EDA Report

Team Lead:

Recorder:

Spokesperson:

# Background and Introduction

Understanding and anticipating the investment decisions of influential figures has long been a subject of interest in financial markets. Warren Buffett, renowned for his disciplined value investing approach, focuses on identifying fundamentally strong yet undervalued companies (Hagstrom, 2013). Investors and financial professionals closely track his moves through SEC filings and media reports, but these sources provide insights only in retrospect. Team Lambda aspires to bridge that gap by leveraging machine learning to predict which stocks Buffett is likely to buy next, based on historical data and company-level financial indicators. By deploying predictive modeling and analysis, we aim to create a decision-support tool that empowers individual investors with insights traditionally available only after the fact.

The intersection of behavioral finance and predictive analytics presents a compelling opportunity to model Buffett’s investment patterns. While previous research has explored stock price prediction and portfolio performance modeling, few studies have specifically attempted to emulate the buy decision-making process of a known value investor using explainable machine learning models (Fischer & Krauss, 2018). Our approach is novel in that it does not merely assess market trends but seeks to identify the financial characteristics that make a stock appealing to Buffett before he purchases.

We hypothesize that historical financial metrics such as a low price-to-earnings (P/E) ratio, strong revenue growth, and high dividend yield are significant predictors of his stock selections. This is because Buffett prioritizes companies that exhibit financial strength, are undervalued relative to their intrinsic worth, and demonstrate steady, long-term growth potential (Buffett & Cunningham, 2020). If successful, we predict this work could enhance traditional financial analysis by providing predictive insights into high-profile investment strategies, offering a practical tool for investors who seek to align their decisions with Buffett’s time-tested approach.

# Methods

Our dataset consists of three data sources for individual stocks, economic metrics, and investor purchasing activity. Individual stocks were selected based on investor’s purchases and the remaining S&P 500 was added to represent the pool of available stocks throughout the 2007 - 2024 timeframe. The stock selection may have missing, unavailable, or incomputable metrics that will be addressed in cleaning. Examples include P/E ratios for companies with negative earnings and companies unavailable through Bloomberg.

Individual stock data, sourced from Bloomberg, was pulled as individual files representing quarterly snapshots from Q1 2007 to Q4 2024. These files were merged on column headers for a dataset containing all stock data. Warren Buffett’s portfolio activity is sourced from Dataroma, filtered on activity representing a buy (initial purchases and additional purchases of existing holdings), then a label was created to indicate 1 for all purchase activity. Economic data, sourced from Bloomberg, was pulled in three month increments to match the quarterly label for a merge. The dates were changed from short form (eg. 03/31/24) into two columns, Quarter (Q1) and Year (2024). These three datasets were then combined. Stock data and purchase activity merged on Ticker Quarter and Year. Stock data and economic data merged on Quarter and Year.

# Results

# Discussion

# Appendix

References:

Buffett, W., & Cunningham, L. A. (2020). *The Essays of Warren Buffett: Lessons for Corporate America*. The Cunningham Group.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270(2)*, 654-669.

Hagstrom, R. G. (2013). *The Warren Buffett Way*. Wiley.