University of BRISTOL

Data Analytics in Business
EFIMM0141

Stock Price Prediction

Instructor: Dr. Aniekan Essien

Group 3

Note: Image adapted from "How Options Traders Played Apple Stock Before Big Event".

October 31st, 2024

Agenda

Introduction

Part 1

Data Preparation

Part 2

Models

Part 3

Results and Analysis

Part 4

Conclusion

Part 5

01 Introduction

Company Background, Project Overview, Market Insights

Company Background & Project Overview

University of BRISTOL

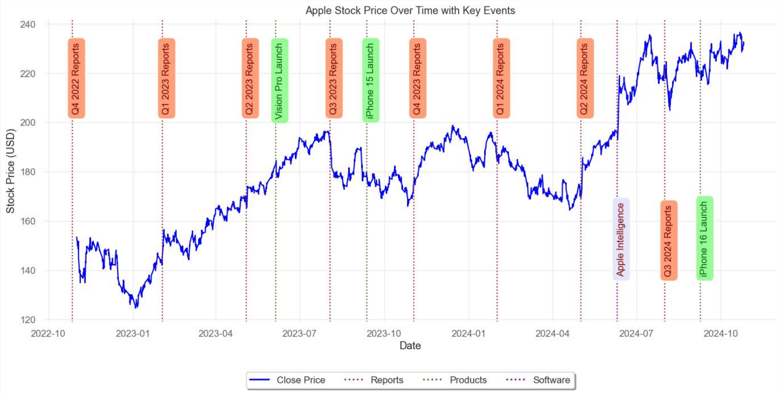
Apple Inc.

- **Overview of Apple Inc.**
  - Founded in 1976, Cupertino, California
  - Known for iPhone, iPad, Mac
- **Business Challenge**
  - Forecast stock prices to guide financial planning
- **Project Objective**
  - Predict Apple Inc. (AAPL) stock prices using historical data and technical indicators

Image Source: The Brand Hopper, 2023. Innovate, Integrate, Dominate: Success Factors of Apple Inc. Available at: https://thebrandhopper.com/2023/11/27/innovate-integrate-dominate-success-factors-of-apple-inc/ (Accessed: 29 October 2024)

Methodology, Data & Market Insights

- **Methodology**
  - Models used- LSTM, ARIMA, MoE
- **Data Source**
  - Yahoo Finance- 2 years of hourly stock data
- **Evaluation Metrics**
  - Accuracy- MAE, RMSE, MAPE
- **Market Position**
  - Stock price rose from $151 to $231
  - iPhone sales = 50% of revenue; services grew 10% annually



02 Data Preparation

Data Preparation, EDA, Data Processing

Technical Indicators

Variable Name	Indicator Name	Purpose	Calculation
RSI	Relative Strength Index	Overbought/Oversold	0-100; avg gains/losses over 14 periods
ADX	Average Directional Index	Trend Strength	0-100; higher values = stronger trend
ATR	Average True Range	Volatility	Based on the "True range" of price changes over 14 periods
BB_width	Bollinger Bands Width	Volatility/Breakouts	(Upper - Lower Band) / Middle Band
VWAP	Volume Weighted Average Price	Fair Price by Volume	Cumulative sum of (Volume * Avg Price) / Cumulative Volume
IMI	Intraday Momentum Index	Intraday Momentum	Intraday gains vs. losses over 14 periods

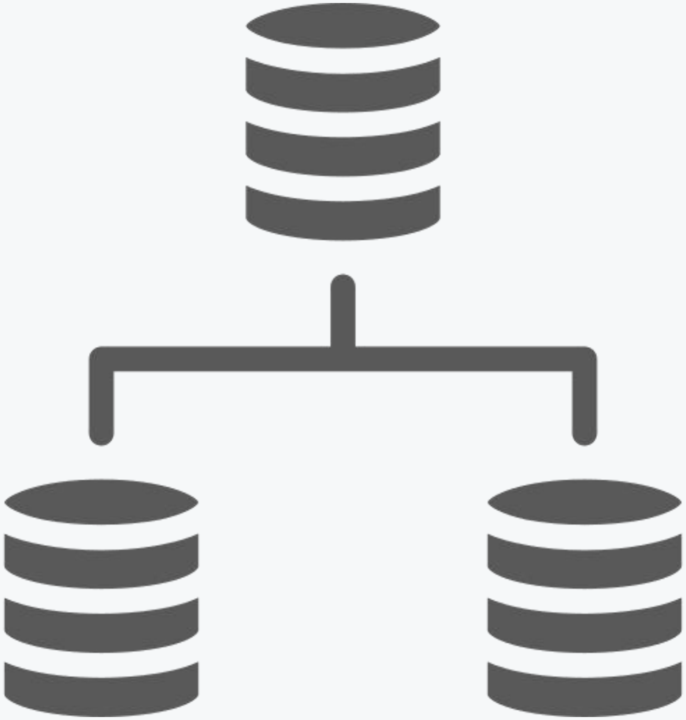
Note: True Range is the max of: 1) High - Low, 2) High - Previous Close, 3) Low - Previous Close

Environmental Variables

Variable Name	Indicator Name	Purpose
NASDAQ_close	Nasdaq-100	Technology Market Representative
XLK_close	The Technology Select Sector SPDR® Fund	Professional Investment Guidance
TSMC_close	Taiwan Semiconductor Manufacturing Company Limited	Upstream Supplier
VIX_close	Chicago Board Options Exchange's Volatility Index	Fear Gauge
IRX_close	13-week Treasury Bill Yield	Risk-free Rate

Data Preparation

Final Dataset



Training

90%

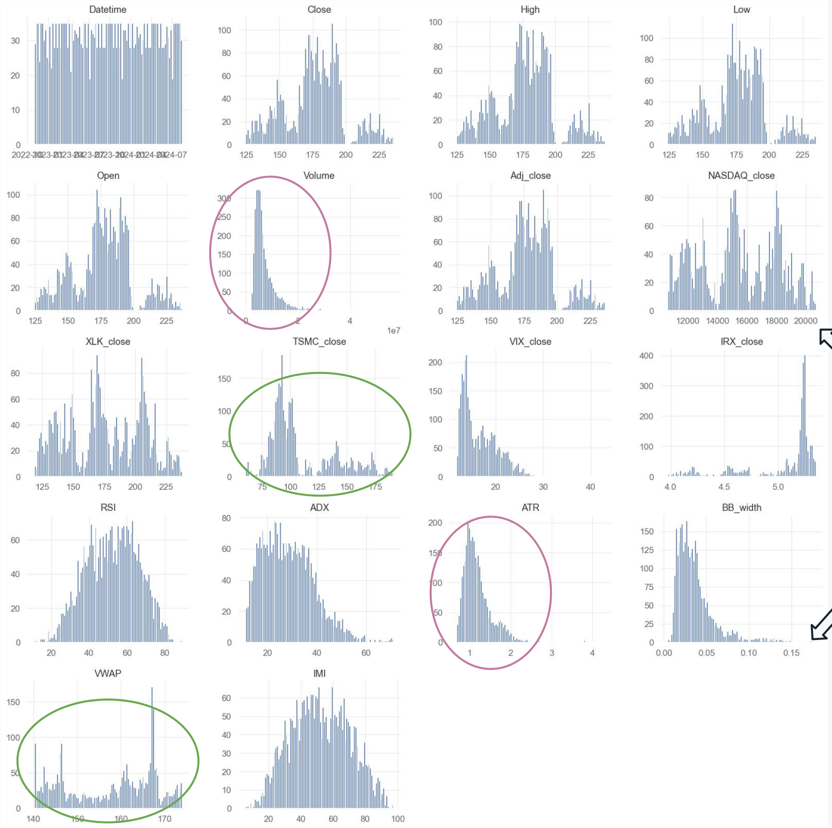
(3130 hours)

Testing

10%

(384 hours)

Exploratory Data Analysis

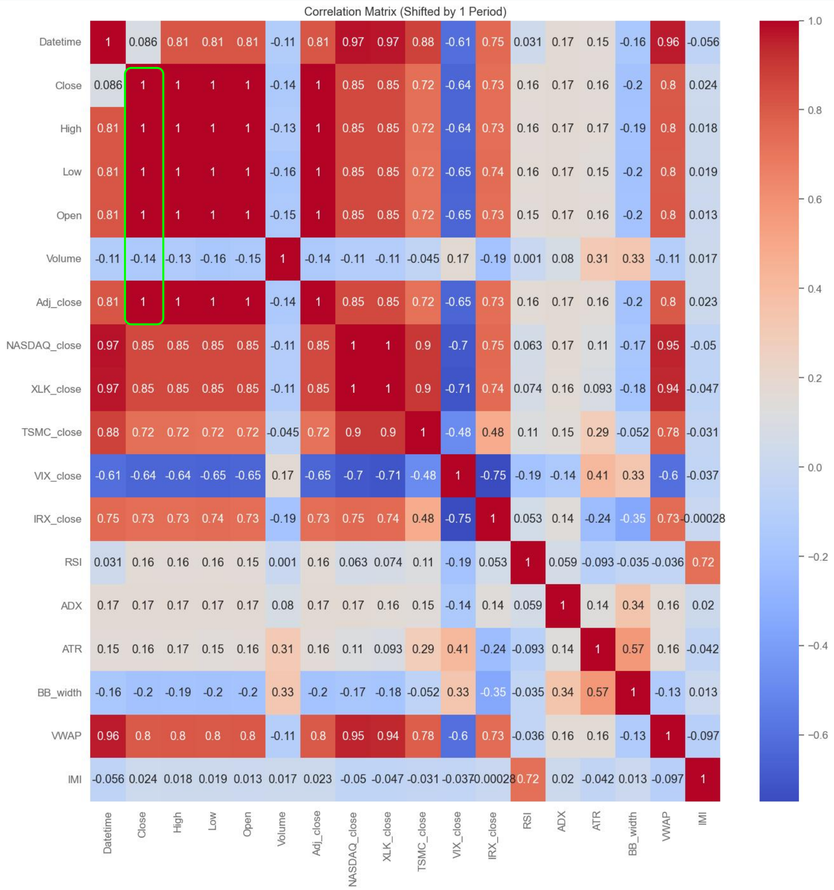


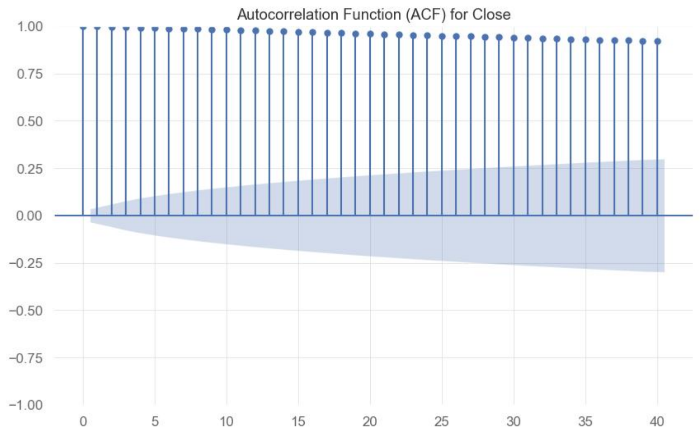
Exponential span

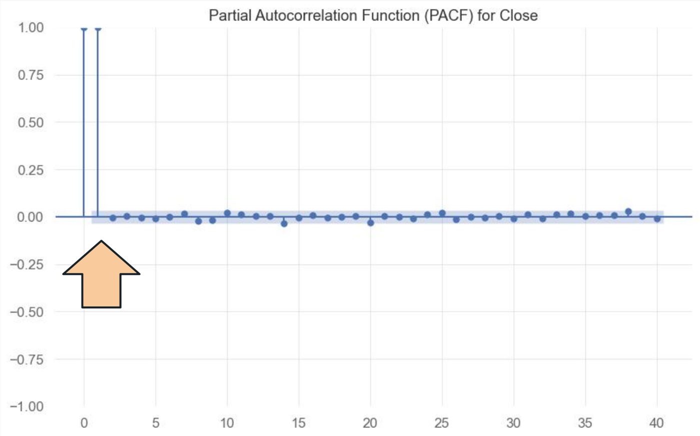
Non-normal distribution

Exploratory Data Analysis

University of BRISTOL







Data Processing

- **Removed Variables**:
  Raw values & variables with correlation > 0.95
- **Added Time-based Variables**:
  Hour_sin, Hour_cos, DayOfWeek_sin, DayOfWeek_cos
- **Final Variables:**
  'NASDAQ_close', 'XLK_close', 'TSMC_close', 'VIX_close', 'IRX_close', 'RSI', 'ADX', 'ATR', 'BB_width', 'VWAP', 'IMI', 'Hour_sin', 'Hour_cos', 'DayOfWeek_sin', 'DayOfWeek_cos'
- **Scaling:**
  RobustScaler
- **Validation:**
  Most recent 10% of training set

03 Models

ARIMA, LSTM, MoE

Model Oveview

1

ARIMA (Autoregressive Integrated Moving Average)

A time series forecasting model

2

LSTM (Long Short-Term Memory)

A type of recurrent neural network (RNN)

3

MoE (Mixture of Experts)

A machine learning model that divides a single task into subtasks using a gating mechanism

University of BRISTOL

1

Stationarity Testing

- An ADF test was conducted to check for stationarity:
  - **p-value > 0.05** indicates non-stationarity
  - **ADF statistic** higher than critical values confirms non-stationarity

Model	ADF Statistic	P-value	Critical Values		
			1%	5%	10%
Training Dataset	-0.71	0.85	-3.43		
ARIMA_1D	-20.42	0.0	-2.86		
ARIMA_2D	-10.11	1.01e-17	-2.57		

2 Initial Model Development

- The initial ARIMA model inputs:
  - Closing prices
  - First-order differencing (**d=1**) was applied for stabilization.
  - The optimal ARIMA order determined was **(0, 1, 0)** by minimizing AIC.

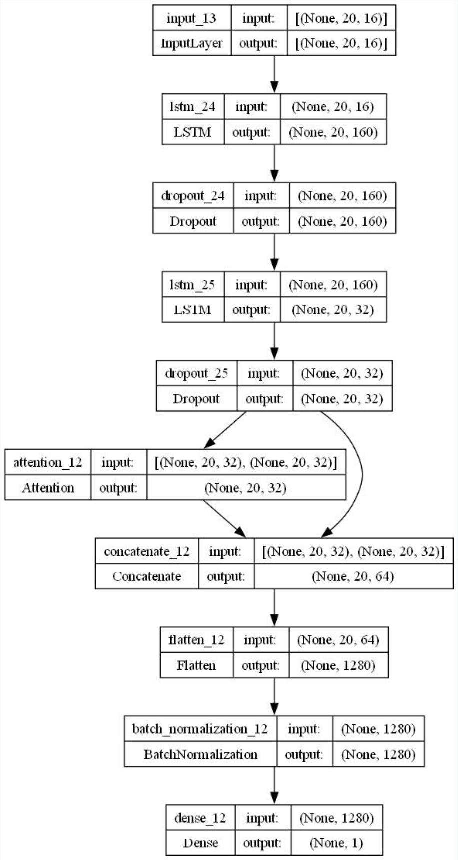
3 Model Refinement

- To enhance stability, second-order differencing (**d=2**) was applied:
  - This yielded an improved optimal order of **(5, 2, 0)**.
  - The refined model (with **d=2**) was chosen as the final outcome due to better performance.

LSTM Model

1 Model Architecture:

- **Layers:** Two recurrent neural network layers with dropout.
- **Normalization**: Batch normalization to stabilize training
- **Attention Mechanism**: Focus on key time steps
- **Optimization**: Adam optimizer with cosine annealing for smoother convergence



University of BRISTOL

2 Hyperparameter Tuning

- **Tool Used:** Keras Tuner with Bayesian Optimization
- **Parameters Tuned:** LSTM units, dropout rates, L2 regularization, learning rate

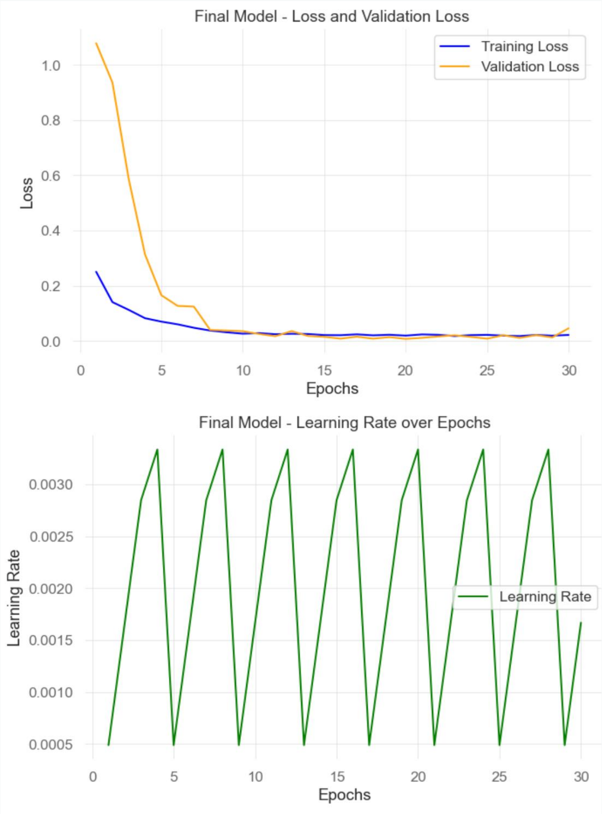
Name	Value
lstm_units	160
lstm_units_2	32
dropout_rate	0.1
lstm_look_back	20
l2_reg	0.0006
learning_rate	0.003
batch_size	16

LSTM Model

3

Model Testing and Evaluation

- **Testing Process:** Full training dataset with callbacks for early stopping
- **Evaluation:** Inverse transform predictions for accuracy against actual prices



University of BRISTOL

1

Gating Network Mechanism:

- **Additional LSTM Layer:** Captures temporal dependencies to adapt to data changes
- **Blending Weights:** Two dense layers with softmax activation, constrained between 20% and 80% for balanced influence.

2 Training Process:

- **Real-Time Adjustment:** Gating network sed actual stock prices as targets to adjust weights.
- **Momentum Smoothing:** Applied for smoother transitions in weight values across forecasts

04 Results

Model Evaluation and Financial Performance

Model Evaluation

- **Key Metrics:**

- **MSE**: Mean Squared Error
- **RMSE**: Root Mean Squared Error
- **MAPE:** Mean Absolute Percentage Error

Model	MSE	RMSE	MAPE(%)	Training Time(s)
ARIMA	30.87	5.56	1.98	0.19
LSTM	4.06	2.01	0.67	48.54
MoE	7.89	2.81	0.99	11.17

- **ARIMA**

  Limited gains in accuracy, despite increased complexity.

- **LSTM**

  A trade-off between accuracy and computational demand

- **MoE**

  Viable alternative for scenarios where both accuracy and efficiency are required.

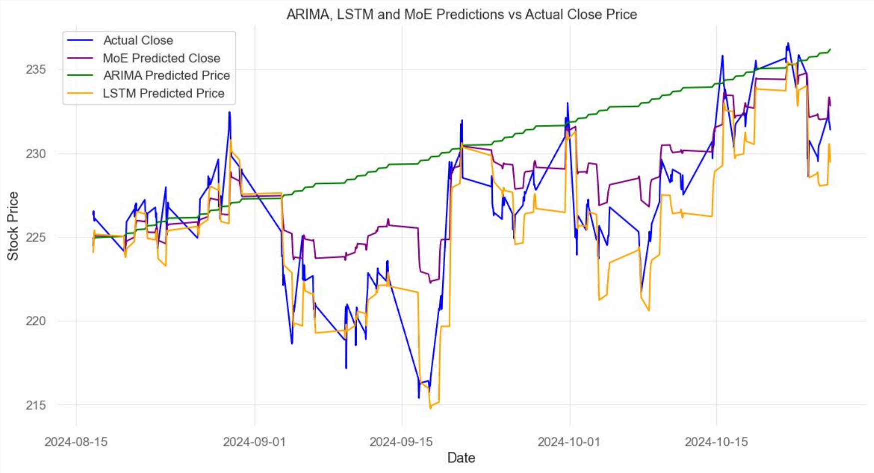
Model Evaluation

➢ **Summary**:

LSTM provides the most accurate predictions

MoE suitable for robust prediction with moderate computational demands

ARIMA optimal for rapid, broad trend analyses but may lack the precision necessary

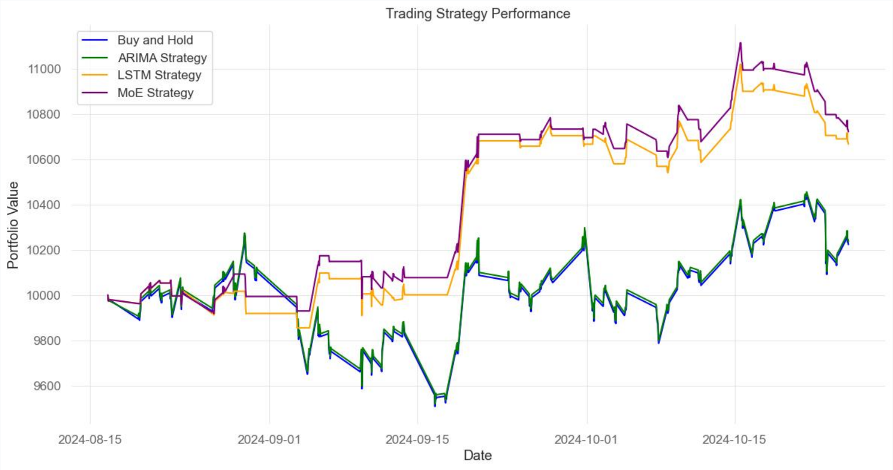


Financial Performance

➢ **Trading Strategy:**

- **Executing a buy order**
  If the model predicts a price increase the following day.

- **Executing a sell order**
  If a price decrease is forecasted.



Financial Performance

University of BRISTOL

On Sept. 15—Apple‘s product launch event: as a critical time marker for examining model performance.

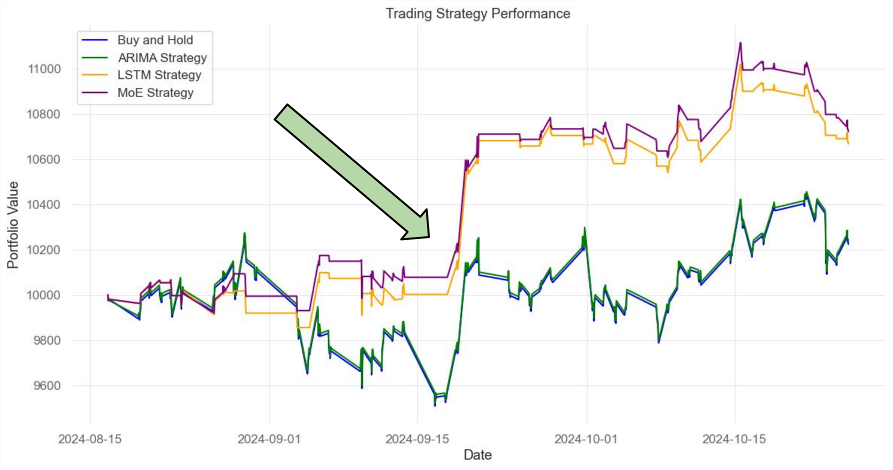
- Prior-Event Performance Result:
- ARIMA(green line): Struggled to adapt to short-term changes, missed the downward trend.
- LSTM & MoE(orange and purple) : Accurately identified the downward trend, effectively reduced losses.

University of BRISTOL

➢ Performance Result:

- Post-Event Performance: All models showed portfolio value growth, capturing positive price movements.

- Conclusion: LSTM and MoE outperformed ARIMA in real-time responsiveness and around key events, supporting more effective trading and portfolio growth.



University of BRISTOL

➢ Backtesting and Metrics:

- **MoE model**
  Outperforms in cumulative returns, highest Sharpe Ratio.
  synthesize various forecasting advantages

- **LSTM model**
  Significant strengths, substantial returns, comparatively high Sharpe Ratio, minimal drawdown
  leverage sequential dependencies

- **ARIMA model**
  Lower overall returns,more cautious performance.

Strategy	Win/Loss Ratio	Cumulative Return(%)	Annualized Volatility(%)	Max Drawdown(%)	Sharpe Ratio
Benchmark	0.13	2.24	9.07	-7.33	-0.28
ARIMA	0.14	2.37	9.06	-7.33	-0.27
LSTM	0.59	6.67	5.84	-3.21	0.05
MoE	0.67	7.22	5.86	-3.54	0.11

Overall: machine learning models demonstrate a higher sensitivity to short-term fluctuations and a capacity for rapid adaptation.

05 Conclusion

Investment Recommendations,

Investment Recommendations

MoE Strategy

Effective for active investors seeking to maximize returns, manage risk through data-driven trading decisions.

LSTM Model

Balancing substantial returns, lower volatility suitable for those desiring both growth and strong risk control.

ARIMA Model

Identifying broader market trends long-term price movements, economic cycles, shifts influenced by macroeconomic indicators.

.

Above all,

- Buy and Hold strategy may be appropriate for conservative investors, though forfeits potential gains observed in model-driven approaches.

- Diversified investment strategy combining MoE and LSTM models, can optimize returns while addressing market risks.

Thanks for Watching!

University of BRISTOL