Utilizing Artificial Intelligence for Algorithmic Trading

James Pavlicek, Jack Burt, Andrew Hocher
QMST 4320 Data Analytics, McCoy College of Business, Texas State University

Emmett & Miriam McCoy College of Business Administration

Introduction

History of Efficient Markets
Introduced by economist Eugene Fama in the 1960s, suggests that financial markets are "efficient," meaning they fully reflect all available information in prices at any given time. According to this theory, consistently achieving higher returns through stock market prediction or timing is impossible, as stock prices always incorporate and reflect all relevant information.

Active Investing vs Passive Investing
Active investing aims to surpass market benchmarks by engaging in frequent trading and strategic asset selection, based on thorough research and forecasts. In contrast, passive investing focuses on achieving long-term growth through a strategy that mirrors market indices, prioritizing lower costs and minimizing transactions.

Algorithmic Trading
Algorithmic trading utilizes complex computer algorithms to place trades based on predefined instructions. Incorporating natural language processing and machine learning, these algorithms detect signals in financial disclosures to anticipate market shifts, minimizing human error and biases for systematic trading decisions.

Natural Language Processing
Natural Language Processing (NLP) in artificial intelligence enables computers to understand, interpret, and generate human language, fostering natural dialogue between humans and machines.

Objective

Research Question
Can artificial intelligence be used to interpret 10-K financial reports and predict stock market movements?

Null Hypothesis (Ho)
NLP analysis of 10-K financial reports does not provide predictive insights that can be used to achieve returns that outperform the market, consistent with the Efficient Market Hypothesis.

Alternative Hypothesis (Ha)
NLP analysis of 10-K financial reports provides predictive insights that can be used to achieve returns that outperform the market, challenging the Efficient Market Hypothesis.

Significance
The significance our research is multi-layered, impacting both the field of finance and the advancement of artificial intelligence. Our research is crucial, using NLP to uncover investment insights and market trends not easily seen through conventional analysis.

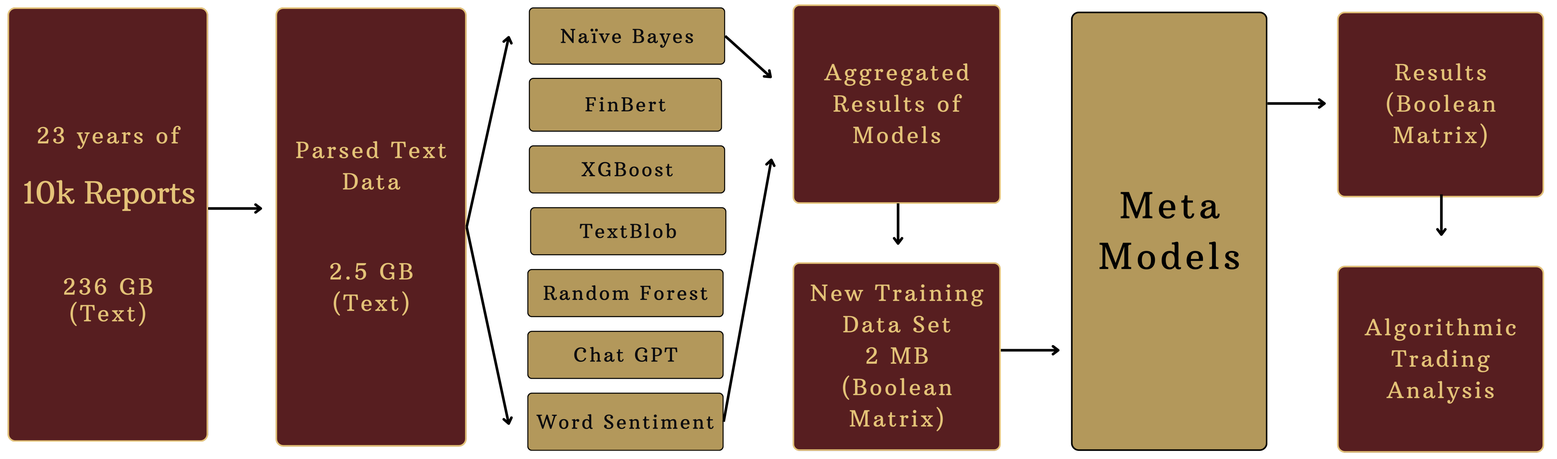
Data & Methodology

10k: Item 7
Item 7 of the Form 10-K, “Management's Discussion and Analysis of Financial Condition and Results of Operations.” offers management’s perspective on the company's financial health, operational results, and cash flows over the last year. This is the portion of the 10k we extracted to use for our analysis.

Data Source
Our data source is a collection of 10-Ks from 2000-2023. The dataset was built by Professor Bill McDonald (ND) from the SEC EDGAR website for public download.

Data Preprocessing
The preprocessing pipeline prepares text data for classification by removing HTML tags, replacing contractions, removing special characters and numbers, converting text to lowercase, tokenizing, removing stop words, lemmatizing words and encoding categories. Then, it applies vectorizing techniques (Count Vectorizer and TF-IDF) to convert text into numerical features.

Models



Model Development Process

NLP Sentiment Model

FinBert
A natural language processing model specifically trained on financial texts to analyze sentiment and understand context in financial documents.

TextBlob
A predictive model utilizes NLP techniques to analyze and predict text sentiment by evaluating the polarity and subjectivity of a piece of text.

Word Sentiment
A comparison of a list of positive and negative finance words to the 10k text. (Frequency analysis).

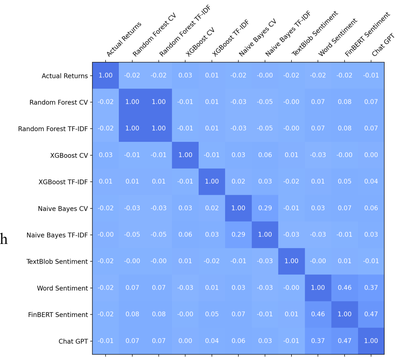
Chat GPT API
An advanced language model used to classify text sentiment considering context and language nuances.

Machine Learning Model

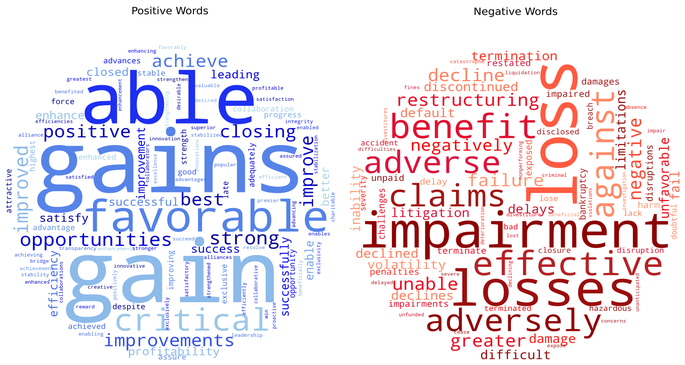
Naïve Bays
Classification algorithm based on Bayes' Theorem that assumes independence among predictors to predict the category of a given sample.

Random Forest
A versatile machine learning algorithm that builds multiple decision trees and merges them together to get a more accurate and stable prediction, widely used for both classification and regression tasks.

XGBoost
A highly efficient and scalable machine learning algorithm that uses gradient boosting frameworks for building fast and accurate predictive models.



Correlation Matrix of Models



Frequency Distribution of Positive and Negative Words

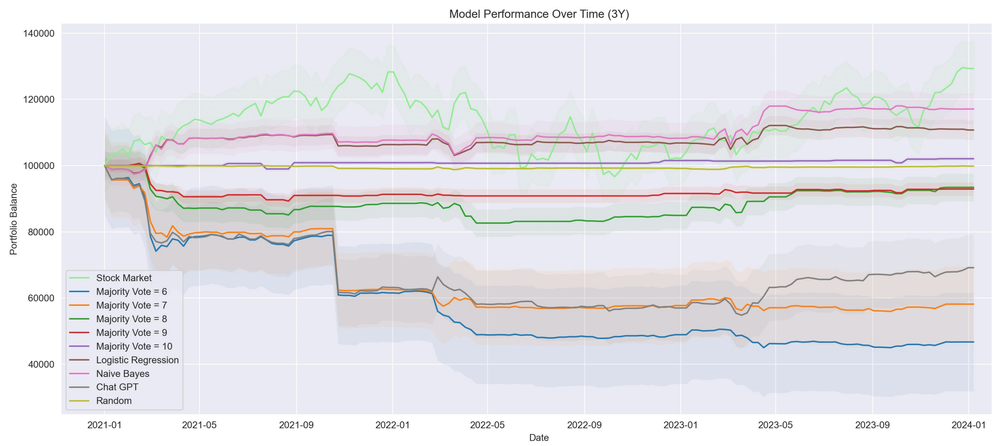
Results

Models
Out of the selected models there were notable observations. The first, traditional machine learning models outperformed the sentiment models. Out of these models XGBoost preformed the best, followed by Naïve Bayes and then Random Forest. Additionally, there was not a clear vectorizing technique that stood out from the other. Our second observation was the two models that had context of Financial documents (FinBERT and Word Sentiment) were two out of three of the bottom performers.

Model	Accuracy	Sensitivity	Specificity
XGBoost CV	51.50%	18.80%	84.21%
XGBoost TF-IDF	50.84%	39.32%	62.37%
Naive Bayes TF-IDF	50.28%	79.49%	21.05%
Naive Bayes CV	50.08%	24.98%	75.20%
Random Forest CV	49.75%	68.77%	30.72%
Random Forest TF-IDF	49.75%	68.77%	30.72%
Chat GPT	49.72%	51.55%	47.89%
Word Sentiment	48.80%	52.33%	45.26%
FinBERT Sentiment	48.80%	44.38%	53.22%
TextBlob Sentiment	48.04%	52.33%	43.75%

Model Performance

Meta Model
After building initial models, we aggregated them into a "meta-model" to explore potential predictive power beyond individual performance. This involved combining outputs using techniques like majority voting and ensemble optimization. Majority voting considers outputs from models, executing a trade if a certain number agree on the same decision. We tested five thresholds: ≥6, ≥7, ≥8, ≥9, or ≥10 models predicting the same output. Ensemble optimization converted initial model outputs into a binary matrix. Then, we processed every combination of columns (1023 possibilities) into logistic regression or Naïve Bayes models to find the best combination. This best combination was tested over separate 50 train/test seed iterations to determine average model accuracy.



Algorithmic Trading Performance Backtest of Meta-Models and Benchmarks

Model	Ending Value	Returns	Standard Devation	Sharpe Ratio
Stock Market	$127,500	8.44%	17.70%	0.42
Naive Bayes	$117,105	1.59%	29.82%	0.02
Logistic Regression	$110,747	0.25%	29.63%	-0.03
Random	$99,838	-2.54%	26.64%	-0.13
Majority Vote = 10	$102,072	0.68%	2.32%	-0.14
Majority Vote = 8	$93,427	-2.38%	9.30%	-0.36
Chat GPT	$69,148	-13.27%	29.38%	-0.49
Majority Vote = 9	$92,919	-2.47%	5.48%	-0.63
Majority Vote = 7	$58,142	-17.81%	26.07%	-0.72
Majority Vote = 6	$46,703	-23.79%	28.62%	-0.87

Meta-Models Portfolio Performance Comparison

Risk vs Return of Meta-Models and Benchmarks

Meta Model Performance
The models' performance was graphed over a 3-year period to illustrate their performance on out-of-sample data. We conducted 1000 different seed iterations of training/testing splits to mitigate "lucky" streaks. The graphed data represents the average portfolio value for a given time, along with a confidence interval band of +/- 1 standard deviation (62.8% of all samples). These meta-models were compared against three benchmarks: the total stock market index (Ticker: VTI), a random 50/50 chance of buying or shorting the stock, and a Chat GPT API call with the query "What is the sentiment of the following text?" along with some additional context.

Findings
In our research, we discovered that, on average, the stock market outperformed all models, with Naïve Bayes and Logistic Regression ensemble models following closely behind. However, ensemble optimization methods, while promising, resulted in significantly higher investment risk, measured by standard deviation of portfolio value over time. For majority vote methods, we observed that as trade frequency increased, returns decreased and standard deviation increased. The random benchmark showed similar starting and ending portfolio balances. Lastly, the Chat GPT model performed poorly when used alone. Utilizing “Sharpe Ratio” (a finance measure of risk vs return) we found that the stock market significantly outperformed any of our models due to its high return for relative volatility. Naïve Bayes and Logistic Regression follow behind due to their positive returns. Finally, for every other method you were better off doing a coin flip for your trading decisions.

Conclusion

Analysis Overview
Our study evaluated the Efficient Market Hypothesis (EMH) by applying Natural Language Processing (NLP) techniques on 10-K financial reports. We developed predictive models aimed at gauging stock market movements from financial sentiments in these reports. The models surpassed the no-information rate, showcasing modest predictive capability.

Efficient Market Validation
Despite our efforts, the research did not outperform the Total Stock Market Index (VTI), reinforcing the EMH. This outcome underlines the market's efficiency in incorporating public information into stock prices and challenges the potential of achieving superior returns through 10-K report analysis alone.

Recommendations

Computational Advancements
Future research should leverage more sophisticated computational techniques. The incorporation of deep learning algorithms and advanced statistical models may unveil patterns not discernible through traditional NLP methodologies.

Diverse Data Sources
Broadening the dataset to include various unstructured data like social media, news articles, and analyst reports might enrich the data pool for more accurate market predictions.

Cross-disciplinary Approaches
Statistics: Future research should leverage advanced statistical methods for evaluating NLP model performance. Error reduction and model optimization through statistical methodologies like cross-validation and ensemble techniques will enhance prediction accuracy.
Finance: Research must take in to account subject matter expertise of financial markets and theories, especially in relation to the EMH. This involves examining the economic viability of exploiting predictive insights from NLP models within market frameworks. Additionally, merging NLP discoveries with behavioral finance theories could illuminate the impact of investor sentiment on market trends, offering a refined perspective on market efficiency that incorporates behavioral insights.

References

Ke, Z. T., Kelly, B., & Xiu, D. (2019). Predicting Returns with Text Data. Becker Friedman Institute for Economics at The University of Chicago.

Loughran, T., & McDonald, B. Loughran-McDonald Master Dictionary. Retrieved from https://sraf.nd.edu/loughranmcdonald-master-dictionary/

U.S. Securities and Exchange Commission. EDGAR Company Filings. Retrieved from https://www.sec.gov/edgar/searchedgar/companysearch.html

Acknowledgements

This presentation is done as part of a project for QMST 4320 Data Analytics class, which is supervised by Dr. Tahir Ekin.

TEXAS STATE UNIVERSITY
The rising STAR of Texas