Update: A Guided Investment Platform for First-Time Investors: Current Progress

Summary of Progress

Data-Related

- Found dataset categories to support our investment problem
- Set up first version of data pipeline to create first versions of training sets, and test sets
- (cont here)

Model-Related

- Found an investment strategy to base our model on
- Settled on the TensorFlow as the ML Framework to build model with
- Currently doing more research to investigate the relationship between metrics used in Value Investing and the magnitude of bias the Model should employ

System Integration-Related

- Decided to create a web app using the following technology stack to support our ML Application: Flask, ReactPy, SQL Cloud Database, TensorFlow .
- Began testing using Edge ML with the intention to transfer to Cloud ML soon

Ideation for ML Model

Data Engineering

- Identifying the metrics used to exercise Value Investing
- Identifying reputable data sources to scrape data from
- Programmatically defining the preprocessing phase to support the automation of data processing using data analysis tools

Initial ML Solution

- Initial idea: Creating a ML model bias to a subset of metrics. I.E. Influencing the model to prefer one or more metric over other metrics.
-

Data Validation & Cleaning

How are we collecting quality Datasets?

- Collecting relevant data that includes our benchmarks
- Using publicly available data of Fortune 500 Companies
- Analyze historical data for the past year

Metrics

Price to Earning (P/E) Ratio

- How the stock price compares to earnings:
- A low P/E suggests that the stock is undervalued which is what we are looking for.

Price to Book (P/B) Ratio

- How the market price compares to assets
- < 1 is ideal (means a company is trading for less than its net assets

Debt to Equity (D/B) Ratio

- Shows whether the company is financially stable or not
- <1 is less risk, safer investment

Free Cash Flow

- How much money the company truly makes
- Positive and gradually growing over time (means the company is making money and has enough to spend after expenses

Benchmarks

S&P 500

- Represents the overall U.S. stock market
- Acts as a baseline to compare our strategy to
  - Essentially if we can't beat it then the strategy does not add value
- Broad baseline

Russell 1000 Value Index

- Focuses on U.S. value companies (cheap in relation to what their actually worth
- Tracks U.S. companies that look "cheap" based on value measures (like low P/E and P/B)
- More focused on value stocks than the S&P 500.
- Direct comparison

HML Factor

- (High minus Low) return of value stocks - return of growth stocks
- Compares returns of value stocks (high book to market) vs. growth stocks (low book to market)
- Negative HML means growth stocks are leading the market
- Positive HML means value stocks are doing better

Benchmarks vs. Metrics: What we want.

- The Strategy: Value investing: buying solid companies when they're undervalued
- P/E & P/B: Lower = cheaper
    - We want below the market AND below Russell Value.
- D/E: Lower = safer
    - We want less than 1.0.
- FCF Yield: Higher = better
    - We want above 6%.
- S&P 500 = overall market test
- Russell 1000 Value = value stock competition
- HML = proof of whether value beats growth historically

Current Goals

Model Training & Evaluation

- Developing an automated system for Model Training
- Undergoing a deeper Investigation of the process of getting maximum ROI using Value Investing to improve training set and testing set
- Streamlining Model Training and Model Evaluation Process
- Creating an experiment involving the tradeoff between memory and compute

System-Integration & Model Monitoring

- Finding a reliable medium for ML Application Deployment
- Creating pipeline for preprocessed data to storage using ETL procedures
- Conducting more research on Model Monitoring