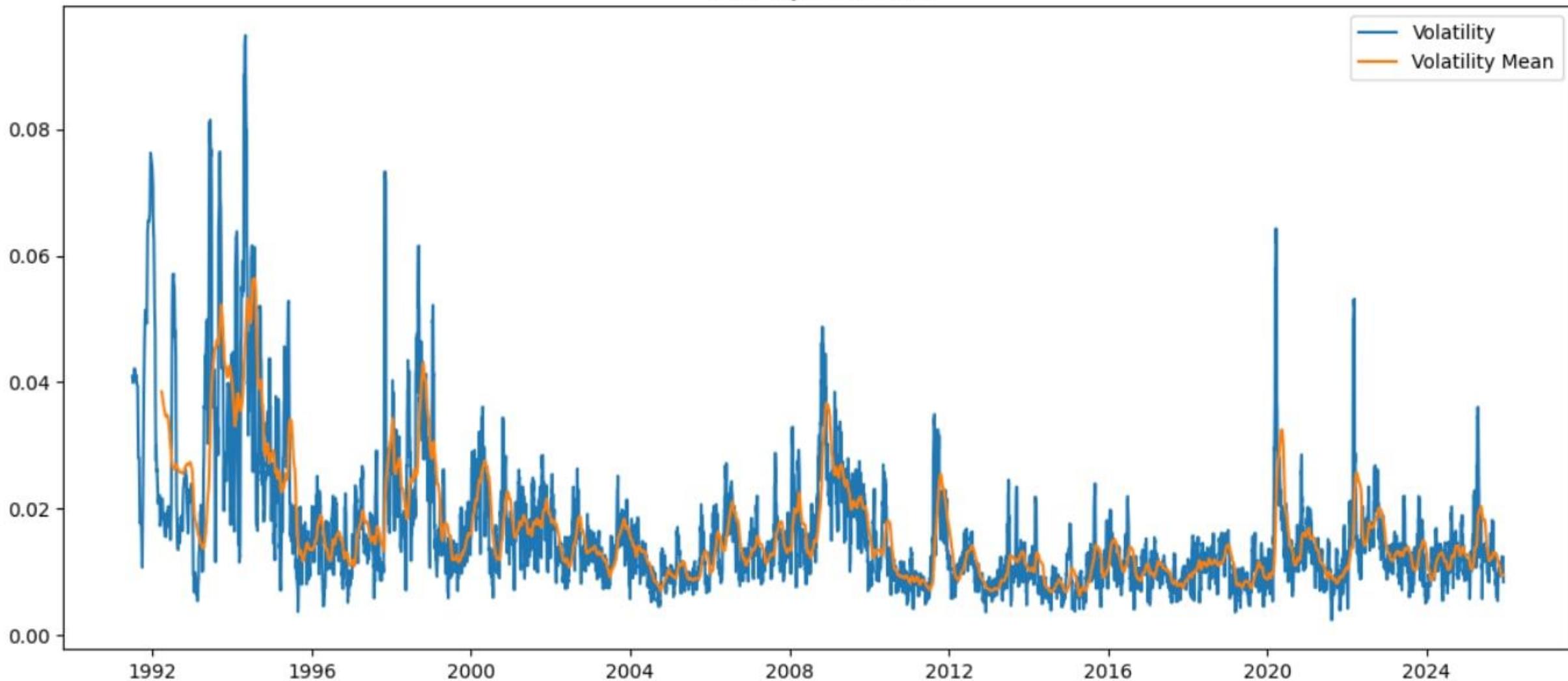
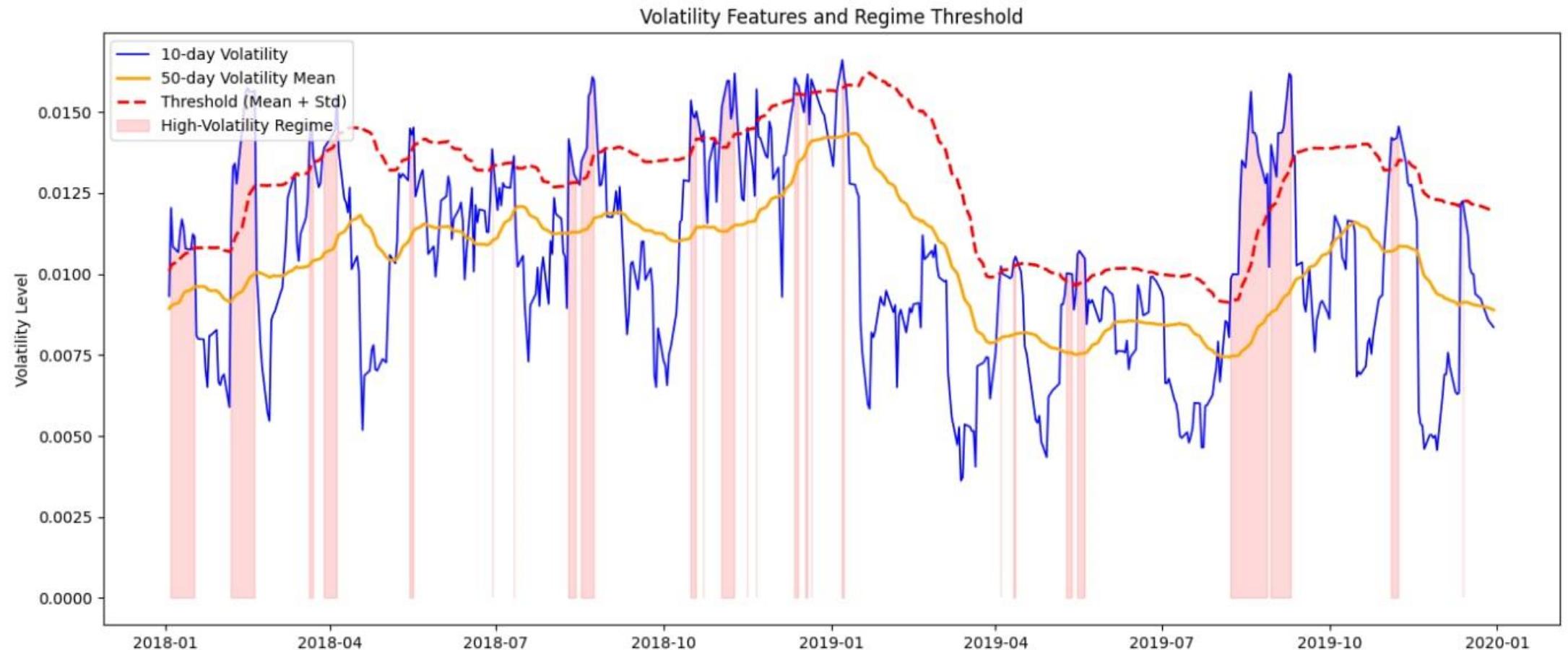


Identifying volatility regimes using supervised learning

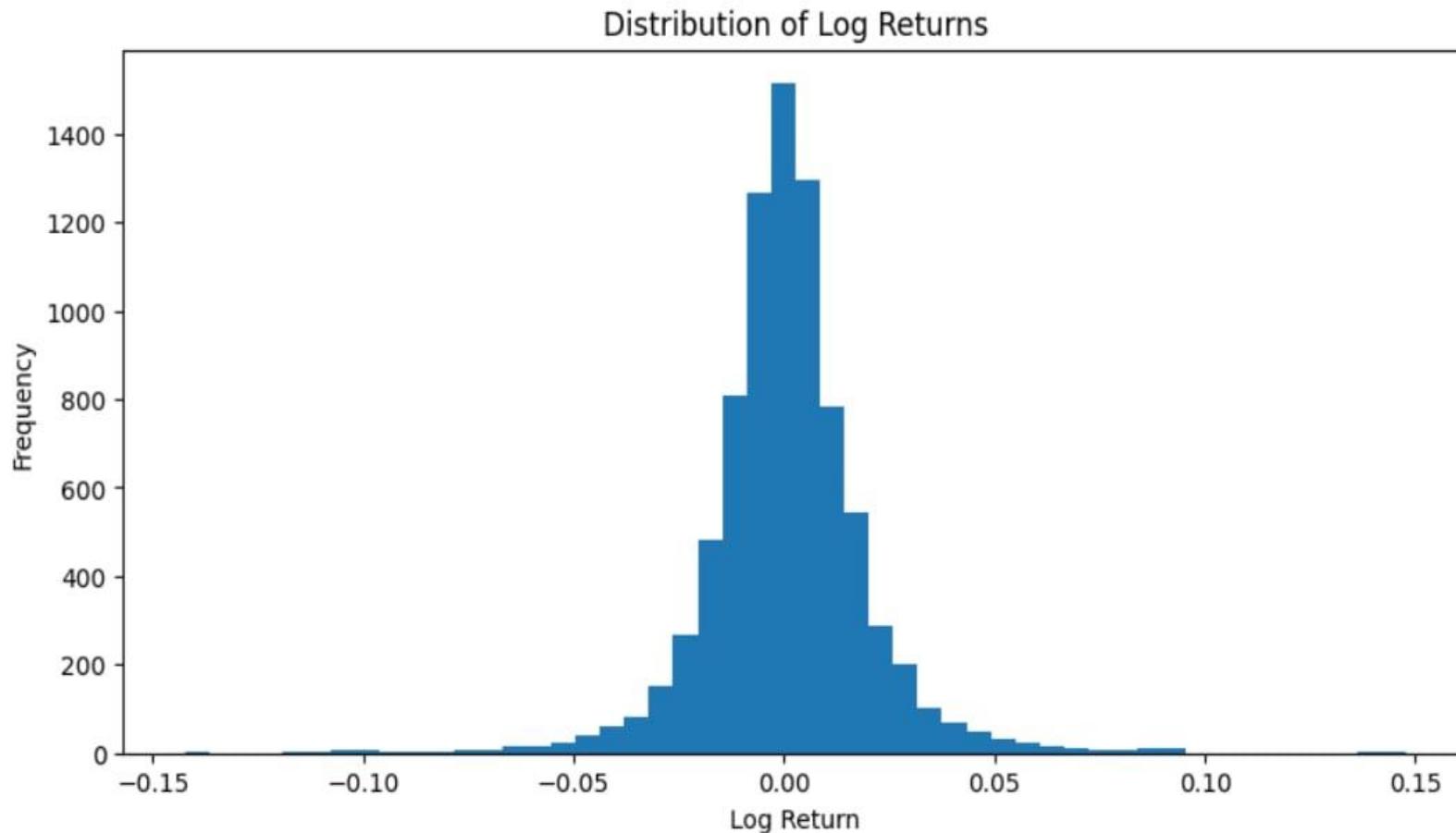
Volatility Over Time



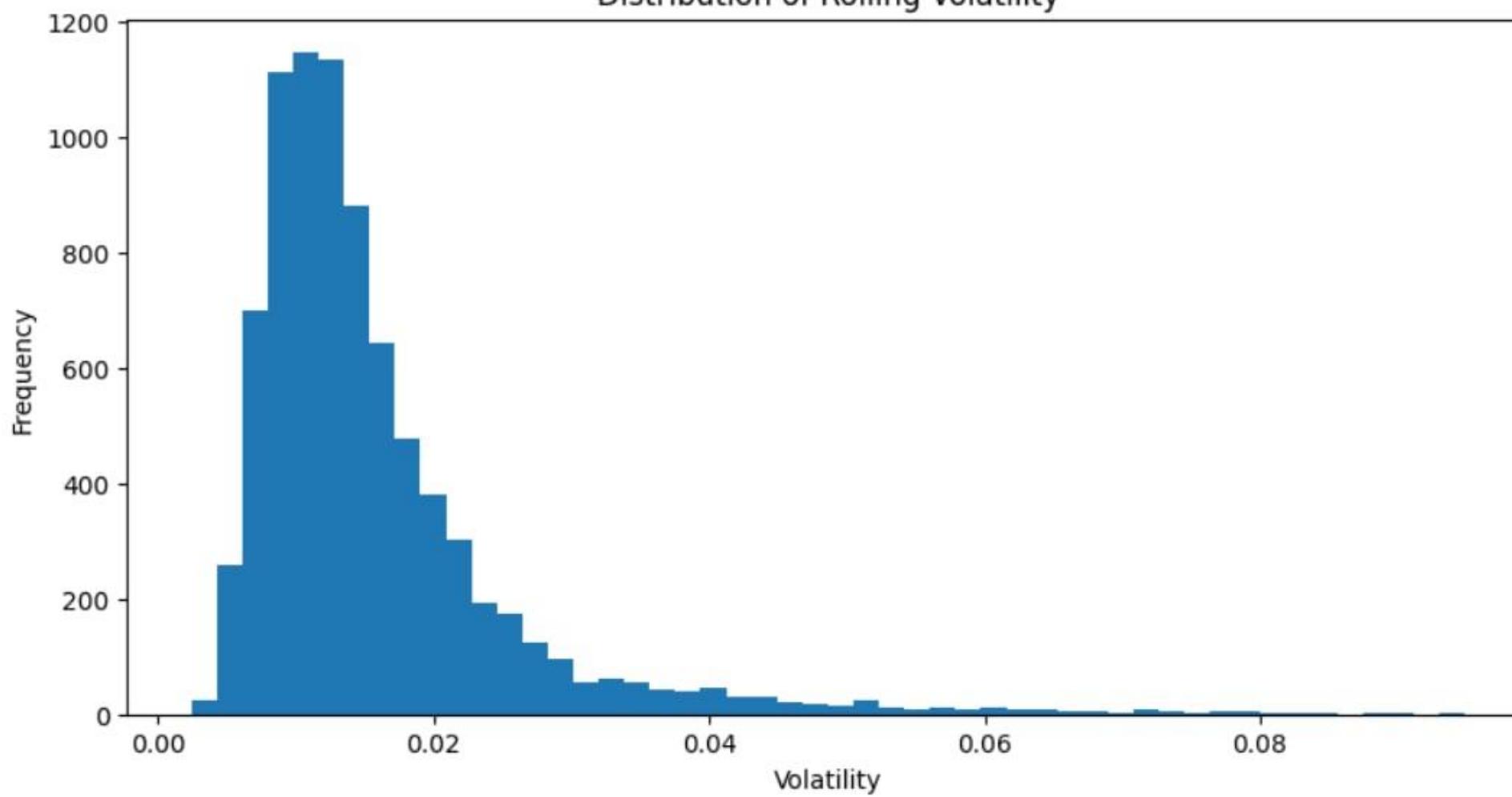
Data and feature engineering



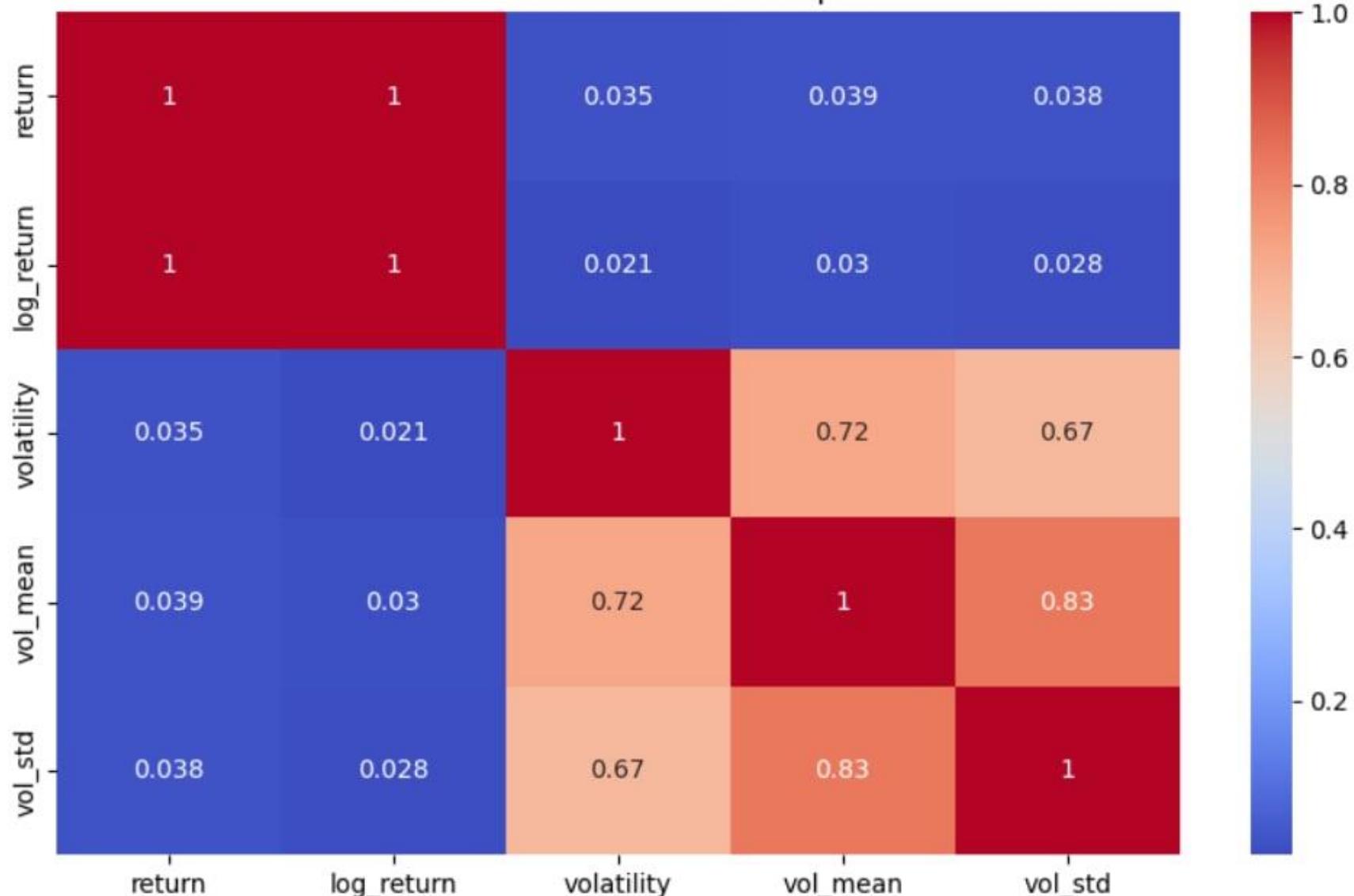
EDA highlights



Distribution of Rolling Volatility



Feature Correlation Heatmap



Models used

- Logistic Regression
- Random Forest
- Gradient Boosting

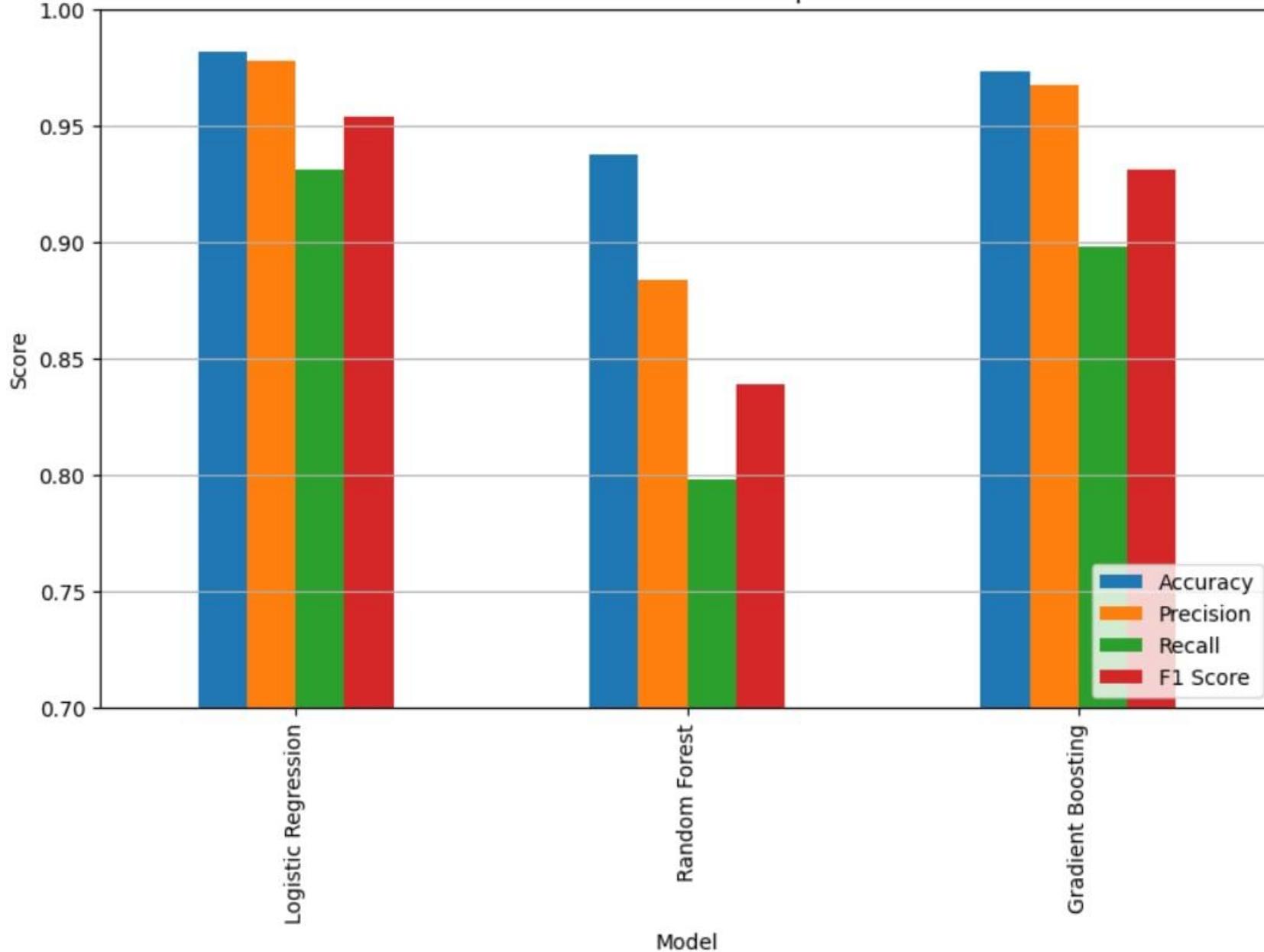
Evaluation metrics

--- Logistic Regression ---				
	precision	recall	f1-score	support
0	0.98	0.99	0.99	1300
1	0.98	0.93	0.95	332
accuracy			0.98	1632
macro avg	0.98	0.96	0.97	1632
weighted avg	0.98	0.98	0.98	1632
--- Random Forest ---				
	precision	recall	f1-score	support
0	0.95	0.97	0.96	1300
1	0.88	0.80	0.84	332
accuracy			0.94	1632
macro avg	0.92	0.89	0.90	1632
weighted avg	0.94	0.94	0.94	1632
--- Gradient Boosting ---				
	precision	recall	f1-score	support
0	0.97	0.99	0.98	1300
1	0.97	0.90	0.93	332
accuracy			0.97	1632
macro avg	0.97	0.94	0.96	1632
weighted avg	0.97	0.97	0.97	1632

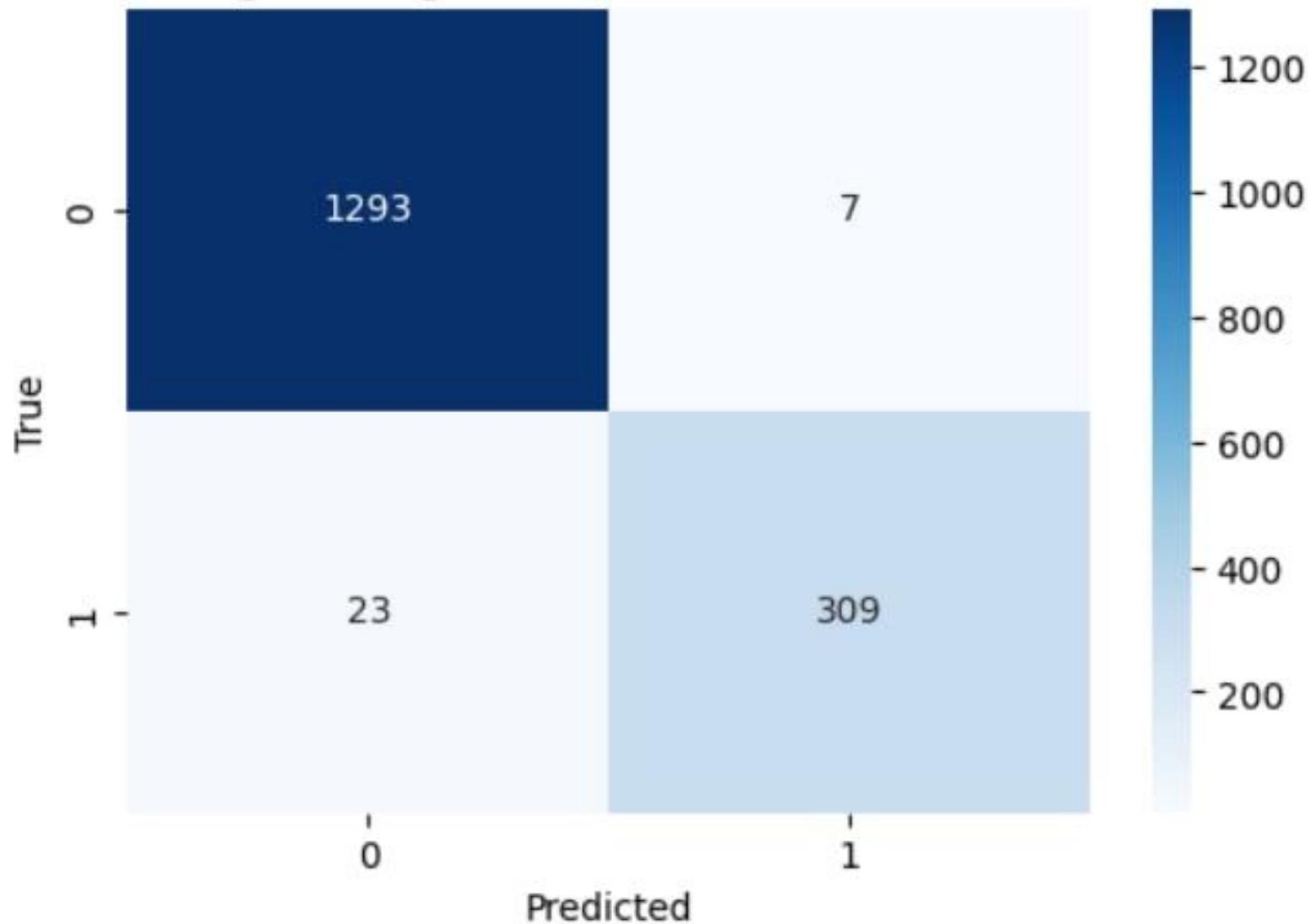
Results table

	Model	Accuracy	Precision	Recall	F1 Score
0	Logistic Regression	0.981618	0.977848	0.930723	0.953704
1	Random Forest	0.937500	0.883333	0.798193	0.838608
2	Gradient Boosting	0.973039	0.967532	0.897590	0.931250

Model Performance Comparison



Logistic Regression — Confusion Matrix



Code and notebook

- <https://github.com/patryklyszyk-ctrl/wse-volatility-regime-ml>