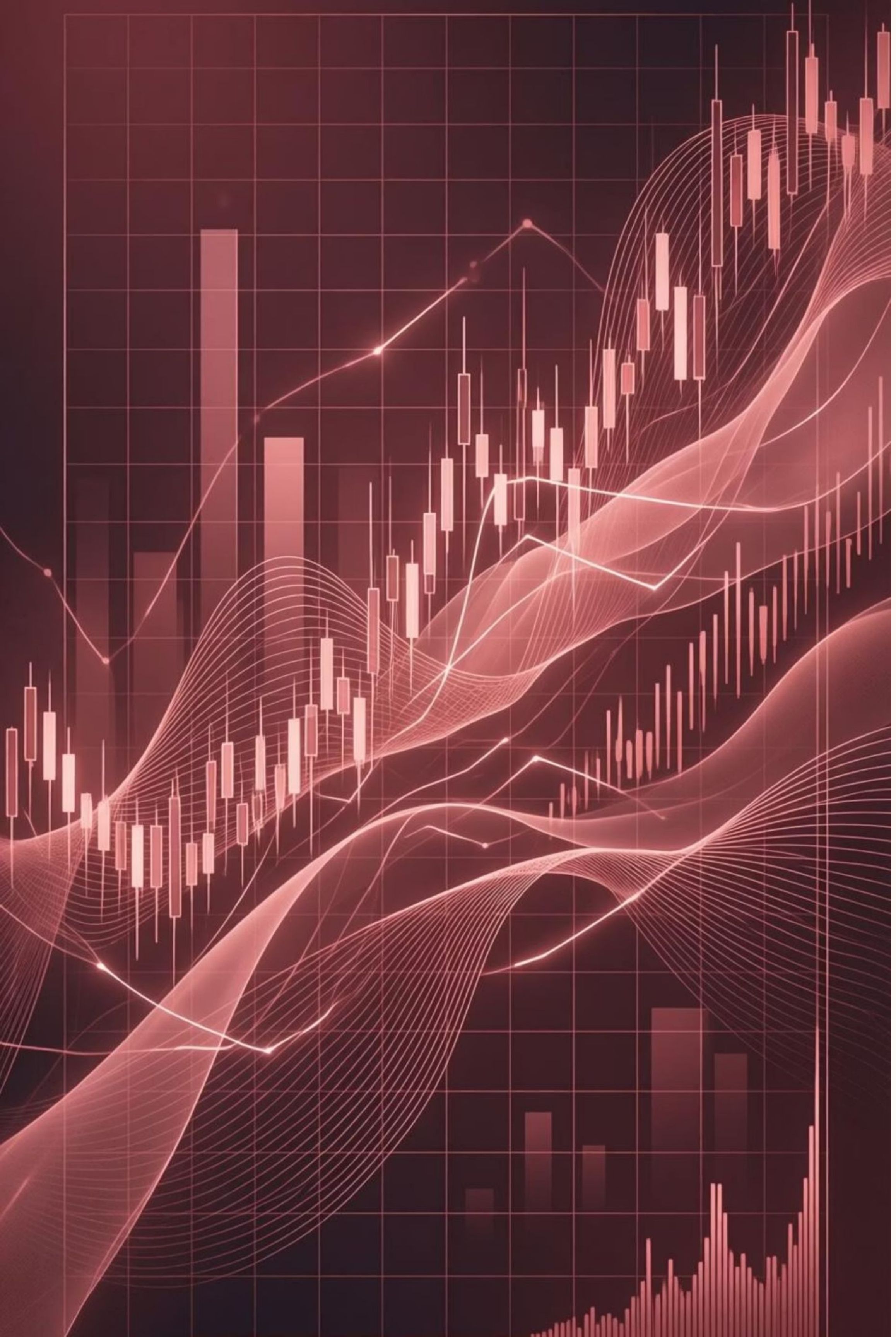


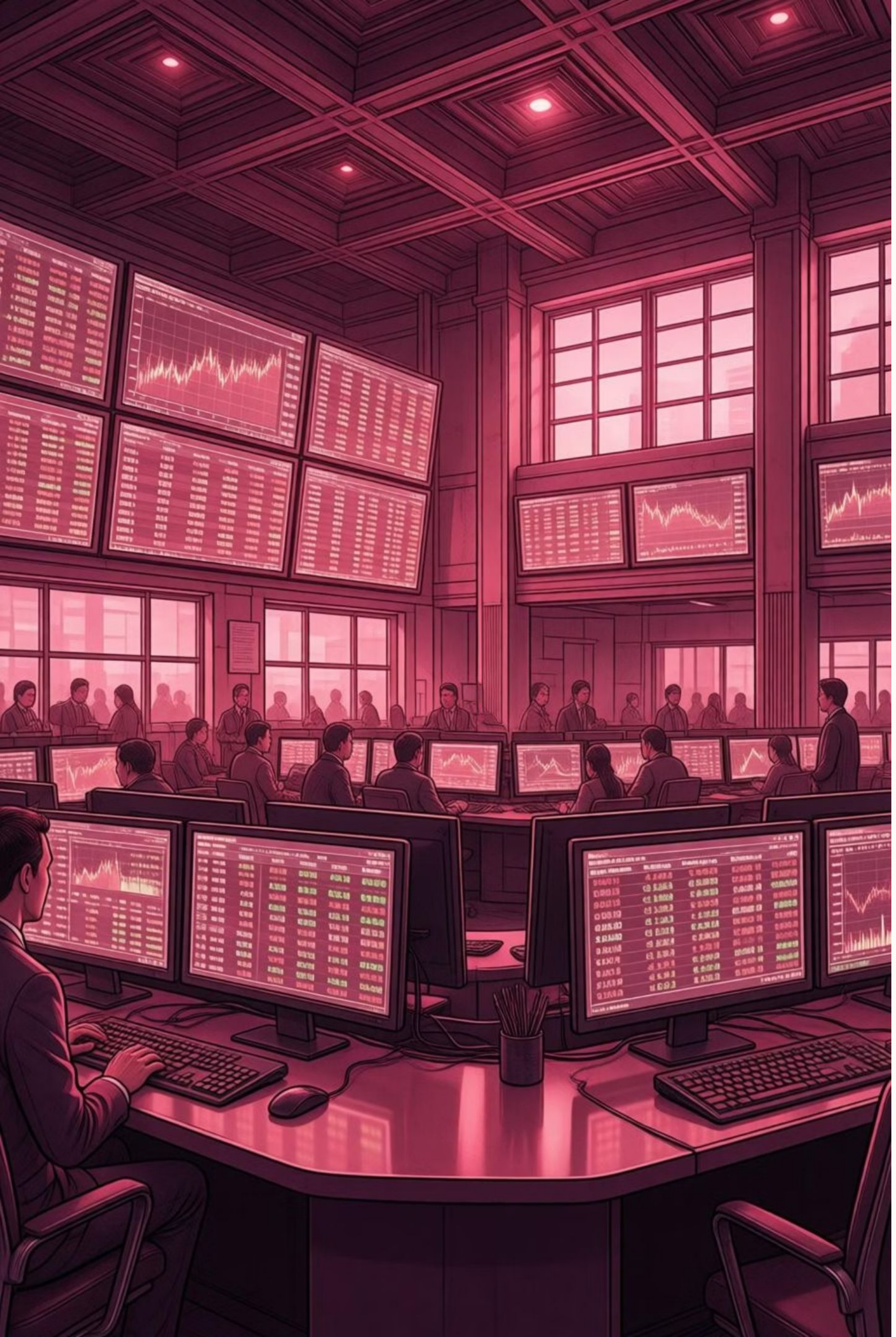
AI for Market Trend Analysis

Minor in Artificial Intelligence (Module E)

Student: Ananya Chandra | Mentor: Dr. Niranjan Deshpande

Indian Institute of Technology Ropar | 2026

Made with GAMMA



PROBLEM STATEMENT

The Challenge of Financial Market Analysis

The Data Volume Problem

Financial markets continuously generate massive volumes of time-series data across thousands of securities. Traditional manual analysis methods struggle to process this information efficiently, leading to missed opportunities and delayed insights.

Research Objective

This project aims to predict next-day stock closing prices using historical market data through machine learning techniques. The academic focus is on demonstrating a complete AI pipeline from data processing to model evaluation.

Made with GAMMA

Dataset Composition

Data Scope

Synthetic S&P 500-style dataset with approximately 8,000 records spanning 2020–2023

Coverage

Eight representative stocks with complete OHLCV features (Open, High, Low, Close, Volume)

Reproducibility

Synthetic data enables consistent offline execution and eliminates market dependencies

The dataset was specifically designed to mirror realistic market patterns while maintaining academic integrity and reproducibility standards for educational purposes.

Made with GAMMA

End-to-End System Architecture

Data Generation & Loading

Synthetic market data creation and import into analysis environment

Exploratory Data Analysis

Statistical analysis and visualization of market patterns and distributions

Feature Engineering

Transformation of raw data into predictive technical indicators

Train/Test Split

Chronological separation maintaining temporal integrity

Model Training & Evaluation

Supervised learning regression models with performance metrics

Made with GAMMA

Feature Engineering Strategy

Trend Indicators

Simple Moving Averages (SMA) and Exponential Moving Averages (EMA) to identify price momentum and directional trends over multiple timeframes

Temporal Features

Lag features capturing previous day prices and returns, enabling the model to recognize sequential patterns and autocorrelation in price movements

Volatility Metrics

Daily return calculations and volatility measures quantifying price fluctuation intensity and market uncertainty

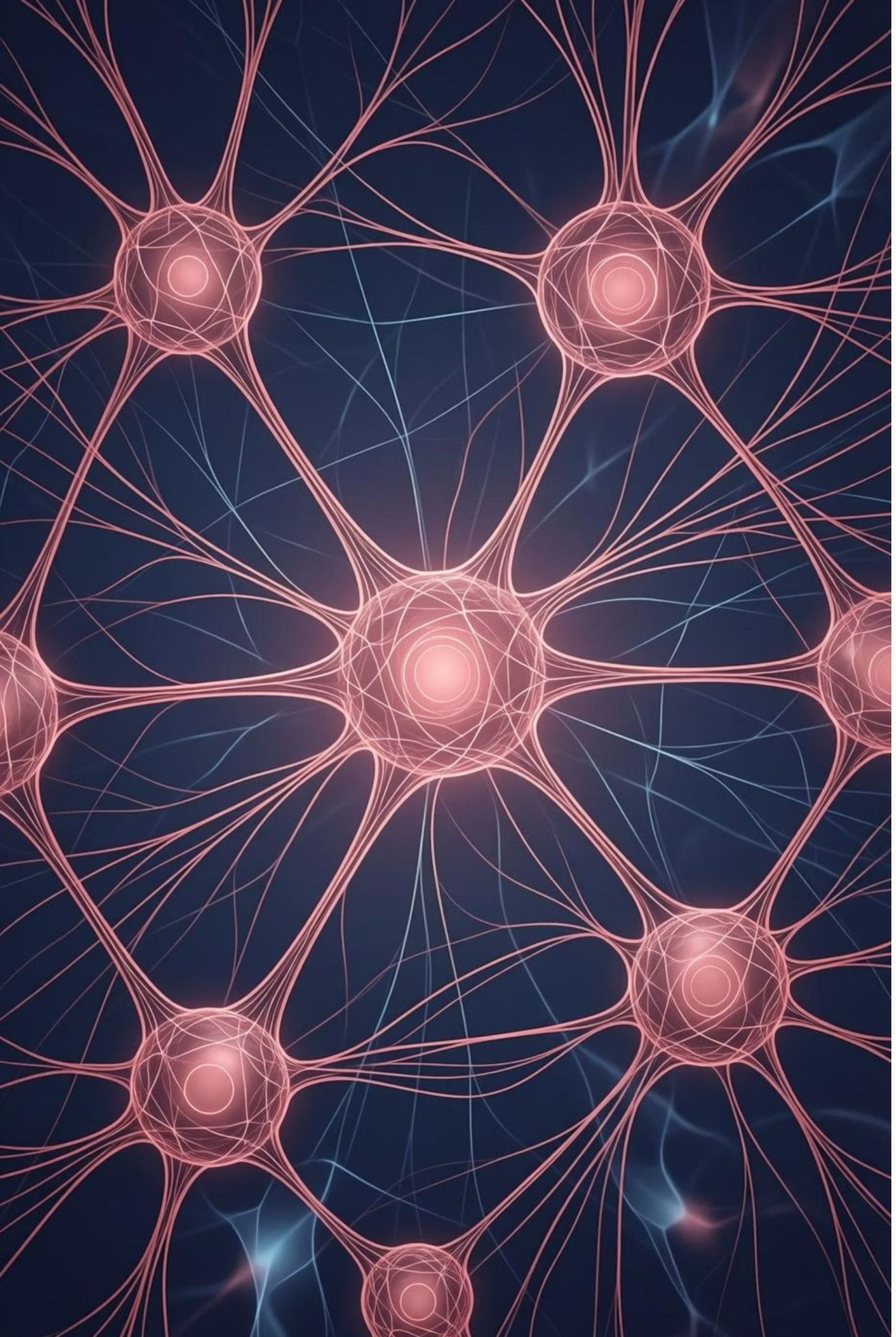
Technical Indicators

RSI (Relative Strength Index) and MACD (Moving Average Convergence Divergence) for momentum analysis and trend confirmation

Volume Analysis

Volume-based ratios revealing trading activity patterns and market participation strength

Made with GAMMA



Machine Learning Model Selection

Linear Regression

Purpose: Baseline model

Strengths: Interpretable coefficients, fast training, captures linear relationships

Use case: Understanding feature importance and establishing performance benchmarks

Random Forest

Purpose: Ensemble learning

Strengths: Captures non-linear patterns, resistant to overfitting, provides feature rankings

Use case: Modeling complex interactions between technical indicators

Gradient Boosting

Purpose: Advanced ensemble

Strengths: Sequential error correction, high predictive accuracy, handles feature interactions

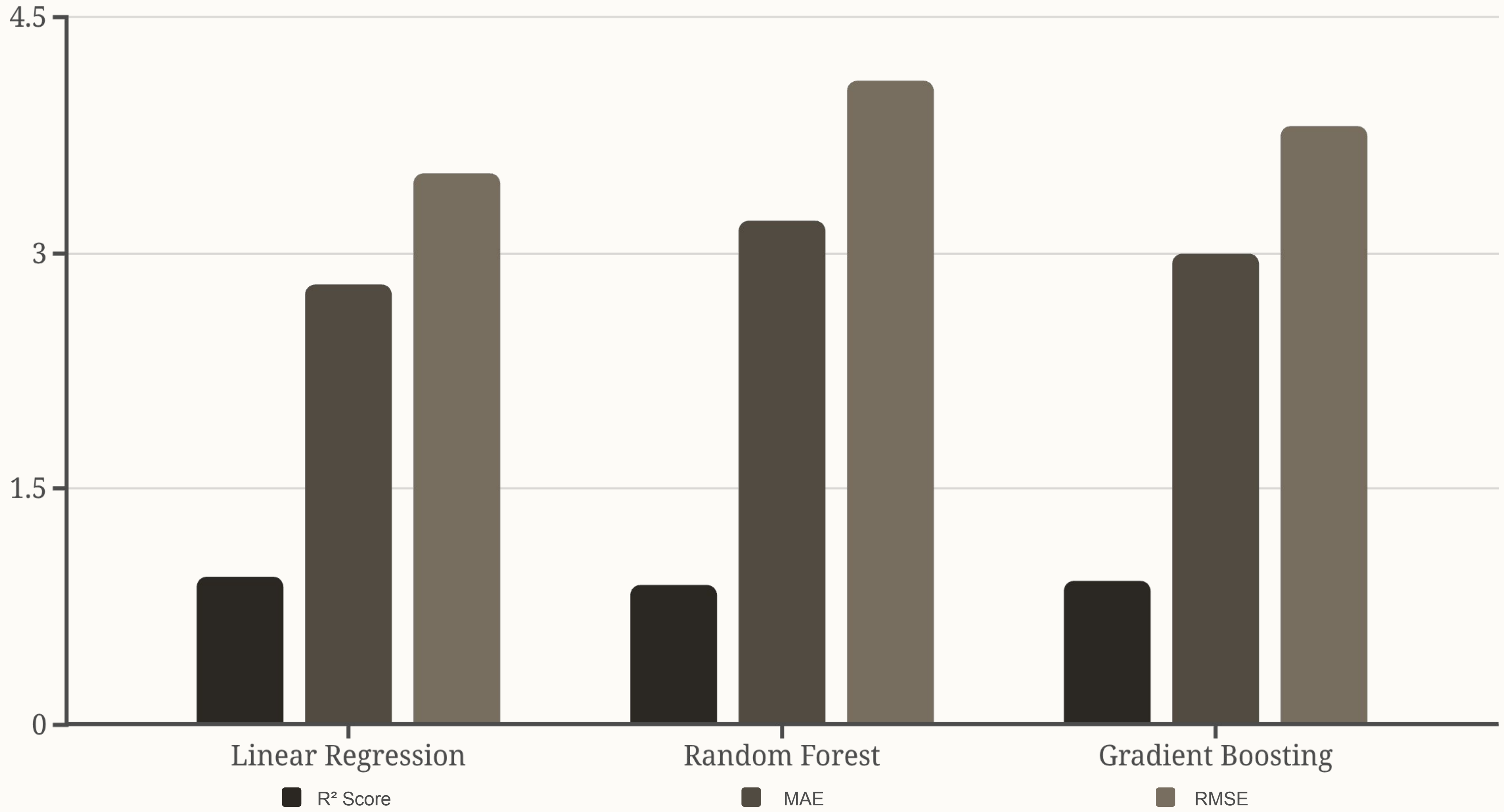
Use case: Maximizing prediction performance through iterative refinement

Critical Design Choice: Time-based train/test splitting was implemented to prevent data leakage, ensuring that models learn only from past data when predicting future prices—mirroring real-world trading constraints.

Made with GAMMA

Model Performance Results (Linear Regression – AAPL)

0.99	2.8	3.5	1.2%
R^2 Score	MAE ($)	RMSE ($)	MAPE
Linear Regression explained 99% of price variance	Mean Absolute Error indicating average prediction accuracy	Root Mean Squared Error penalizing larger deviations	Mean Absolute Percentage Error showing relative accuracy

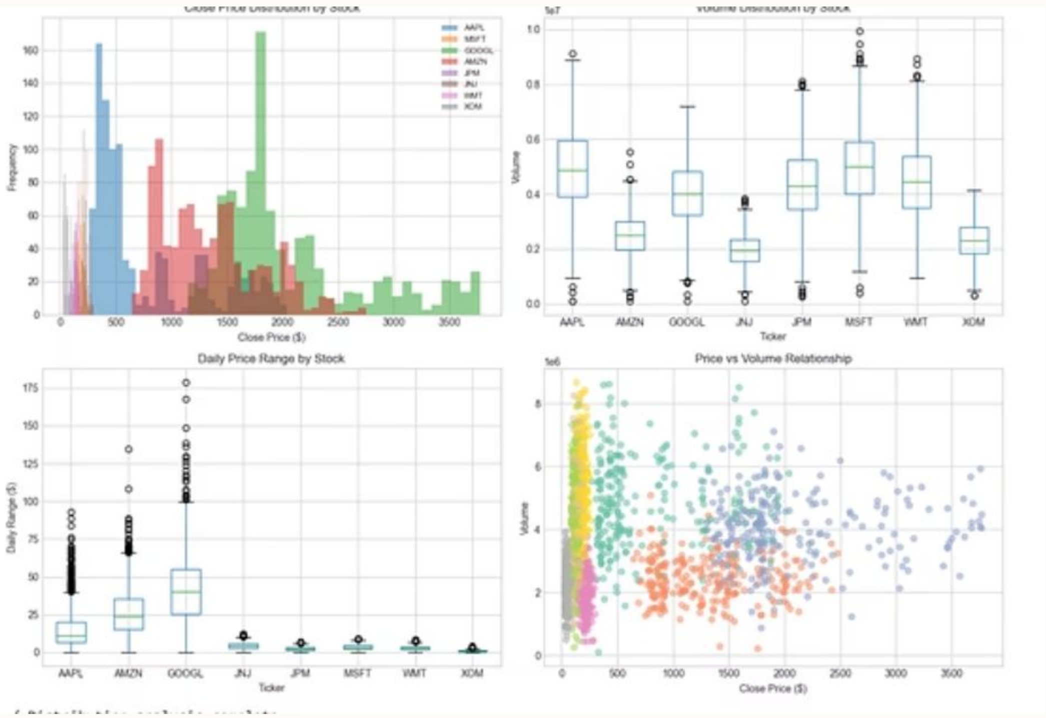


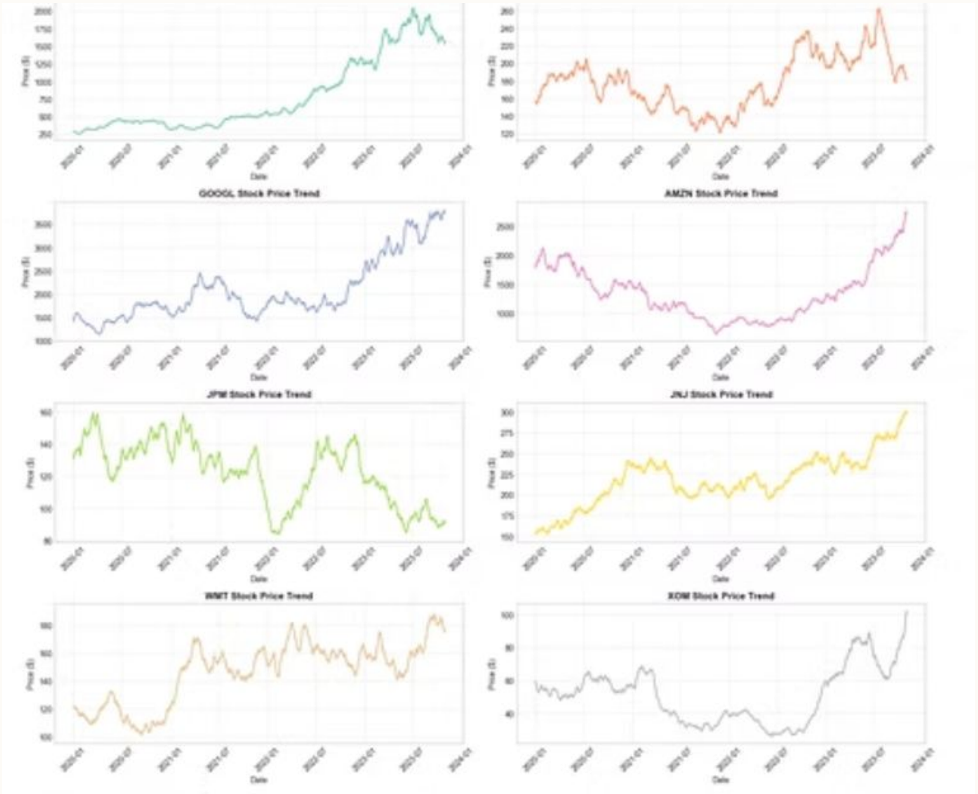
Linear Regression achieved superior performance due to strong autocorrelation in time-series data. Ensemble models provided valuable feature importance rankings, revealing that lagged prices and moving averages were the strongest predictors.

Made with GAMMA

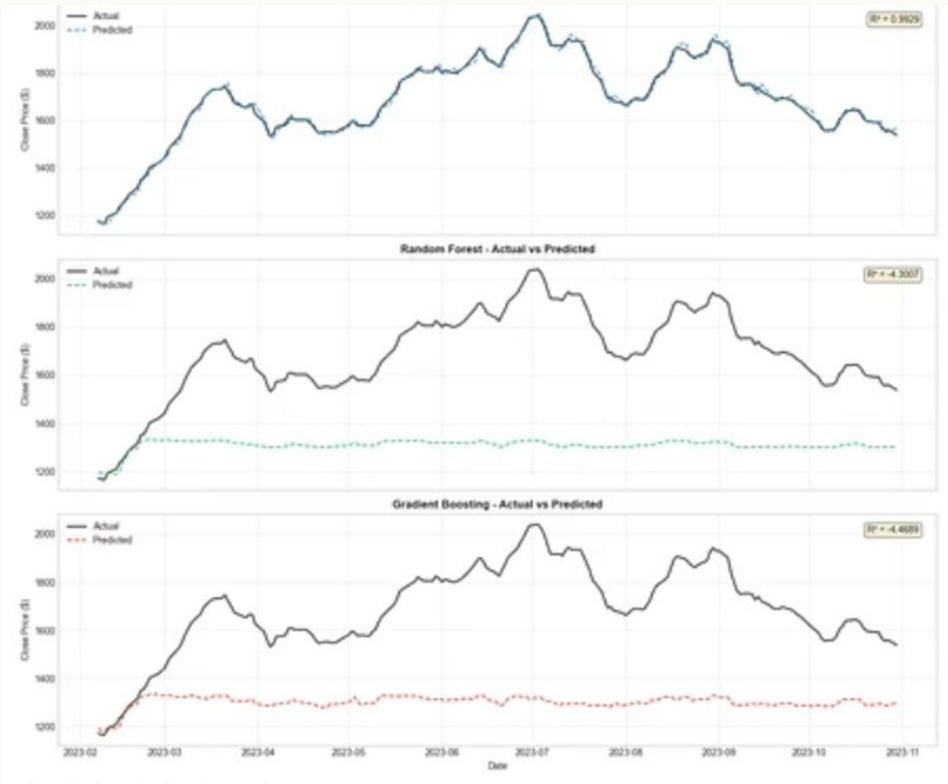
Implementation Demonstration

The demonstration captures key stages of the project workflow: initial data processing in Jupyter Notebook, engineered feature outputs displaying calculated technical indicators, model prediction accuracy visualized against actual prices, and final execution confirmation showing successful completion of the AI pipeline.






```
PROJECT SUMMARY: AI FOR MARKET TREND ANALYSIS
==========================================================================

📊 DATASET
  • Synthetic S&P 500-style data: 8,000 records
  • Stocks analyzed: 8 companies
  • Date range: 2020-01-01 to 2023-10-31

🔧 FEATURE ENGINEERING
  • 25 features created
  • Includes: Moving averages, volatility, RSI, MACD, lag features

🤖 MODELS TRAINED
  • Linear Regression: R² = 0.9929, MAPE = 0.72%
  • Random Forest: R² = -4.3007, MAPE = 21.74%
  • Gradient Boosting: R² = -4.4689, MAPE = 22.12%

🏆 BEST MODEL: Linear Regression
  • R² Score: 0.9929
  • Mean Absolute Error: $12.17
  • MAPE: 0.72%

==========================================================================
```



Key Insights and Learnings

1. **Feature Engineering Dominance**

   The quality and relevance of engineered features—particularly moving averages and lag features—had a more significant impact on model performance than algorithm complexity.

2. **Temporal Data Integrity**

   Time-series data requires strict chronological splitting to avoid look-ahead bias. Random shuffling would artificially inflate performance metrics and produce unrealistic results.

3. **Simplicity vs. Complexity**

   Linear models can outperform sophisticated ensemble methods when data exhibits strong linear patterns, demonstrating that model selection should be data-driven rather than complexity-driven.

4. **Visual Diagnostics**

   Numeric metrics alone are insufficient. Residual plots and prediction vs. actual visualizations reveal model behavior patterns not captured by summary statistics.

5. **Interpretability Matters**

   In financial AI applications, stakeholders require understanding of model decisions. Transparent feature relationships are essential for building trust and regulatory compliance.

Made with GAMMA

ETHICS & CONCLUSION

Responsible AI and Project Summary

Ethical Considerations

No Trading Intent: This project is purely academic and not designed for real-world trading decisions

Synthetic Limitations: Artificial data lacks market complexity, volatility shocks, and external factors

Generalization Risks: Historical patterns may not predict future performance; markets are non-stationary

LLM Usage Disclosure

Large Language Models were **not used** in model training, feature engineering, or prediction tasks. LLMs served exclusively as learning aids and coding assistants during development—maintaining the integrity of the machine learning methodology.

Project Outcome

Successfully demonstrated an end-to-end AI pipeline for market trend analysis, from raw data to actionable predictions. The project achieved strong predictive accuracy while reinforcing fundamental principles of time-series modeling, feature engineering, and responsible AI development.

Made with GAMMA