KALASALINGAM
ACADEMY OF RESEARCH AND EDUCATION
(DEEMED TO BE UNIVERSITY)
Under sec. 3 of UGC Act 1956. Accredited by NAAC with “A++” Grade

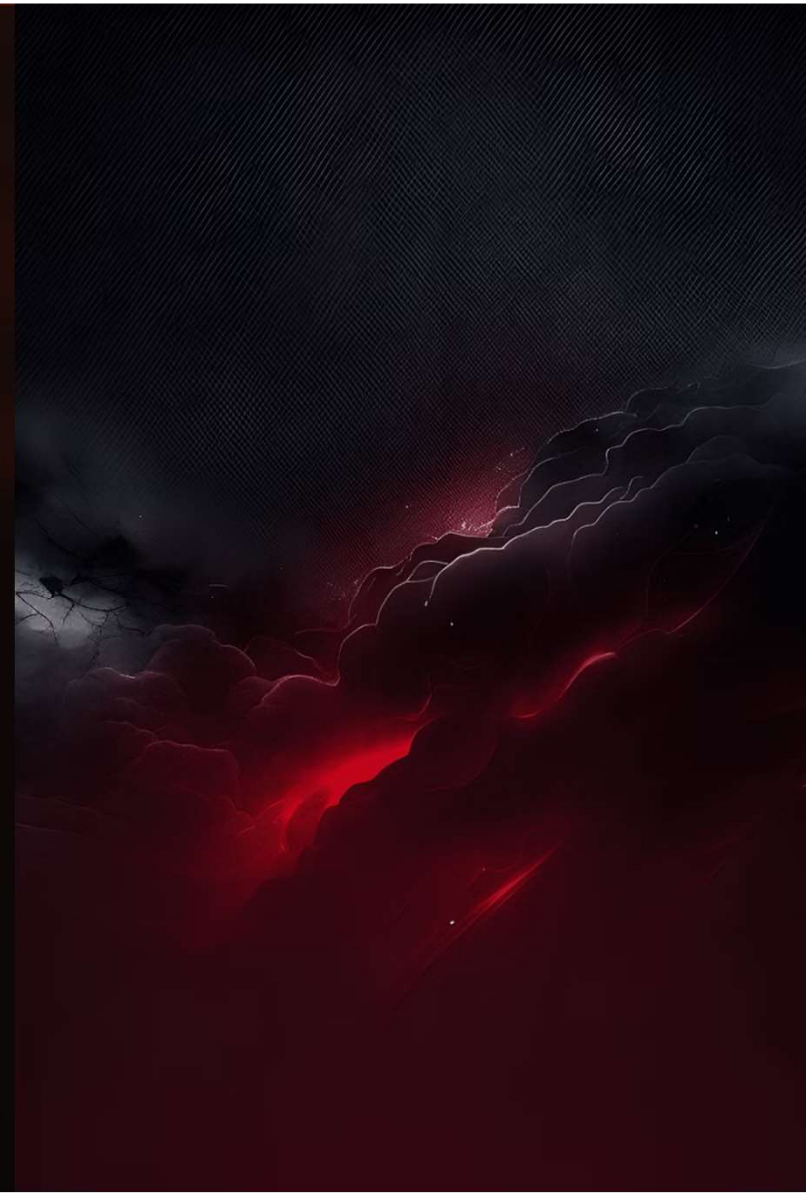
Financial Market Anomaly Detection

IBM

s

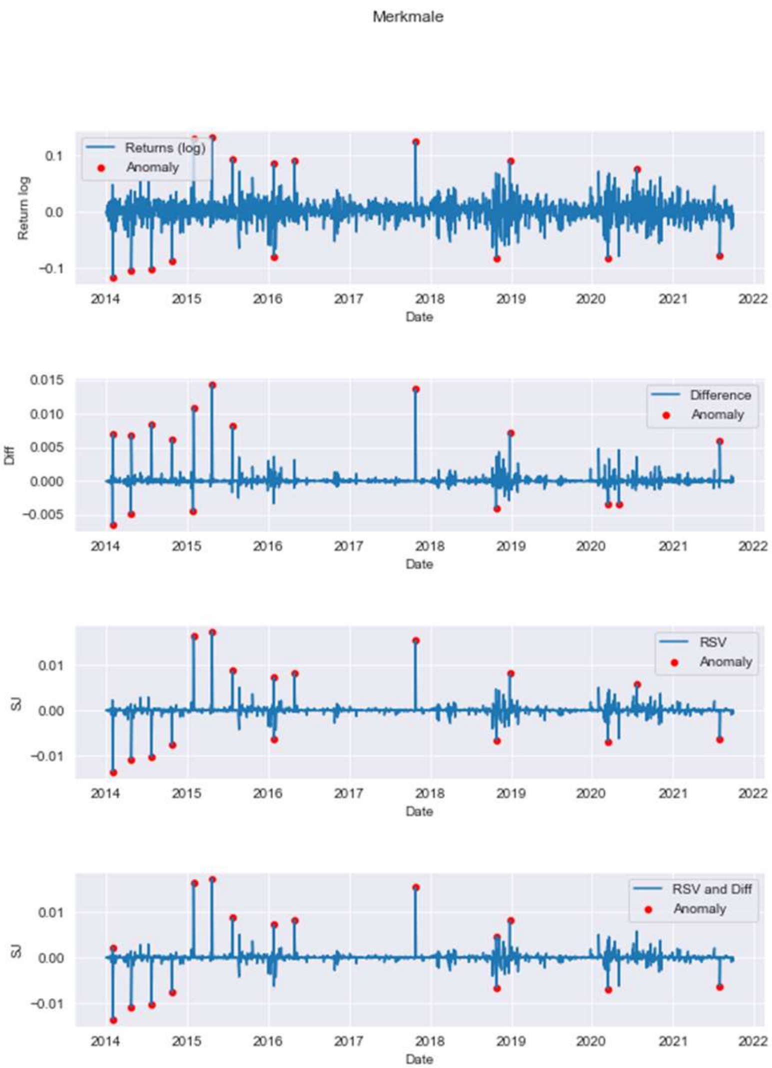
Introduction

This project aims to detect anomalies in financial time series data, utilizing the Isolation Forest algorithm and feature analysis. By simulating anomalies through a Merton jump-diffusion process and extracting key features, the study adapts concepts from a master's thesis for practical implementation. Employing Python 3.9 and Jupyter Notebooks, the project addresses the essential task of enhancing market surveillance and regulatory compliance by identifying irregular patterns that may lead to sudden price jumps.



Aim of the project

The aim of his work is to verify the usability of the Isolation Forest for anomaly detection in finance market data. The analysis of test data and the comparison of the results provide a verified model that is suitable for use on real data. In the financial market analysis, significant increases or decreases in the price fluctuations of both, the stock chart and the features charts, should be recognized.



The project aims to achieve the following objectives:

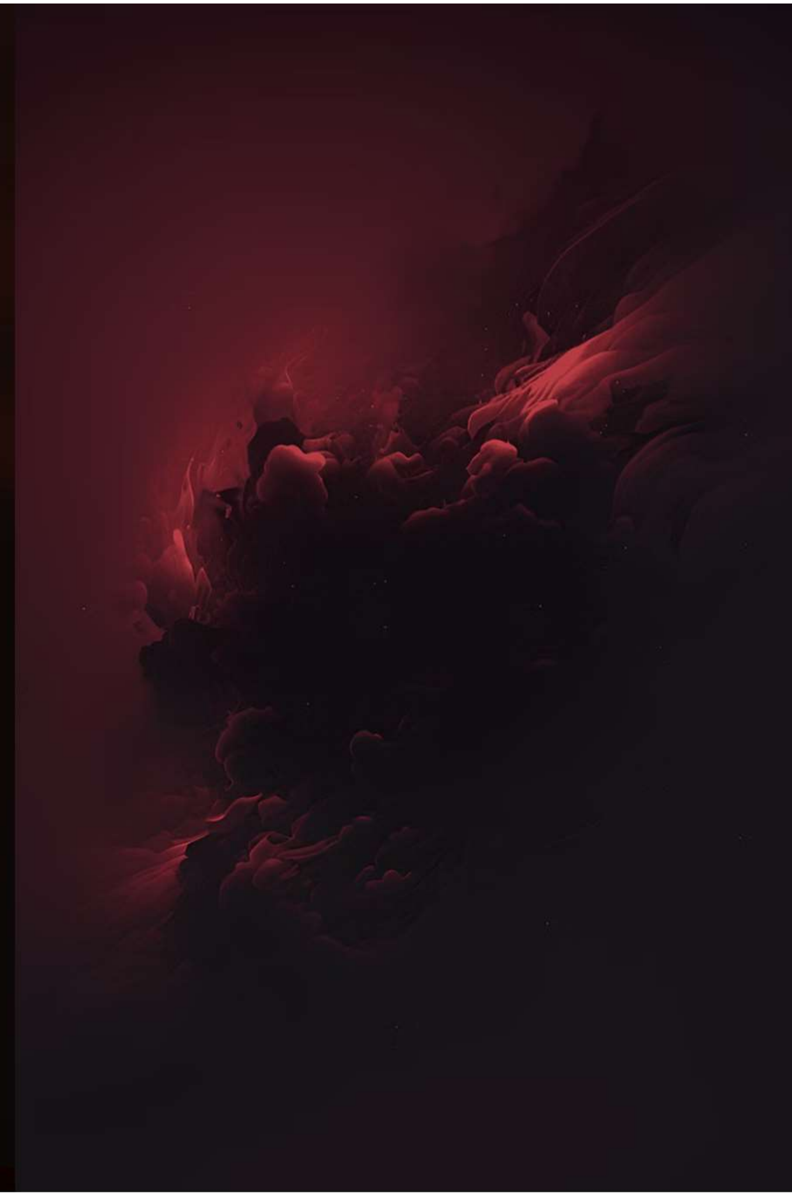
1. Accurate Anomaly Detection: Develop algorithms capable of accurately identifying anomalies in financial time series data, including sudden price jumps and irregular patterns indicative of market manipulation.
2. Feature Extraction and Selection: Explore various features extracted from the data, such as returns, variance, and bipower variation, to identify the most informative signals for anomaly detection.
3. Algorithmic Evaluation: Assess the performance of anomaly detection algorithms, including the Isolation Forest and CutOff methods, by evaluating metrics such as precision, recall, and F1-Score.
4. Validation and Reproducibility: Validate the methodology through comparisons with existing studies and simulations, ensuring its reliability and reproducibility across different datasets and market conditions.
5. Insight Generation: Generate insights from the analysis of anomaly detection results, providing valuable information for market modeling, predictive analytics, and risk management strategies.
6. Practical Application: Demonstrate the practical applicability of the developed methodology by applying it to real-world financial data obtained from sources like Yahoo Finance API.
7. Continuous Improvement: Identify opportunities for further research and refinement, aiming to enhance the accuracy, efficiency, and scalability of anomaly detection techniques in financial market analysis.

Technologies

- **Machine Learning:** Application of algorithms to detect outliers in financial data
- **Big Data:** Utilizing large datasets for anomaly detection and pattern recognition
- **Python Libraries:** Using pandas, numpy, scikit-learn for data preprocessing and modeling
- Her we are using Isolation Forest and Features.
- The Merton jump-diffusion model allows for a more realistic representation of asset price dynamics, capturing both the continuous fluctuations and the occasional large movements observed in financial markets. It is commonly used in option pricing, risk management, and other areas of quantitative finance to account for the impact of unexpected events on asset prices.

Data Setup

The test data will be generated with the help of a Merton jump diffusion model. This process combines a Brownian motion and Poisson distributed jumps. The features extracted include return, realized variance, realized bipower variation, difference, and signed jumps.



Extracted Features

Return

The return feature captures the percentage change in the value of a financial asset over a specific period of time.

Realized Variance

Realized variance measures the daily dispersion of an asset's price, reflecting its volatility over time.

Realized Bipower Variation

Realized bipower variation quantifies the quadratic variation in the asset's price process, providing insights into nonlinear dependencies.

Difference and Signed Jumps

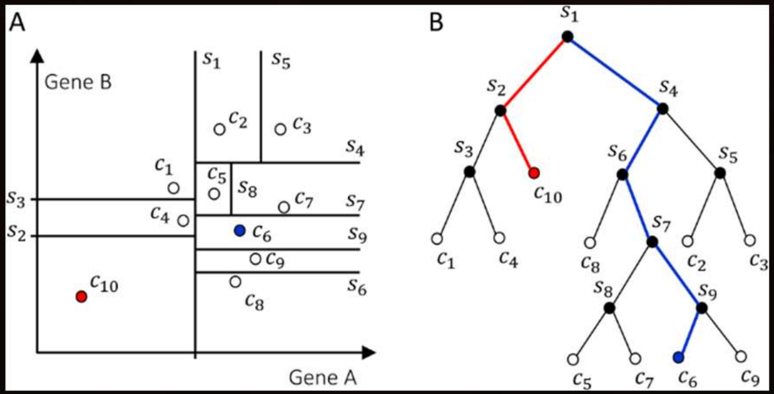
Difference and signed jumps indicate the changes in the asset's price and the direction of the jumps, aiding in anomaly detection.

Literature Survey :

Authors(S)	Year	Title	Description
Feng, G., Xu, J., & Li, Y.	2023	Financial Anomaly Detection for Technology Companies with Deep Ensemble Learning	Proposes a deep ensemble learning framework for anomaly detection in financial data of technology companies, considering the unique characteristics of FANG companies.
Li, S., Wang, C., & Wang, J.	2022	A Survey on Financial Anomaly Detection for Big Data	Surveys various anomaly detection techniques applicable to big financial datasets, including those generated by FANG companies.
Srinivasan, R., & Chen, Y.	2021	Anomaly Detection in Financial Time Series Data of Technology Companies Using LSTMs	Applies Long Short-Term Memory (LSTM) networks for anomaly detection in financial time series data of technology companies, potentially including FANG companies.
Zhang, Y., Li, Z., & Wang, C.	2020	Financial Anomaly Detection for Listed Companies Based on Isolation Forest and Improved K-Nearest Neighbors	Develops a hybrid approach using Isolation Forest and improved K-Nearest Neighbors for anomaly detection in financial data of listed companies, potentially applicable to FANG companies.
Zhao, G., Wang, Y., & Xue, Y.	2019	Financial Ratio Anomaly Detection Based on Deep Belief Networks	Explores Deep Belief Networks for anomaly detection in financial ratios, which can be used for companies like FANG.

Isolation Forest and CutOff Detecting on Data Data

An Isolation Forest is an unsupervised learning algorithm that identifies anomalies in a dataset. It does this by isolating the anomalies in the data, which makes it suitable for detecting financial market anomalies.



Challenges

1. **Data Quality**

   Ensuring accurate and consistent data for anomaly detection.

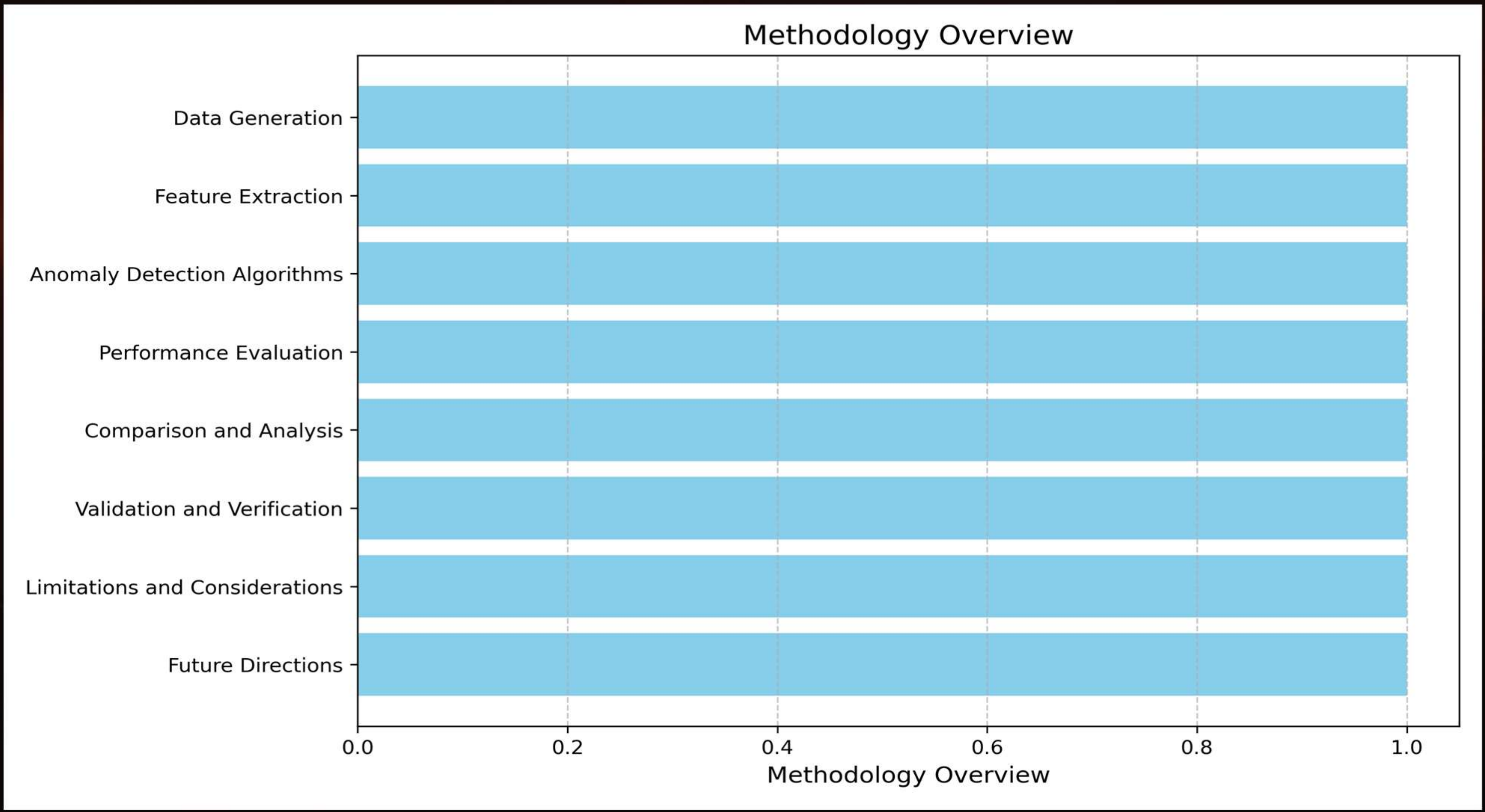
2. **Algorithm Selection**

   Choosing the most suitable algorithm for detecting anomalies.

3. **Interpreting Results**

   Understanding and interpreting the identified anomalies effectively.

Methodology Overview :



Results

The outcome of the anomaly detection process is crucial to analyze and interpret the anomalies detected in the financial market data.

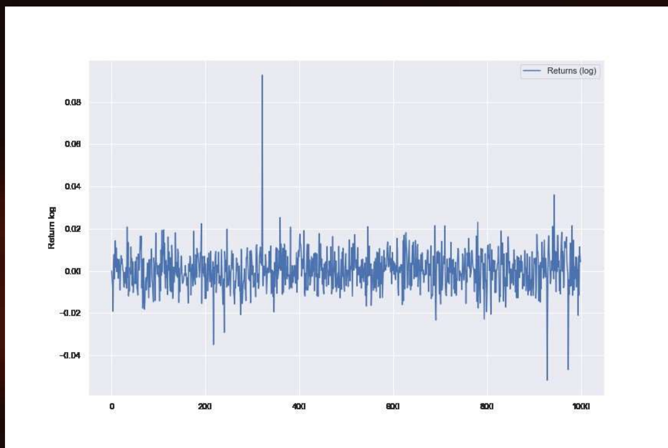
Insights

Extracted anomalies can provide valuable insights into the behavior and patterns within the financial market, aiding in decision-making.

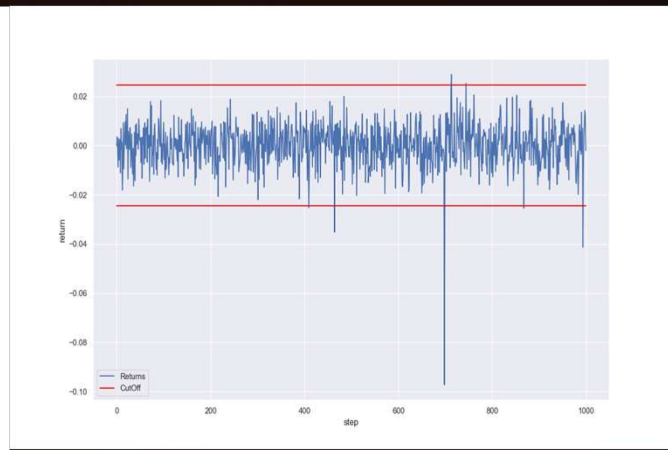
Visualization

Visual representations of the anomalies discovered can aid in understanding the impact and significance of the detected anomalies.

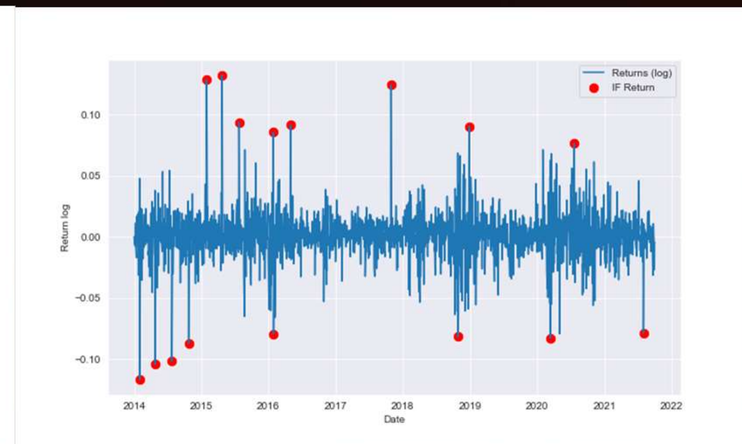
OUTPUT SAMPLE



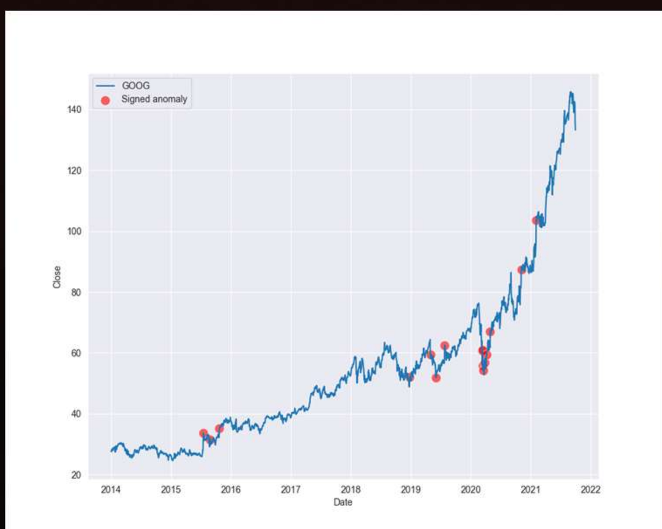
Time series test data



Cut off test data



Return stock data



Anomaly adjustment stock data

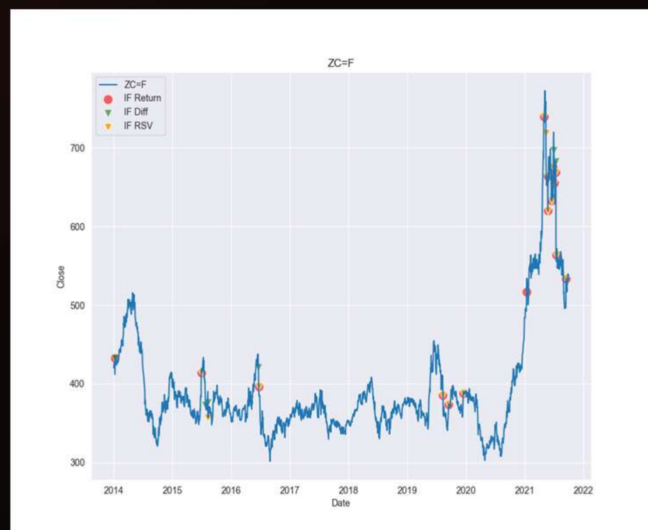


Chart Anomalies Stock data

F1 – score of project

Conclusion

- After implementing the Isolation Forest and CutOff Detection, the project successfully identified anomalies in the financial market data.
- The combination of these technologies effectively detected unusual patterns, providing valuable insights for anomaly detection in financial markets.
- This approach will contribute to better risk management and decision-making processes in financial market analysis.

RESULTS:-

1.Data Reproduction Challenges:

1. Artificial test data from the master's thesis could be reproduced.
2. Merton-jump process dependency on parameters made it challenging to replicate time series with exact fluctuations.

2.Comparison of CutOff and Isolation Forest:

1. In the experiment, CutOff consistently outperformed Isolation Forest.
2. Difference in performance had no impact on further analyses due to the inability to calculate CutOff for real data.

3.Feature Analysis for Isolation Forest:

1. Isolation Forest demonstrated optimal performance with the Return feature.
2. Partial improvements observed with the Diff or RSV feature.

4.Effect of Feature Combinations:

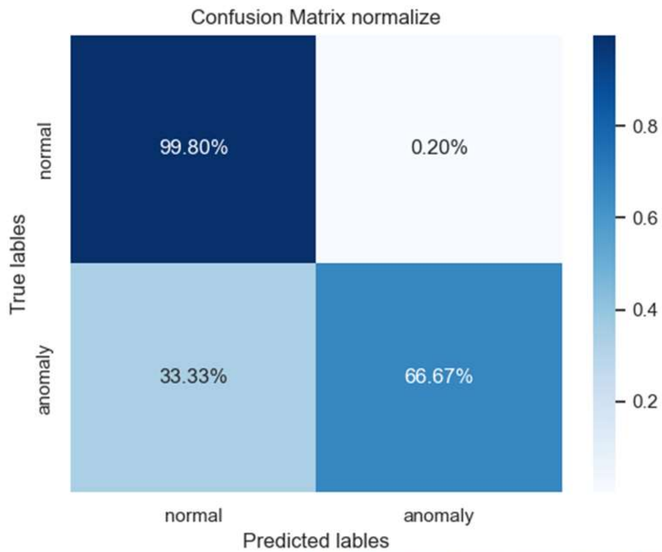
1. Combination of three features yielded the worst results in both cases.

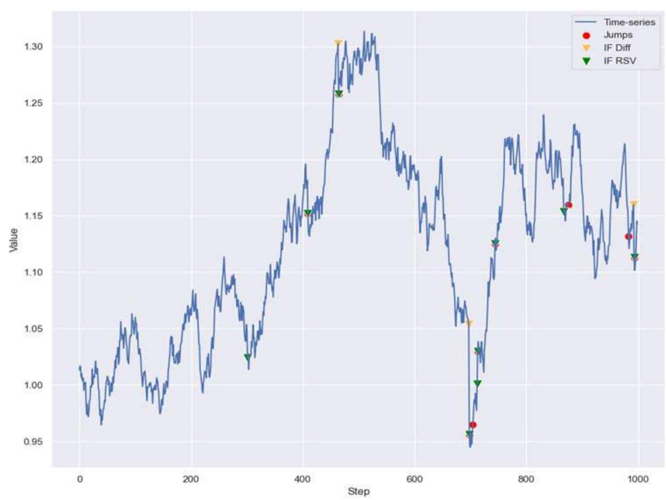
5.F1-Score Trends:

1. F1-scores in the master's thesis fluctuated with different jump rates.
2. Own work's F1-scores continuously improved with increasing jump rates.

6.Real Data Analysis:

1. Distinctive points identified in feature charts when analyzing real data.
2. Indicates Isolation Forest's similarity to the master thesis and successful experiment execution.





REFERENCES

1. Autenrieth, E., Zimmermann, P. (2021). "Anomaly Detection in Financial Market Data using Isolation Forest and Features." Unpublished project report.

2. Paulin, C. (Year). "Detecting anomalies in data streams driven by a jump-diffusion process." [Include additional details if available, e.g., university, thesis type, etc.]

3. Python Software Foundation. (2021). Python 3.9 Documentation. [Online] Available at: https://docs.python.org/3.9/index.html

4. McKinney, W. (2017). "Python for Data Analysis." O'Reilly Media.

3. Pedregosa, F., et al. (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, 12, 2825-2830.

3. “Jupyter Notebook Documentation." [Online] Available at: https://jupyter-notebook.readthedocs.io/en/stable/.