GenAI-NN3 Long-Short Investment Model

McGill-FIAM Asset Management Hackathon

October 2024

Teerthraj Patel, McGill University

Jasem Khelifi, École de technologie supérieure

Isha Sharma, McGill University

Agenda

Executive Summary

Brief overview of our investment strategy

Investment Strategy

1

A Long-Short Portfolio Strategy

Each month, we select the top 30 stocks with the highest predicted returns for long positions and the bottom 30 for short positions, using a long-short portfolio strategy driven by machine learning predictions of stock returns.

2

Usage of NN3 Model & Genetic Algorithm

The selection process is driven by a neural network model (NN3) for stock return predictions, enhanced by a genetic algorithm for optimal feature selection.

3

Incorporation of Economic Factors Using LLM

We further enhance our predictions by leveraging a large language model (LLM), which analyzes key economic factors to predict sector performance. Based on the LLM's output, we adjust the stock return predictions, slightly tweaking long or short positions in sectors expected to rise or fall.

Performance Metrics

Key Performance Metrics

- **Sharpe Ratio:** 3.86
- **CAPM Alpha:** 0.0244
- **t-statistic:** 11.87
- **Information Ratio:** 3.34
- **Max 1-Month Loss:** -4.67%
- **Maximum Drawdown:** -4.95%
- **Long Portfolio Turnover:** 0.8147
- **Short Portfolio Turnover:** 0.8935
- **Sortino Ratio** : 8.32

Investment Strategy

- Stock selection using an NN3 model
- Gen AI for optimal feature selection
- Incorporating lag features (1, 3, and 6 months)
- narrowing down the feature set from 151 to 22.

Bottom Up

Investment Strategy

Long-Short (30-30)

Our Top 10 Holdings

1. CHENIERE ENERGY
2. ROSETTA RESOURCE
3. AMERN INTL GROUP
4. WEST ALLNCE
5. WATSON PHARMACEU
6. SOFI TECH
7. EQUITRANS MID
8. TELADOC HEALTH
9. APPLE
10. GROCERY OUTLET

Investment Strategy

Long-Short (30-30)

Top Down

- Leveraging a Large Language Model (LLM) to generate economic forecasts, allowing us to anticipate different phases of economic cycles.
- Dynamically adjust sector weights by favoring sectors that are expected to outperform in specific economic conditions

Data and Methodology

Data Description

•**Financial Data**: Monthly stock returns and fundamental data, spanning 2000 to 2023.

•**Features**: 116 total features, including original and lagged versions of 29 fundamental, technical and sentimental indicators

•**Analyst recommendation data**, retrieved from the **I/B/E/S (Institutional Brokers' Estimate System)** database in the **WRDS (Wharton Research Data Services) library**

•**Macroeconomic factors**: 10 features based on the research by Amit Goyal (A Comprehensive Look at the Empirical Performance of Equity Premium Prediction).

•**Market Data**: S&P 500 returns and risk-free rates.

Statistical Tests

- **R^2 OOS : 0.001**
  This suggests that our strategy is market neutral and hence relies on idiosyncratic factors rather than market trends.
- **Diebold-Mariano Test : -1.4737**
  The negative value suggests that the NN3 model has smaller forecast errors compared to the Linear Regression model.
- **White's Reality Check Test : 1.5489**
  A positive test statistic indicates that the model (NN3) outperforms the benchmark in terms of predictive accuracy.

Methodology

Genetic Algorithm

Efficiently search the feature space to select the most predictive subset of features for each time window.

Limiting the number of features (subset size of 22), we reduce the risk of overfitting.

Model Architecture & Training Approach

NN3 Model
o Input Layer: Accepts selected features.
o Hidden Layers: 2 hidden layers using ReLU activation & L2 regularization.
o Output Layer: For regression output

L2 regularization & early stopping to prevent overfitting.
Lagged features capture momentum & temporal effects.

Neural Networks

To capture complex Nonlinear Relationships exhibited by Financial Markets effectively.

Models interactions between features, improving predictive accuracy.

Training Structure

Rolling windows with a 10-year training period and a 2-year testing period, advancing by 2 years each time.

Within each training window, we perform time-series cross-validation to evaluate model performance before testing on out-of-sample data.

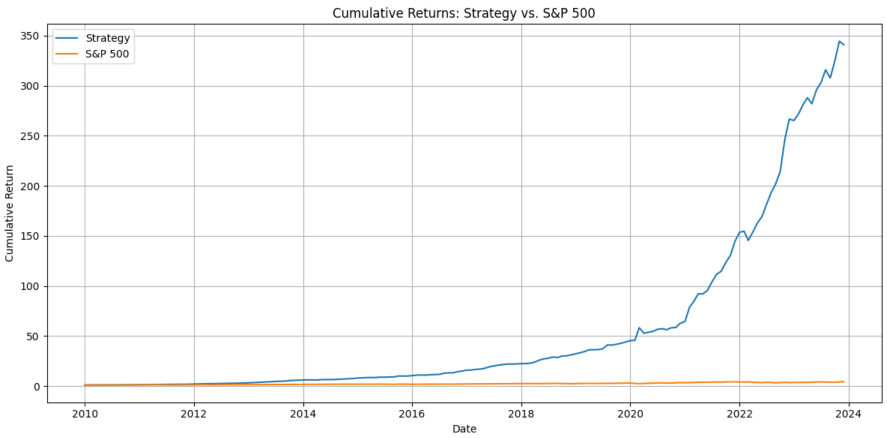
Incorporation of Economic Factors Using LLM

Portfolio Performance Statistics

Performance Metrics

Metrics	Our Portfolio	S&P 500
Average Annualized Return	52.56%	12.15%
Annualized Standard Deviation	12.71%	14.82%
Annualized Alpha	42.54%	-
Sharpe Ratio (Annualized)	3.86	0.76
Information Ratio (Annualized)	3.34	-
Maximum Drawdown	-4.95%	-24.77%
Maximum One-Month Loss	-4.67%	-12.51%
Long Portfolio Turnover	0.8147	-
Short Portfolio Turnover	0.8935	-

Our Portfolio vs S&P 500



Top 3 features

Analyst Recommendations	**Captures market sentiment and expert opinions on a stock’s future performance, often influencing price movements and investor behavior.**
Sale_gr1_lag1	**Measures the growth rate of a firm's sales over the last year, providing insights into its revenue momentum and future performance potential, adjusted with a one-period lag.**
B_me	**The book-to-market equity ratio, reflecting a company's valuation by comparing its book value to its market value, helps identify undervalued or overvalued stocks.**

Strategy Discussion & Future Improvements

Strategy Discussion

We grounded our strategy in a blend of established financial and machine learning literature while keeping computational efficiency in mind. Based on insights from Kelly, Gu, and Xiu's "Empirical Asset Pricing via Machine Learning," we determined that a shallow machine learning model like NN3 would best optimize our Sharpe ratio. The next critical step was identifying which covariates to include. The works of Goenko and Zhang guided us in selecting key factors, leading us to leverage the predictive power of analyst recommendations—a key driver of stock prices. Through experimentation, we incorporated lagged features, which dramatically improved our performance metrics. This refinement was enabled by a Genetic Algorithm that helped us reduce over 160 initial features to a more focused set of relevant factors. To further enhance predictive accuracy, we integrated an LLM as an economic forecaster, drawing on the research of Tang and Lopez-Lira in their paper, "Can ChatGPT Forecast Stock Price Movements?" Finally, we implemented cross-validation and a rolling window approach, leading to the development of our robust, final model.

Main Fundamental Signals

Analyst Recommendations: Features like **'Number Up'**, **'Number Down'**, **'Mean Recommendation'**, and **'Median Recommendation'** were frequently selected.

Ownership Percentages: **'Buy Percent'** and **'Hold Percent'**

Market Capitalization: The feature **'Market_Equity'** and its lagged versions were selected multiple times,

Volatility Measures: **'Standard Deviation'**
Financial Growth Metrics: Variables such as **'be_gr1a'** (book equity growth), **'inv_gr1a'** (inventory growth), **'sale_gr1'** (sales growth), and **'noa_at'** (net operating assets)

Profitability Ratios: **'ni_me'** (net income to market equity) and **'op_at'** (operating profitability).

Most Profitable Positions

- GROCERY OUTLET
- EQUITRANS MIDSTREAM
- ROSETTA RESOURCE
- WATSON PHARMACEUTICALS:
- SOFI TECHNOLOGIES:

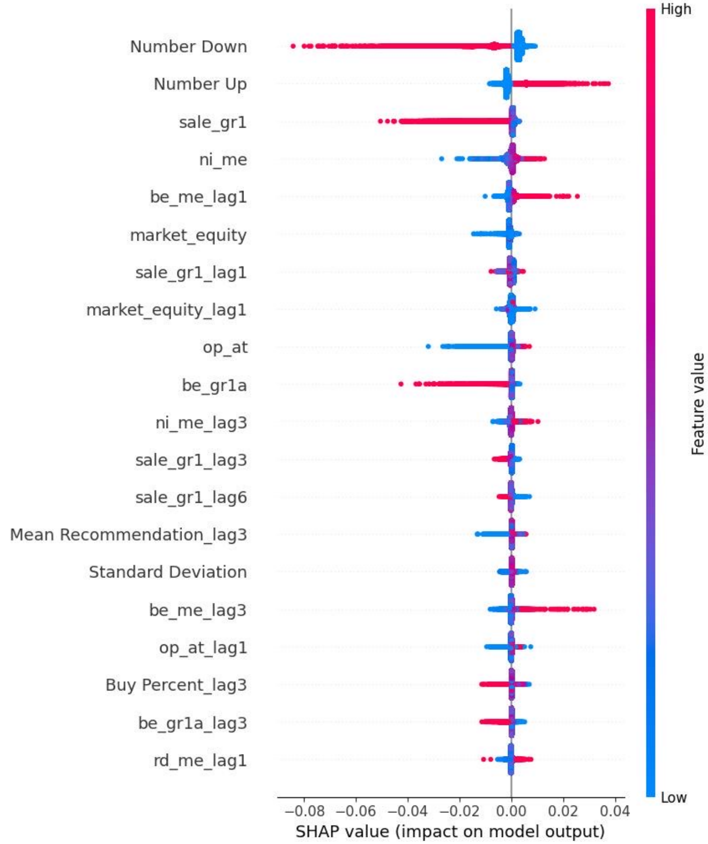
Contributing Macroeconomic Events

Technological Advancements: Rapid innovation in technology sectors could have positively impacted companies like APPLE and SOFI TECHNOLOGIES, leading to higher stock valuations.
Energy Market Fluctuations: Changes in oil and natural gas prices could have significantly affected energy companies like CHENIERE ENERGY and EQUITRANS MIDSTREAM, influencing their profitability and stock performance.
Healthcare Industry Developments: The pharmaceutical and healthcare sectors, represented by companies like WATSON PHARMACEUTICALS and TELADOC HEALTH, may have been influenced by regulatory changes, pandemic-related factors, or advancements in medical technology.

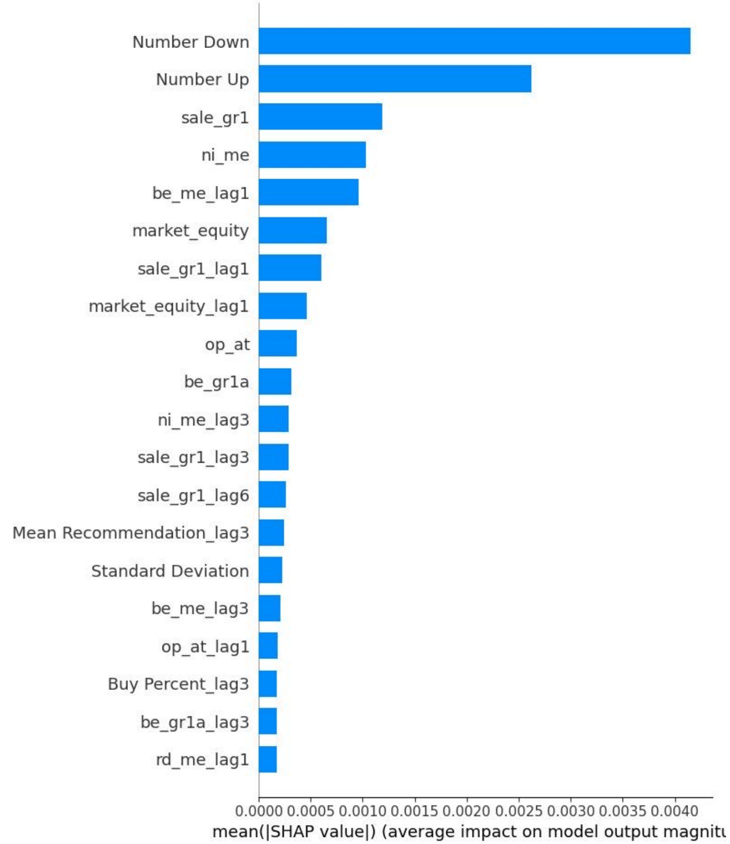
Appendix

Section VI

Which Covariates Matter?

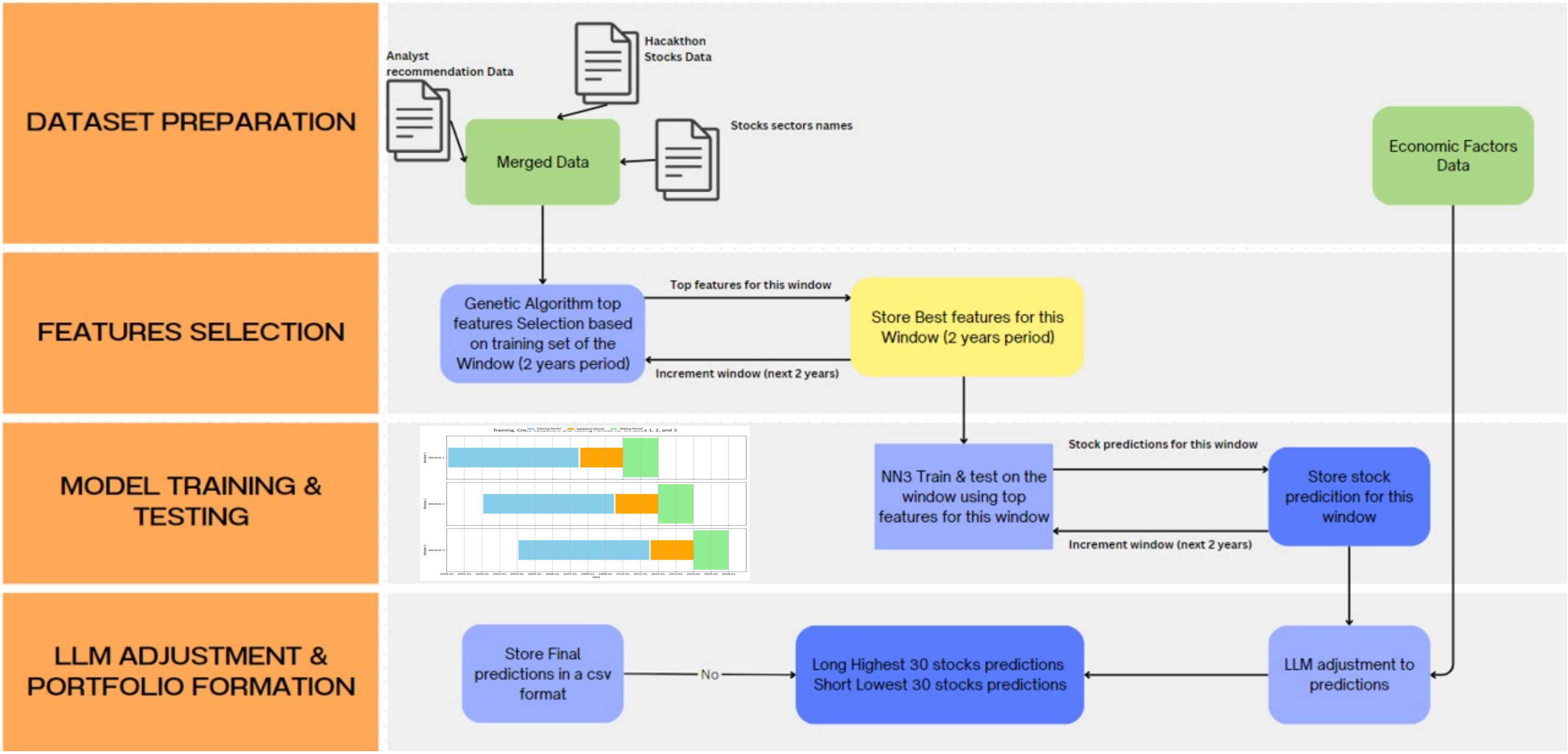


Overall SHAP Summary



Overall Features Importance

Model Architecture



Workflow of Data Processing, Feature Selection, and Model Training for Stock Prediction

Full Program working process

GPT3.5 Turbo API example prompt input and response for 1 month

```
TERMINAL

---- GPT Prompt ----

Based on the following economic factors:
{'vwretd': 0.015977, 'vwretx': 0.014957, 'ewretd': 0.028049, 'ewretx': 0.027344, 'totval': 11007604400.0,
 'totcnt': 500, 'usdval': 10843493500.0, 'usdcnt': 500, 'spindx': 1186.69, 'sprtrn': 0.014759}
Assess the following sectors:
['Information Technology', 'Utilities', 'Financials', 'Health Care', 'Industrials', 'Consumer Staples', '
Materials', 'Energy', 'Consumer Discretionary', 'Communication Services', 'Real Estate']
Return a mapping for each sector: 1 if the sector is safe and expected to rise, -1 if the sector is expec
ted to fall based on the provided economic factors. Format the response as JSON with sector names as keys
 and 1 (expected to rise) or -1 (expected to fall) as integer values, and make sure your response contain
s only json data, no other words or explanations.

---- GPT Response ----

{
    "Information Technology": 1,
    "Utilities": -1,
    "Financials": 1,
    "Health Care": 1,
    "Industrials": 1,
    "Consumer Staples": -1,
    "Materials": 1,
    "Energy": -1,
    "Consumer Discretionary": 1,
    "Communication Services": 1,
    "Real Estate": -1
}
```

Example of Processing predictions for 2010-04

Investment Strategy Development Timeline

1 Basic NN3 Model

We began with a basic NN3 model using 20 features derived from financial literature, achieving a **Sharpe Ratio of 1.08** after remodeling the portfolio formation technique

2 Gradient Boost Model

Next, we experimented with a gradient boost model using **50 features** from the original 151 based on literature reviews, but the performance declined, with the **Sharpe Ratio dropping to 0.69**.

3 Analyst Recommendations in NN3

We then added sentimental data, specifically **analyst recommendations**, to the initial NN3 model, which significantly improved performance, achieving a **Sharpe Ratio of 1.88**, our best result at that stage.

4 Ensemble Methods

Afterward, we explored an ensemble method combining **Random Forests and Gradient Boost Models**, but due to inconsistent financial metrics and one model dragging down the others, we decided to abandon this approach.

5 Feature Selection & Gen AI

We turned to various feature selection techniques, including Random Forests and SHAP analysis, but ultimately, **Gen AI** proved to be the most effective for selecting features, leading to strong results in our model.

6 Lag Features in NN3

Next, we integrated **lag features (1 month)** into the NN3 model to capture inter-temporal dependencies, which boosted the **Sharpe ratio to 1.96.**

7 Narrowing Features

Finally, we refined our model by narrowing down the feature set from 151 to 80 and then to 22, incorporating lag features, and achieved **our best Sharpe ratio between 3.5 and 4.5**.

8 Top-Down Approach with LLM

We further enhanced the strategy by using a **Large Language Model** for **economic forecasting** to adjust sector weights based on economic cycles.

We finalized our model with NN3, using Gen AI for feature selection, lag features, and LLM-based economic forecasting, achieving a Sharpe ratio of 3.5 to 4.5.

Reference

Section VII

References

1. Multi (Horizon) Factor Investing with AI by Russ Goyenko and Chengyu Zhang(2023).
2. Forest Through the Trees: Building Cross-Sections of Stock Returns by Svetlana Bryzgalova, Markus Pelger and Jason Zhu(2023).
3. Empirical Asset Pricing via Machine Learning by Shihao Gu, Bryan Kelly and Dacheng Xiu(2019).
4. Can ChatGPT Forecast Stock Price Movements? Return predictability and Large Language Models by Alejandro Lopez-Lira and Yuehua Tang(2024)
5. The Joint Cross Section of Options and Stock Returns Predictability with Big Data and Machine Learning by Ruslan Goyenko and Chengyu Zhang(2022)
6. Machine Learning for Continuous-Time Finance by Victor Duarte, Diogo Duarte, Dejanir Silva(2024).
7. Analyst Recommendation Data – I/B/E/S Database from WRDS Library.
8. A Comprehensive Look at the Empirical Performance of Equity Premium Prediction by Amit Goyal and Ivo Welch(2008) – Data updated 2023

List of our references to used research and data sources