SENTIMENT-DRIVEN STOCK TRADING USING NLP AND OPTIMAL POLICY TREES

15.095 – Machine Learning under a Modern Optimization Lens

MIT SLOAN SCHOOL OF MANAGEMENT

MIT MANAGEMENT SLOAN SCHOOL

MEET THE TEAM

Fiona Aga

Pranav Girish

MIT MANAGEMENT SLOAN SCHOOL

AGENDA

- MOTIVATION & METHODOLOGY
- DATA COLLECTION
- SENTIMENT ANALYSIS
- PRESCRIPTIVE MODELING
- RESULTS & DISCUSSION

MIT MANAGEMENT SLOAN SCHOOL

AGENDA

- MOTIVATION & METHODOLOGY
- DATA COLLECTION
- SENTIMENT ANALYSIS
- PRESCRIPTIVE MODELING
- RESULTS & DISCUSSION

MIT MANAGEMENT SLOAN SCHOOL

MOTIVATION

Effectively quantify market sentiment obtained from financial news articles

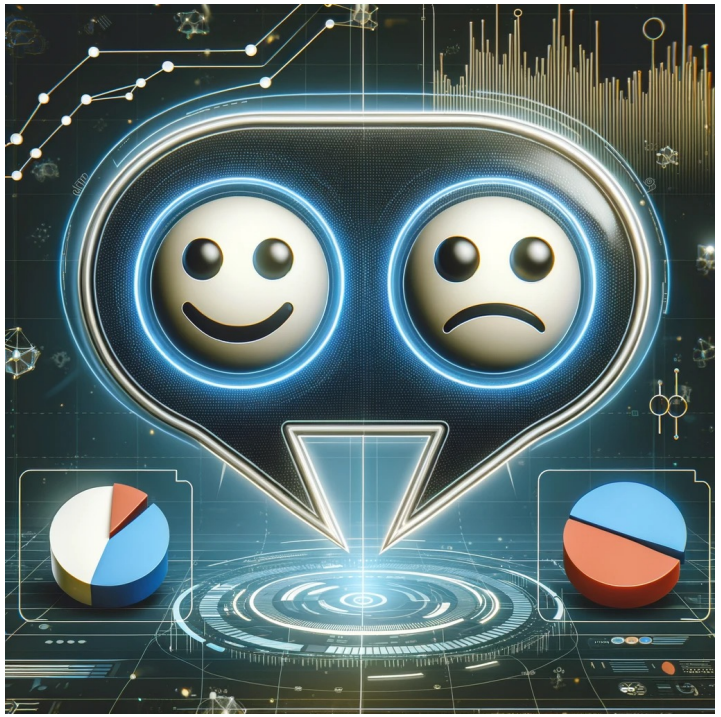
Build an intelligent stock trading agent leveraging sentiments and economic trends

- Trading decisions are influenced by a various factors
- There exists a vast array of existing research of utilizing analytics in finance
- A bulk of this research only considers economic trends

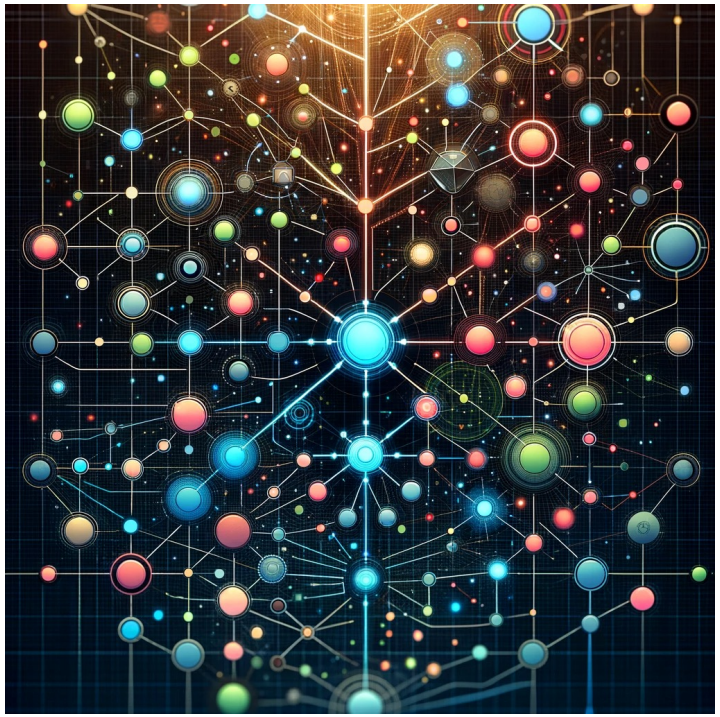
METHODOLOGY

DATA & WEBSCRAPING

SENTIMENT ANALYSIS



OPTIMAL POLICY TREES



AGENDA

- MOTIVATION & METHODOLOGY
- DATA COLLECTION
- SENTIMENT ANALYSIS
- PRESCRIPTIVE MODELING
- RESULTS & DISCUSSION

MIT MANAGEMENT SLOAN SCHOOL

DATA COLLECTION & WEBSCRAPING

1. Initial Data Collection: AlphaVantage

- Encountered API limitations with AlphaVantage
- Restricted data collection timeframe

2. Transition to Yahoo Finance

- Shifted to a more comprehensive data source
- Gained access to extensive stock coverage and financial articles

3. Webscraping Process

- Employed advanced scraping techniques using Beautiful Soup Python library
- Integrated article collection & sentiment generation through built-in textual analysis libraries (TextBlob)

4. Data Integrity and Reliability

- Yahoo Finance ensured accuracy and relevance
- Data collected over 6 months and 20 stock symbols

AGENDA

- MOTIVATION & METHODOLOGY
- DATA COLLECTION
- **SENTIMENT ANALYSIS**
- PRESCRIPTIVE MODELING
- RESULTS & DISCUSSION

MIT MANAGEMENT SLOAN SCHOOL

SENTIMENT ANALYSIS

Model Selected	Evaluation	Inference Time	Training Time	Computational Cost
Logistic Regression	F1 Score: 0.73	Fast	Short	Low
Fine-tuned Logistic Regression	F1 Score: 0.77	Fast	Short	Low
Random Forest	F1 Score: 0.71	Moderate	Longer	Medium
Fine-tuned Random Forest	F1 Score: 0.77	Moderate	Longer	Medium
Neural Network	Test Accuracy: 0.813	Variable	Long	High
Zero-shot LLM	Val Accuracy: 0.85	Slow	None (pre-trained)	High

SENTIMENT MODEL SELECTED

Model Selected	Evaluation	Inference Time	Training Time	Computational Cost
Logistic Regression	F1 Score: 0.73	Fast	Short	Low
Fine-tuned Logistic Regression	F1 Score: 0.77	Fast	Short	Low
Random Forest	F1 Score: 0.71	Moderate	Longer	Medium
Fine-tuned Random Forest	F1 Score: 0.77	Moderate	Longer	Medium
Neural Network	Test Accuracy: 0.813	Variable	Long	High
Zero-shot LLM	Val Accuracy: 0.85	Slow	None (pre-trained)	High

AGENDA

- MOTIVATION & METHODOLOGY
- DATA COLLECTION
- SENTIMENT ANALYSIS
- PRESCRIPTIVE MODELING
- RESULTS & DISCUSSION

MIT MANAGEMENT SLOAN SCHOOL

FEATURE ENGINEERING

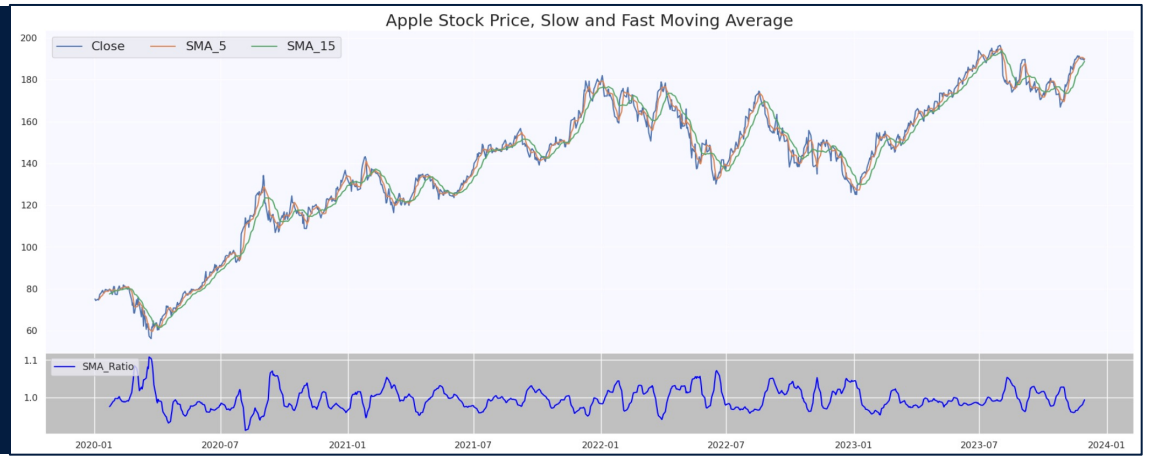
Lagged Features

- Previous Day Values (Open, Close, High, Low Prices)
- Rolling Window Benchmarks (5 and 15 days)
- Range of Prices encountered

Stock Price Simple Moving Average

- 5-day window
- 15-day window
- SMA Ratio

Using Close Price of Stock



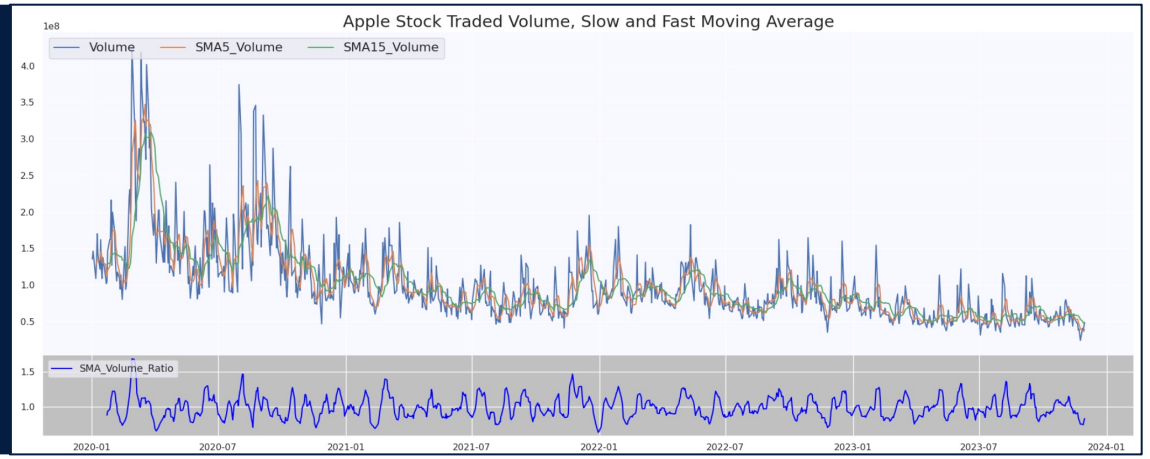
Sentiment Scores

- Aggregated sentiment scores for stocks for each date
- Positive, negative, and neutral scores combined through weighted average

Traded Volume Simple Moving Average

- 5-day window
- 15-day window
- SMA Volume Ratio

Using Close Price of Stock



MIT MANAGEMENT SLOAN SCHOOL

PRESCRIPTIVE MODELING

COMMON RULE-BASED STRATEGIES

- Investing based on short-term stock upticks
- Investing based on simple moving average ratio

Naïve (short-term) rule-based strategies result in adverse results, reporting low overall profits, signaling the need for a more holistic approach

OPTIMAL POLICY TREE

- Prescribes trading decisions (buy or not) for a set of stocks over a given period, based off of expected rewards
- Outperforms rule-based strategies, reporting higher overall profits in the testing period
- Effectively utilizes sentiment scores alongside economic trends
- Achieved 80% accuracy in reward estimation and policy determination, resulting in over 2% return on investment in just a week

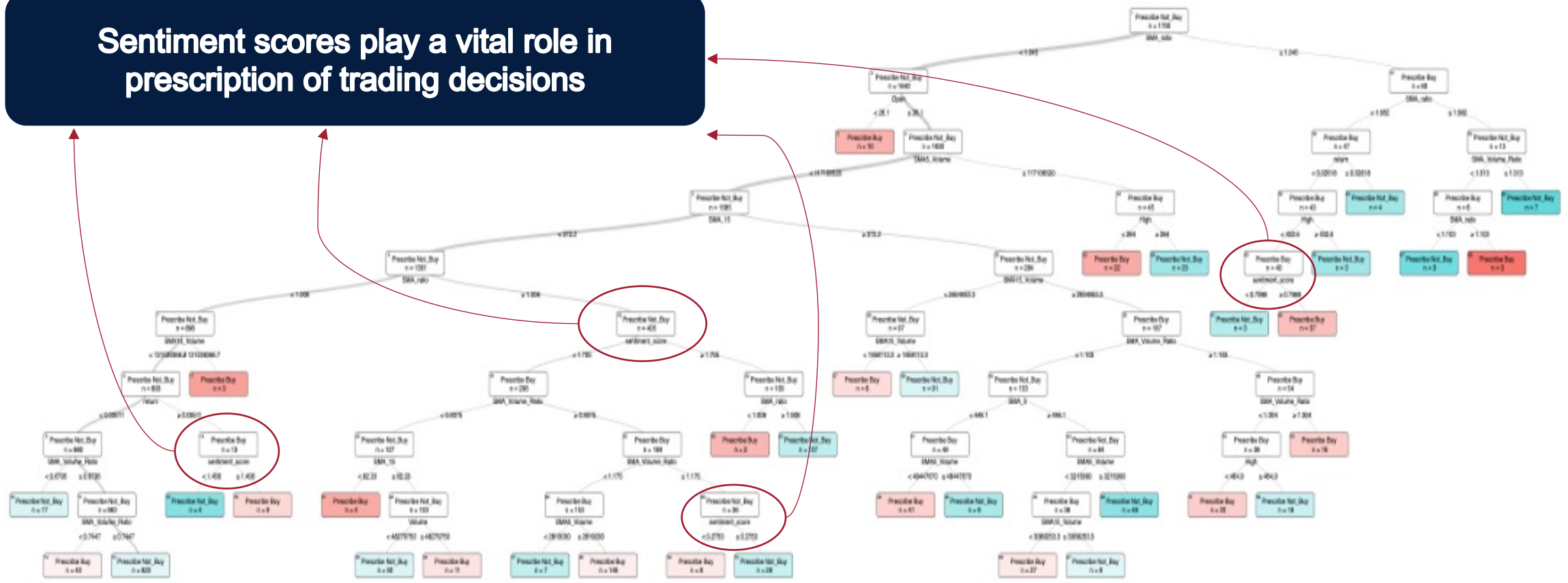
AGENDA

- MOTIVATION & METHODOLOGY
- DATA COLLECTION
- SENTIMENT ANALYSIS
- PRESCRIPTIVE MODELING
- **RESULTS & DISCUSSION**

MIT MANAGEMENT SLOAN SCHOOL

OPTIMAL POLICY TREE (WITH SENTIMENTS)

Sentiment scores play a vital role in prescription of trading decisions



MIT MANAGEMENT SLOAN SCHOOL

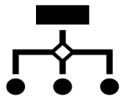
INSIGHTS AND NEXT STEPS

Insights Gained

- Validated effectiveness of sentiment analysis in stock selection and trading
- Demonstrated power of optimal models over conventional strategies

Enhancing Model Precision

- Implement clustering algorithms to categorize similar stocks
- Adapt the OPT model for tailored cluster-wise suggestions



Enhancing the Framework

- Explore additional data sources for a broader market perspective
- Test the model against different market states for better robustness

Long-term Goals

- Add portfolio optimization capabilities to the system
- Improve return estimation using financial features
- Pipeline for sequential model to enable real-time decisions

MIT MANAGEMENT SLOAN SCHOOL

THANK YOU

MIT SLOAN SCHOOL OF MANAGEMENT

MIT MANAGEMENT SLOAN SCHOOL