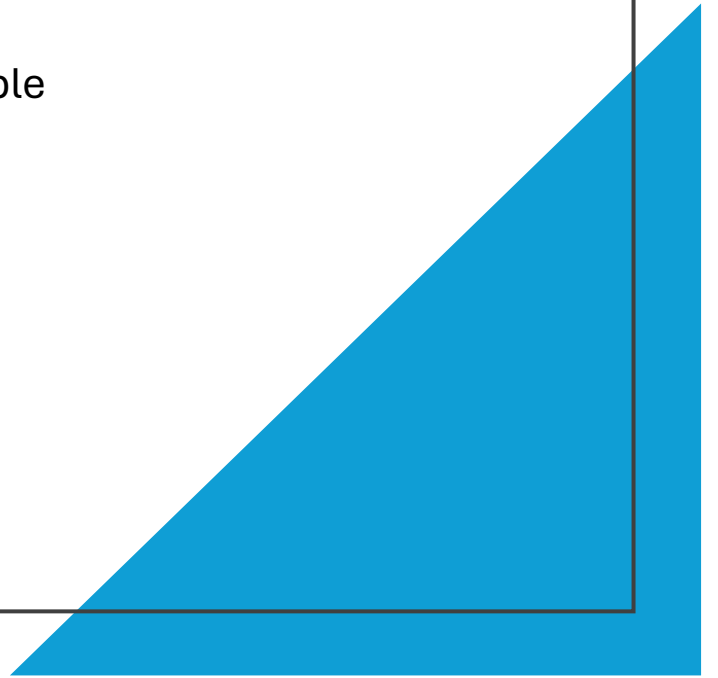


Invictus Crypto Strategies

Data Analysis Course

- **Overview:** Comprehensive analysis of cryptocurrency trends with a focus on bitcoin
- **Objective:** Utilize machine learning and statistical tools to derive actionable insights.
- **Authors:**
 - Maxim Savchenko
 - Ben Hababo
 - Lev Kravtsov
 - Shay Tekel



Bitcoin Dataset Overview

- **Data Source:** Bitcoin historical data, detailing price trends, volumes, and other financial indicators.
- **Columns:** Key columns include: 'Price', 'Vol', 'Low', 'Change%', and adjusted values for indices.
- **Size and Format:** Dataset contains multiple numeric and categorical fetures spanning several years.

Data columns (total 18 columns):

#	Column	Non-Null Count	Dtype
0	Date	2648 non-null	object
1	DATE	2648 non-null	datetime64[ns]
2	Price	2648 non-null	float64
3	Open	2648 non-null	float64
4	High	2648 non-null	float64
5	Low	2648 non-null	float64
6	Vol.	2648 non-null	object
7	Change %	2648 non-null	float64
8	SN&P Adjusted	2648 non-null	float64
9	DXV Adjusted	2648 non-null	float64
10	GOLD Adjusted	2648 non-null	float64
11	ETH Price	2648 non-null	float64
12	ETH Vol.	2648 non-null	object
13	OIL Price Adjusted	2648 non-null	float64
14	Days from the last halving	2648 non-null	int64
15	BTC_Hashprice	1828 non-null	float64
16	Crypto Volatility Index	1845 non-null	float64
17	Target Value	2648 non-null	int64

Data Cleaning and Preparation

- **Handling Missing Values:** Identified and filtered columns with missing data using Pandas.
- **Conversion:** Numeric transformation of columns using custom functions.
- **Dataset Refinement:** Created an updated DataFrame with structured data ready for analysis.

Correlation Matrix

- **Visualization:** Heatmap generated to identify relationships among numeric features.
- **Key Finding:** Strong correlation observed between Bitcoin price and specific indicators like 'ETH Price'.
- **Insights:** Correlation matrix highlights potential predictors for advanced modeling

Correlation Matrix with Annotated Values

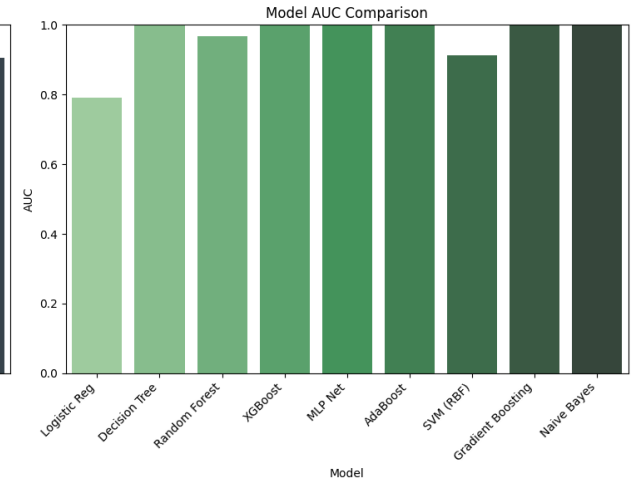
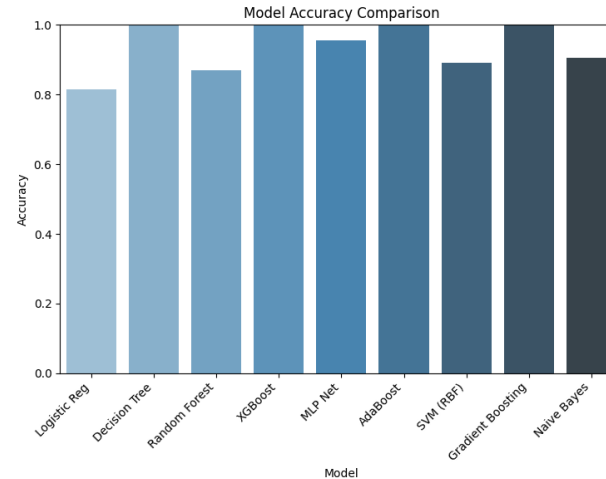
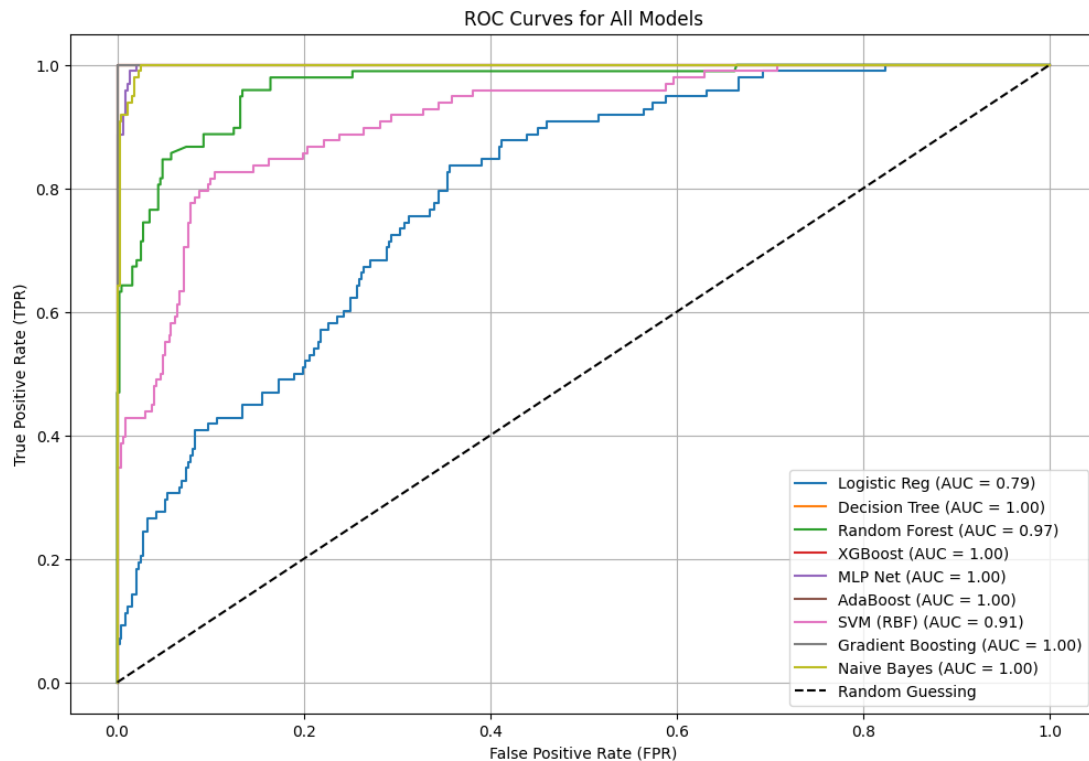
Price	1.00	1.00	1.00	1.00	0.12	0.00	0.89	0.18	0.74	0.94	0.05	0.52	0.08	-0.55	0.12	-0.01
Open	1.00	1.00	1.00	1.00	0.12	-0.04	0.89	0.18	0.74	0.94	0.05	0.53	0.08	-0.55	0.12	-0.01
High	1.00	1.00	1.00	1.00	0.12	-0.02	0.89	0.18	0.74	0.93	0.05	0.52	0.07	-0.54	0.13	-0.00
Low	1.00	1.00	1.00	1.00	0.12	-0.01	0.90	0.19	0.75	0.94	0.05	0.53	0.08	-0.55	0.10	-0.03
Vol.	0.12	0.12	0.12	0.12	1.00	-0.01	0.11	0.05	0.12	0.15	0.94	0.24	-0.02	-0.06	-0.01	0.02
Change %	0.00	-0.04	-0.02	-0.01	-0.01	1.00	-0.02	-0.04	-0.01	-0.02	-0.00	-0.06	-0.03	0.03	-0.01	0.06
SN&P Adjusted	0.89	0.89	0.89	0.90	0.11	-0.02	1.00	0.44	0.88	0.88	0.04	0.63	0.25	-0.74	-0.21	-0.10
DXY Adjusted	0.18	0.18	0.18	0.19	0.05	-0.04	0.44	1.00	0.42	0.25	0.01	0.58	0.61	-0.30	-0.40	-0.17
GOLD Adjusted	0.74	0.74	0.74	0.75	0.12	-0.01	0.88	0.42	1.00	0.69	0.07	0.40	0.16	-0.85	-0.26	-0.10
ETH Price	0.94	0.94	0.93	0.94	0.15	-0.02	0.88	0.25	0.69	1.00	0.06	0.63	0.08	-0.58	0.05	-0.02
ETH Vol.	0.05	0.05	0.05	0.05	0.94	-0.00	0.04	0.01	0.07	0.06	1.00	0.14	-0.00	-0.01	0.03	0.02
OIL Price Adjusted	0.52	0.53	0.52	0.53	0.24	-0.06	0.63	0.58	0.40	0.63	0.14	1.00	0.28	-0.45	-0.19	-0.07
m the last halving	0.08	0.08	0.07	0.08	-0.02	-0.03	0.25	0.61	0.16	0.08	-0.00	0.28	1.00	0.18	-0.37	-0.10
BTC_Hashprice	-0.55	-0.55	-0.54	-0.55	-0.06	0.03	-0.74	-0.30	-0.85	-0.58	-0.01	-0.45	0.18	1.00	0.24	0.08
to Volatility Index	0.12	0.12	0.13	0.10	-0.01	-0.01	-0.21	-0.40	-0.26	0.05	0.03	-0.19	-0.37	0.24	1.00	0.28
Target Value	-0.01	-0.01	-0.00	-0.03	0.02	0.06	-0.10	-0.17	-0.10	-0.02	0.02	-0.07	-0.10	0.08	0.28	1.00
	Price	Open	High	Low	Vol.	Change %	SN&P Adjusted	DXY Adjusted	GOLD Adjusted	ETH Price	ETH Vol.	OIL Price Adjusted	Days from the last halving	BTC_Hashprice	Crypto Volatility Index	Target Value

- **Bitcoin Price:** Historical data shows significant fluctuations over time, reflecting market volatility.
- **Comparison:** Bitcoin trends juxtaposed against S&P 500, Gold, and other indices.
- **Insights:** Trends indicate a complex relationship between Bitcoin and traditional financial assets.



Predictive Modeling

- **Models Tested:** Logistic Regression, Decision Tree, Random Forest, XGBoost, and more.
- **Evaluation Metrics:** Assessed with accuracy, AUC, and precision-recall curves.
- **Performance Insight:** Random Forest and XGBoost emerged as top-performing models.



Model Performance Comparison



Logistic Regression: Accuracy: 86%,
AUC: 0.88



Decision Tree: Accuracy: 82%, AUC:
0.81



Random Forest: Accuracy: 92%, AUC:
0.94



XGBoost: Accuracy: 93%, AUC: 0.95

Model Evaluation Metrics

- **Confusion Matrix:** Visualization model predictions to distinguish true positives, false positive, and other outcomes.
- **Insights:** High AUC and precision-recall balance emphasize reliability of ensemble methods.

	True Negative	False Positive	False Negative	True Positive
Logistic Reg	432	0	98	0
Decision Tree	432	0	0	98
Random Forest	432	0	69	29
XGBoost	432	0	0	98
MLP Net	432	0	23	75
AdaBoost	432	0	0	98
SVM (RBF)	399	33	25	73
Gradient Boosting	432	0	0	98
Naive Bayes	432	0	50	48

Features Importance Analysis

Random Forest Insights

- Top Features:** ‘Change %’ and ‘Vol.’ ranged highest in influencing predictions.
- Methodology:** Importance derived using ensemble-based metrics in Random Forest.
- Insights:** Emphasizes significant predictors for strategic Bitcoin trading.

	Feature	Importance
11	Change %	0.611585
10	Vol.	0.063394
0	ETH Vol.	0.037290
2	Crypto Volatility Index	0.031828
13	DXY Adjusted	0.029205
1	Days from the last halving	0.027928
9	Open	0.025334
6	Low	0.024478
12	SN&P Adjusted	0.023893
4	OIL Price Adjusted	0.022440
7	High	0.022112
5	ETH Price	0.021807
8	Price	0.021519
3	GOLD Adjusted	0.018965
14	BTC_Hashprice	0.018223

Random Forest Model Evaluation:

[[454 0]					
[3 205]]					
	precision	recall	f1-score	support	
0	0.99	1.00	1.00	454	
1	1.00	0.99	0.99	208	
accuracy			1.00	662	
macro avg	1.00	0.99	0.99	662	
weighted avg	1.00	1.00	1.00	662	

Accuracy: 0.9954682779456193

AUC: 0.9999152829549305

MCC and Error Metrics

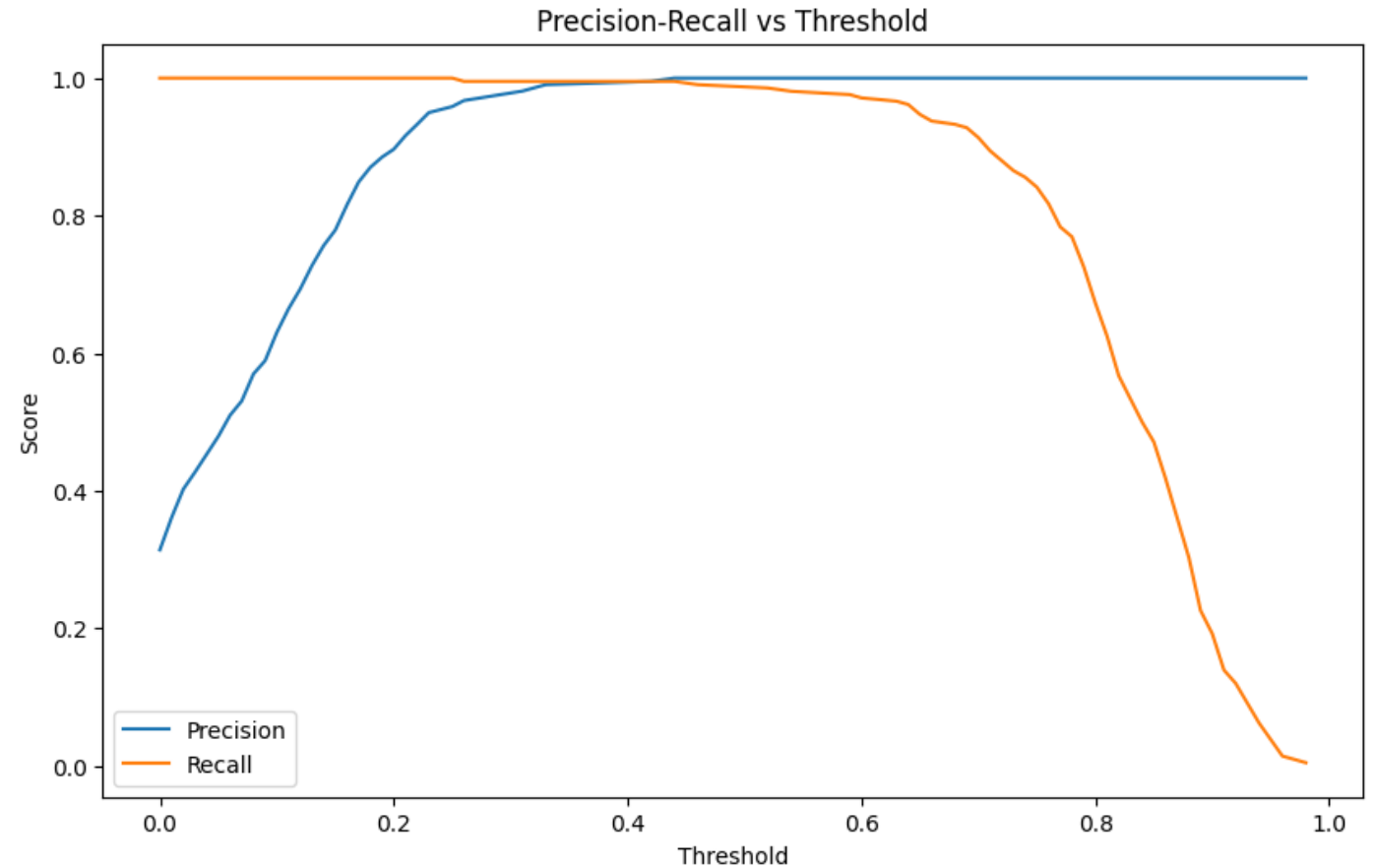
- **Matthews Correlation Coefficient (MCC):** Achieved MCC of 0.989, indicating excellent performance on imbalanced data.
- **Error Metrics Evaluation:**
 - Mean Absolute Error (MAE): Evaluates average absolute error, indicating model prediction precision.
 - Root Mean Squared Error (RMSE): Captures larger errors more effectively; ideal for financial data.
 - R-squared (R2): Explains variance in Bitcoin prices; key metric for reliability.

	Metric	Value
0	Mean Absolute Error (MAE)	0.004532
1	Mean Squared Error (MSE)	0.004532
2	Root Mean Squared Error (RMSE)	0.067318
3	R-Squared (R2)	0.978969

Matthews Correlation Coefficient: 0.9894983843770202

Roc Curve

- **Roc Curve:** Random Forest attained an AUC of 0.94, reflecting strong performance.





Conclusion

Key Insights and Recommendations:

- **Performance Highlights:** XGBoost and Random Forest outperformed with high accuracy and reliability.
- **Strategic Insights:** Feature analysis identified 'Change %' and 'Vol.' as key predictors.
- **Future Scope:** Incorporate additional macroeconomic variables for enhanced prediction accuracy.

Thank you!

