Data Science Project

Intelligent Trading Agent

Chaima Gharbi

Aziz Khalsi

Sondes Tlili

Houssem Jabally

Amal Rhimi

Mohamed Ben Hmida

Hadil Amamou

4 DS 3

2023/2024

Acknowledgement

We'd like to thank everyone who helped make this intelligent trading platform report a success.

First and foremost, we’d want to thank our project supervisors for their experience, direction, and steadfast support throughout the project's development and conclusion. Your wise suggestions and support have been quite beneficial.

We are grateful to VALUE for providing us with the ressources and assistance we needed during this endeavor and for their assistance and sharing their extensive expertise and experience.

We thank our colleagues and team members for their hard work, teamwork, and devotion. Your collaboration and common vision were vital to attaining our goals.

In addition, we would like to thank our family and friends for their encouragement and patience, which helped us stay focused and motivated.

Finally, we are grateful for the important input and ideas from the academic and professional communities, which helped us refine and improve our work.

Thank you for your contributions and support.

Summary

Summary **2**
Table of Figures **4**
Introduction **1**
Chapter 1: Business Understanding **2**
I. Business Overview **2**
1. Problematic 2
2. Portfolio 2
II. Project Scope **2**
1. Business Objective 2
2. Data Science Objective 2
3. Added Value 3
4. Metrics 3
5. Methodology 4
6. Conclusion 4
Chapter 2: Data Acquisition and Understanding **5**
1. Introduction 5
2. Data Acquisition & Understanding 5
1.1 Data Acquisition 5
1.1.1 Web Scrapping 6
1.2 Data Understanding 7
3. Conclusion 7
Chapter 3: Data Preparation **8**
1. Introduction 8
2. Data Cleaning 8
2.1 Market Data 8
2.1.1 Correlation 8
2.1.2 Checking for Missing Data 9
2.1.3 Handling Missing Data Using Linear Imputation 9
2.2 Sector Data 11
2.2.1 Correlation 11
2.2.3 Handling Missing Values using Interpolate with Linear Method 12
3. Feature Engineering 12
3.1 Calculate Features 12
3.2 Combining Each Company with each Sector 14

4. Time Series Preprocessing 14
5. Normalization and Scaling 15
6. Handling Categorical Data 16
7. Data Splitting 16
8. Feature Selection 16
9. Conclusion 17
Chapter 4: Features Extraction 18
1. Introduction 18
2. Relative Performance Metrics 18
2.1 Relative Performance of Company to Sector 18
2.2 Relative Performance of Sector to Company 18
3. Technical Analysis Indicators 18
4. Lag Features 20
5. Support and Resistance Levels 21
6. Conclusion 22
Chapter 5: Modeling 23
1. Introduction 23
2. Data Collection Model 23
3. Prediction Model 25
3.1 Model Architecture 26
3.2 Training Process 26
3.3 Evaluation and Metrics 27
4. Strategies 28
4.1 Optimal Growth Strategy 28
4.2 Volatility Minimization Strategy 30
4.3 Maximising Returns Strategy 32
4.4 Trend Following Strategy 34
4.5 Positive Reputation Strategy 36
4.6 Investment by sector strategy 38
5. Conclusion 40
Chapter 6: Deployment 41
1. Introduction 41
2. Web deployment 41
2.1 Home page 41
2.2 Login page 42
2.3 Dashboard page 42
2.4 Strategies page 44
2.5 Profile page 45
3. Conclusion 46
Conclusion 46
References 48

Table of Figures

Figure 1: TDSP 4
Figure 2: Data Sources 6
Figure 3: web scrapping schema 7
Figure 4: correlation database 8
Figure 5: correlation matrix 9
Figure 6: missing values 9
Figure 7: handling missing values 10
Figure 8: sector data correlation 11
Figure 9: sector correlation matrix 11
Figure 10: handling missing values for sector 12
Figure 11: data with extracted features 13
Figure 12: extracted features' correlation matrix 13
Figure 13: Combining Each Company with each Sector 14
Figure 14: Time Series Preprocessing 14
Figure 15: Normalization and Scaling 15
Figure 16: PCA 16
Figure 17: ANOVA 16
Figure 18: Simple Moving Average (SMA) 18
Figure 19: Exponential Moving Average (EMA) 18
Figure 20: Relative Strength Index (RSI) 19
Figure 21: Bollinger Bands 19
Figure 22: Lag Features 20
Figure 23: Pivot Points 20
Figure 24: Fibonacci Retracement 21
Figure 25: Chart Patterns 21
Figure 26: Sentiment analysis Data 22
Figure 27: vader model 23
Figure 28: XLNet model 23
Figure 29: Robert model 23
Figure 30: Bert Model 24
Figure 31: Sentiment Analysis models comparison 24
Figure 32: Training and validation MAE 26
Figure 33: Calculated Sharpe Ratio for each company 27
Figure 34: portfolio mertics' values for optimal growth strategy 27

Figure 35: annualized return and volatility values for optimal growth strategy 27
Figure 36: investment allocation for optimal growth strategy 28
Figure 37: portfolio performance for optimal growth strategy 28
Figure 38: annual volatility for each company 29
Figure 39: portfolio mertics' values for volatility Minimization Strategy 29
Figure 40: annualized return and volatility values for volatility Minimization Strategy 29
Figure 41: investment allocation for volatility Minimization Strategy 30
Figure 42: portfolio performance for volatility Minimization Strategy 30
Figure 43: annualized Return for each company 31
Figure 44: top 20 companies 31
Figure 45: portfolio mertics' values for Maximizing Returns Strategy 31
Figure 46: annualized return and volatility values for Maximizing Returns Strategy 32
Figure 47: investment allocation for Maximizing Returns Strategy 32
Figure 48: portfolio performance for Maximizing Returns Strategy 32
Figure 49: companies where the signal equals 1 33
Figure 50: portfolio mertics' values for Trend Following Strategy 33
Figure 51: annualized return and volatility values for Trend Following Strategy 34
Figure 52: investment allocation for Trend Following Strategy 34
Figure 53: portfolio performance for Trend Following Strategy 34
Figure 54: companies exhibiting notably high scores in 'positive sentiment_bert.' 35
Figure 55: portfolio mertics' values for Positive Reputation Strategy 35
Figure 56: annualized return and volatility values for Positive Reputation Strategy 35
Figure 57: investment allocation for Positive Reputation Strategy 36
Figure 58: portfolio performance for Positive Reputation Strategy 36
Figure 59: sector selection 36
Figure 60: portfolio mertics' values for Investment by sector strategy 37
Figure 61: annualized return and volatility values for Investment by sector strategy 37
Figure 62: investment allocation for Investment by sector strategy 37
Figure 63: portfolio performance for Investment by sector strategy 38
Figure 64: home page 39
Figure 65: login page 40
Figure 66: dashboard 41
Figure 67: strategies 42
Figure 68: strategy form 43
Figure 69: portfolio allocation 44
Figure 70: profile page 45

Introduction

The Tunisian stock market presents particular hurdles due to its small liquidity limits, making it difficult for investors to make educated decisions.

In this context, our paper analyzes the fundamental difficulties confronting investors in the Tunisian market and presents a comprehensive solution based on data science and AI.

We hope to provide actionable insights to investors by developing a complex platform that incorporates data collecting, indicator calculation, portfolio creation, and AI-powered sentiment analysis.

Our multidimensional strategy solves not just market limits but also coincides with business and data science goals, eventually adding value through novel features and collaborative functions.

Chapter 1: Business Understanding

I. Business Overview

1. Problematic

The Tunisian stock market is a relatively small and illiquid market, with a limited number of listed companies and a low trading volume. This makes it difficult for investors to make informed decisions about their investments, and can lead to higher risk and lower returns.

2. Portfolio

In the trading market, a portfolio refers to a collection or combination of financial assets held by an individual, institution, or entity. The primary purpose of creating a portfolio is to diversify investments to spread risk and potentially maximize returns.

Key aspects of a portfolio include:

- Asset Allocation
- Risk Management
- Return Objectives
- Monitoring and Rebalancing

II. Project Scope

1. Business Objective

- *Maximize returns*: Build portfolios exceeding benchmark performance.
- *Minimize risk*: Ensure diversification and control volatility within user-defined limits.
- *Personalization*: Cater to individual risk tolerance and investment goals.
- *Improve decision-making*: Provide data-driven insights for confident investment choices.

2. Data Science Objective

- *Predict market trends*: Analyse historical and real-time data to forecast future performance.

- *Uncover hidden patterns*: Identify relationships between assets and risk factors.
- *Optimize diversification*: Build resilient portfolios across diverse asset classes and sectors.
- *Convenience*: Automate data collection, analysis, and portfolio construction.
- *Reduce model bias*: Ensure transparency and fairness in algorithmic recommendations.
- Create reports and dashboards for each company.

3. Added Value

- Investment by sector
- *AI-powered Sentiment Analysis*: Integrate real-time sentiment analysis from news and social media to anticipate market reactions and adjust portfolios accordingly.
- *Personalized Risk Tolerance Optimization*: Go beyond static risk profiles by dynamically adjusting portfolio risk based on individual behavior and market volatility. This could involve using gamified simulations to assess real-time risk tolerance and automatically adapt the portfolio.
- *Comprehensive Reporting:* Deliver comprehensive reporting features that offer detailed insights into users' trading performance, portfolio composition, and risk exposure, enabling them to track their progress and make informed adjustments.
- *Integration with External Platforms:* Integrate your AI trading application with external financial platforms and APIs to give users access to additional data sources, liquidity pools, and trading opportunities.

4. Metrics

- Model Quality Metrics:
    - *MSE/RMSE: it measures the average squared difference between the actual values and the predicted values in a regression problem.*
    - *Portfolio Volatility: It is a measure of the risk associated with the portfolio.*
    - *Efficient Frontier: represents a set of optimal portfolios that offer the highest expected return for a given level of risk.*
    - *Accuracy: a metric used to evaluate the performance of classification models.*
    - *Sharpe Ratio: a measure of risk-adjusted return that evaluates the excess return generated by an investment per unit of volatility or risk.*

5. Methodology

Although the lifecycle graphic looks quite different, TDSP's project lifecycle is like CRISP-DM and includes five iterative stages:

1. *Business Understanding*: define objectives and identify data sources
2. *Data Acquisition and Understanding*: ingest data and determine if it can answer the presenting question (effectively combines *Data Understanding* and *Data Cleaning* from CRISP-DM)
3. *Modeling*: feature engineering and model training (combines *Modeling* and *Evaluation*)
4. *Deployment*: deploy into a production environment
5. *Customer Acceptance*: customer validation if the system meets business needs.

figure1: TDSP's lifecycle

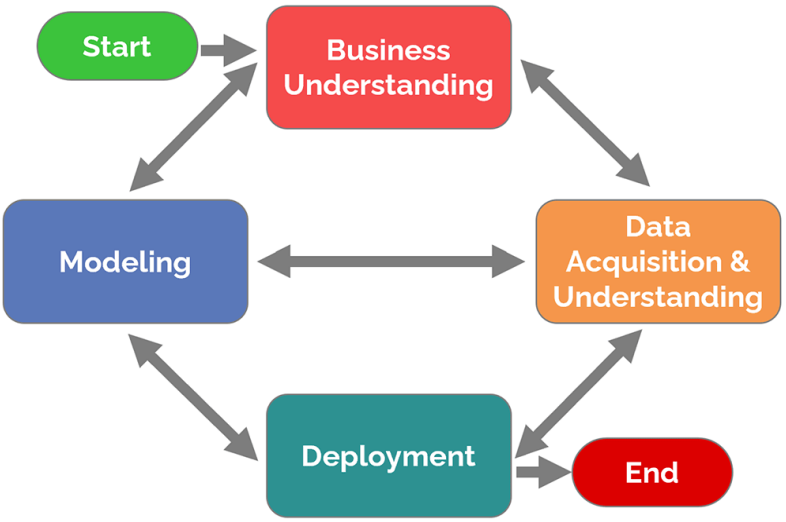


Figure 1: TDSP

6. Conclusion

This project addresses the challenge of achieving optimal investment outcomes in the Tunisian market by proposing an intelligent trading agent. Leveraging the power of machine learning and mathematical models, this agent has the potential to maximize portfolio returns, minimize volatility, and achieve effective risk management.

Chapter 2: Data Acquisition and Understanding

1. Introduction

This section focuses on data preparation for our machine-learning model. We'll clean and process the data relevant to the target price variable. Additionally, we'll perform feature engineering and create data visualizations to aid in the subsequent modeling phase.

2. Data Acquisition & Understanding

In this phase, we'll focus on preparing the data for our machine-learning model. First, we'll clean and pre-process the data to ensure it's high quality and relevant to our target. Next, we'll create new features through feature engineering. These features are designed to capture insights from the data that might be missed by the raw data points alone.

1.1 Data Acquisition

To build strong predictive models for the Intelligent Trading Agent, acquiring a comprehensive dataset was essential. We employed various methods to gather data, including structured financial information from CSV files, relevant textual data from text files, and even sentiment analysis extracted by scraping financial news articles. This diverse approach ensures a rich dataset that will be used in the following stages of the project to analyze trends and ultimately enhance the agent's performance.

- ***Market Data:*** Financial data encompasses real-time and historical information on financial instruments, exchanges, and trading activity. This data is vital for informed investment decisions and accurate financial analysis. It reflects the dynamic landscape of markets, where companies and indexes constantly influence and shape data trends.
- ***Sector Data:*** Sector data delves deeper into specific segments of the financial market. It provides detailed information on the performance of industries and groups of companies. This granular data allows investors to assess trends within a particular sector, identify potential opportunities, and make informed investment choices that align with their risk tolerance and investment goals. By analyzing sector data, investors can gain insights into the economic climate and how it might affect specific industries.
- ***Marco-economic Data:*** To investigate potential correlations between the Tunisian market and major global markets, we're gathering macroeconomic data from the European Union (EU), the United States (US), and Japan (JP). These regions are key players in the Tunisian market.
- ***Scrapped Data:*** **Financial news articles** provide the Intelligent Trading Agent with a rich source of qualitative data. These articles offer valuable insights into company performance, market trends, industry developments, and investor sentiment.

This information is crucial for enhancing market awareness, informing investment strategies, and ultimately facilitating data-driven decision-making within the project. **Sentiment analysis** of the collected articles will be employed to gain a deeper understanding of market psychology. This analysis will provide valuable insights for optimizing investment strategies and portfolio management within the Intelligent Trading Agent framework.

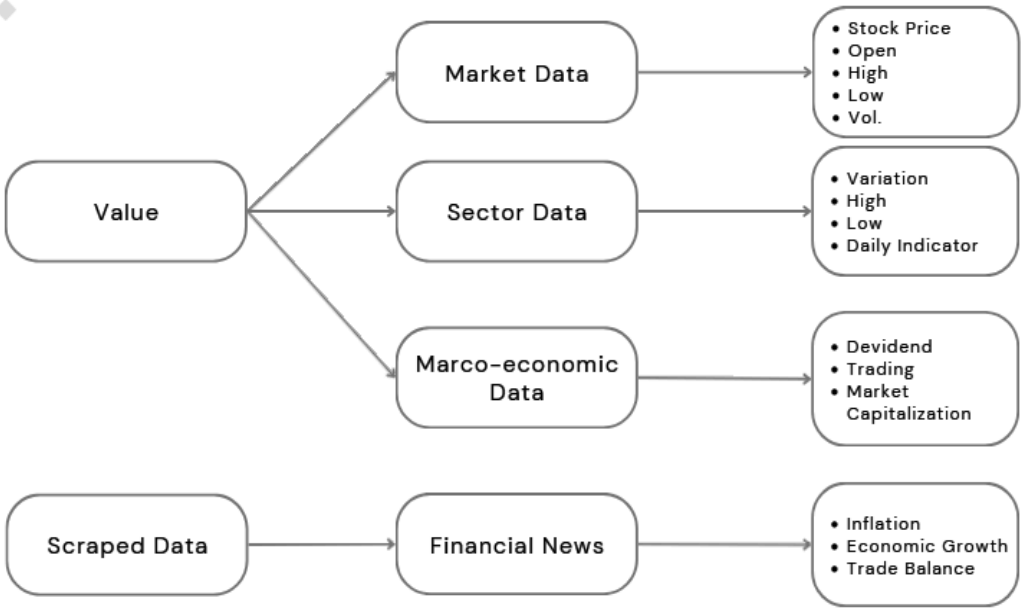


Figure 2: Data Sources

1.1.1 Web Scrapping

In the data acquisition phase, the project prioritized collecting financial news articles for sentiment analysis. We utilized web scraping techniques facilitated by the Selenium library. Custom scripts were developed for each target website (Africa News, IBource, BVMT, Investing.com) to ensure efficient extraction of relevant articles. These scripts automated browser interactions, including launching Chrome drivers, navigating to financial news websites, querying company names, and extracting article links. This targeted approach ensured a rich dataset of financial news relevant to sentiment analysis. By extracting the content of each identified article, along with its title, publication date, and source, we built a comprehensive dataset for further analysis aimed at understanding market sentiment and its impact on investment decisions.

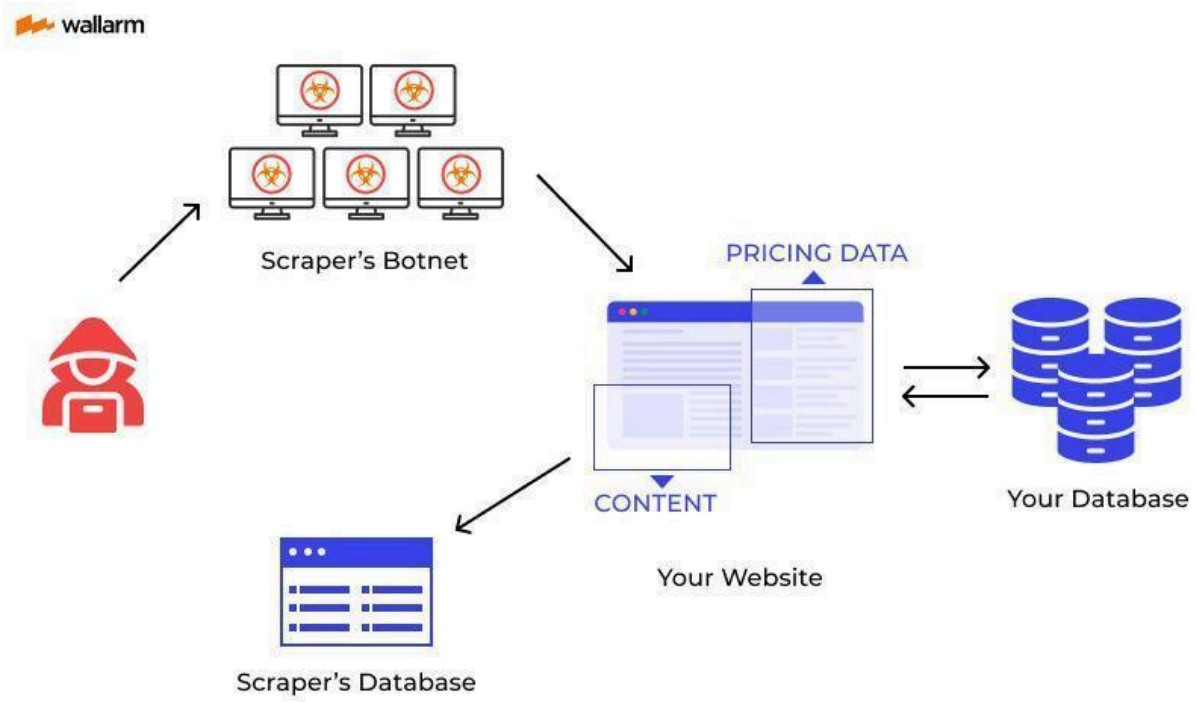


Figure 3: web scrapping schema

1.2 Data Understanding

Spanning diverse sectors like banking, insurance, and leasing, it offers a granular view of company performance within these key economic drivers.

Organized into meticulously curated categories (A, B, and S), the data empowers in-depth analysis of company characteristics and performance across each sector. This allows us to not only grasp the overall dynamics of the financial market but also to pinpoint the unique forces shaping individual sectors.

Our dataset contains seven key data points for each trading day:

- **Date:** Represents the specific date of the data.
- **Price:** The closing price of the stock at the end of the trading session.
- **Open:** The opening price of the stock at the beginning of the session.
- **High:** Highest price reached by the stock during the session.
- **Low:** The lowest price reached by the stock during the session.
- **Volume (Vol.)**: Total number of shares traded during the session.
- **Change:** Percentage change in the stock price compared to the previous day.

3. Conclusion

The Data Acquisition and Understanding phase laid the groundwork for our intelligent trading agent on the Tunisian stock market. This chapter delved into the meticulous processes of gathering, cleaning, and comprehending diverse datasets – the foundation for our comprehensive financial analysis.

Chapter 3: Data Preparation

1. Introduction

Financial projects hinge on meticulous data preprocessing. This critical phase ensures data used for analysis and modeling is accurate, reliable, and specifically tailored to the financial domain. This section will outline the key data preparation steps, focusing on the unique considerations of financial data.

2. Data Cleaning

Financial data often lacks complete entries. We address this by strategically imputing missing values, interpolating within a series, or carefully removing incomplete data points. The chosen method depends on the missing data's nature and analysis goals.

2.1 Market Data

2.1.1 Correlation

Aggregating the Feature of all other companies into one series and then computing the correlation between that series and the Feature of the current company. This results in a single correlation value representing the relationship between the Features of the current company and the Features of all other companies for that day.

	Date	VALEUR	Price_Correlation	Open_Correlation	High_Correlation	Low_Correlation	Change_Percent_Correlation
0	2021-01-04	ADWYA	NaN	-1.639095e-16	1.668728e-17	-6.712406e-17	NaN
1	2021-01-05	ADWYA	2.658885e-16	-1.639095e-16	NaN	NaN	NaN
2	2021-01-06	ADWYA	1.783216e-16	-1.639095e-16	4.833872e-17	-4.152486e-17	6.914521e-17
3	2021-01-07	ADWYA	-1.353420e-16	-1.639095e-16	-5.520786e-17	1.362742e-17	6.305864e-17
4	2021-01-08	ADWYA	NaN	-1.639095e-16	NaN	NaN	-2.353475e-17
...	...	...	...	...	...	...	...
52988	2023-12-25	WIFACKINTBANK	-8.267146e-17	1.059498e-16	2.379215e-16	3.900908e-17	1.310230e-17
52989	2023-12-26	WIFACKINTBANK	-1.340815e-17	-1.760429e-16	-1.775766e-16	-1.248570e-16	9.818739e-18
52990	2023-12-27	WIFACKINTBANK	-2.438620e-17	-7.087338e-17	-5.988699e-17	5.546017e-17	3.869691e-18
52991	2023-12-28	WIFACKINTBANK	-1.583576e-16	-2.438620e-17	-1.153667e-16	-5.173950e-17	NaN
52992	2023-12-29	WIFACKINTBANK	7.861442e-17	-1.187667e-16	5.759104e-17	9.291884e-17	NaN

Figure 4: correlation database

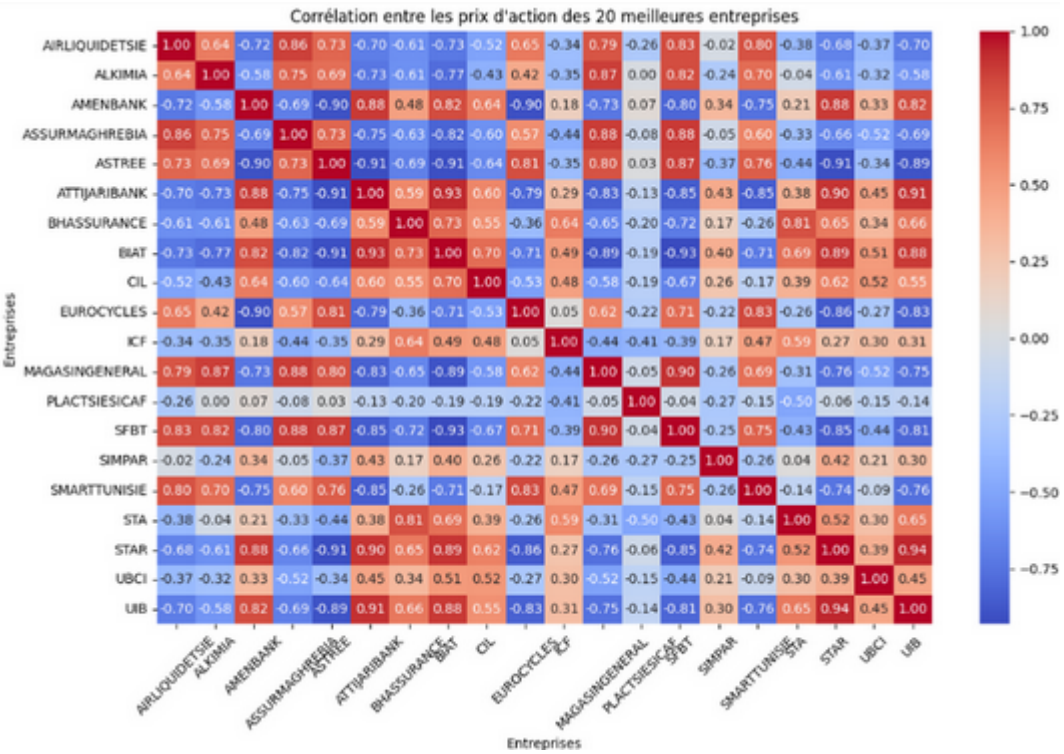


Figure 5: correlation matrix

2.1.2 Checking for Missing Data

```
Percentage of NaN values in each column after filling with 0:
Date                          0.000000
VALEUR                        0.000000
Open                          0.000000
Price                         0.000000
Low                           0.000000
High                          0.000000
Vol.                          0.000000
Change %                      0.000000
Price_Correlation            34.021475
Open_Correlation             34.163003
High_Correlation             30.904082
Low_Correlation              31.005982
Change_Percent_Correlation   23.044553
```

Figure 6: missing values

2.1.3 Handling Missing Data Using Linear Imputation

- **Last Observation Carried Forward (LOCF):**

  Pros: Simple and intuitive, preserves trends in the data.

Cons: May not be suitable for all types of data, especially if the last observed value is not a good representation of the missing value.

- **Next Observation Carried Backward (NOCB):**

  Pros: Similar to LOCF but uses the next observation.

  Cons: Similar to LOCF, may not be suitable for all types of data.

- **Moving Average:**

  Pros: Smooth out noise in the data, and preserves trends.

  Cons: May introduce lag in the data, especially with large window sizes.

- **Time-based Mean/Median:**

  Pros: Uses information from similar time periods, robust to outliers.

  Cons: May not capture changes in the data over time, assumes stationarity.

- **K-Nearest Neighbors (KNN):**

  Pros: Uses neighboring data points, flexible.

  Cons: Requires careful selection of K, sensitive to noise, may not work well with sparse data.

- **Interpolation using linear:**

  Pros: Can capture linear patterns, flexible.

  Cons: Requires careful selection of interpolation method and parameters, may overfit the data.

  In this figure, we present you the percentage of missing values after using **Interpolation using linear**

```
Percentage of NaN values in each column
Date                          0.000000
VALEUR                        0.000000
Open                          0.000000
Price                         0.000000
Low                           0.000000
High                          0.000000
Vol.                          0.000000
Change %                      0.000000
Price_Correlation             0.809786
Open_Correlation              8.312275
High_Correlation              0.428823
Low_Correlation               0.403936
Change_Percent_Correlation    0.335018
```

Figure 7: handling missing values

2.2 Sector Data

2.2.1 Correlation

Aggregating the Feature of all other Sectors into one series and then computing the correlation between that series and the feature of the current company. This results in a single correlation value representing the relationship between the features of the current Sector and the features of all other companies for that day.

	Date	High_INDICE	Close_INDICE	Low_INDICE	Open_INDICE	VALEUR_INDICE	Close_INDICE_Correlation	Open_INDICE_Correlation	High_INDICE_Correlation	Low_INDICE_Correlation	CODE_INDICE
6803	2021-01-04	6884.176	6807.968	6798.975	6879.654	TUNINDEX	2.496583e-16	2.044612e-16	-1.181535e-16	5.437932e-17	905001
6804	2021-01-05	6814.154	6765.164	6751.427	6813.717	TUNINDEX	8.703534e-17	-8.682380e-17	2.170519e-17	-8.698438e-17	905001
6805	2021-01-06	6773.142	6743.705	6723.942	6765.859	TUNINDEX	-4.395084e-17	4.380091e-17	4.392310e-17	-6.616430e-17	905001
6806	2021-01-07	6743.290	6719.628	6711.851	6738.375	TUNINDEX	-3.295312e-17	-1.846402e-17	6.614101e-17	-5.547054e-17	905001
6807	2021-01-08	6752.147	6743.144	6727.231	6729.262	TUNINDEX	-2.195540e-17	-8.072895e-17	1.706089e-19	6.588961e-17	905001

Figure 8: sector data correlation

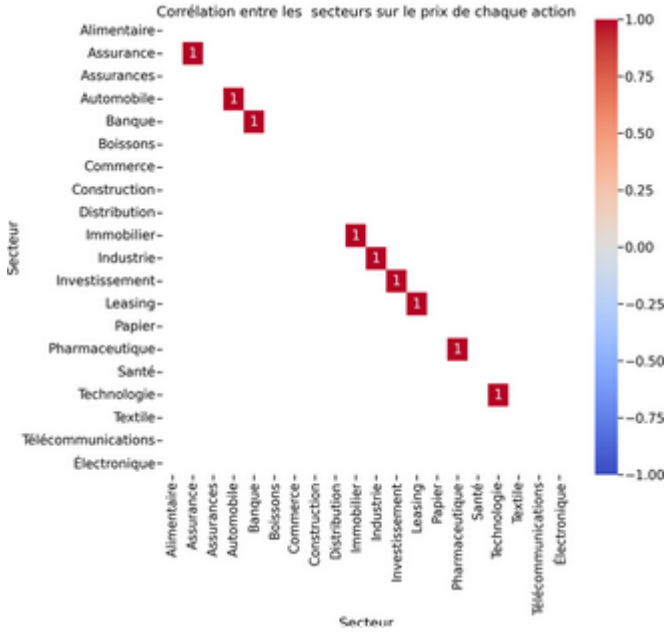


Figure 9: sector correlation matrix

2.2.3 Handling Missing Values using Interpolate with Linear Method

```
Percentage of NaN values in each column
Date                        0.000000
High_INDICE                 0.000000
Close_INDICE                0.000000
Low_INDICE                  0.000000
Open_INDICE                 0.000000
VALEUR_INDICE               0.000000
Close_INDICE_Correlation    0.152641
Open_INDICE_Correlation     0.325633
High_INDICE_Correlation     0.162817
Low_INDICE_Correlation      0.050880
CODE_INDICE                 0.000000
```

Figure 10: handling missing values for sector

3. Feature Engineering

In working with data, especially high-dimensional data, two key practices can improve your modeling results. First, you can create new features that might be more relevant to your task. In finance, for instance, this could involve calculating technical indicators or financial ratios from the existing data. Second, you can use dimensionality reduction techniques to reduce the number of features in your data. This can be helpful because it can simplify your models and improve their performance, especially when you're dealing with a lot of data.

3.1 Calculate Features

- **VaR at 95% confidence level:** This means that we are interested in estimating the potential loss of the investment with 95% confidence. In other words, we are saying that there is a 95% chance that the actual loss will not exceed the estimated VaR. The 95% confidence level is widely used in financial risk management as it provides a balance between being conservative and practical. It implies a high level of confidence while still acknowledging that there is a small chance of exceeding the estimated loss.

- **CVaR at 95% confidence level:** Conditional Value at Risk (CVaR), also known as Expected Shortfall, is a measure of the expected loss given that the loss exceeds the VaR. It provides further insight into the severity of losses beyond the VaR threshold. Choosing 95% for CVaR ensures consistency with the VaR measure and maintains the same level of confidence. It indicates that the average loss beyond the VaR threshold is estimated with 95% confidence.

- **Sharpe Ratio:** The Sharpe Ratio is a measure of risk-adjusted return. It quantifies the excess return generated by an investment per unit of risk taken. It's calculated by subtracting the risk-free rate (typically the rate of return on a safe asset like government bonds) from the average return of the investment and then dividing the result by the standard deviation of the investment's returns.

- **Skewness:** Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. It indicates whether the data is skewed to the left (negative skewness) or the right (positive skewness) relative to the mean.
- **Kurtosis:** Kurtosis measures the "tailedness" of the probability distribution of a real-valued random variable. It indicates the propensity of the data to produce outliers (extreme values) relative to the mean.
- **Skewness and kurtosis:** Provide additional information about the distribution of returns, allowing investors to better understand and manage the risks associated with their investment decisions.

	Date	VALEUR	Open	Price	Low	High	Vol.	Change %	Price_Correlation	Open_Correlation	...	avg_return	std_dev	Liquidity	Volatility	Risk	Sharpe Ratio	Skewness	Kurtosis	VaR	CVaR
0	2021-01-04	ADWYA	4.47	4.50	4.350000	4.540000	8161.00000	-0.004444	0.000000e+00	-1.639095e-16	...	0.000056	0.000056	0.19	0.000000	0.000000	-893.074243	0.351320	-0.861199	4.29	4.142609
1	2021-01-05	ADWYA	4.47	4.48	4.380000	4.500000	2457.00000	-0.004444	2.658885e-16	-1.639095e-16	...	0.000056	0.000056	0.12	0.000000	0.000000	-893.074243	0.351320	-0.861199	4.29	4.142609
2	2021-01-06	ADWYA	4.47	4.42	4.350000	4.450000	14723.00000	-0.013393	1.783216e-16	-1.639095e-16	...	0.000056	0.000056	0.10	0.000000	0.000000	-893.074243	0.351320	-0.861199	4.29	4.142609
3	2021-01-07	ADWYA	4.47	4.45	4.360000	4.450000	2647.00000	0.006787	-1.353420e-16	-1.639095e-16	...	0.000056	0.000056	0.09	0.000000	0.000000	-893.074243	0.351320	-0.861199	4.29	4.142609
4	2021-01-08	ADWYA	4.47	4.68	4.380000	4.680000	68713.00000	0.051685	4.686633e-17	-1.639095e-16	...	0.000056	0.000056	0.30	0.025857	0.101882	-893.074243	0.351320	-0.861199	4.29	4.142609
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
52231	2023-12-25	WIFACKINTBANK	8.54	8.54	8.545847	8.545847	5393.30369	0.001036	-8.267146e-17	1.059498e-16	...	0.001376	0.001376	0.00	0.006979	0.043359	-35.331615	0.979955	0.093372	6.50	6.421154
52232	2023-12-26	WIFACKINTBANK	8.54	8.60	8.600000	8.600000	200.00000	0.007026	-1.340815e-17	-1.760429e-16	...	0.001376	0.001376	0.00	0.007467	0.043359	-35.331615	0.979955	0.093372	6.50	6.421154
52233	2023-12-27	WIFACKINTBANK	8.60	8.60	8.600000	8.600000	3564.00000	0.001167	-2.438620e-17	-7.087338e-17	...	0.001376	0.001376	0.00	0.007213	0.043359	-35.331615	0.979955	0.093372	6.50	6.421154
52234	2023-12-28	WIFACKINTBANK	8.60	8.70	8.700000	8.700000	200.00000	0.011628	-1.583576e-16	-2.438620e-17	...	0.001376	0.001376	0.00	0.004429	0.065422	-35.331615	0.979955	0.093372	6.50	6.421154
52235	2023-12-29	WIFACKINTBANK	8.70	8.70	8.700000	8.700000	250.00000	0.005112	7.861442e-17	-1.187667e-16	...	0.001376	0.001376	0.00	0.004423	0.070143	-35.331615	0.979955	0.093372	6.50	6.421154

Figure 11: data with extracted features

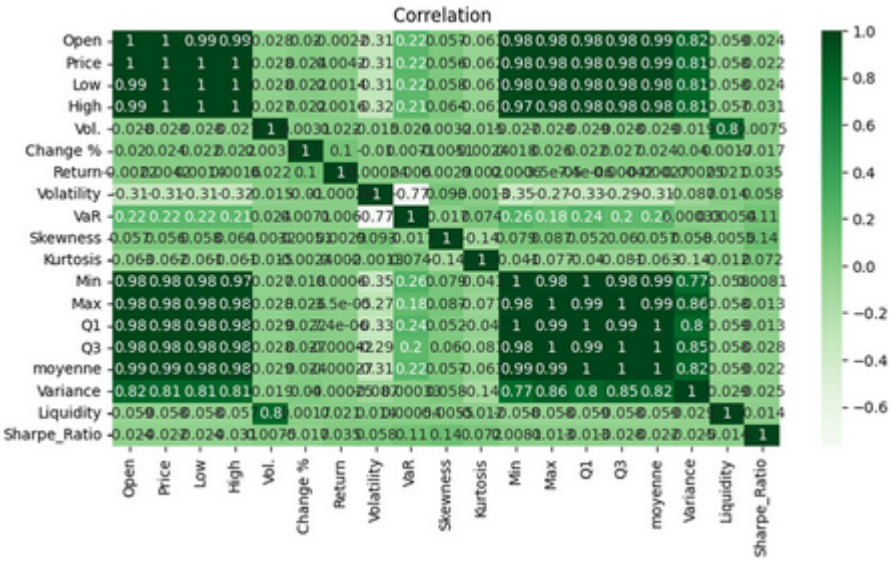


Figure 12: extracted features' correlation matrix

3.2 Combining Each Company with each Sector

	Date	VALEUR	Open	Price	Low	High	Vol.	Change %	Price_Correlation	Open_Correlation	...	CODE_INDICE	High_INDICE	Close_INDICE	Low_INDICE	Open_INDICE	VALEUR_INDICE	Close_INDICE_Correlation	Open_INDICE_Correlation	High_INDICE_Correlation	Low_INDICE_Correlation
0	2021-01-04	ADWYA	4.47	4.50	4.35	4.54	8161.0	-0.004444	0.000000e+00	-1.639095e-16	...	905001	6884.176	6807.968	6798.975	6879.654	TUNINDEX	2.496583e-16	2.044612e-16	-1.181535e-16	5.437932e-17
1	2021-01-04	AIRLIQUIDETSIE	77.85	77.82	77.82	77.85	76.0	0.000000	-3.350482e-17	-4.731744e-18	...	905001	6884.176	6807.968	6798.975	6879.654	TUNINDEX	2.496583e-16	2.044612e-16	-1.181535e-16	5.437932e-17
2	2021-01-04	ALKIMIA	47.80	47.80	0.00	0.00	0.0	0.001358	2.547967e-16	5.678300e-17	...	905001	6884.176	6807.968	6798.975	6879.654	TUNINDEX	2.496583e-16	2.044612e-16	-1.181535e-16	5.437932e-17
3	2021-01-04	ASSAD	4.44	4.44	4.40	4.50	1195.0	0.009317	0.000000e+00	0.000000e+00	...	905001	6884.176	6807.968	6798.975	6879.654	TUNINDEX	2.496583e-16	2.044612e-16	-1.181535e-16	5.437932e-17
4	2021-01-04	ASSMULTITTIHAD	2.00	1.99	1.99	1.99	10.0	-0.011765	1.884590e-16	0.000000e+00	...	905001	6884.176	6807.968	6798.975	6879.654	TUNINDEX	2.496583e-16	2.044612e-16	-1.181535e-16	5.437932e-17
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
52157	2023-12-27	TPR	4.90	4.97	4.90	5.00	975759.0	0.014286	-5.913454e-17	1.875496e-16	...	905020	3917.932	3909.201	3879.893	3878.421	TUNBASE	7.463533e-17	-3.992546e-16	-2.343354e-16	6.481040e-17
52158	2023-12-28	SOTIPAPIER	5.65	5.65	5.65	5.68	1130.0	0.000110	-2.661532e-17	9.103367e-17	...	905020	3920.982	3883.006	3867.473	3869.003	TUNBASE	5.330462e-17	1.065078e-17	-5.615723e-17	2.130364e-16
52159	2023-12-28	TPR	4.97	4.93	4.82	4.98	17696.0	-0.008048	-2.228637e-17	-5.913454e-17	...	905020	3920.982	3883.006	3867.473	3869.003	TUNBASE	5.330462e-17	1.065078e-17	-5.615723e-17	2.130364e-16
52160	2023-12-29	SOTIPAPIER	5.65	5.63	5.54	5.63	7056.0	-0.003540	8.564299e-17	8.851029e-17	...	905020	3907.075	3903.266	3871.706	3882.483	TUNBASE	3.197390e-17	1.065078e-17	1.366874e-16	-2.991827e-16
52161	2023-12-29	TPR	4.93	4.97	4.90	4.98	20778.0	0.008114	6.471853e-17	0.000000e+00	...	905020	3907.075	3903.266	3871.706	3882.483	TUNBASE	3.197390e-17	1.065078e-17	1.366874e-16	-2.991827e-16

Figure 13: Combining Each Company with each Sector

4. Time Series Preprocessing

Financial data analysis often necessitates temporal aggregation to align with investment horizons or trading frequencies. This strategic process involves summarizing data points into larger timeframes, such as converting daily prices to monthly averages.

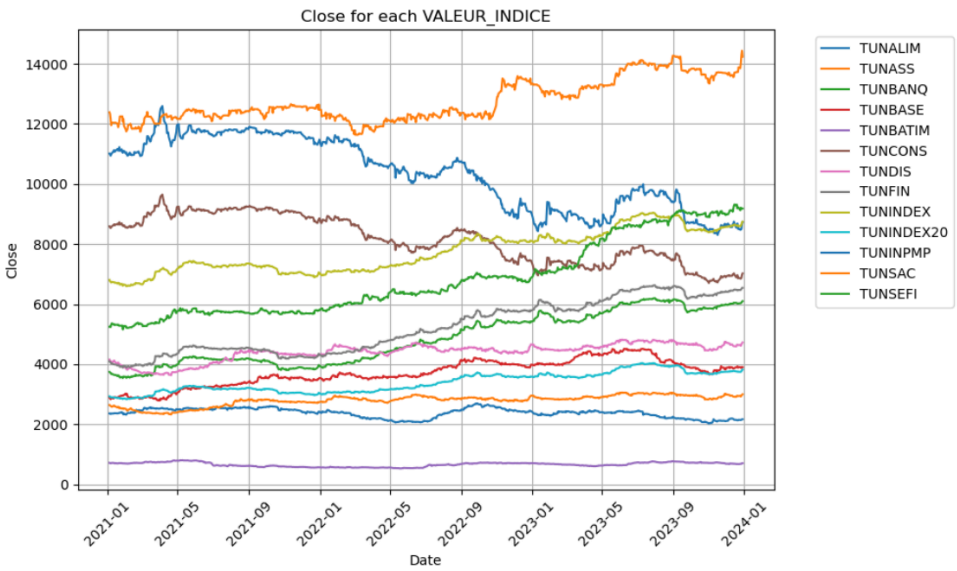


Figure 14: Time Series Preprocessing

5. Normalization and Scaling

Financial data often encompasses variables measured in different units and with varying magnitudes. To address this heterogeneity and optimize model performance, a common practice is to employ feature scaling techniques like MinMaxScaler. MinMaxScaler transforms all features to a pre-defined range, typically between 0 and 1.

This mitigates the undue influence of features with larger scales on model outputs, ensuring all features contribute equally during model training.

```
(array([[[0.        , 0.        , 0.00688276, ..., 0.47338424,
          0.54707429, 0.        ],
         [0.00116009, 0.        , 0.00605918, ..., 0.47338424,
          0.54707429, 0.00116009],
         [0.00232019, 0.        , 0.01723631, ..., 0.47338424,
          0.54707429, 0.00232019],
         ...,
         [0.19257541, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.19257541],
         [0.1937355 , 0.        , 0.00294135, ..., 0.47338424,
          0.54707429, 0.1937355 ],
         [0.19837587, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.19837587]],

        [[0.00116009, 0.        , 0.00605918, ..., 0.47338424,
          0.54707429, 0.00116009],
         [0.00232019, 0.        , 0.01723631, ..., 0.47338424,
          0.54707429, 0.00232019],
         [0.00348028, 0.        , 0.00811812, ..., 0.47338424,
          0.54707429, 0.00348028],
         ...,
         [0.1937355 , 0.        , 0.00294135, ..., 0.47338424,
          0.54707429, 0.1937355 ],
         [0.19837587, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.19837587],
         [0.19953596, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.19953596]],

        [[0.00232019, 0.        , 0.01723631, ..., 0.47338424,
          0.54707429, 0.00232019],
         [0.00348028, 0.        , 0.00811812, ..., 0.47338424,
          0.54707429, 0.00348028],
         [0.00464037, 0.        , 0.02053062, ..., 0.47338424,
          0.54707429, 0.00464037],
         ...,
         [0.19837587, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.19837587],
         [0.19953596, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.19953596],
         [0.20417633, 0.        , 0.58821107, ..., 0.47338424,
          0.54707429, 0.20417633]],
```

Figure 15: Normalization and Scaling

6. Handling Categorical Data

Many financial datasets contain categorical variables (e.g., asset class, credit rating). To facilitate their use in machine learning models, these features require transformation into numerical representations. A common approach is label encoding, which assigns a unique integer value to each distinct category.

7. Data Splitting

To evaluate a model's effectiveness on unseen financial data, we must subdivide the dataset into training, validation, and test sets. Due to the inherent time-ordered nature of financial data, it's crucial to maintain temporal ordering during this process.

8. Feature Selection

Feature selection, employing **PCA** and **ANOVA**, plays a critical role in optimizing algorithmic trading strategies. PCA identifies key patterns in complex data, while ANOVA selects impactful features related to the target variable. This process enhances model performance, reduces overfitting, and improves interpretability for more effective decision-making.

```
Feature Importance Ranking:
Variance: 1.942288052686901
VaR: 1.882804924439438
Min: 1.8742521184093404
Price: 1.7682041022786514
Kurtosis: 1.7423152276335268
Skewness: 1.6610923150188879
Sharpe_Ratio: 1.6506298099601133
VALEUR_encoded: 1.3787849653962032
Max: 1.1579485652226915
Return: 1.059357884766063
Q1: 0.9923875809738287
moyenne: 0.985069520105437
Q3: 0.9718483842649343
Liquidity: 0.07722050621496654
Vol.: 0.05873400284954429
```

Figure 16: PCA

```
Feature Importance Ranking:
Company: 2.469734065772789
Liquidity: 2.452174778143771
VaR: 2.3276024228113075
av-Q2: 2.219605379825391
av-Q1: 2.197883622848693
Kurtosis: 2.1633514343715814
Skewness: 2.008811273718778
Vol.: 1.9192889169598362
Date: 1.4750203590684596
Min: 0.7522103930733262
Q1: 0.7408604424407835
Max: 0.7232117953585745
Q2: 0.7203464919410273
Price: 0.6921645275331599
Return: 0.06973345171510321
```

Figure 17: ANOVA

9. Conclusion

In conclusion, this chapter highlights the critical role of data preparation in any analytical endeavor. Whether it's financial analysis or any other field, the meticulous process of cleaning, engineering features, and preprocessing time series data is essential. These steps ensure the accuracy, reliability, and relevance of our models to the specific domain we're working in. By following these practices, we set a strong foundation for our project's success, empowering us to draw meaningful insights and make informed decisions.

Chapter 4: Features Extraction

1. Introduction

To prepare our algorithmic trading models, this phase focuses on two crucial steps: feature engineering and model selection. First, we'll meticulously extract relevant features from the financial data, capturing historical price movements, technical patterns, and support/resistance levels. These features will serve as valuable inputs for our models. Next, we'll explore the application of various machine learning models, including both regression and classification algorithms, to generate robust trading signals and optimize portfolio construction.

2. Relative Performance Metrics

2.1 Relative Performance of Company to Sector

- **Pros**:
  - Provides insights into how individual companies perform relative to their respective sectors.
  - Helps identify companies that are outperforming or underperforming compared to the broader market or sector.
- **Cons**:
  - May overlook broader market trends or sector-wide movements if focusing solely on individual company performance.
  - Less effective for understanding the overall health or direction of the market as a whole.

2.2 Relative Performance of Sector to Company

- **Pros**:
  - Offers insights into how sectors perform relative to individual companies, which can reveal broader market trends or sector-specific movements.
  - Useful for gauging the overall direction or health of the market by analyzing sector-level performance.
- **Cons**:
  - May not capture the nuances of individual company performance within a sector.
  - Less focused on identifying specific company outliers or trends.

3. Technical Analysis Indicators

Technical indicators are crucial tools in financial analysis, we are gonna use just some of the most famous ones.

- **Simple Moving Average (SMA):** Averages the closing prices over a specified window, smoothing out price fluctuations and highlighting trends.

Figure 18: Simple Moving Average (SMA)

- **Exponential Moving Average (EMA):** Similar to SMA, EMA gives more weight to recent prices, making it sensitive to short-term price movements.

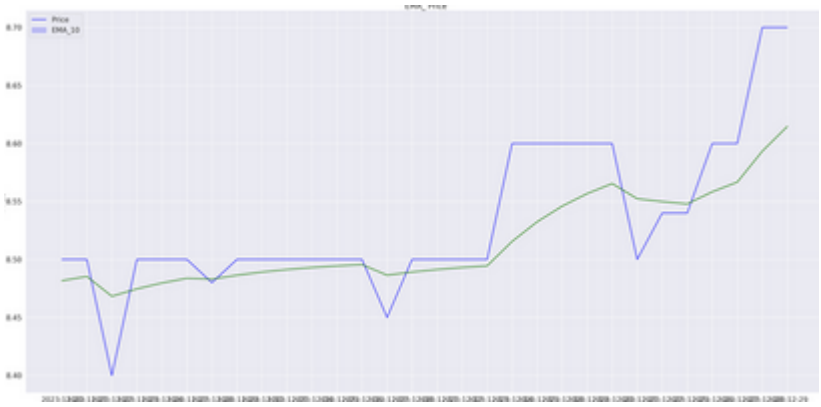


Figure 19: Exponential Moving Average (EMA)

- **Moving Average Convergence Divergence (MACD):** This represents the difference between short-term EMA and long-term EMA, providing insights into the strength and direction of a trend.
- **Relative Strength Index (RSI):** Measures the speed and change of price movements, indicating overbought or oversold conditions in the market.

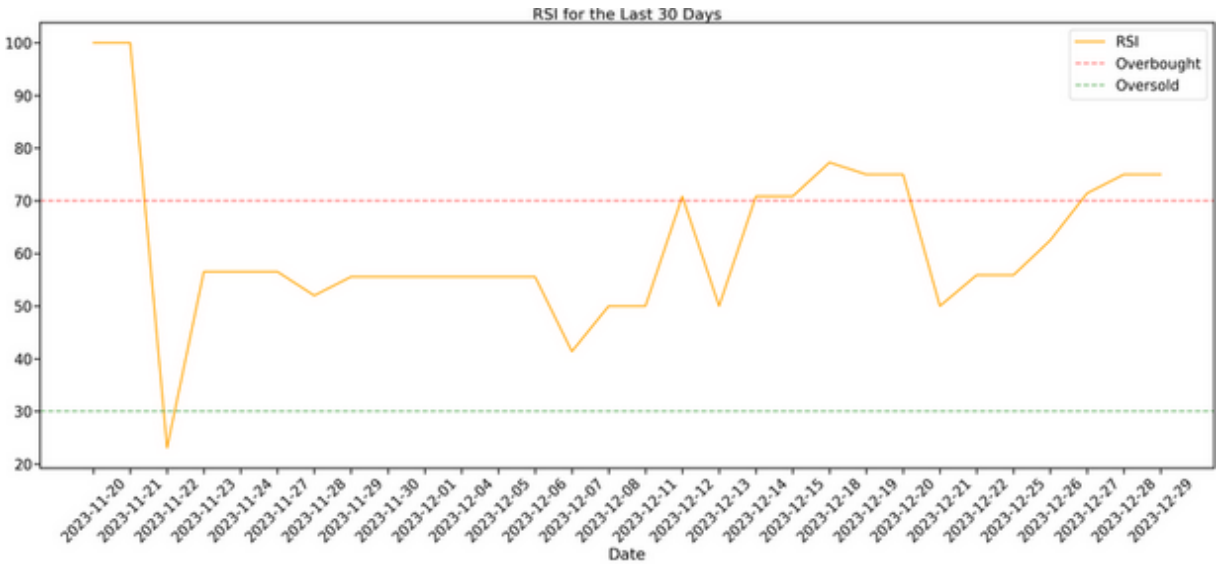


Figure 20: Relative Strength Index (RSI)

- **Bollinger Bands: Consists of a middle band (SMA)** and upper/lower bands representing price volatility. They help identify price extremes and potential reversal points.

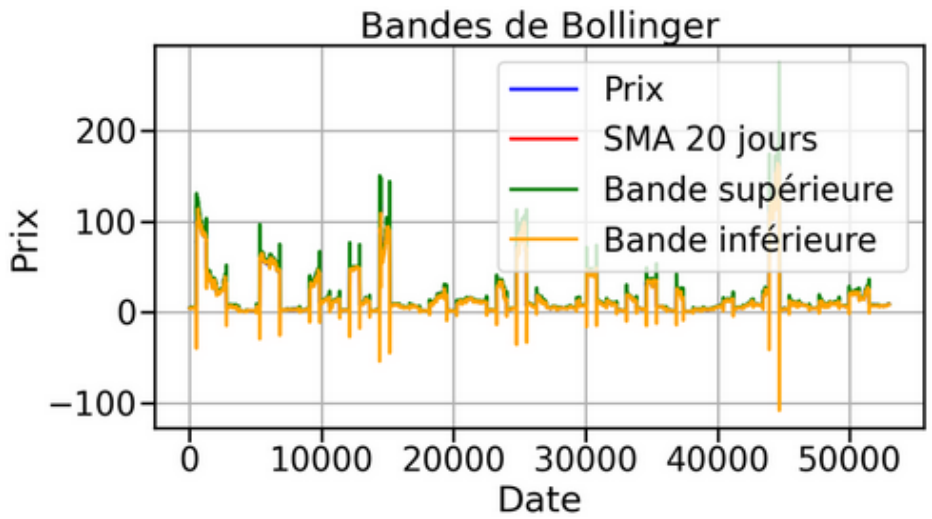


Figure 21: Bollinger Bands

4. Lag Features

- **Price Lag:** This involves creating lag features by shifting historical prices either forward or backward in time. These features provide insight into past price behaviors at different intervals, such as one-hour lag, one-day lag, and beyond.
- **Return Lag**: Lagged returns are calculated to capture historical return patterns. These features are valuable for identifying signals related to momentum or mean reversion in the data.

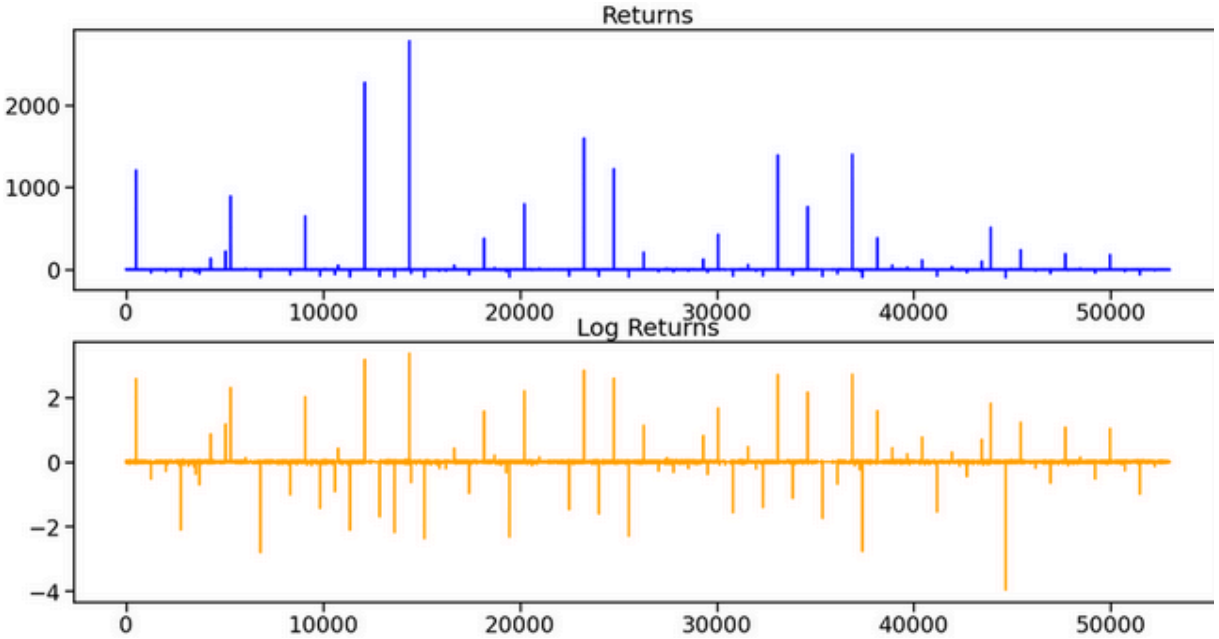


Figure 22: Lag Features

5. Support and Resistance Levels

- **Pivot Points**: These are calculated daily, weekly, or monthly based on previous high, low, and closing prices. Pivot points are valuable for identifying crucial support and resistance levels in the market.

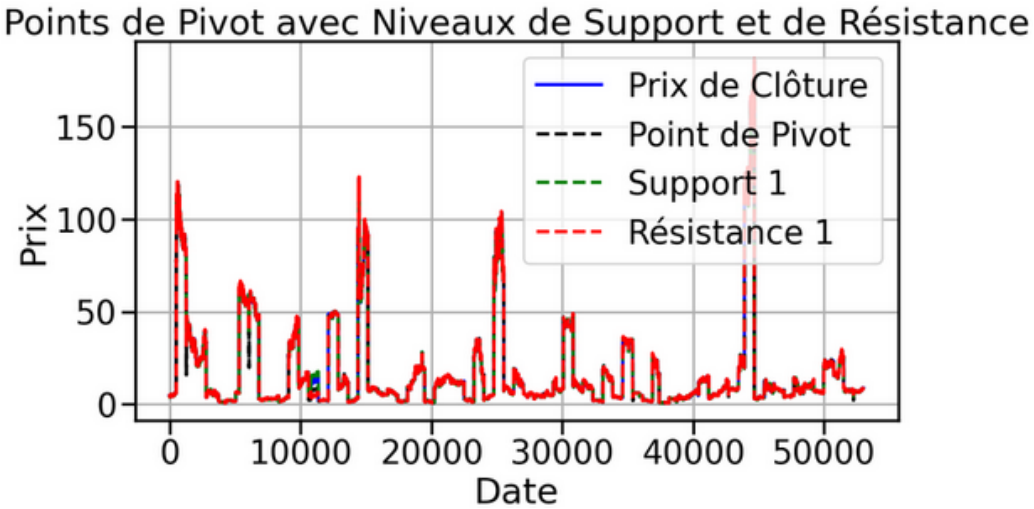


Figure 23: Pivot Points

- **Fibonacci Retracement**: This technique utilizes Fibonacci retracement levels to pinpoint potential support and resistance areas, following the Fibonacci sequence's proportions.

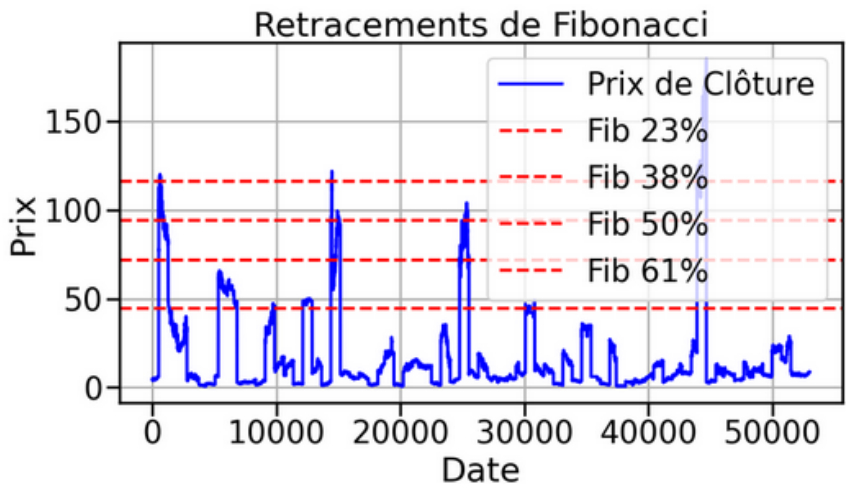


Figure 24: Fibonacci Retracement

- **Chart Patterns**: Recognizing and labeling common chart patterns like head and shoulders, double tops, and double bottoms is crucial. These patterns often indicate potential reversals or continuations in price movements, offering valuable insights for traders and analysts alike

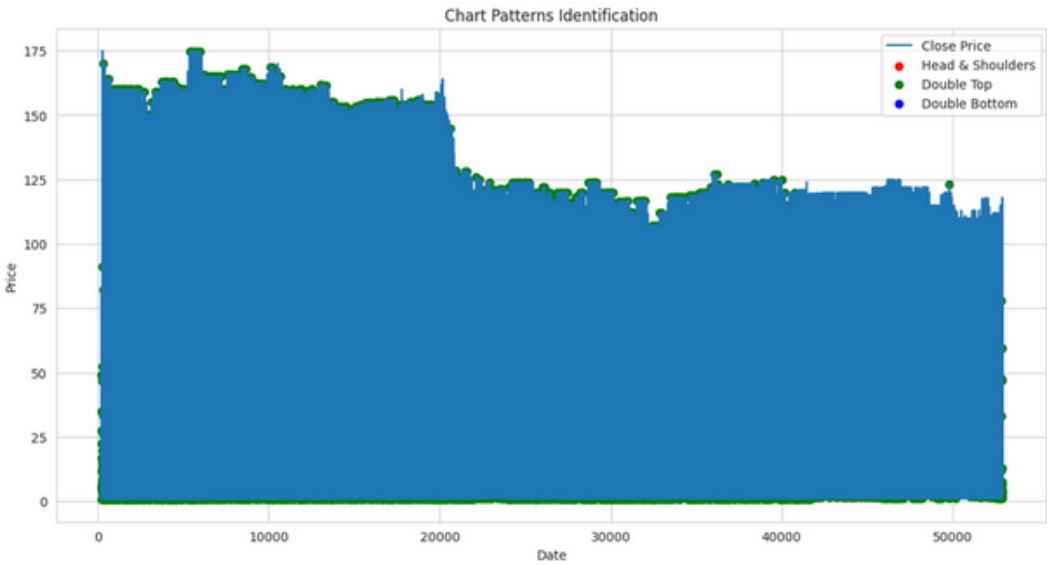


Figure 25: Chart Patterns

6. Conclusion

In this chapter, we've covered the key steps in feature extraction and modeling for algorithmic trading. We explored how to extract valuable features from financial data, including historical price movements, technical indicators, and support/resistance levels, we aim to generate effective trading signals and optimize portfolio construction. These features and indicators serve as powerful tools for traders and analysts to navigate the complexities of financial markets.

Chapter 5: Modeling

1. Introduction

This chapter dives into our intelligent trading system, powered by machine learning (ML) and deep learning (DL). We'll explore how it forecasts market prices, manages risk, and optimizes portfolios. By leveraging data collection, predictive models for both indices and stocks, and sentiment analysis, we aim to equip investors with practical tools for navigating volatile markets.

2. Data Collection Model

- ***Sentiment analysis:***

Sentiment analysis acts as a compass for gauging investor sentiment within financial news articles. It employs a two-pronged approach.

First, it analyzes the inherent positivity or negativity of individual words. Second, it considers the broader context and structure of sentences to capture the overall emotional tone. By leveraging machine learning algorithms trained on vast datasets of labeled text, sentiment analysis can effectively classify the emotional outlook expressed in financial news. This understanding of market sentiment provides valuable insights for investors, allowing them to factor in the psychological influences that can impact financial markets.

	date_publication	titre	contenu	url	Positive Sentiment	Neutral Sentiment	Negative Sentiment	polarity
0	21 février 2024 19:02	Le Tunindex grignote 22,1 points, la SFBT gagn...	TUNIS UNIVERSNEWS La Bourse de Tunisterminé lé...	https://universnews.tn/le-tunindex-grignote-de...	0.000001	1.000001	0.000001	0.000000
1	20 février 2024 18:16	Deuxième séance consécutive haussière à la bou...	TUNIS UNIVERSNEWS Le marché de la bourseenchaî...	https://universnews.tn/deuxieme-seance-consecu...	0.013001	0.987001	0.000001	0.000000
2	19 février 2024 17:43	Volume d'échanges modeste à la bourse de Tunis...	TUNIS UNIVERSNEWS Le marché de la Bourseentamé...	https://universnews.tn/volume-dechanges-modest...	0.020001	0.980001	0.000001	0.113636
3	18 février 2024 13:00	Bourse de Tunis: De quelle couleur sera la nou...	TUNIS UNIVERSNEWS SEF La Bourse de Tunis conti...	https://universnews.tn/bourse-de-tunis-de-quel...	0.009001	0.991001	0.000001	0.000000
4	16 février 2024 17:44	La Bourse quasiment stable ce vendredi, ICF re...	TUNIS UNIVERSNEWS La Bourse de Tunisterminé qu...	https://universnews.tn/la-bourse-quasiment-sta...	0.031001	0.969001	0.000001	0.166667
...	...	...	...	...	...	...	...	...
215	08 avril 2019 17:14	Bourse de Tunis 08-04-2019 : Le Tunidex clôtur...	La bourse de Tunisclôturé en baisse la séance ...	https://universnews.tn/bourse-de-tunis-08-04-2...	0.000001	1.000001	0.000001	0.000000
216	01 avril 2019 16:47	Bourse de Tunis 01-04-2019 : La Bourse de Tuni...	La bourse de Tunisclôturé la séance de lundi s...	https://universnews.tn/bourse-de-tunis-01-04-2...	0.015001	0.985001	0.000001	0.200000

Figure 26: Sentiment analysis Data

In the Sentiment Analysis, we utilized 4 models: **Vader**, **Bert**, **Robert**, and **XLNet**.

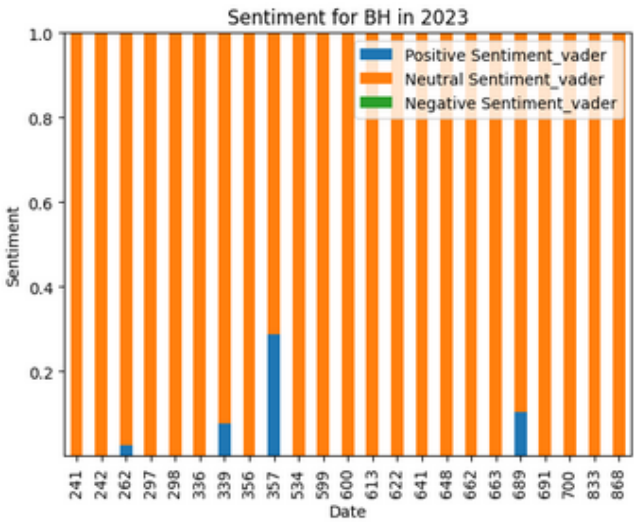


Figure 27: vader model

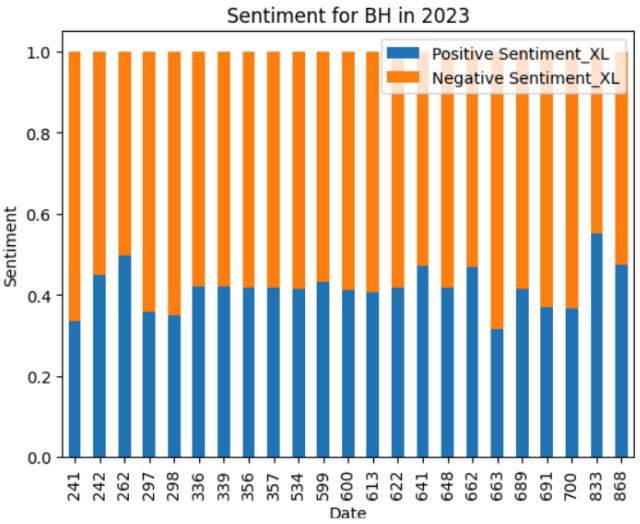


Figure 28: XLNet model

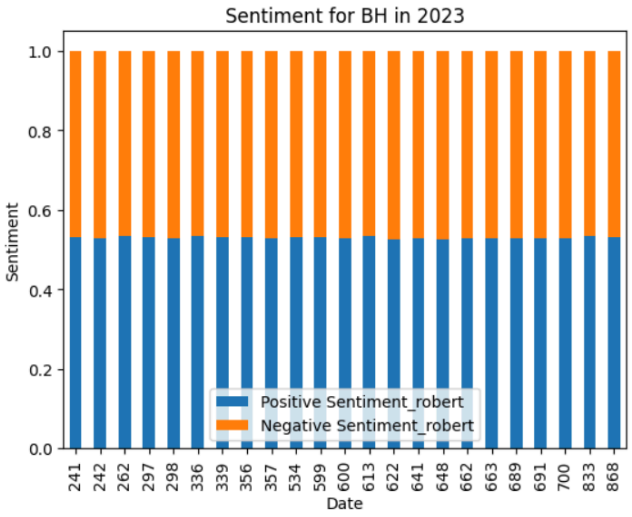


Figure 29: Robert model

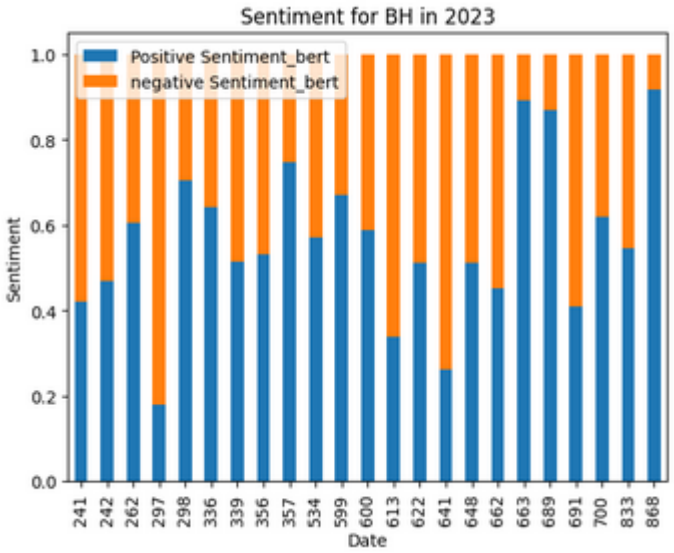


Figure 30: Bert Model

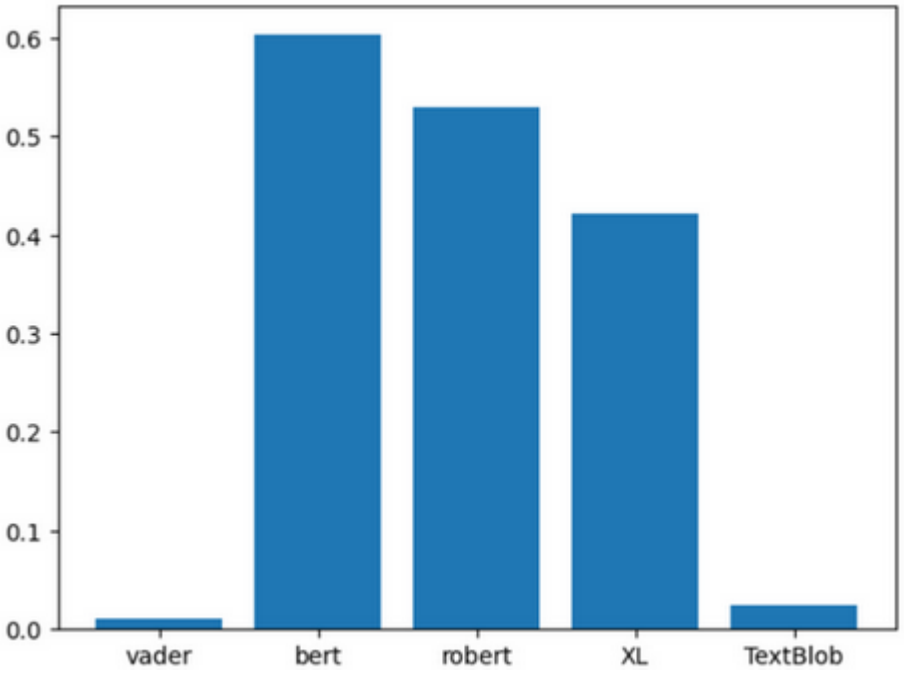


Figure 31: Sentiment Analysis models comparison

3. Prediction Model

In our project, we aimed to develop a robust predictive model capable of understanding and forecasting complex time-series data. The model leverages advanced neural network architectures, including convolutional layers, Temporal Convolutional Networks (TCN), LSTM (Long Short-Term Memory) units, and attention mechanisms, to effectively capture temporal dependencies and nuances in the data.

3.1 Model Architecture

The architecture of our predictive model is designed to process time-series input data efficiently, making it suitable for forecasting tasks that require handling large volumes of data with varying temporal patterns. Here's a breakdown of the model layers:

- **Input Layer:** Handles inputs of shape ($None, features$), where "None" can be any number of timesteps, and "features" is the number of variables per timestep.
- **Dense Layer:** Serves as Shortcut (using Residual Connection to avoid Input sequence that doesnt need to be handled by Convolutional Layers)
- **Convolutional Layers**: Two 1D convolutional layers with 64 filters, a kernel size of 3, 'same' padding, and ReLU activation function to extract high-level features from the input data.
- **Temporal Convolutional Network (TCN):** A TCN layer with multiple dilation rates helps the model learn at various temporal resolutions, enhancing its ability to predict long sequences effectively.
- **Bidirectional LSTM Layer:** Incorporates two LSTM layers wrapped in a Bidirectional layer, allowing the model to learn from both past (backward) and future (forward) dependencies.
- **Unidirectional LSTM Layer:** Incorporates two LSTM layers Unidirectional layers, allowing the model to to focus on future (forward) dependencies.
- **Attention Mechanism:** A MultiAttention Mechanism (5heads) layer follows the LSTM to focus the model on relevant parts of the input sequence from different points, improving the precision of the output predictions.
- **Output Layer**: A dense layer with a linear activation function predicts continuous values, which are crucial for forecasting tasks.

Regularization techniques such as Batch Normalization and Dropout (20% rate) are used to prevent overfitting, ensuring the model generalizes well on unseen data.

3.2 Training Process

Our model's training process is meticulous, employing a comprehensive grid search strategy to optimize hyperparameters such as the number of layers, type of layers, learning rates, and dropout rates. This approach ensures that we systematically explore a wide range of configurations, leading to the selection of the best-performing model based on validation loss.

- **Custom Loss Function Description :**

Introduction:

The custom loss function is designed to improve the predictive accuracy of financial models by incorporating penalties for violating logical market constraints alongside the standard Mean Absolute Error (MAE).

Components:

Extracting Individual Components

- **True and Predicted Prices:** The actual and forecasted prices.
- **True Low and High Values:** The minimum and maximum prices within a specific period.
- **Market Indices:** Various market values, including closing, high, low, and opening indices.

Defining Condition Penalties :

- **Penalty for Price Below True Low:** Ensures the predicted price is not lower than the actual minimum price.
- **Penalty for Price Above High Index:** Ensures the predicted price does not exceed the high index value.
- **Penalty for Low Index Above Close Index:** Ensures the low index value is not higher than the closing index.
- **Mean Absolute Error (MAE) Calculation**

This measures the average absolute difference between the true and predicted values.

Penalty Weights :

Weights are assigned to each penalty to control their influence on the total loss, reflecting their relative importance.

Total Loss Calculation

The total loss is computed by combining the MAE with the weighted penalties, ensuring the model minimizes errors and adheres to logical market constraints.

Returning the Total Loss

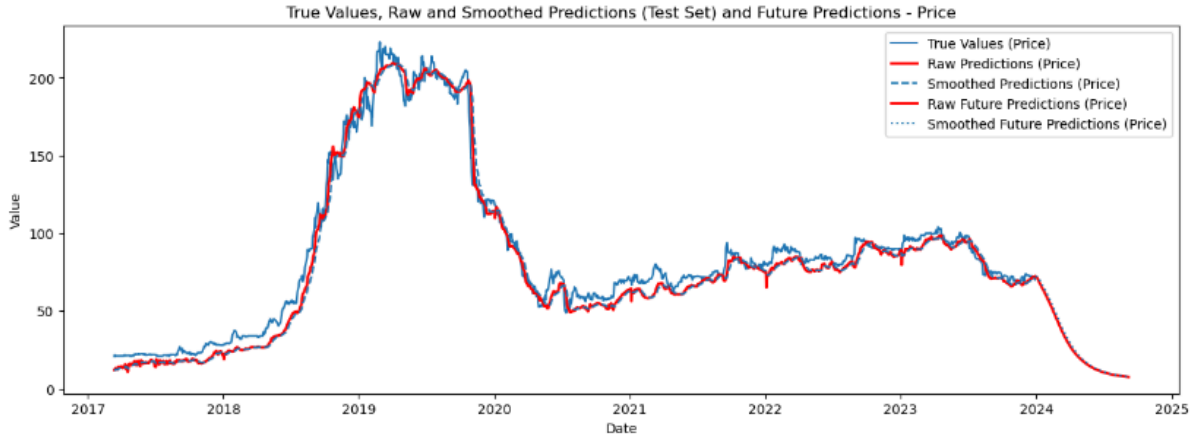
The final loss value is used during model training to optimize both accuracy and logical consistency.

- **Learning Rate Scheduler:** Implements an exponential decay strategy, which reduces the learning rate as the number of epochs increases, allowing for finer adjustments in weights as training progresses.
- **Callbacks:** Key callbacks used include:
  - ➢ **EarlyStopping:** Monitors the validation loss and stops training if no improvement is seen, preventing overfitting.
  - ➢ **ModelCheckpoint:** Saves the best model based on the lowest validation loss, ensuring we can retrieve the best version post-training.
  - ➢ **History:** Logs training and validation metrics for each epoch, aiding in the visualization and analysis of the training process.
  - ➢ **LearningRateScheduler:** Adjusts the learning rate according to a predefined schedule, enhancing the training dynamics.

3.3 Evaluation and Metrics

The model's performance is rigorously evaluated using a comprehensive set of metrics to assess its accuracy and reliability in predicting stock market prices for multiple companies. Here are the key metrics and their respective outcomes from the validation phase:

```
Average Metrics Across All Companies:
Mean Absolute Error (MAE): 0.1218
Mean Squared Error (MSE): 0.0237
Directional Accuracy: 66.04%
Mean Absolute Percentage Error (MAPE): 39.81%
```




```
54/54 ━━━━━━━━━━━━━━━━━━━━ 0s 6ms/step
Metrics for company ICF and feature Price:
Mean Absolute Error (MAE): 0.028187450161745894
Mean Squared Error (MSE): 0.00120265642672799
Directional Accuracy: 43.955399061032864%
```

- **Mean Absolute Error (MAE):** The MAE for our model stands at 0.1934, indicating the average absolute difference between the predicted and actual stock prices. This metric helps in understanding the typical error magnitude without considering the error direction.
- **Mean Squared Error (MSE):** The MSE recorded is 0.0694, reflecting the average squared difference between the predicted and actual values. MSE is useful for highlighting larger errors as it squares the error values.
- **Directional Accuracy:** This metric, standing at 68.26%, measures the percentage of predictions that correctly forecast the direction of stock price movements (i.e., up or down). It is crucial for trading strategies that depend on correctly predicting the movement direction rather than the exact prices.
- **Mean Absolute Percentage Error (MAPE):** At 44.80%, the MAPE provides insight into the accuracy of predictions in terms of their deviation from the actual values, expressed as a percentage. This metric is particularly useful for comparing the performance across different data scales or stock prices.

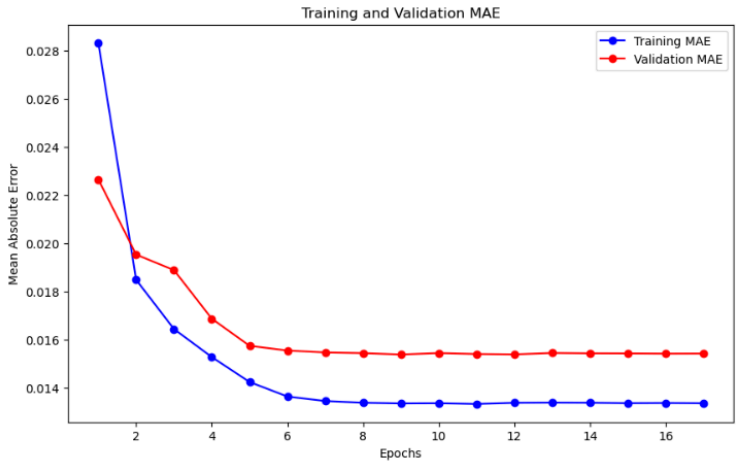


Figure 32: Training and validation MAE

4. Strategies

4.1 Optimal Growth Strategy

This strategy aims to achieve superior risk-adjusted returns by maximizing the Sharpe ratio, which measures the return earned above the risk-free rate per unit of volatility.

Steps :

- **Calculate Sharpe ratio** : We began by calculate the Sharpe ratio for each company and provide an interpretation for each value based on the ratios :

	Entreprise	sharpe Ratio	Interpretation_Sharpe
0	BIAT	1.387512	Bon placement (rendement / risque équilibré)
1	ADWYA	0.066315	Rendement faible(Risque plus élevé que le rend...
2	AIRLIQUIDETSIE	-0.212243	Sous-performance (ne pas investir)
3	ALKIMIA	-3.865470	Sous-performance (ne pas investir)
4	AMENBANK	0.649242	Rendement faible(Risque plus élevé que le rend...
...	...	...	...
67	TUNISIELEASINGF	0.054264	Rendement faible(Risque plus élevé que le rend...
68	TUNISRE	-0.588291	Sous-performance (ne pas investir)
69	UBCI	0.711957	Rendement faible(Risque plus élevé que le rend...
70	UIB	-0.561948	Sous-performance (ne pas investir)
71	UNIMED	1.891358	Bon placement (rendement / risque équilibré)

Figure 33: Calculated Sharpe Ratio for each company

  - If ratio < 0: Underperformance (do not invest)
  - If ratio is between 0 and 1: Low return (risk higher than return)
  - If ratio > 1: Good investment
  - If ratio > 2: Excellent investment

- **Filtering of Dataset:** filtering the dataset to include only companies with interpretations indicating "Good investment" or "Excellent investment."

- **Application of Markowitz Portfolio Optimization:** Employing the Markowitz portfolio optimization technique, we aimed to construct a portfolio that maximizes the Sharpe ratio while minimizing risk.
- **Calculation of highest and lowest Annualized Return and Annualized Volatility of this strategy :**we determine both the highest and lowest annualized returns and volatilities to provide investors with insights into potential performance extremes and help them decide whether or not to choose this strategy.

```
Portfolio Metric                    Annualized Value
-----------------------------  -----------------
Highest Annualized Return                 0.227035
Lowest Annualized Return                  0.0486933
Highest Annualized Volatility             0.146581
Lowest Annualized Volatility              0.0465209
```

Figure 34: portfolio mertics' values for optimal growth strategy

- **User Input for Portfolio Customization:** In order to tailor the portfolio to investor preferences, we asked users to indicate the number of companies they wished to invest in out of 13 for this strategy, and the total amount they intended to allocate.

- **Calculation of Annualized Return and Annualized Volatility:** Following the execution of the Markowitz portfolio optimization technique, we obtained the annualized return and annualized volatility for the constructed portfolio.

```
Allocation de portefeuille avec le ratio de Sharpe maximal

Rendement annualisé: 0.17
Volatilité annualisée: 0.06
```

Figure 35: annualized return and volatility values for optimal growth strategy

- **Investment Allocation Analysis:** Subsequently, we analyzed the distribution of our portfolio and determined the allocation of funds for each company. This step provided

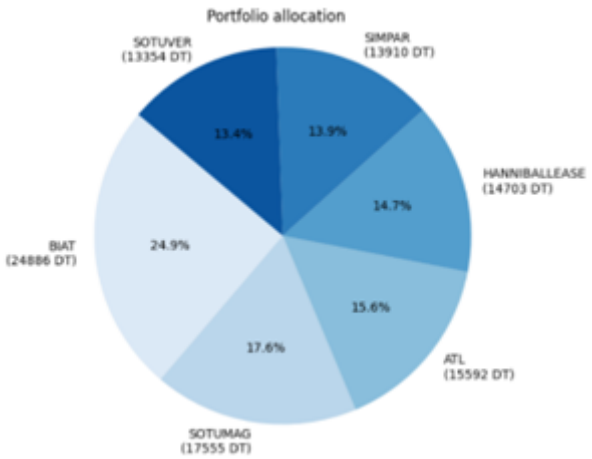


Figure 36: investment allocation for optimal growth strategy

- **Return Percentage of Each Company:** Additionally, we computed the percentage of returns contributed by each company relative to others. This analysis shed light on the relative performance of individual companies within the portfolio.

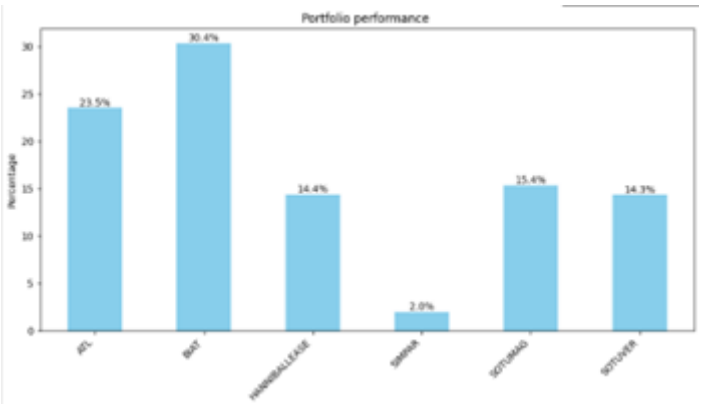


Figure 37: portfolio performance for optimal growth strategy

4.2 Volatility Minimization Strategy

This Strategy aims to construct an investment portfolio with the objective of reducing or minimizing the level of volatility, which refers to the fluctuations in returns.

Steps :

- calculate the annualized volatility for each company

Entreprise	Annual volatility
SIMPAR	0.004530
BHASSURANCE	0.006590
ASTREE	0.008851
SANIMED	0.009773
CITYCARS	0.009829
...	...
TPR	0.022494
CIL	0.022569
ATTIJARILSROMPU	0.022878
ASSAD	0.023640
BHLEASING	0.027448

Figure 38: annual volatility for each company

- **Filtering of Dataset:** filtering the data to include only companies with volatility below 0.0178 and returns above 0.2, where 0.0178 is considered the minimum volatility threshold and 0.2 the minimum return threshold.

- **Application of Markowitz Portfolio Optimization:** Use the Markowitz portfolio optimization technique to construct a portfolio that minimizes volatility
- **Calculation of highest and lowest Annualized Return and Annualized Volatility of this strategy :**we determine both the highest and lowest annualized returns and volatilities to provide investors with insights into potential performance extremes and help them decide whether or not to choose this strategy.

```
Portfolio Metric                    Annualized Value
-----------------------------  ------------------
Highest Annualized Return                 0.204696
Lowest Annualized Return                  0.067847
Highest Annualized Volatility             0.0987671
Lowest Annualized Volatility              0.0412839
```

Figure 39: portfolio mertics' values for volatility Minimization Strategy

- **User Input for Portfolio Customization:** In order to tailor the portfolio to investor preferences, we asked users to indicate the number of companies they wished to invest in out of 13 for this strategy, and the total amount they intended to allocate.
- **Calculation of Annualized Return and Annualized Volatility:**Following the execution of the Markowitz portfolio optimization technique, we obtained the annualized return and annualized volatility for the constructed portfolio.

```
Portfolio allocation with minimum volatility

Annualized return: 0.1
Annualized volatility: 0.05
```

Figure 40: annualized return and volatility values for volatility Minimization Strategy

- **Investment Allocation Analysis:** Subsequently, we analyzed the distribution of our portfolio and determined the allocation of funds for each company. This step provided insights into how much the investor should allocate to each company within the portfolio.

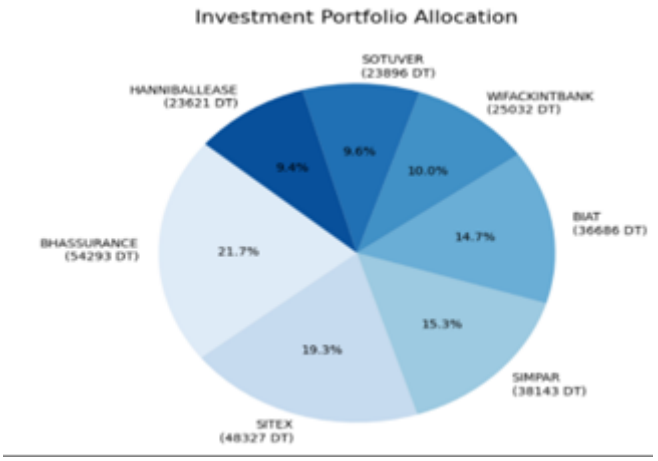


Figure 41: investment allocation for volatility Minimization Strategy

- **Return Percentage of Each Company:** Additionally, we computed the percentage of returns contributed by each company relative to others. This analysis shed light on the relative performance of individual companies within the portfolio.

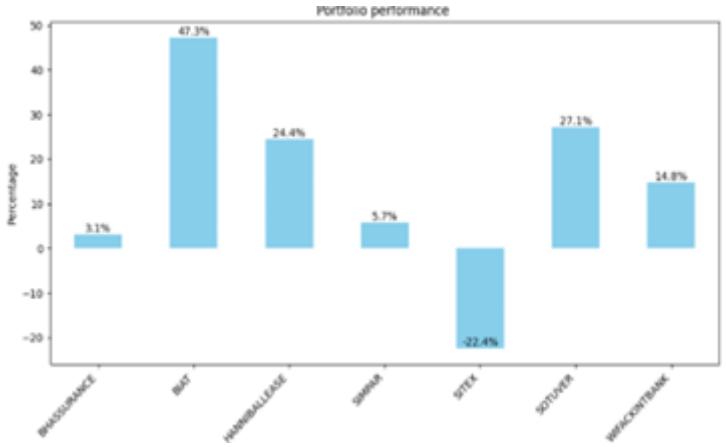


Figure 42: portfolio performance for volatility Minimization Strategy

4.3 Maximising Returns Strategy

This strategy involves selecting an investment portfolio with the aim of achieving the highest possible return, without necessarily taking into account the associated level of risk

Steps :

- calculate the annualized Return for each company

Date	VALEUR	Annual Return
2023-12-29	ATELIERMEUBLEINT	0.192682
2023-12-29	ATTIJARIBANK	0.154000
2023-12-29	PLACTSIESICAF	-0.585162
2023-12-29	POULINAGPHOLDING	-0.054079
2023-12-29	UNIMED	-0.125094
...	...	...
2021-01-05	SITEX	0.578343
2021-01-05	SIMPAR	1.564422
2021-01-05	ASTREE	-1.623075
2021-01-05	SIAME	-0.009900
2021-01-05	WIFACKINTBANK	0.348153

Figure 43: annualized Return for each company

- **Filtering of Dataset:** Filtering the data to include the top 20 companies.

```
[SIMPAR,STETUNDUSUCRE,HANNIBALLEASE,CEREALIS,SITEX,BHLEASING,BHASSURANCE,STAR,ATL,ATTIJARILSROMPU,WIFACKINTBANK,ARTES,BIAT,
SOTUMAG,SOTUVER,MPBS,NEWBODYLINE,ATELIERMEUBLEINT,CITYCARS,ATTIJARIBANK]
```

Figure 44: top 20 companies

- **Application of Markowitz Portfolio Optimization:** Use the Markowitz portfolio optimization technique to construct a portfolio that maximizes return.
- **Calculation of highest and lowest Annualized Return and Annualized Volatility of this strategy :**we determine both the highest and lowest annualized returns and volatilities to provide investors with insights into potential performance extremes and help them decide whether or not to choose this strategy.

```
Portfolio Metric                    Annualized Value
----------------------------  -----------------
Highest Annualized Return                0.223871
Lowest Annualized Return                 0.0955341
Highest Annualized Volatility            0.0884491
Lowest Annualized Volatility             0.0469016
```

Figure 45: portfolio mertics' values for Maximizing Returns Strategy

- **User Input for Portfolio Customization:** In order to tailor the portfolio to investor preferences, we asked users to indicate the number of companies they wished to invest in out of 20 for this strategy, and the total amount they intended to allocate
- **Calculation of Annualized Return and Annualized Volatility:** Following the execution of the Markowitz portfolio optimization technique, we obtained the annualized return and annualized volatility for the constructed portfolio.

```
Portfolio allocation with maximum return
Annualized return: 0.22
Annualized volatility: 0.08
```

Figure 46: annualized return and volatility values for Maximizing Returns Strategy

- **Investment Allocation Analysis:** Subsequently, we analyzed the distribution of our portfolio and determined the allocation of funds for each company. This step provided insights into how much the investor should allocate to each company within the portfolio.

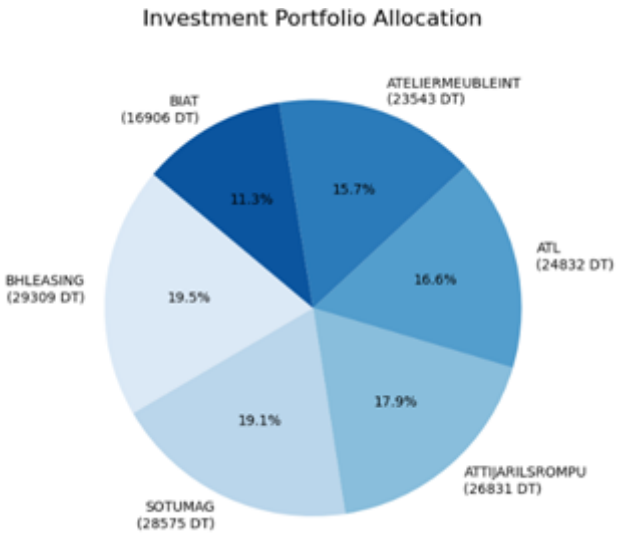


Figure 47: investment allocation for Maximizing Returns Strategy

- **Return Percentage of Each Company:** Additionally, we computed the percentage of returns contributed by each company relative to others. This analysis shed light on the relative performance of individual companies within the portfolio.

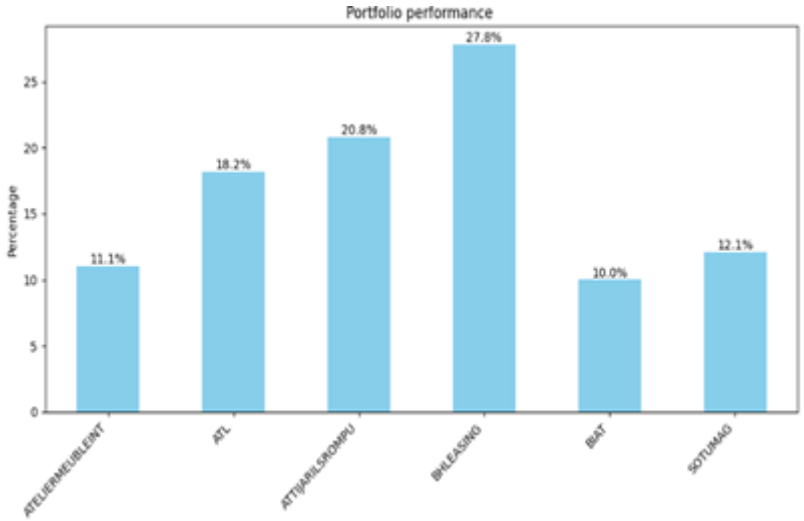


Figure 48: portfolio performance for Maximizing Returns Strategy

4.4 Trend Following Strategy

This strategy aims to capitalize on directional price movements by following existing short-term market trends.

Steps :

- Calculate the 10-day Simple Moving Average (SMA_10) and the 200-day Simple Moving Average (SMA_200) for the prices for each company.
- Assign a signal to indicate trading actions based on the comparison of short-term and long-term moving averages. A value of 1 indicates a buy signal for short-term

investment when the short-term moving average exceeds the long-term moving average.

- **Filtering of Dataset:**Filtering the data to include only the companies where the signal equals 1.

```
['ALKIMIA, ASSAD, ASTREE, BESTLEASE, BHASSURANCE, BHBANK, BNA, BTEADP, CIMENTSDEBIZERTE,
 ENNAKLAUTOMOBILES, ESSOUKNA, EUROCYCLES, ICF, POULINAGPHOLDING, SITS,SMARTTUNISIE, STA,
 TELNETHOLDING, TUNINVESTSICAR, TUNISRE, UBCI]
```

Figure 49: companies where the signal equals 1

- **Application of Markowitz Portfolio Optimization:** Use the Markowitz portfolio optimization technique to construct a portfolio that maximize return and minimize risk
- **Calculation of highest and lowest Annualized Return and Annualized Volatility of this strategy :**we determine both the highest and lowest annualized returns and volatilities to provide investors with insights into potential performance extremes and help them decide whether or not to choose this strategy.

Portfolio Metric	Annualized Value
Highest Annualized Return	-0.0421212
Lowest Annualized Return	-0.559875
Highest Annualized Volatility	0.0988721
Lowest Annualized Volatility	0.0361992

Figure 50: portfolio mertics' values for Trend Following Strategy

- **User Input for Portfolio Customization:** In order to tailor the portfolio to investor preferences, we asked users to indicate the number of companies they wished to invest in out of 21 for this strategy, and the total amount they intended to allocate.
- **Calculation of Annualized Return and Annualized Volatility:** Following the execution of the Markowitz portfolio optimization technique, we obtained the annualized return and annualized volatility for the constructed portfolio.

```
Maximum Sharpe ratio portfolio allocation
Annualized return: -0.04
Annualized volatility: 0.05
```

Figure 51: annualized return and volatility values for Trend Following Strategy

- **Investment Allocation Analysis:** Subsequently, we analyzed the distribution of our portfolio and determined the allocation of funds for each company. This step provided insights into how much the investor should allocate to each company within the portfolio.

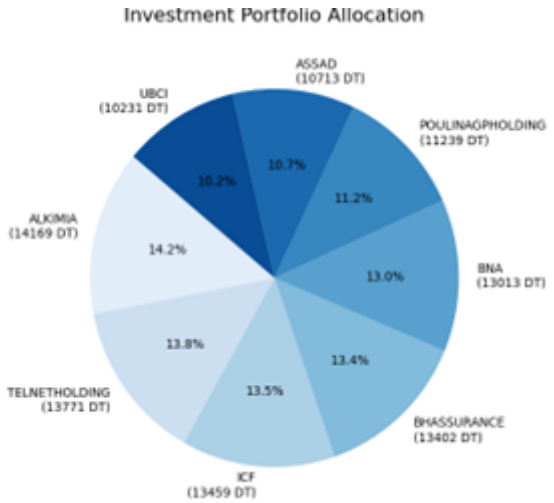


Figure 52: investment allocation for Trend Following Strategy

- **Return Percentage of Each Company:** Additionally, we computed the percentage of returns contributed by each company relative to others. This analysis shed light on the relative performance of individual companies within the portfolio.

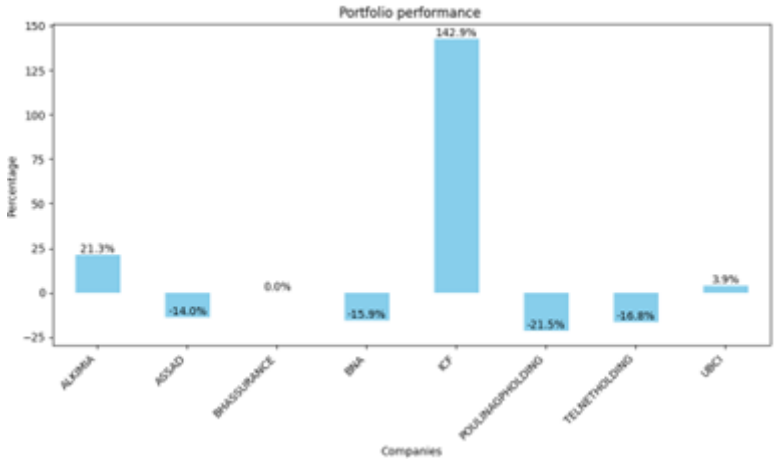


Figure 53: portfolio performance for Trend Following Strategy

4.5 Positive Reputation Strategy

This strategy involves selecting companies based on positive sentiment analysis coverage, aiming to leverage their favorable reputation for potential investment opportunities.

Steps:

- **Filtering of Dataset:**Filtering the data through sentiment analysis using BERT, we identified companies exhibiting notably high scores in 'positive sentiment_bert.' This process allowed us to curate a list of companies characterized by favorable sentiment in their market discourse. Building upon this analysis, we propose an investment strategy centered around these companies, leveraging their positive reputation as reflected in market articles. Investors intrigued by this premise may find value in exploring opportunities within this subset of companies.

```
['BIAT', 'ATB', 'UBCI', 'SOTUVER', 'ADWYA', 'STB', 'ASTREE', 'UIB',
 'ASSAD', 'ARTES'], dtype=object)
```

Figure 54: companies exhibiting notably high scores in 'positive sentiment_bert.'

- **Application of Markowitz Portfolio Optimization:** Use the Markowitz portfolio optimization technique to construct a portfolio that maximize return and minimize risk
- **Calculation of highest and lowest Annualized Return and Annualized Volatility of this strategy :**we determine both the highest and lowest annualized returns and volatilities to provide investors with insights into potential performance extremes and help them decide whether or not to choose this strategy.

```
Portfolio Metric                    Annualized Value
----------------------------  ------------------
Highest Annualized Return                 0.170423
Lowest Annualized Return                 -0.167821
Highest Annualized Volatility             0.20013
Lowest Annualized Volatility              0.070367
```

Figure 55: portfolio mertics' values for Positive Reputation Strategy

- **User Input for Portfolio Customization:** In order to tailor the portfolio to investor preferences, we asked users to indicate the number of companies they wished to invest in out of 10 for this strategy, and the total amount they intended to allocate.
- **Calculation of Annualized Return and Annualized Volatility:** Following the execution of the Markowitz portfolio optimization technique, we obtained the annualized return and annualized volatility for the constructed portfolio.

```
--------------------------------------------------
Maximum Sharpe ratio portfolio allocation
Annualized return: 0.16
Annualized volatility: 0.09
```

Figure 56: annualized return and volatility values for Positive Reputation Strategy

- **Investment Allocation Analysis:** Subsequently, we analyzed the distribution of our portfolio and determined the allocation of funds for each company. This step provided insights into how much the investor should allocate to each company within the portfolio.

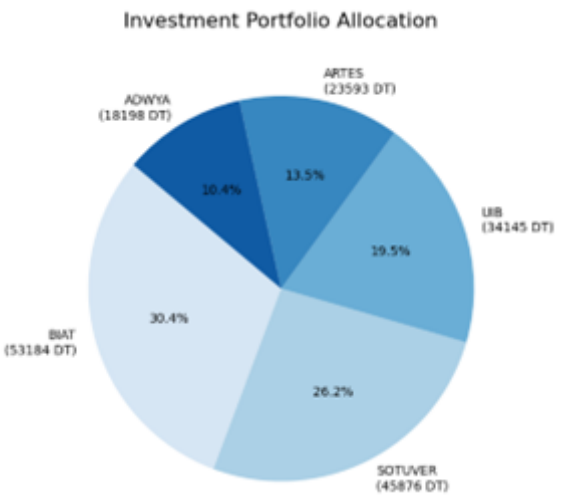


Figure 57: investment allocation for Positive Reputation Strategy

- **Return Percentage of Each Company:** Additionally, we computed the percentage of returns contributed by each company relative to others. This analysis shed light on the relative performance of individual companies within the portfolio.

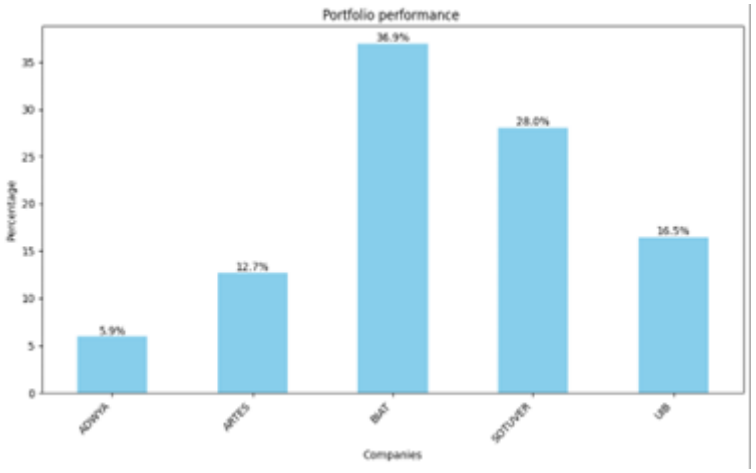


Figure 58: portfolio performance for Positive Reputation Strategy

4.6 Investment by sector strategy

This strategy allows users to tailor their investments based on specific sector , enabling them to focus on sectors they believe will perform well.

Steps:

- **Sector Selection:** Investors can begin by selecting specific sectors they wish to focus on for their investments.

```
Secteurs disponibles :  TUNBANQ, TUNDIS, TUNINPMP, TUNINDEX, TUNBATIM, TUNSEFI, TUNALIM, TUNBASE, TUNASS
Choisissez un secteur parmi les suivants : TUNBANQ
```

Figure 59: sector selection

- **Filter the dataset:** filtering the dataset to include only the companies operating within the selected sector.
- **Application of Markowitz Portfolio Optimization:** Use the Markowitz portfolio optimization technique to construct a portfolio that maximize return and minimize risk.
- **Calculation of highest and lowest Annualized Return and Annualized Volatility of this strategy :**we determine both the highest and lowest annualized returns and volatilities to provide investors with insights into potential performance extremes and help them decide whether or not to choose this strategy.
- **User Input for Portfolio Customization:** In order to tailor the portfolio to investor preferences, we asked users to indicate the number of companies they wished to invest in out of 11 for this strategy, and the total amount they intended to allocate.

```
Portfolio Metric                    Annualized Value
-----------------------------  ------------------
Highest Annualized Return                 0.169958
Lowest Annualized Return                  0.0243065
Highest Annualized Volatility             0.12391
Lowest Annualized Volatility              0.0729392
```

Figure 60: portfolio mertics' values for Investment by sector strategy

- **Calculation of Annualized Return and Annualized Volatility:** Following the execution of the Markowitz portfolio optimization technique, we obtained the annualized return and annualized volatility for the constructed portfolio.

```
Maximum Sharpe ratio portfolio allocation
Annualized return: 0.14
Annualized volatility: 0.08
```

Figure 61: annualized return and volatility values for Investment by sector strategy

- **Investment Allocation Analysis:** Subsequently, we analyzed the distribution of our portfolio and determined the allocation of funds for each company. This step provided insights into how much the investor should allocate to each company within the portfolio.

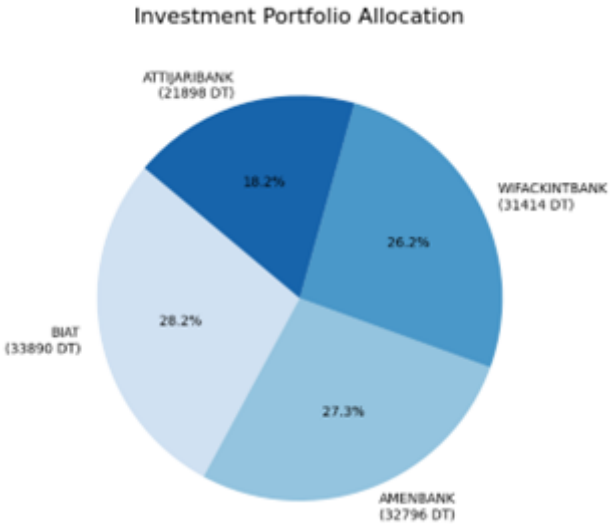


Figure 62: investment allocation for Investment by sector strategy

- **Return Percentage of Each Company:** Additionally, we computed the percentage of returns contributed by each company relative to others. This analysis shed light on the relative performance of individual companies within the portfolio.

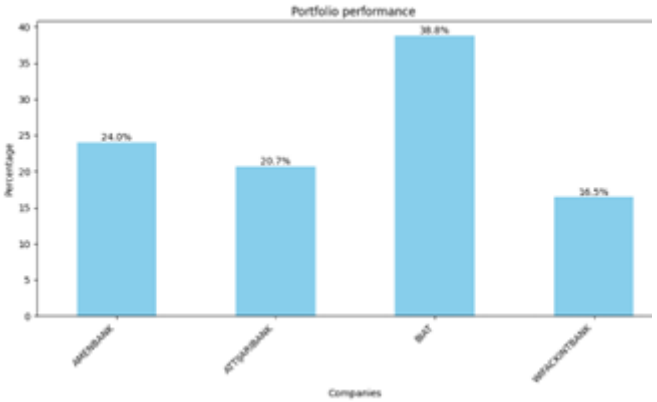


Figure 63: portfolio performance for Investment by sector strategy

5. Conclusion

The predictive model demonstrates significant capability in forecasting complex time-series data, attributed to its sophisticated architecture and the rigorous process of hyperparameter optimization. Challenges such as selecting the appropriate model configuration were addressed through extensive grid search, highlighting the model's adaptability and robustness. Future work will focus on exploring deeper architectures and integrating more diverse data sources to further enhance model accuracy.

Chapter 6: Deployment

1. Introduction

In the previous chapter, we explored various machine learning algorithms for predicting future stock prices based on historical data. We achieved a best performing model . This chapter delves into deploying the models as a web service to make predictions accessible to users.

We will utilize Flask, a lightweight Python web framework, to create a user-friendly API for interacting with the model. Additionally, we will leverage a MySQL database to store the data and facilitate efficient model training and prediction serving.

2. Web deployment

2.1 Home page

The "Home Page" serves as the central access point for the web application. It provides a user-friendly interface for navigating different functionalities, such as accessing the profile, portfolio, and dashboard.

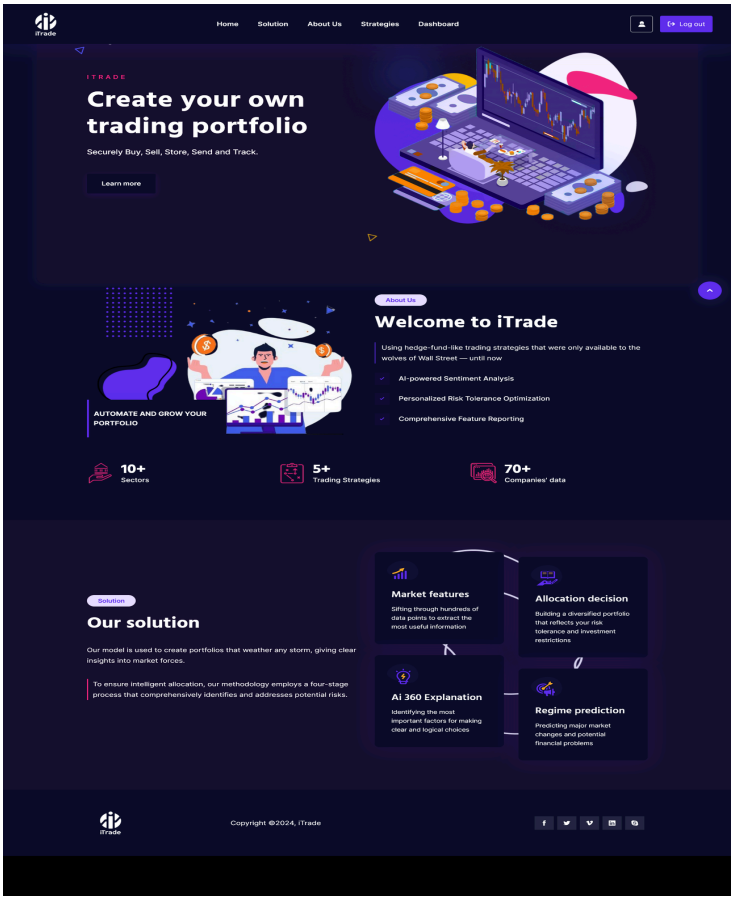


Figure 64: home page

2.2 Login page

The "Sign In" and "Sign Up" functionalities are crucial for user access. Existing users can log in using their credentials, while new users can register to create an account and explore the application's features.

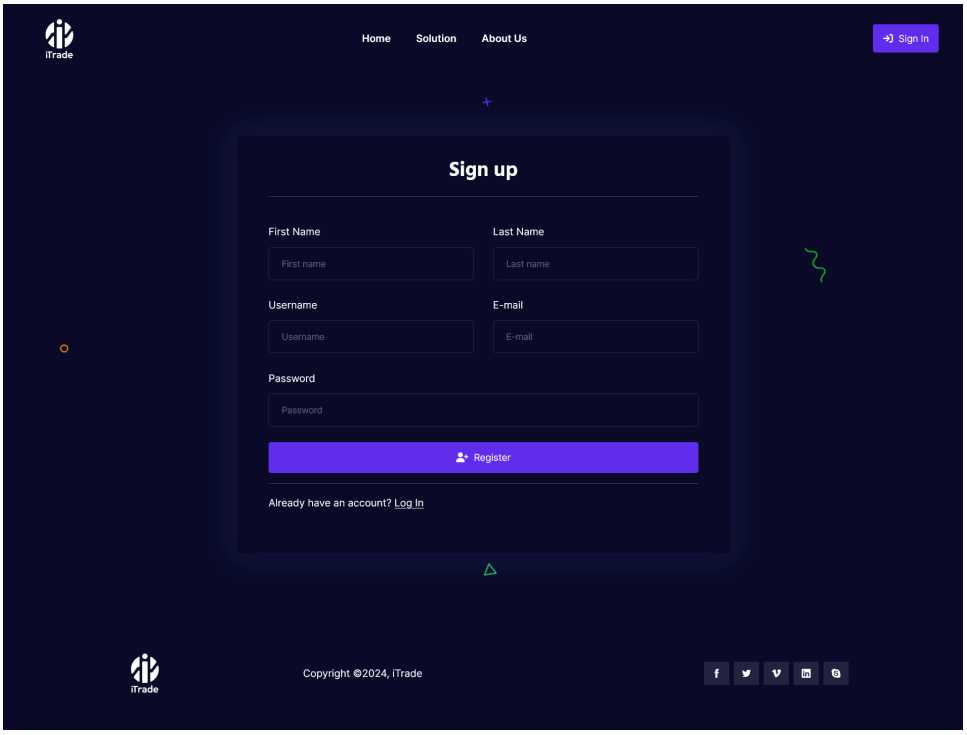


Figure 65: login page

2.3 Dashboard page

The "Dashboard" section caters to data visualization. Here, users can explore various charts and graphs that visually represent the data, offering insights into market trends. This visual representation can be instrumental for informed investment decisions.

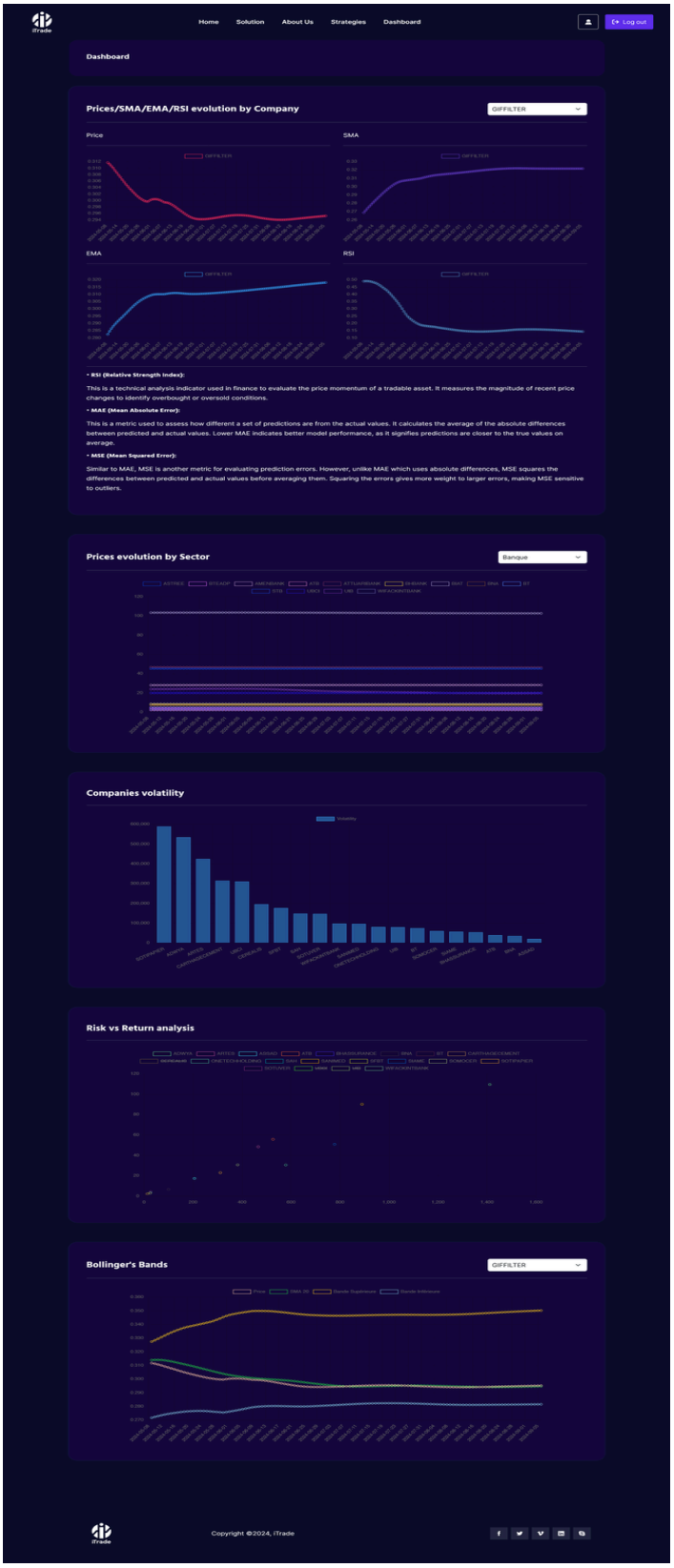


Figure 66: dashboard

2.4 Strategies page

The "Strategies" section acts as a central hub for users to track their investments. This is where users can view a comprehensive list of strategies, allowing them to monitor performance and make informed investment decisions.

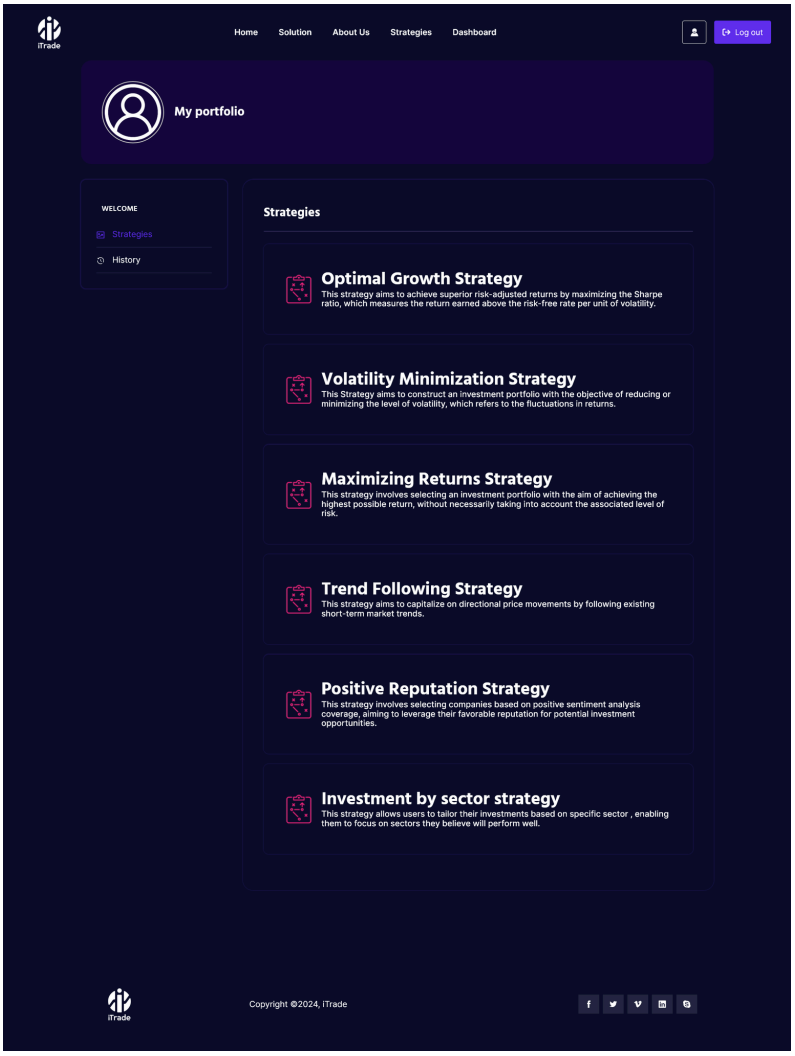


Figure 67: strategies

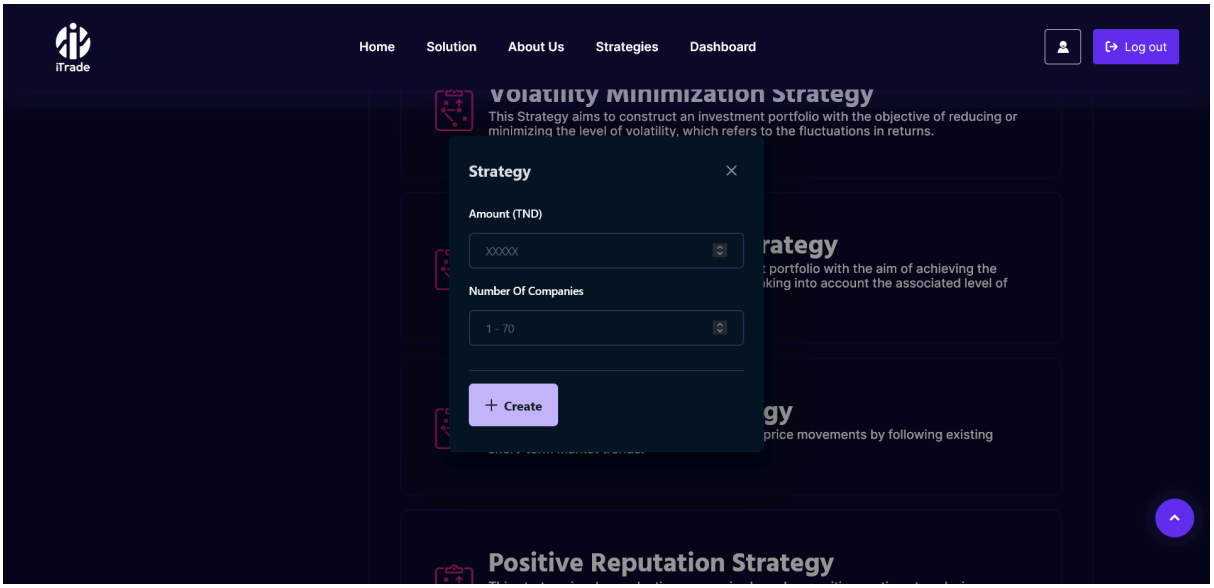


Figure 68: strategy form

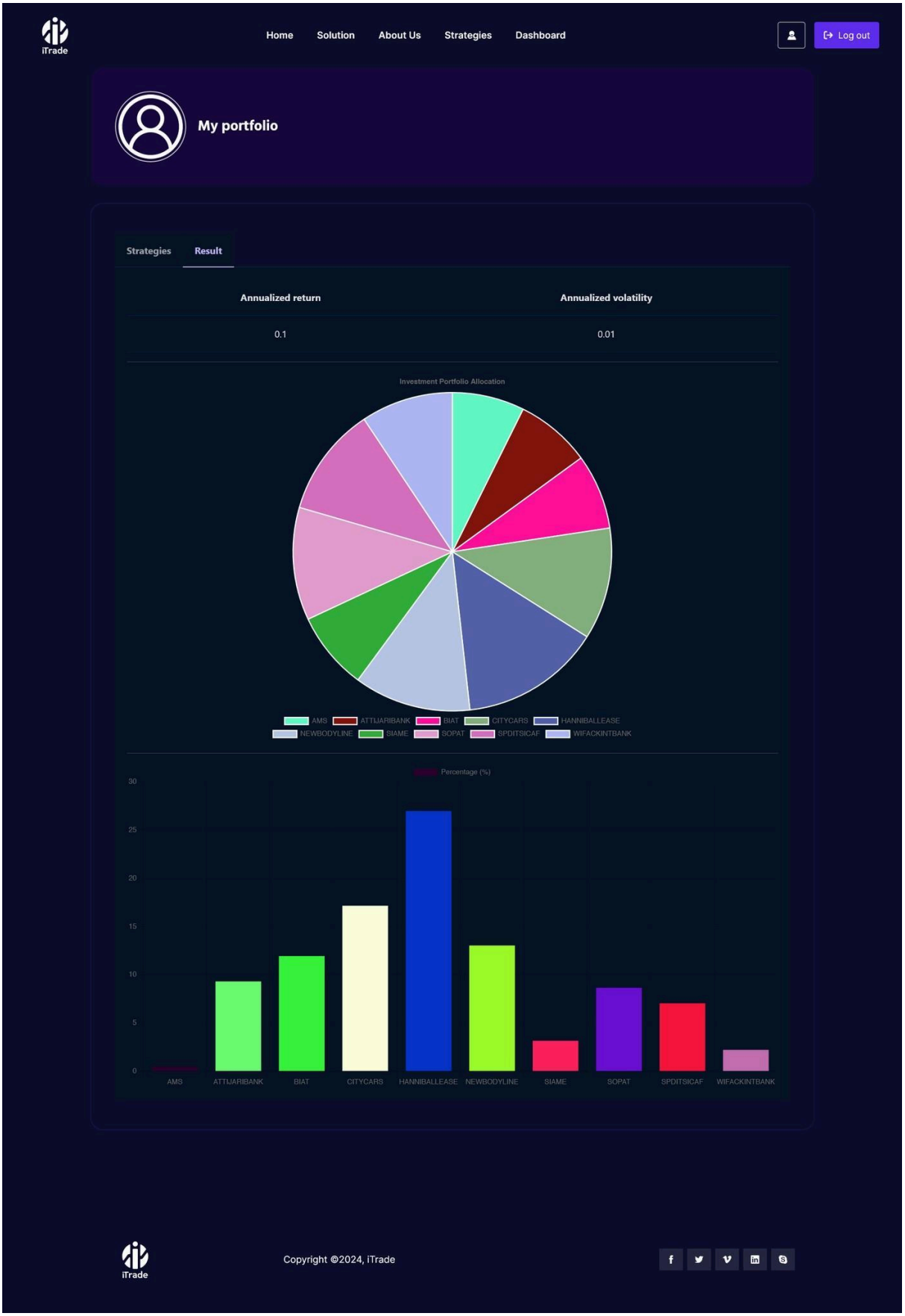


Figure 69: portfolio allocation

2.5 Profile page

The "Profile" section empowers users to manage their account details. They can update personal information and modify their login credentials.

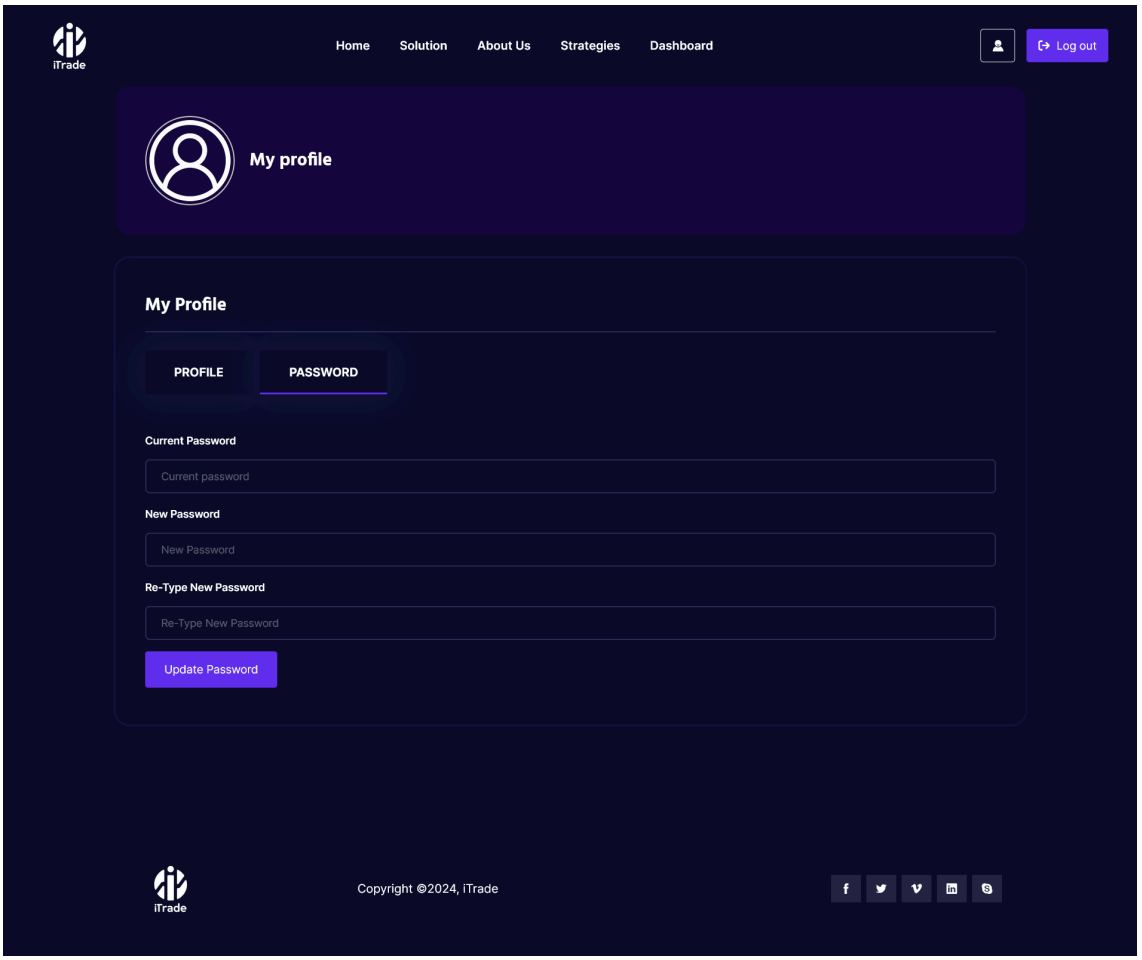


Figure 70: profile page

3. Conclusion

This chapter, we successfully deployed the prediction models as a web service using Flask and MySQL. Users can now interact with the models through the API to obtain predictions based on historical data. This paves the way for further development, such as integrating the API into a user interface or dashboard for a more comprehensive analysis experience.

Conclusion

Our suggested solution for the Tunisian stock market aims to transform investment decision-making by merging data science approaches with a user-centric platform.

By concentrating on optimizing returns, limiting risk, and improving customization, we want to improve the investment experience in a challenging market.

AI-powered sentiment analysis, individual risk tolerance optimization, and community-driven portfolio discovery increase sophistication while encouraging investors to collaborate.

Furthermore, the use of comprehensive model quality measures promotes transparency and reliability. As we execute this solution, we are committed to offering a comprehensive, efficient, and user-friendly tool that enables Tunisian stock market investors.

References

- Keras API documentation [https://keras.io/api/]
- TensorFlow documentation [https://www.tensorflow.org/api_docs]