



U.S. STOCK MARKET ANALYSIS

NOT TO BE TAKEN AS FINANCIAL ADVICE

**Amy Yucus, Annie Kittendorf,
Dalton Bode, Theodore Korby**

OVERVIEW

Introduction

Exploratory Questions

Research Process

Machine Learning

Results and Report Demo

Recommendations

Q&A

The background is a complex, abstract composition. It features deep blue and vibrant orange colors. The blue areas are filled with fine, wavy, brushstroke-like lines that create a sense of movement. The orange areas are more solid but also contain some of the wavy patterns. Interspersed throughout are intricate, fractal-like structures that resemble branching, organic forms or perhaps stylized flames. These structures are primarily orange with some blue highlights, adding to the overall complexity and visual interest of the background.

INTRODUCTION

MEET THE TEAM



Theodore Korby
Twin Cities Cohort
Hamline University
BS Applied Physics



Dalton Bode
Milwaukee Cohort
Arizona State University
MS Nanophysics



Annie Kittendorf
Milwaukee Cohort
University of Michigan
BS Biopsychology,
Cognition and
Neuroscience &
Evolutionary Anthropology



Amy Yucus
Milwaukee Cohort
Texas Woman's University
MLS Library Science



PROJECT FOCUS

- EXAMINE THE US DOMESTIC STOCK MARKET AND ITS TOP 5 SECTORS AS REPRESENTED BY VANGUARD INDEX ETFS
 - PREDICTIVE ANALYSIS
 - TIME SERIES ML MODELS
 - DATA ANALYSIS – FEATURE TRENDS

EXPLORATORY QUESTIONS

- How do historical market prices compare to historical national revenue?
- Which economic features are the most correlated to index prices?
- How does investment risk vary across sectors?
- How does return on investment vary across sectors?
- How do different machine learning algorithms compare when modeling the domestic stock market?

VANGUARD TOTAL STOCK MARKET INDEX FUND(VTI)

- **Why VTI?**
 - Vast majority of portfolio is US-based stocks (>99%)²
 - Contains stocks in numerous industries
- **What is VTI?**
 - Index Fund started in 2001 by Vanguard Equity Index Group
 - Currently encompasses over 4,100 stocks in portfolio
 - Total Stock Market ETF for stocks of all sizes on NYSE and Nasdaq¹



The background of the slide features a blue-tinted image of a financial data screen. It displays a line graph with a white line showing fluctuations, overlaid on a grid of numerical data. The numbers are arranged in rows and columns, with some values like '18775.8', '18771.56', and '18765.39' visible. The overall aesthetic is professional and data-oriented.

SECTOR FUNDS

- **VANGUARD INFORMATION TECHNOLOGY INDEX FUND (VGT)**
- **VANGUARD CONSUMER DISCRETIONARY INDEX FUND (VCR)**
- **VANGUARD HEALTH CARE INDEX FUND (VHT)**
- **VANGUARD FINANCIALS INDEX FUND (VFH)**
- **VANGUARD INDUSTRIALS INDEX FUND (VIS)**

A person wearing an orange long-sleeved shirt is sitting at a desk. They are holding a light blue pen and writing in a dark-colored notebook. In the background, a laptop is visible, and the person's other hand is resting on the keyboard. The scene is softly lit, creating a focused and professional atmosphere.

RESEARCH PROCESS

A dark background with a blurred financial chart on the left side, showing blue bars and lines. The text "DATA SOURCES" is prominently displayed in the center in a large, white, bold, sans-serif font.

DATA SOURCES

**Alpha Vantage Stock
APIs**

**Alpha Vantage
Economic Indicators
APIs**

**United States Census
Bureau**

METHODS & TECHNOLOGIES



Microsoft Azure



Apache Kafka



SQL



Python



Power BI



MACHINE LEARNING

MODELS TESTED

Linear
Regression

```
graph TD; A[Linear Regression] --> B[LASSO CV]; B --> C[ARIMAX]; C --> D[SARIMAX]; D --> E[LSTM];
```

LASSO CV

ARIMAX

SARIMAX

LSTM

LINEAR REGRESSION & LASSO CV

LINEAR REGRESSION

- Least Squares Method
- Univariate or Multivariate
- **Poor performance** with time-dependent features

LASSO CV

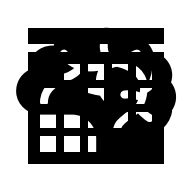
- Uses cross-validation
- Multiplies weight to each feature
- **Poor performance** with time-dependent features



ARIMAX

**"Auto Regressive Integrated
Moving Average with
eXogenous features"**

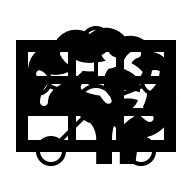
**Handles Non-
Stationary Data**



SARIMAX

**"Seasonal Auto Regressive
Integrated Moving Average
with eXogenous features"**

Seasonality



LSTM

“Long Short-Term Memory”

**Univariate or Multivariate
Capabilities**

**Artificial
Recurrent Neural
Network (RNN)**



RESULTS AND DASHBOARD DEMO



RECOMMENDATIONS

CONCLUSION

Vanguard's VTI Index can be used as a reliable US economic indicator

- With over 4,100 stocks, more than 99% of stocks are US based [2]

SARIMAX, ARIMA, and LSTM are all excellent models to use for predicting Vanguard's VTI stock prices [3]

- Reliably predicts index price trend
- SARIMAX has the most accurate short-term predictions (<1 year) between the tested models

The Vanguard ETFs representing the top 5 sectors in the US

- Low-risk, low-reward
- Industry with highest returns on investment:
 - Technology and Information (VGT)
- Industry with lowest risk:
 - Healthcare (VHT)

Correlations between top 5 sectors and economic features [3]

- Unemployment rate and 10-Year Treasury Constant Maturity Rate have strong inverse correlations
- Retail Trade had a strong correlation with all index funds



QUESTIONS

Visit our GitHub repository for more information:

[akittendorf/Domestic-Stock-Market-Analysis: This is a basic stock market analysis created as a capstone project for Dev10. \(github.com\)](#)

CITATIONS

1. *Vanguard Total Stock Market ETF (VTI) - Portfolio*. Vanguard. (Jan. 31, 2022). Retrieved Feb. 15, 2022, from "<https://investor.vanguard.com/etf/profile/portfolio/vti>"
2. *Vanguard Total Stock Market ETF (VTI)*. U.S. News - Money. (2022). Retrieved Feb. 16, 2022, from "<https://money.usnews.com/funds/etfs/large-blend/vanguard-total-stock-market-etf/vti>"
3. Project Executive Summary. (Feb. 17, 2022). Retrieved Feb. 17, 2022, from "<https://github.com/akittendorf/Domestic-Stock-Market-Analysis/blob/main/ProjectExecutiveSummary.pdf>"