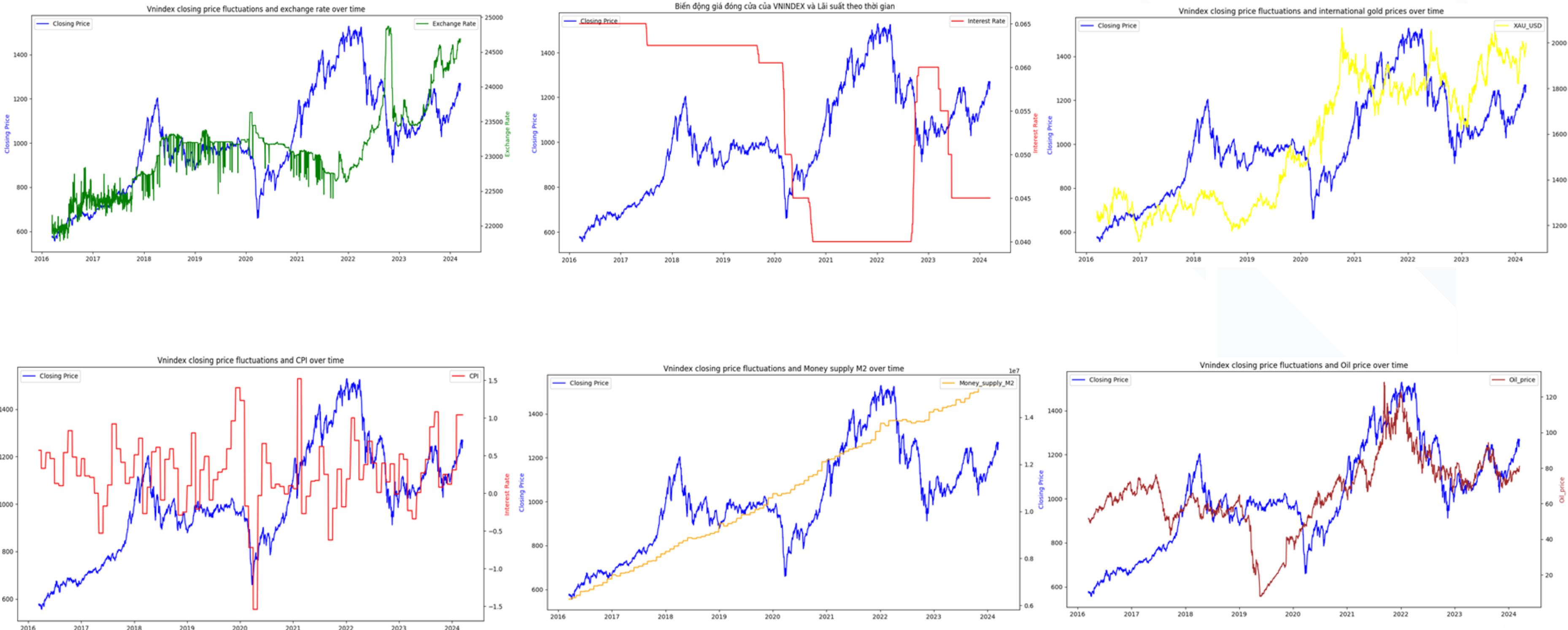


CORRELATION MATRIX BETWEEN VARIABLES

## 4.2. Results of the Performance Metrics

- **Strong positive correlations between Closing\_price** with both **S&P\_500** and **Money\_supply\_M2**, suggest the influential role of investor sentiment and the money supply in driving stock prices.
- In contrast, **a strong negative correlation exists between interest** and both **Closing\_price** and **S&P\_500**, highlighting how rising **interest rates**, by increasing borrowing costs, tend to depress stock and gold prices.
- **The correlation between S&P\_500 and Money\_supply\_M2** underscores the significant impact of monetary policy on stock market fluctuations.
- **Gold prices**, moderately correlated with **Closing\_price** and **S&P\_500**, affirm **gold's status as a haven or hedge in volatile markets**.
- Conversely, **oil prices** show **a weaker correlation**, reflecting the complex dynamics affecting oil market volatility, such as geopolitical factors and global demand, independent of broader economic trends.
- **CPI's ambiguous correlation** with market variables suggests inflation's impact on stock or gold prices may be indirect or delayed, pointing to the multifaceted influences on market movements.

# 4.1. Results of the Exploratory Data Analysis



VOLATILITY OF CLOSING PRICES WITH INDEPENDENT VARIABLES FROM 2016 - 2024

## 4.2. Results of the Performance Metrics

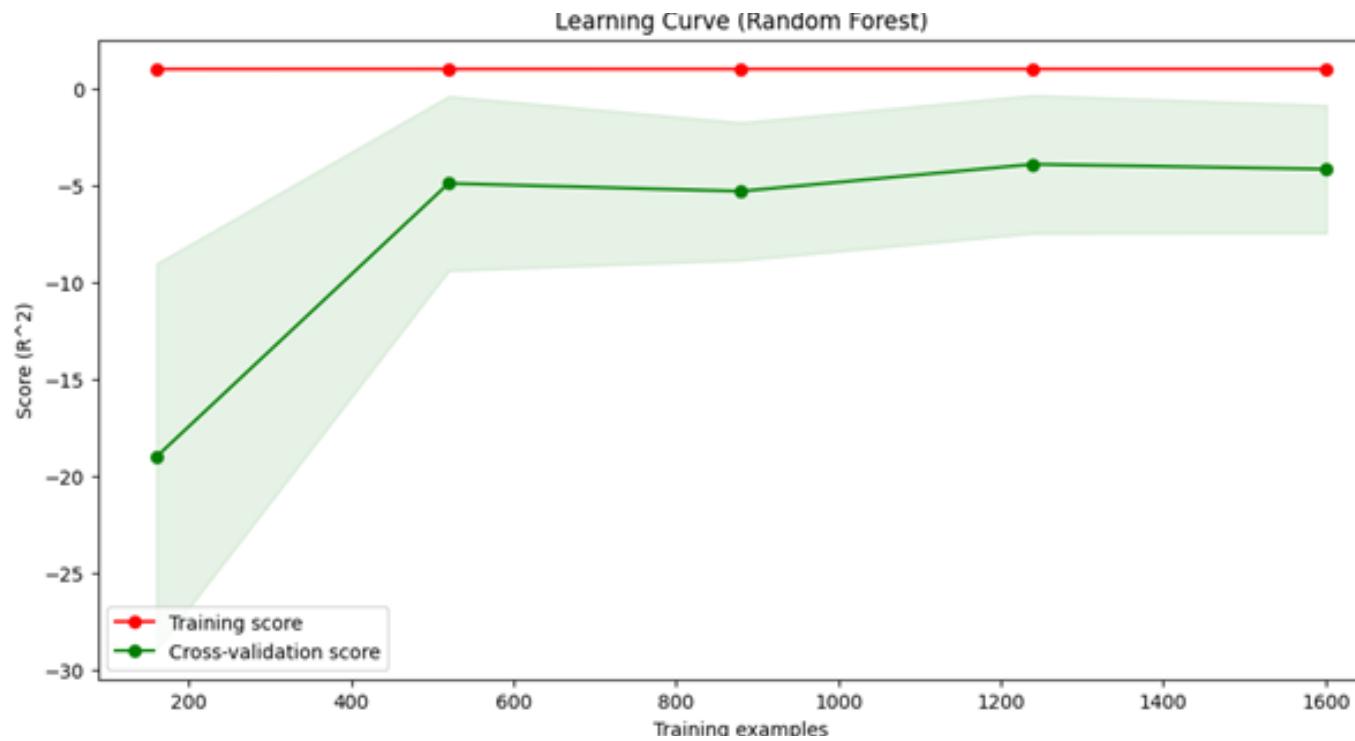
	Model	MSE	RMSE	MAE	MAPE	R-Squared
1	Random Forest	4.197507e+02	20.487818	12.420641	1.209722	0.991750
2	XGBoost	4.959086e+02	22.269005	14.214132	1.391971	0.990253
3	Linear Regression	7.935919e+03	89.083775	66.476096	6.860058	0.844021
4	Decision Tree	7.247153e+02	26.920537	15.114913	1.489434	0.985756
5	KNN	4.476300e+02	21.157269	13.684214	1.350528	0.991202
6	ANN	5.519568e+06	2349.376165	1844.781844	168.855271	-107.485870
7	SVM	3.122309e+04	176.700570	136.190948	14.281501	0.386317

CORRELATION MATRIX BETWEEN VARIABLES

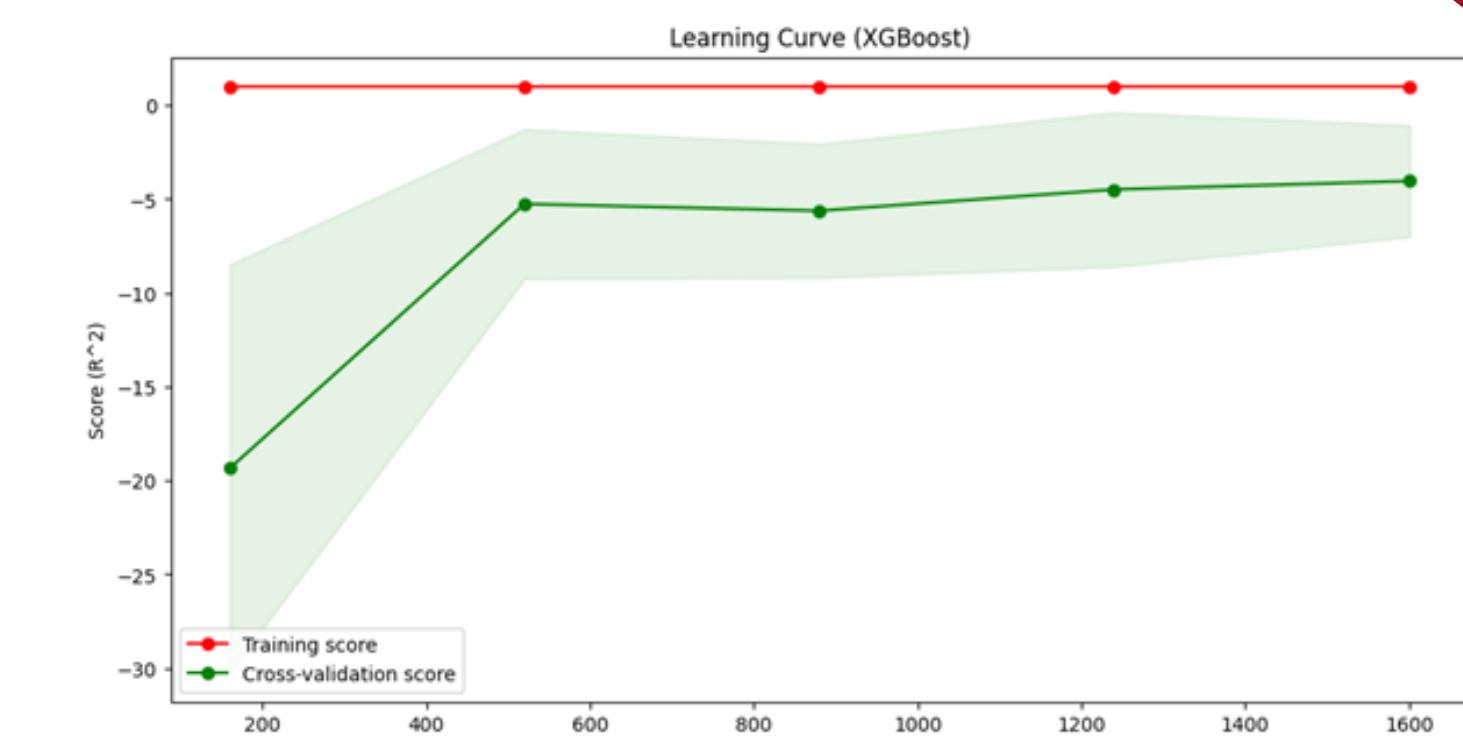
## 4.2. Results of the Performance Metrics

Tree-based models (Random Forest, XGBoost, and Decision Tree) show good predictive performance, with Random Forest emerging as the best model. KNN has a relatively good performance Squared, but has a slightly higher MSE and RMSE, indicating a certain degree of inaccuracy. Linear Regression is not suited for this data type and the problem at hand. ANN and SVM do not suite with the data and current approach and need to be adjusted or replaced with alternative models.

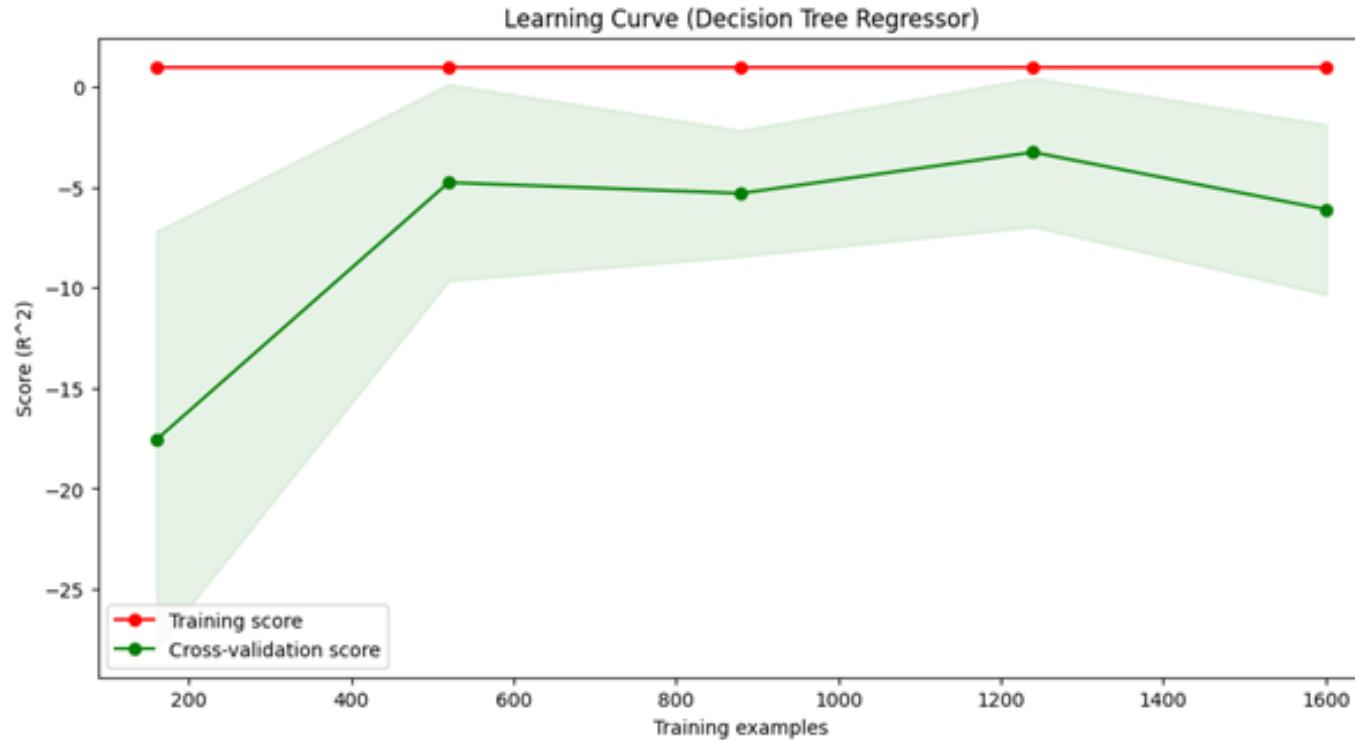
## 4.3. Results of Model Comparison



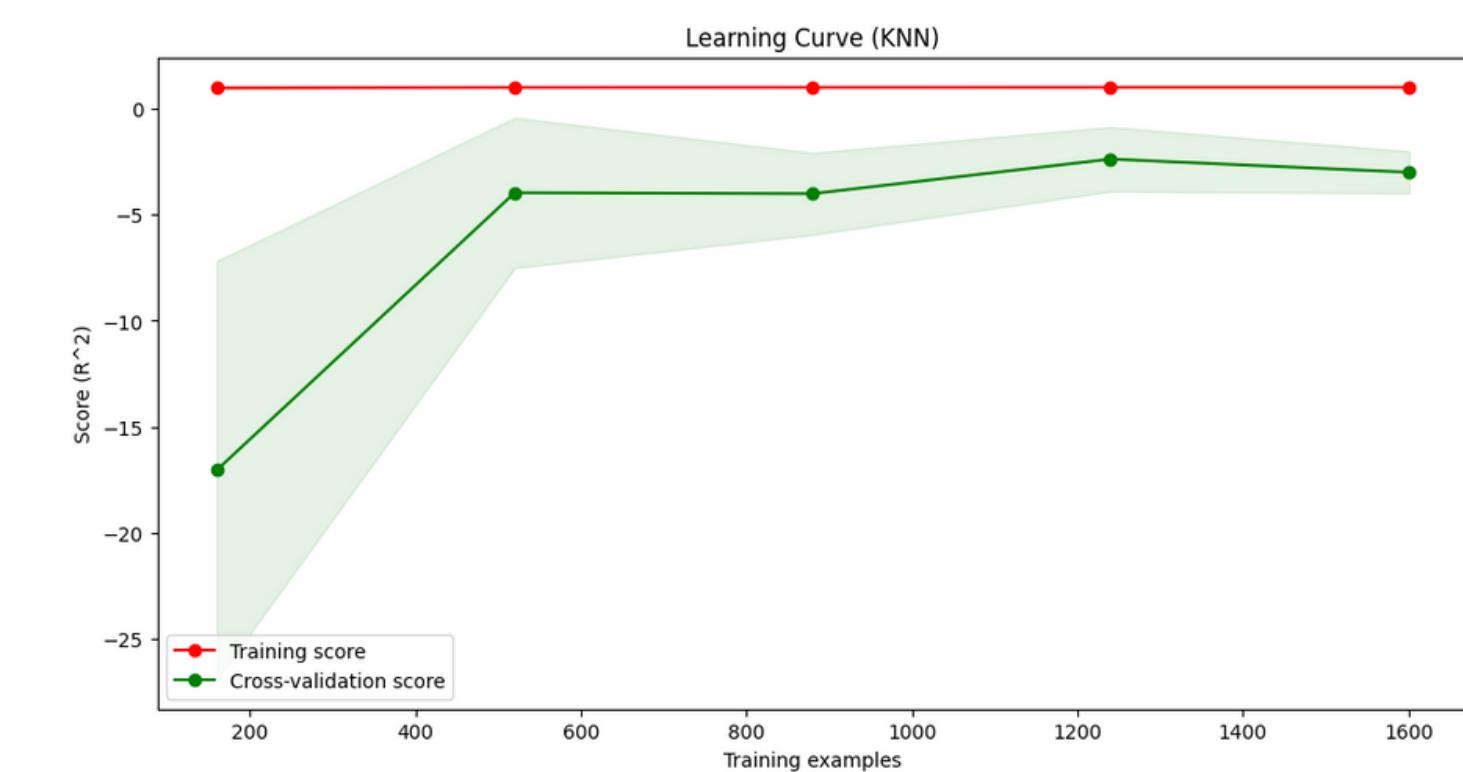
LEARNING CURVE OF RANDOM FOREST



LEARNING CURVE OF XGBOOST

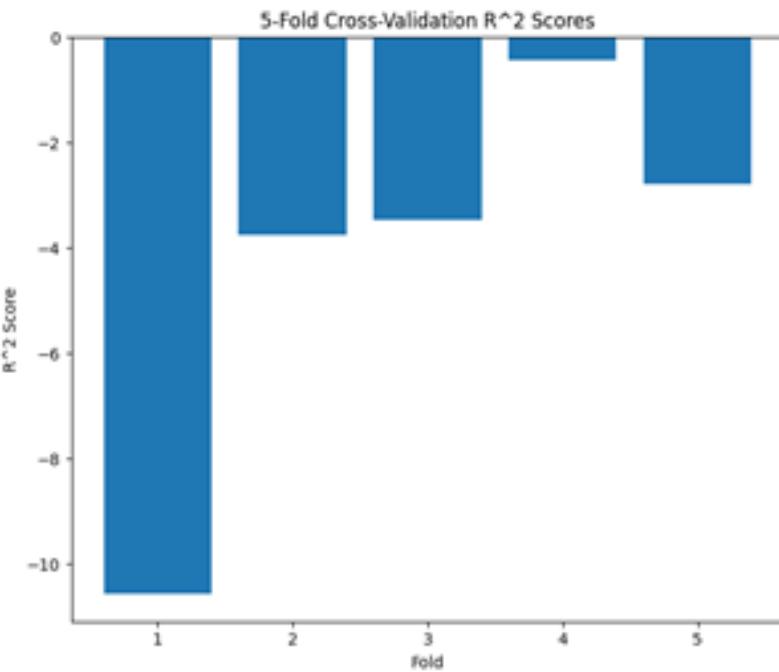


LEARNING CURVE OF DECISION TREE REGRESSION

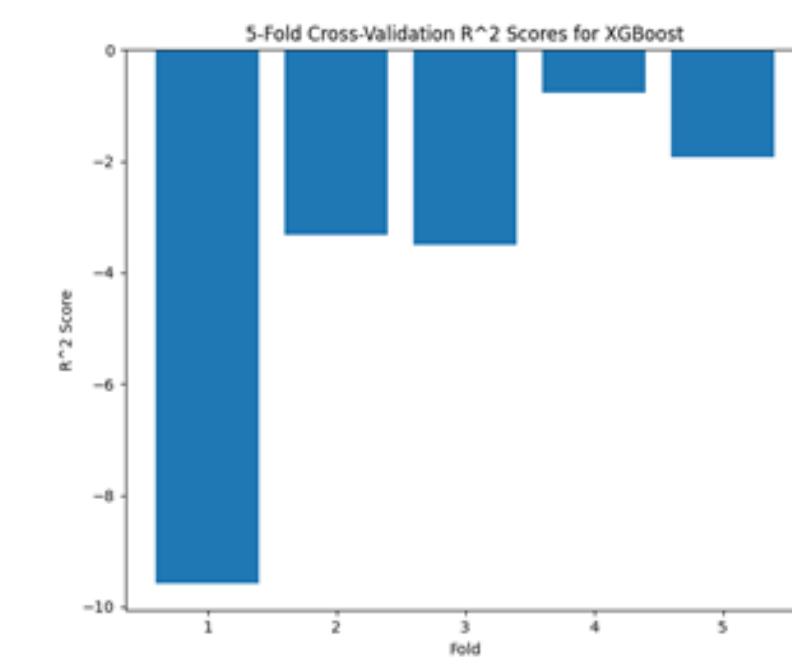
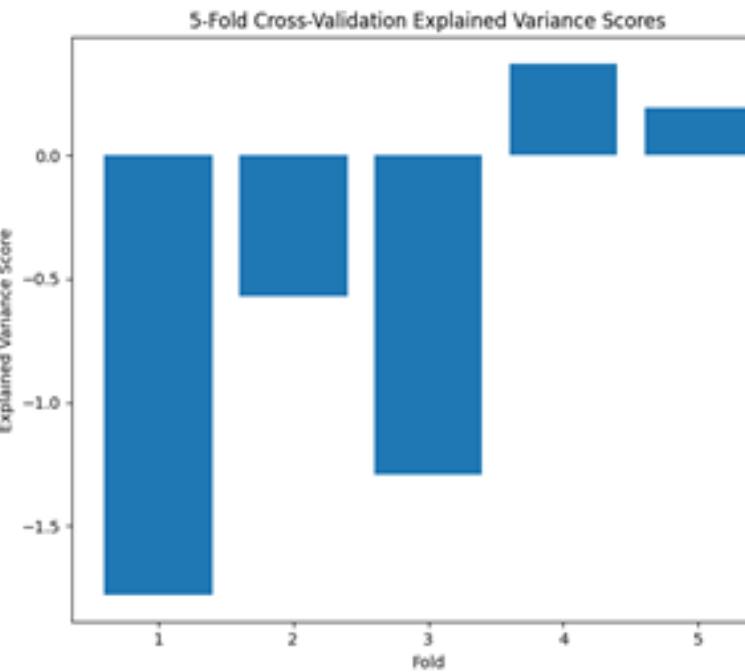


LEARNING CURVE OF KNN

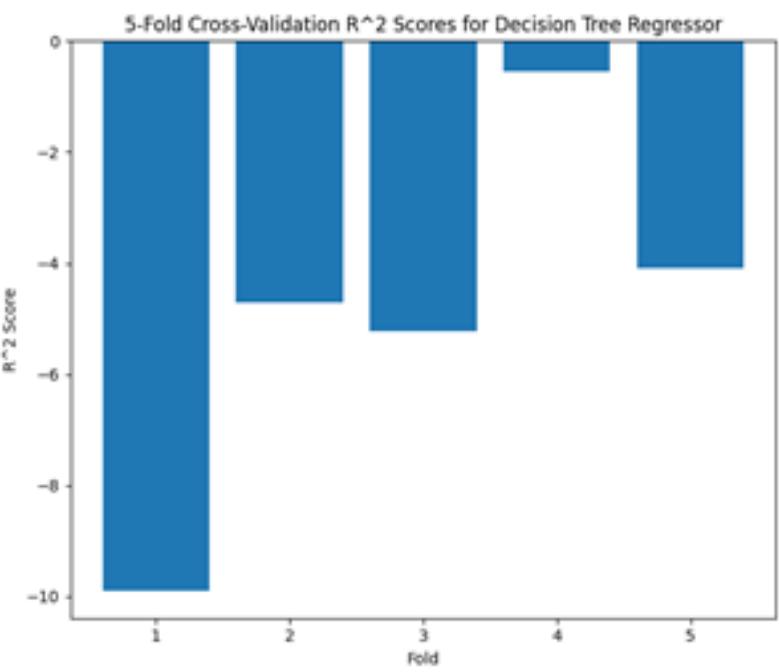
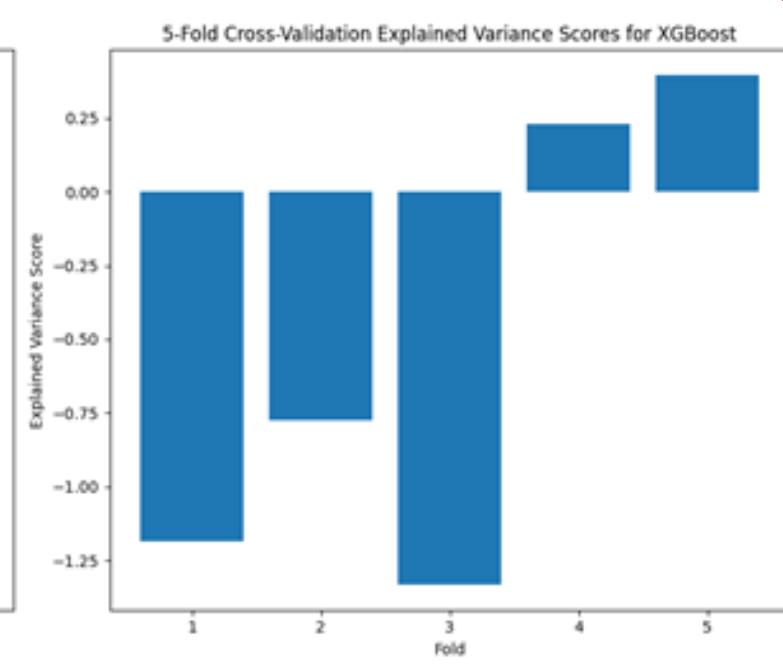
## 4.3. Results of Model Comparison



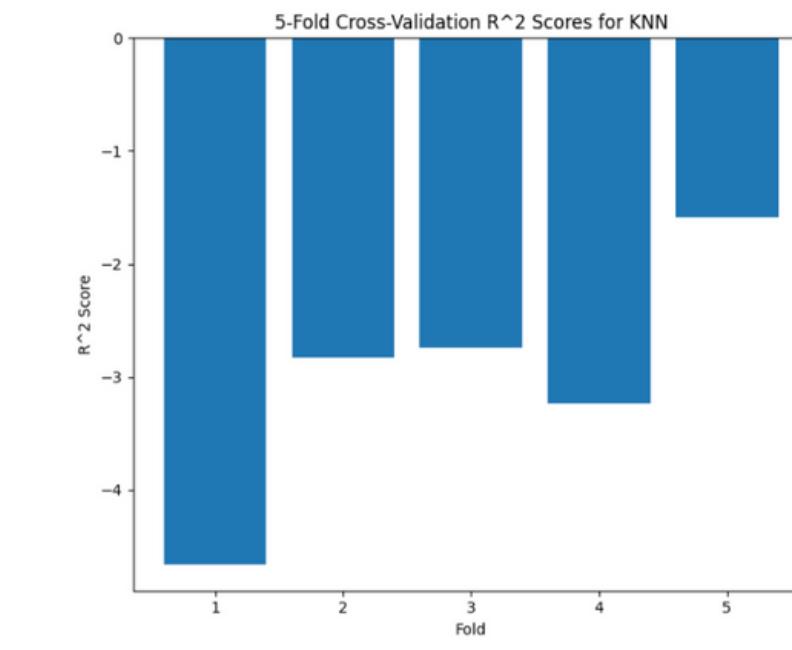
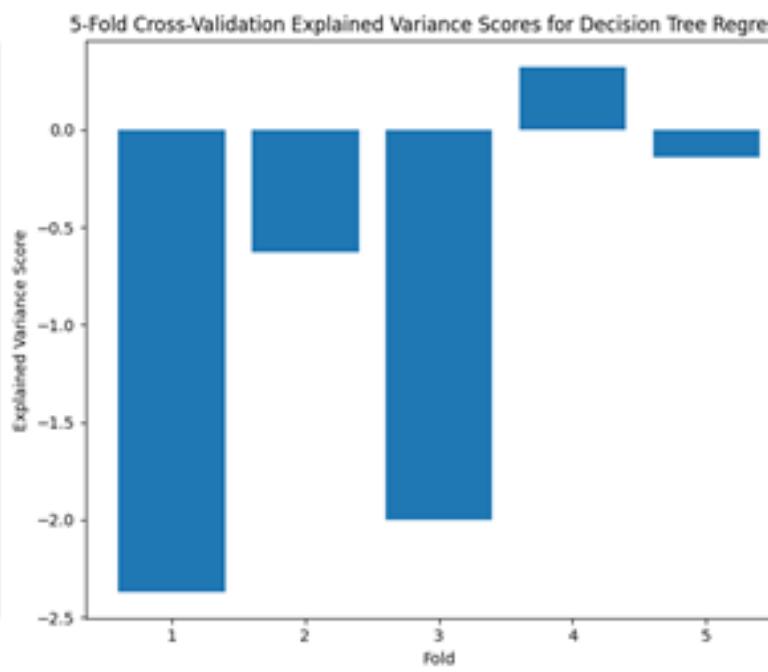
CROSS-VALIDATION SCORES OF RANDOM FOREST



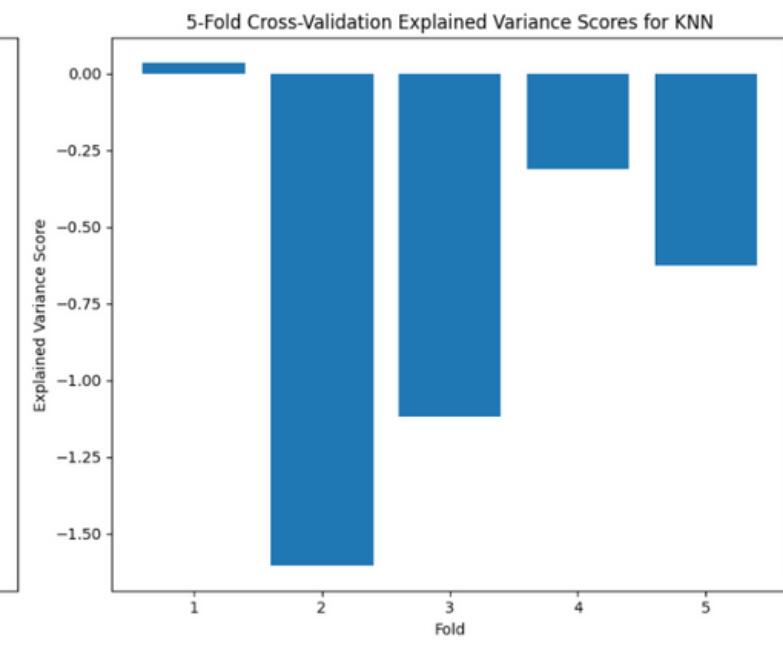
CROSS-VALIDATION SCORES OF XGBOOST



CROSS-VALIDATION SCORES OF DECISION TREE REGRESSION



CROSS-VALIDATION SCORES OF KNN



## 4.3. Results of Model Comparison

All four models show a good learning curve, but the Random Forest and XGBoost models display better generalization capabilities with less variance in the validation scores, suggesting they would perform better on new, unseen data compared to the Decision Tree Regressor. Finally, summarize the results, although both 2 models show excellent predictive capabilities, but the Random Forest seems to perform slightly better across most metrics and displays a more stable learning curve, suggesting better generalization. The R-squared values from the cross-validation (as seen in the bar charts) are also consistently higher for Random Forest than for XGBoost, which further supports the conclusion that Random Forest is the better model out of the two for this particular dataset.

In summary, for this particular set of data, the **Random Forest model** is the recommended choice due to its superior ability to capture the underlying patterns in the data, making it the most effective model among those tested. The remaining **two tree models, XGBoost and Decision Tree**, also show good performance, but **not as good as Random Forest**. Additionally, SVM and KNN, **did not show good performance**. Linear Regression is not suited for this data type. Meanwhile, the Deep Learning model ANN does not seem to be suitable for the data set and approach applied in this study.

## 5. CONCLUSIONS AND RECOMMENDATIONS

5.1

Conclusions

5.2

Recommendations

5.3

Limitations  
and further  
research

## 5.1. Conclusions

This study focuses on **identifying key factors influencing the VN-Index** using machine learning techniques, based on **macroeconomic news** and **technical analysis** from **March 2016 to March 2024**. Machine learning models used include Decision Tree Regression, Random Forest, XGBoost, KNN, ANN, SVM, and Linear Regression. The results show that **Random Forest** has superior prediction performance compared to other models, highlighting its superiority in predicting VN-Index movements based on macroeconomic variables. **SHAP** analysis indicates that the **S&P 500 index, M2 currency quantity, exchange rate and interest rate** have a significant influence on the Vietnamese stock market, while global oil prices, CPI and global gold prices have less impact. The study highlights the importance of global stock market trends, especially **the S&P 500**, and **domestic monetary policy**, especially M2 currency, in **shaping stock market dynamics**, thereby **emphasizing the close connection of global financial markets and the important role of monetary stability in market health**.



## First Recommendation

For **policymakers and regulatory bodies**, they should **enhance the financial market's resilience** by strengthening legal frameworks to ensure transparency and stability. Furthermore, they need to **implement policies** to maintain an optimal money supply, thereby preventing inflation or deflation that could destabilize the market. Simultaneously, they should **promote more monitoring and analysis of global economic trends** to predict their impact on the domestic market.



## Second Recommendation

For **Investors**, they should **incorporate advanced analytical tools and machine learning models** in investment strategies to better predict market movements. Diversify investment portfolios to **hedge against potential risks** posed by **global economic volatility** and **domestic monetary policy adjustments**.



## Third Recommendation

For **researchers and academics**, they should continue the exploration of machine learning applications in financial market analysis, with an **emphasis on developing models** that **can adapt to the rapid changes** in economic indicators. Encourage **interdisciplinary studies** that combine economics, finance, and data science to uncover deeper insights into market dynamics.

# LIMITATION

## The first limitation

The data span from March 2016 to March 2024 might not fully represent market behaviors, particularly during rare economic events.

## The second limitation

while models like Random Forest and XGBoost are effective, their ability to generalize across different contexts may be limited

## The third limitation

The focus on selected macroeconomic variables overlooks other potential influencers within the complex global financial system

## The fourth limitation

Rapid advancements in machine learning could outdated the current methodologies

## The fifth limitation

The models may not accurately predict unforeseeable Black Swan events, highlighting the need for adaptable forecasting techniques.



## 5.3. FURTHER RESEARCH

Future research should address these limitations by expanding the dataset to cover a broader time frame, including a wider range of variables, and exploring advanced modeling techniques. Investigating the impact of qualitative factors and developing models capable of adapting to unforeseen events could also enhance the robustness and applicability of future studies. This expanded approach will contribute to a more holistic understanding of the dynamics influencing the VN-Index and other financial indices.

**I WOULD LIKE TO SINCERELY  
THANK THE TEACHERS FOR  
LISTENING**



**Q&A SESSION**

**Class:** K20414C

**Student:** Nguyen Thanh Phuong Uyen

**Student Code:** K204141939

**Instructor:** Dr. Pham Thi Thanh Xuan

