A Study on Quantifying ESG Alpha as an access return on portfolio using ESG Sentiment in the Banking Sector with an Automated Machine Learning Approach

Authors: Maria Babaeva
Supervisor: Joakim Westerlund

Data Analytics and Business Economics
Course: Economics Master Essay I, VT24

CONTENTS

ABSTRACT

This research delves into the burgeoning field of incorporating Environmental, Social, and Governance (ESG) criteria into investment strategies, spurred by a growing interest in sustainability and the demonstrated influence of ESG-related events on stock prices. Sanctify Financial Technologies is at the forefront of quantifying firms' ESG performance, which plays a pivotal role in this study. Various machine learning models are applied to forecast long-term stock returns for companies in the banking, financial services, and investment sectors. These models are trained on time series data spanning from 2015 to 2019, encompassing both financial metrics (ratios) and Sanctify's ESG scores, to capture comprehensive insights.

Simultaneously, this study aims to predict long-term returns for investments spanning over three years, leveraging machine learning models that are trained on historical data from 2015 to 2019. The models integrate both ESG scores from Sanctify Financial Technologies and financial metrics specific to the banking, financial services, and investment sectors. The predictions are then performed on 2020 financial data and corresponding ESG scores, allowing for a robust evaluation of the models' predictive accuracy and the impact of ESG factors on long-term investment performance.

The significance of the ESG variable in the model is rigorously assessed using various statistical tests, demonstrating a high level of significance in influencing long-term investment returns. Furthermore, the proposed portfolio constructed based on a model incorporating ESG scores achieves a cumulative return higher than a portfolio built solely on financial ratios, indicating the added value of ESG considerations in investment decision-making processes.

However, despite these advantages, both portfolios, with and without ESG incorporation, do not outperform a benchmark S&P 500 portfolio return over the three-year period studied. This comparative analysis provides valuable insights into the relative performance of ESG-informed investment strategies within the context of broader market benchmarks.

Through this comprehensive evaluation, stakeholders can gain a nuanced understanding of the impact of ESG factors on investment performance and the potential trade-offs between sustainability considerations and market benchmark performance. These insights contribute to ongoing discussions in sustainable finance and investment strategy development.

CCS CONCEPTS Computing methodologies → Machine learning; Semi-supervised learning settings;

KEYWORDS ESG alpha, scholar data, alternative data, AI in finance, quantitative investment

1 INTRODUCTION

The intersection of environmental, social, and governance (ESG) factors with investment strategies has garnered significant attention in recent years.

In tandem with advancements in quantitative finance, there has been a shift in societal expectations regarding corporate behavior. Back in 1970, prominent economist Milton Friedman argued that a business's sole responsibility was to maximize profits [12]. However, in recent years, this perspective has largely fallen out of favor, and it has become standard for corporations to pursue both financial and non-financial objectives [30].

The reasons behind this shift are complex, but one key factor is the realization among managers and CEOs that integrating social and environmental goals can also contribute to financial success. For instance, companies catering to ESG-conscious customers risk losing market share due to poor sustainability practices [20].

The increasing adoption of corporate social responsibility (CSR) practices has led to growing investor interest in these companies. A company's perceived performance in Environmental, Social, and Governance (ESG) metrics now influences its stock prices [3]. Moreover, the evolution of ESG factors within the financial sector has transformed them from mere buzzwords into essential components of regulatory frameworks and investment strategies [13]. The global growth of sustainable assets underscores the increasing relevance of ESG considerations in investment decisions [29]. This shift is reflected in the disclosure of Corporate Social Responsibility information by most leading companies, highlighting the widespread adoption of ESG practices [21].

Within this context, banks occupy a pivotal role due to their exposure to major ESG risks highlighted by regulatory bodies such as the Basel Committee on Banking Supervision [4]. As institutions vital to financial flows and sustainable activities, banks face scrutiny regarding their ESG risk exposure and are encouraged to integrate ESG principles into their risk management frameworks [13].

However, a challenge arises when attempting to incorporate ESG performance into investment decisions because measuring the ESG impact of a company in objective numerical terms is difficult [17].

To address this challenge, a new wave of fintech companies is leveraging advancements in machine learning and vast amounts of sentiment data to provide quantitative measurements of company ESG performance. These efforts aim to make ESG factors more tangible and actionable for investors in today's data-rich environment.

Drawing from the insights of Qian Chen and Xiao-Yang Liu's research [7], which showed the long-term impact of ESG efforts on business development and stock performance, we adopt a strategic focus on long-term investment returns spanning over three years. Our study aims to capture the gradual but substantial influence of ESG policies and practices on investment outcomes, contrasting traditional financial metrics with ESG scores. This approach aligns with the findings that while ESG alpha strategies may initially lag behind traditional benchmarks, they demonstrate steady growth and ultimately outperform, reflecting sustainable business trends.

Additionally, we adopt a strategic focus on capturing alpha—excess return over a benchmark like the S&P 500. Alpha represents the sought-after measure of performance that investors aim to achieve through strategic investment decisions. One hypothesis we aim to test is whether incorporating ESG scores into predictive models allows us to capture this alpha effectively, reflecting sustainable business trends and responsible investment strategies.

Through machine learning models trained on sector-specific data encompassing ESG scores and financial metrics from 2015 to 2019, our study assesses the predictive power of models incorporating ESG scores against those using financial metrics alone. The ultimate goal is to construct and compare portfolios based on high share return strategies derived from these models. Alongside considerations of the Capital Asset Pricing Model (CAPM) and portfolio return access knowledge, our study aims to quantify and compare the alpha generated by investment strategies incorporating ESG scores against those using traditional financial metrics alone.

By evaluating portfolio performance against benchmarks such as the S&P 500 return over a three-year period, we seek to elucidate the added value of ESG integration in achieving long-term investment objectives and outperforming market benchmarks. This comprehensive approach addresses the evolving landscape of investment decision-making, highlighting the importance of ESG integration for long-term value creation and risk-adjusted returns within the banking and financial sectors.

Summary: The ESG alpha is a quantitative measurement of companies' ESG commitment, which is closely related with their stock prices. Therefore, in this paper, we propose a novel approach to quantitatively capturing this ESG alpha using an automatic machine learning method on a combination of ESG data provided from one of fintech companies called Sanctify as an alternative data source and traditional financial indicators. In addition, our study investigates the significance of ESG variables within predictive models for long-term investments in the banking, financial services, and investment sectors.

2 THEORETICAL FINANCIAL BACKGROUND

2.1 Capital Asset Pricing model

The foundational description of the Capital Asset Pricing Model (CAPM) and its significance within modern finance, including its relationship to portfolio theory and its applications in portfolio risk management, fund performance measurement, and security valuation, can be attributed to Zabarankin, Pavlikov, and Uryasev (2014) [37]. They outlined how the CAPM, developed by Sharpe (1964) [34], Lintner (1965a,b) [24,25] and Mossin (1966) [26], can be understood from two key perspectives:

(i) The CAPM can be seen as a reimagining of the essential optimality conditions essential to Markowitz's mean-variance portfolio problem, thereby linking its foundations to the concept of risk as variance.

(ii) Additionally, the CAPM functions as a single-factor linear model known as the security market line, which establishes a relationship between the expected returns of an asset and those of a market portfolio. Here, the slope of the line, termed asset beta, acts as a metric for asset non-diversifiable (systematic) risk.

Here are the key components and concepts of the CAPM model:

1. Expected Return: The expected return of an asset is calculated using the risk-free rate of return and a risk premium based on the asset's beta (systematic risk).

$$ER = R_f + \beta \times (R_m - R_f) \tag{2.1}$$

- ER: Expected return of the asset.
- R_f: Risk-free rate of return (e.g., return on government bonds).
- β: Beta of the asset, representing its systematic risk relative to the market.
- R_m: Expected return of the market portfolio.
- $R_m - R_f$: Market risk premium, the excess return of the market over the risk-free rate.

Beta (β): Beta measures an asset's sensitivity to market movements. A beta of 1 implies the asset moves in line with the market, while a beta greater than 1 indicates higher volatility, and a beta less than 1 suggests lower volatility.

$$\beta = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)} \tag{2.2}$$

Where:

- β is the beta of the security.
- $\mathrm{Cov}(R_i, R_m)$ is the covariance between the security's returns (R_i) and the market returns (R_m).
- $\mathrm{Var}(R_m)$ is the variance of the market returns (R_m).

Risk-Free Rate (): This is the theoretical return on an investment with zero risk, often approximated using government bond yields.

Market Risk Premium ( ): It represents the excess return investors demand for holding a risky asset over the risk-free rate. It reflects the risk inherent in investing in the overall market.

The CAPM model assumes rational investors who require compensation for both the time value of money (risk-free rate) and the risk they undertake (beta risk). It's a cornerstone in modern finance for estimating the required rate of return for assets, evaluating investment opportunities, and assessing the performance of investment portfolios relative to their risk exposure.

2.2 Alpha (α) as an access return

Alpha (α) represents the prowess of an investment approach to outperform the market or demonstrate its "edge." It is alternatively termed “excess return” or the “abnormal rate of return” when compared to a benchmark, adjusting for risk. Alpha is often discussed alongside beta (β), which quantifies the overall market volatility or risk, known as systematic market risk.

In finance, alpha serves as a metric of performance, indicating when a strategy, trader, or portfolio manager surpasses the market return or another benchmark within a specific timeframe. Considered the active return on an investment, alpha measures investment performance against a market index or benchmark representing overall market movements.

An investment’s alpha is the surplus return relative to the benchmark index return, showcasing the investment’s ability to outperform. Alpha values can be positive or negative, reflecting active investing outcomes. Conversely, beta gains are typical in passive index investing scenarios.

Active portfolio managers aim to generate alpha within diversified portfolios, utilizing diversification to mitigate unsystematic risk. Alpha, reflecting portfolio performance relative to a benchmark, is often seen as the value added or subtracted by a portfolio manager from a fund's overall return.

In this work "Jensen’s alpha" will be considered. Jensen’s alpha incorporates elements of the capital asset pricing model (CAPM) market theory, integrating a risk-adjusted element into its computation [18]. The CAPM model employs beta (or beta coefficient) to estimate an asset's expected return based on its unique beta and the anticipated market returns. Investment professionals utilize alpha and beta together for return calculation, comparison, and analysis. It helps assess whether a portfolio manager has added value beyond what would be expected given the portfolio's systematic risk (beta) and the market's return.

Equation (2.3) in the CAPM model posits that the expected return on any asset equals the risk-free rate plus a risk premium derived from the asset's systematic risk multiplied by the market portfolio's risk premium. The market portfolio's risk premium is the disparity between its expected returns and the risk-free rate.

$$\text{Alpha} = R(i) - (R(f) + B \times (R(m) - R(f))) \tag{2.3}$$

Here's what each component of the formula represents:

- $R(i)$: The actual return of the investment or portfolio.
- $R(f)$: The risk-free rate of return.
- B: Beta of the investment or portfolio.
- $R(m)$: The expected return of the market portfolio.

The CAPM model equation straightforwardly indicates what returns any security (or portfolio) could yield considering its level of systematic risk, beta (β). Should a portfolio manager or security analyst successfully forecast future security prices, they could potentially achieve higher returns than those implied by equation (2.1) and their portfolio's risk level.
Transitioning equation (2.1) to assess any portfolio manager's forecasting ability involves reframing it in terms of objectively measurable realized returns on portfolios and the market portfolio, as opposed to strictly hypothetical expected returns.
A positive alpha suggests that the investment has outperformed expectations after adjusting for risk, while a negative alpha indicates underperformance relative to the CAPM model's predictions. It's a valuable tool for evaluating the skill or effectiveness of portfolio managers in generating excess returns.

2.3 Simple Moving average

A moving average is a statistical calculation used to analyze data points by creating a series of averages of different subsets of the full dataset. In financial analysis and time series analysis, the most commonly used moving averages are Simple Moving Average (SMA) and Exponential Moving Average (EMA).

- Simple Moving Average (SMA) (2.4): It calculates the average of a specified number of data points over a given period by equally weighting each data point. For example, a 10-day SMA calculates the average of the last 10 closing prices.
- Exponential Moving Average (EMA): It gives more weight to recent data points while still considering older data. This is achieved by applying a smoothing factor to the previous EMA value and the current data point. EMAs are more responsive to recent price changes compared to SMAs.

$$\text{SMA} = \frac{\text{Sum of Prices for Last } n \text{ Periods}}{n} \tag{2.4}$$

Where:

- SMA is the Simple Moving Average.
- "Sum of Prices for Last n Periods" refers to adding up the closing prices of an asset over the last n periods (days, weeks, etc.).
- n represents the number of periods used in the calculation of the SMA.

Moving averages are used for various purposes in finance and data analysis

2.4 Sanctify Financial Technologies

This thesis is the result of a collaboration with Sanctify Financial Technologies, a company based in Lund that specializes in generating numerical ESG-performance scores for companies globally. The company utilizes an API to gather data from numerous sources, such as tweets, news, articles etc, and employs Natural Language Processing (NLP) techniques to analyze this data for sentiment, which is then translated into numerical scores. These scores are designed for use by fund managers across different sectors to integrate ESG factors into their investment strategies. These ESG scores cover a wide range of companies globally and span over ten years with daily frequency.

3 RELATED WORK

My thesis was inspired by several works done in finding the relation between ESG scores and financial performance of the company.

Farma et al. investigated the impact of integrating ESG data points on investment profitability and sustainability promotion. Their study revealed a positive correlation between higher ESG scores and improved financial performance. Notably, machine learning (ML) models such as linear regression and random forest regression exhibited superior performance when incorporating both ESG and financial data into their training datasets, highlighting the potential benefits of considering ESG factors in investment decision-making [33].

Margot et al. developed a machine learning algorithm designed to uncover relationships between ESG profiles and companies' financial performances. This algorithm, continually updated with rule sets, maps data points from the high-dimensional ESG feature space to predict excess returns. The resulting predictions are transformed into scores used for investment screening, surpassing traditional strategies reliant solely on ESG ratings. This nonlinear machine learning approach effectively links ESG characteristics with financial outcomes, offering an efficient tool for stock selection [28].

Gupta et al. introduced a framework utilizing statistical analysis and machine learning techniques to assess the significance of ESG parameters in investment decisions and their impact on financial performance. Their findings indicated higher return on equity (ROE) for companies with top-tier ESG ratings compared to others [16].

Guo et al. investigated the influence of ESG issues within financial news and assessed the predictive capability of ESG news in forecasting stock volatility. Bayesian inference methods were utilized for stable learning and prediction during both training and inference phases. The study evaluated stock market volatility using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics [15].

Yu et al. employed machine learning algorithms to explore the relationship between ESG scores and stock returns using CSR Rating Agency data. Their analysis revealed that while ESG-related stocks did not yield excess returns and exhibited typical risk performance during normal periods, they did not underperform either [36].

Chen et al. proposed a machine learning approach to quantify a company's ESG premium and capture ESG alpha using scholarly data as alternative data sources. Their study demonstrated that ESG alpha strategies based on scholarly data more effectively capture ESG premiums than traditional financial indicators [7].

Previous studies have explored various methodologies for analyzing ESG data, including the use of machine learning and deep learning algorithms. However, these studies have not used specific ESG scores calculated using an unique technique provided by Sanctify company. Moreover, they did not consider the assumption of predicting returns for long-term investments and considering only banking and financial sectors for their analysis.

Therefore, in this thesis, we will investigate the hypothesis that a portfolio constructed based on a model incorporating ESG knowledge, particularly from Sanctify's ESG scores, will outperform a portfolio constructed without ESG scores over a three-year investment horizon. Additionally, we aim to demonstrate that this ESG-inclusive model generates excess returns, known as ESG alpha, surpassing the benchmark index S&P 500. This approach not only tests the efficacy of ESG integration in investment strategies but also evaluates its potential as a source of alpha in financial markets. Hence, our research aims to fill these significant gaps in the literature.

4 OBJECTIVE

The objective of this thesis is to explore the validity of AI-based ESG scores as a source of alpha in the investing landscape, particularly by leveraging these scores to generate ESG portfolios that outperform the market. Essentially, there will be an examination whether having access to more detailed and up-to-date information about a company's activities and performance, especially in terms of its ESG practices, can contribute to generating excess returns in a portfolio. By linking news-derived information with established ESG scores, we aim to explore how this combined dataset influences investment decisions and potentially leads to alpha generation in portfolio management strategies.

There will be other valuable predictors of excess returns (alpha) in investment portfolios, several factors beyond ESG scores and news-related information can play significant roles.

Here are some additional predictors that will be used:

Profitability Ratios: Ratios such as Return on Equity (ROE), Return on Assets (ROA), Gross Profit Margin, and Net Profit Margin assess a company's ability to generate profits relative to its equity or assets and its revenue.

Liquidity Ratios: Metrics like Current Ratio and Quick Ratio evaluate a company's ability to meet short-term financial obligations with its liquid assets. They reflect liquidity and short-term solvency.

Efficiency Ratios: Ratios such as Asset Turnover Ratio and Inventory Turnover Ratio gauge how effectively a company utilizes its assets and manages its inventory to generate revenue.

Solvency Ratios: Debt-to-Equity Ratio, Debt Ratio, and Interest Coverage Ratio assess a company's debt levels, financial leverage, and ability to meet interest payments.

Valuation Ratios: Price-to-Earnings (P/E) Ratio, Price-to-Book (P/B) Ratio, and Price-to-Sales (P/S) Ratio compare a company's market value or share price with its earnings, book value, or sales, providing insights into valuation levels.

Growth Ratios: Metrics like Earnings Growth Rate, Revenue Growth Rate, and Dividend Growth Rate analyze the rate at which a company's earnings, sales, or dividends are growing over time.

By integrating these diverse predictors into your analysis alongside ESG scores and news-related data, we can create a comprehensive framework for evaluating and predicting alpha generation in investment portfolios. Each factor contributes unique insights and can help refine investment strategies based on a holistic understanding of market dynamics and company-specific attributes.

The significance of these predictors will be evaluated through various techniques and methods. Subsequently, two portfolios will be constructed: one incorporating an ESG score model, and the other without. These portfolios will then be compared against each other and against the benchmark S&P 500 portfolio. All analyses will be conducted over the period from 2015 to 2024.

This approach allows for a direct comparison between the performance of portfolios incorporating ESG scores and those without, providing insights into the impact of ESG considerations on investment outcomes relative to a widely recognized benchmark. It also covers a substantial time frame, enhancing the robustness of the analysis and capturing potential long-term trends or patterns.

5 DATA COLLECTION

In this section, we introduce a machine learning methodology for quantifying ESG alpha. Initially, we prepare the training dataset by web-scraping from different financial resources. And then we will combine ESG data with financial indicators. Subsequently, we trained six distinct machine learning models and chose the best one for making predictions. Finally, we create two portfolios and compare their real returns over a 3 year period with one from S&P 500.

5.1 Scope

The overarching objective could potentially encompass every stock that Sanctify evaluates and every method used to integrate these scores into stock price prediction models. However, due to feasibility constraints, a more focused scope has been defined.

Given that Sanctify's ESG-score data is limited before 2015, our analysis will be confined to the period from 2015 to 2024, with the year 2020 designated as the test period and the preceding five years utilized for training and validation.

The Sanctify ESG database offers insights into tens of thousands of companies across global exchanges, providing numerical scores derived from NLP-analyzed news media. These scores are based on news relevant to 26 sustainability categories defined by the Sustainability Accounting Standards Board (SASB) and are updated daily. Sentiment analysis assigns scores ranging from [-1, 1] based on news positivity or negativity, which are then integrated into E, S, and G scores for each company. Absolute and relative variants of these scores exist, with our analysis focusing solely on long-term relative ESG scores because by just looking at the Absolute scores, it’s difficult to know if the difference actually means that one company performs better than the other, or if it just depends on the data amount.

So Sanctify scores used in this thesis are composite scores designed to reflect a full picture of a company’s ESG performance both historically and currently in a single numeric. Two main parts go into this score, the Trend scores and the Overall performance scores. A company with historically many controversies or constant bad press will get a lower score to reflect this, while news about positive current ESG work will increase the score. This is our go-to score to instantly get an overview of a company’s ESG performance.

Additionally, our investment horizon will span a three-year period as our aim is to consider an influence of ESG score on a long term investment opportunity. Financial ratios for predictive modeling will be sourced from annual financial statements, including Balance Sheets, Income Statements, and Cash Flow Statements.

To streamline the analysis, we will exclude administrative hurdles and transaction fees from consideration. When selecting attributes for stock price prediction, various data sources can be leveraged, ranging from macroeconomic indicators to company-specific financial disclosures. Given the complexity of attribute selection, financial ratios for predictive modeling will be derived from yearly statements.

The S&P 500 index serves as our benchmark. This index represents the performance of 500 large-cap U.S. stocks, capturing the overall performance of the U.S. stock market and serving as a key indicator of market trends and investor sentiment.

As for the selection criteria, we will consider 56 companies from the banking, financial services, and insurance sectors traded on Nasdaq. Nasdaq is a global electronic marketplace for buying and selling securities, particularly known for listing technology and growth-oriented companies. It provides a platform for trading a wide range of stocks, including those from financial sectors such as banking and insurance.

In addition, to facilitate a meaningful comparison between our portfolios and the benchmark, we selected companies with betas close to 1. This approach aims to eliminate significant deviations based on sensitivities, focusing our analysis on factors other than what we are investigating.

5.2 ESG dataset

To gather data, we utilized various financial web databases. Initially, we received a list of companies from Sanctify's database. From this initial list, we selected approximately 5000 companies that had ESG scores available from 2015 onwards. Using an API, we conducted web scraping on different JSON files to obtain ESG scores.

Subsequently, leveraging the Finnhub Stock API (finnhub.io) company and its API, we obtained tickets for these selected companies. Unfortunately, due to this process, the dataset size decreased to 1000 companies.

After employing the Yahoo Finance Python library to calculate beta for the initial 1000 companies, we narrowed down the dataset to 550 companies whose betas fell within the range of 0.5 to 1.5.

With assistance from finnhub.io, we extracted industry information for each company to facilitate the selection of companies exclusively from the banking, financial services, and insurance sectors. As a result of our selection criteria, we identified a final set of 56 companies with an average beta ranging between 0.7 and 1.3. The average beta for our portfolio stands at 1.03.

Ultimately, utilizing the ESG API, we successfully obtained ESG scores for each of the 56 selected companies on a monthly basis spanning from 2015 to 2020. Subsequently, we computed the moving averages of the ESG scores for both 3-month and 6-month periods for each year.

5.3 Financial data

Initially, we intended to calculate financial ratios directly from financial statements. However, due to limited access to freely available financial information, we opted to acquire datasets from finnhub.io containing pre-calculated financial ratios. Subsequently, we conducted web scraping to obtain selected ratios based on annual reports spanning from 2015 to 2020.

The chosen financial indicators aim to capture a company's fundamental conditions and aid in predicting stock prices [14, 31]. Similar to previous studies [9, 35], we selected the Price-to-Sales ratio (P/S), Price-to-Book ratio (P/B), and Price-to-Equity ratio (P/E) to assess the reasonableness of current stock prices. Additionally, we included the Current Ratio (CR) and Quick Ratio (QR) to evaluate short-term liquidity. For profitability assessment, we considered Earnings-per-Share (EPS), Return-on-Assets (ROA), Return-on-Equity (ROE), and Net Profit Margin (NPM). Finally, the Debt-to-Equity ratio (D/E) was included to evaluate the long-term financial condition of the company.

Stock market returns were sourced from Alfa Vantage (alphavantage.co) using a web scraping method similar to previous data collection procedures. Utilizing the Stock Market API provided by this financial service, we gathered daily open and close prices for all selected companies. Following this, we performed data engineering by calculating the three-year return for each company, defined as the logarithmic ratio of stock return (log return) using the formula.

To calculate the expected market return based on financial ratios gathered for time t we will use the formula:

- r_t represents the return at time t.
- $P_{\text{close},t+3}$ represents the closing stock price at time $t+3$.
- $P_{\text{open},t+1}$ represents the opening stock price at time $t+1$.

So, the formula for the return r_t can be described as:

$$r_t = \log\left(\frac{P_{\text{close},t+3}}{P_{\text{open},t+1}}\right) \tag{5.1}$$

This equation calculates the logarithmic ratio of the closing stock price at time $t+3$ to the opening stock price at time $t+1$, providing the return at time t.

For example, the calculation (5.2) represents the logarithmic ratio of the closing price at the end of the third year (e.g., 2018) to the opening price at the beginning of the following year (e.g., 2016), reflecting the expected market return over that period:

$$r_{2015} = \log\left(\frac{P_{\text{close},2018}}{P_{\text{open},2016}}\right) \tag{5.2}$$

The relationship between the log return r_{t+1} of the following period and the feature vector X_{t+1} can be described as:

$$r_{t+1} = f(X_{t+1}) + \epsilon_{t+1} \tag{5.3}$$

Here's a breakdown:

- r_{t+1} represents the log return of the next period.
- X_{t+1} represents the feature vector at time $t+1$ containing relevant information for prediction.
- $f()$ represents the mapping or predictive function relating features X_{t+1} to log return r_{t+1}.
- ϵ_{t+1} represents the error term at time $t+1$, capturing the difference between the actual log return and the predicted log return.

6 MACHINE LEARNING THEORY

Machine learning authorities Tom M. Mitchell and Michael I. Jordan articulated the essence of machine learning in a 2015 publication in Science. They stated that machine learning involves creating computers that enhance their performance automatically with experience [19]. This dynamic field is witnessing rapid growth, positioned at the confluence of computer science and statistics, and forms the foundation of artificial intelligence and data science. Their observation from 2015 is even more relevant today, indicating the escalating interest and relevance of the field [32].

Supervised learning

It's a machine learning paradigm where the algorithm learns from labeled training data, which means the input data is paired with the correct output. The goal of supervised learning is to learn a mapping from input variables to output variables. The labeled data serves as examples for the algorithm to learn from, allowing it to make predictions or decisions when new, unseen data is presented [8].

Unsupervised learning

It's a machine learning paradigm where the algorithm learns patterns and structures from unlabeled data, meaning there are no predefined output labels provided. The goal of unsupervised learning is to explore and discover hidden patterns or representations in the data without explicit guidance [22].
The focus of this thesis are models that all fit under the umbrella-term called supervised learning.

6.1 Linear Regression & Improvements

In regression, the goal is to develop a model capable of generating continuous, real-valued outputs in response to one or multiple inputs [11].

Linear regression is known for its simplicity and interpretability in modeling relationships. We describe the relationship between log return r_{t+1} and feature vector X_{t+1} using the error term ε_{t+1} in the equation:

$$r_{t+1} = \alpha + X_{t+1}\beta + \varepsilon_{t+1} \tag{6.1}$$

The cost function is measured thought the residual sum of squares (RSS):

$$RSS = \sum_{t=1}^{n}(r_{t+1} - f(X_{t+1}))^2 \tag{6.2}$$

To prevent overfitting we introduce regularization term, Lasso (L1) and Ridge (L2), into the cost function:

$$RSS_{Lasso} = RSS + \lambda \sum_{j=1}^{p} |\beta_j| \tag{6.3}$$
$$RSS_{Ridge} = RSS + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{6.4}$$

Here, β represents the coefficient vector, p is the number of features, and λ controls the regularization strength. Lasso helps in feature selection by eliminating insignificant features due to its sparse nature.

6.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) with Radial Basis Function (RBF) kernel is effective for high-dimensional training data [2, 10]. SVR relies on a subset of training data as it disregards data points near the model prediction. The RBF kernel is particularly useful when the decision boundary is nonlinear, and it offers easy parameter tuning. Given that linear regression has already been implemented, SVR is employed to handle non-linearly predictable data.

6.3 Random Forest

Random Forests [5,23], is an ensemble method that combines bagging [6], decision trees and random attribute selection. Regression trees are utilized to build the Random Forest model, employing RSS as the cost function. The primary advantage of Random Forest is its ability to mitigate overfitting, making it suitable for large feature spaces and reducing out-of-sample variance while maintaining flexibility.

6.3.1 Bagging

Bagging [6], short for Bootstrap Aggregating, is an ensemble learning method used in machine learning to improve the stability and accuracy of models, especially decision trees and other high-variance models. Here are the key points about bagging:

Bootstrap Sampling: Bagging works by creating multiple subsets of the original dataset through bootstrap sampling. Bootstrap sampling involves randomly selecting samples with replacement from the original dataset to create new subsets of data. Each subset is of the same size as the original dataset but may contain duplicate instances and miss some original instances due to sampling with replacement.

Model Training: Bagging involves training multiple base models (often the same type of model) on each bootstrap sample. For example, in the case of decision trees, multiple decision tree models are trained on different bootstrap samples of the data.

Parallel Training: The training of these models is typically done in parallel, as each model is independent of the others. This parallelization makes bagging computationally efficient.

Aggregation: Once the base models are trained, bagging aggregates their predictions to make a final prediction. For regression tasks, this aggregation is often done by averaging the predictions of all models.

Reduced Variance: The key benefit of bagging is that it reduces variance and helps to alleviate overfitting, especially in high-variance models. By training multiple models on different subsets of data and aggregating their predictions, bagging reduces the impact of outliers and noisy data points that can lead to overfitting in individual models.

6.4 XGBoost (eXtreme Gradient Boosting)

XGBoost is an advanced implementation of gradient boosting algorithms designed for speed and performance. It is highly efficient, scalable, and widely used in machine learning competitions and industry applications due to its accuracy and flexibility.

XGBoost works by sequentially building an ensemble of decision trees [27], where each subsequent tree corrects errors made by the previous ones. It optimizes a differentiable loss function to minimize prediction errors during training.

6.4.1 Boosting

Boosting is an ensemble learning technique that combines multiple weak learners (often decision trees) sequentially to create a strong learner. Unlike bagging methods like Random Forest where models run independently in parallel, boosting methods incrementally build a strong model by focusing on instances that previous models have misclassified or have high errors.

6.5 Evaluation Metrics

For model evaluation, we utilize Mean Square Error (MSE) as the loss function, which is commonly used in regression tasks:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(r_{t+1} - f(X_{t+1})\right)^2} \tag{6.5}$$

where:

- n is the number of training data points,
- r_{t+1} is the true value of stock return at time step $t+1$,
- $f(X_{t+1})$ is the predicted value of stock return at time step $t+1$ based on the feature vector X_{t+1}.

6.6 Feature Importance techniques

Feature importance refers to the process of determining which features (variables) in a dataset have the most significant impact on the target variable or outcome in a predictive model. It helps in understanding the relative importance of different features in making predictions, identifying key drivers of the target variable, and potentially improving model performance. Also feature selection is crucial for optimizing machine learning models, ensuring they focus on relevant information and avoid overfitting.

Here are some common methods used to assess feature importance:

Correlation Analysis: Assessing the correlation between features and the target variable or between features themselves can provide insights into feature importance. Highly correlated features with the target are often important.

Test p-value and Holm's significance test: used to evaluate the statistical significance of features and determine which ones are most relevant for a predictive model.

- Test p-value: The test p-value is a measure of the probability that the observed data would occur if the null hypothesis were true. In feature selection, it helps determine whether there is a statistically significant relationship between each feature and the target variable.

  Typically, Null Hypothesis (H0): The feature has no significant impact on predicting the target variable, and any observed relationship or importance is due to random chance or noise in the data. And lower p-values suggest stronger evidence against the null hypothesis (i.e., the feature is irrelevant), while higher p-values indicate weaker evidence against the null hypothesis (i.e., the feature may be relevant).

- Holm's significance test: Holm's method is a multiple testing correction technique used to adjust p-values when conducting multiple statistical tests simultaneously, such as in feature selection where you are evaluating many features. The idea behind Holm's method is to control the family-wise error rate (FWER), which is the probability of making at least one false discovery (Type I error) among all the hypotheses tested.
  Features with adjusted p-values below the significance level are considered statistically significant and may be selected for inclusion in the predictive model.

L1 Regularization (Lasso): In linear models, L1 regularization penalizes the absolute size of coefficients, driving less important features' coefficients to zero. The non-zero coefficients indicate important features.

Decision Trees and XGBoost: Decision tree-based models like Random Forest and Gradient Boosting Machines (GBM) provide feature importance scores based on how frequently a feature is used to make decisions across all trees in the ensemble. Features used at higher nodes in the trees or features that result in significant information gain are considered more important.

Recently, in a study by Ubaid Ahmed et al. boosting algorithms were explored for their effectiveness in feature selection applications [1]. The study focused on solar irradiance forecasting, where accurate predictions are essential for efficient solar energy harvesting.

The study compared boosting algorithms against traditional methods using metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Surprisingly, the model trained on features selected by the XGBoost model outperformed conventional models.

This finding highlights the importance of incorporating advanced techniques like boosting algorithms, especially in domains where feature selection plays a critical role. Boosting techniques not only improve predictive accuracy but also help identify the most relevant features for a given task, leading to more efficient and accurate machine learning models.

Each method has its strengths and weaknesses, and the choice of method depends on the dataset, model type, and specific goals of the analysis. It's often beneficial to use multiple methods to gain a robust understanding of feature importance. However, relying heavily on the results of feature importance derived from applying the XGBoost model is a strategic approach to evaluate the significance of our ESG score on market return.

This process helps us understand the impact of ESG considerations on investment outcomes and guides decision-making in constructing portfolios or making investment choices.

6.8 Data Normalization

Data normalization is a preprocessing technique used to rescale data to a standard range or distribution, which is typically done to ensure that the data features have similar scales or distributions. The goal is to bring all features to a common scale without distorting differences in the ranges of values. Normalization helps improve the performance and stability of machine learning models by preventing features with large scales from dominating those with smaller scales. There are two common methods for data normalization:

1. **Min-Max Scaling:** This method scales the data to a fixed range, often between 0 and 1. It is calculated using the formula:

$$X_{\text{normalized}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{6.6}$$

where X is the original value, X_{min} is the minimum value in the feature, and X_{max} is the maximum value in the feature.

2. **Standardization (Z-score Normalization):** This method rescales data so that it has a mean of 0 and a standard deviation of 1. It is calculated using the formula:

$$X_{\text{standardized}} = \frac{X - \mu}{\sigma} \tag{6.7}$$

where X is the original value, μ is the mean of the feature, and σ is the standard deviation of the feature.

Normalization is typically more appropriate when the distribution of the data does not follow a Gaussian (normal) distribution. It scales the data based on the range of values, bringing all features to a similar scale within a specific range, such as [0, 1] or [-1, 1].

Standardization, on the other hand, assumes that your data follows a Gaussian distribution. It transforms the data to have a mean of 0 and a standard deviation of 1, resulting in a distribution that is centered around 0 with unit variance. This technique is beneficial for algorithms that assume normally distributed data, such as linear regression, logistic regression, and support vector machines (SVM).

6.9 Missing data imputation

Mean imputation is suitable when your data is normally distributed or approximately normally distributed. Filling missing values with the mean preserves the overall mean of the dataset but can slightly impact the variance, especially if the data has outliers.

Median Imputation is more robust against outliers compared to mean imputation. It is suitable when your data contains outliers or is skewed, making the mean less representative of the central tendency. Filling missing values with the median preserves the median of the dataset and is less sensitive to extreme values or outliers. If your data is skewed or has a non-normal distribution, median imputation may be more appropriate to maintain the data's inherent distribution.

7 DATA PREPARATION, FEATURE IMPORTANCE AND MODELS CREATION

7.1 Methods

In this section, we initiate our analysis by cleaning and preparing the data. Next, we explore various methods of feature selection to understand the impact of the ESG score on our target variable, which is market return. Subsequently, we apply six distinct machine learning algorithms to model this relationship, drawing insights from [9, 35]. Among these algorithms, three are based on linear regression techniques (including Ridge and Lasso), while the remaining three encompass XGBoost, Random Forest (RF), and Support Vector Regression (SVR). By employing these diverse models, we aim to capture data patterns from different angles, enabling us to identify the most effective model based on the lowest RMSE observed on the validation set. Below are the detailed steps involved in this process:

7.2 Software and libraries

The analysis of this thesis was performed in Python 3.9.7. The Numpy and Pandas libraries were chosen for basic mathematical functions and data structuring. For machine learning, pre-made predictors from Scikit-learn were utilized for all ML algorithms.

7.3 Data Preparation

As long as data cleaning and normalization play pivotal roles in the preparation of data for machine learning tasks. Several important steps were undertaken before initiating any analysis:

7.3.1 Data cleaning and handling missing values

We check for missing values in the financial ratios data for each company and impute missing values using techniques such as median imputation to keep the original distribution and to eliminate a lot of zero values after performing a standardization.

7.3.2 Normalization

Then we normalize the values of financial ratios to a standard scale. After filling in missing values with the median, applying standardization (Z-score normalization) is a suitable choice. It centers the data around 0 and scales it to have a standard deviation of 1, which aligns well with the assumptions of many machine learning algorithms, particularly those that assume normally distributed data. we use standardization (scaling data to have a mean of 0 and standard deviation of 1). Scaling ensures that different features contribute equally to the model training process.

Note that we apply data cleaning and normalization techniques separately for each company. This ensures that data preprocessing steps are tailored to each company's financial data characteristics. Group financial ratios data by company symbol and apply cleaning and normalization techniques within each group.

After those steps we ensure that the data is in a suitable format for training machine learning models. This process helps improve model accuracy and generalization on unseen data.

7.4 Feature Importance

While the primary aim of conducting feature importance analysis is to evaluate the significance of all variables on a target variable and select appropriate variables for our model, the core objective is to determine the presence and strength of the influence that our ESG score has on market return.

7.4.1 Correlation matrix

To initiate the feature importance process, we began by analyzing a correlation matrix to assess the relationships between variables and their impact on our target variable, Log return. Our analysis revealed strong correlations among several features. Consequently, the decision was made to eliminate redundant features to prevent multicollinearity issues. Multicollinearity occurs when two or more features in a dataset are highly correlated, leading to unstable and inaccurate model predictions. By removing redundant features, we can enhance the model's interpretability, stability, and predictive performance.

As a solution, we eliminated, combined, and even derived new variables through summation and multiplication to address multicollinearity.

The correlation matrix indicates that the variables Price-to-Earnings ratio (P/E ratio) and Price-to-Book ratio (P/B ratio) exhibit the strongest correlation with our target variable. The ESG score shows a moderate level of correlation strength in comparison.

7.4.2 Test p-value and Holm's significance test

Features with low p-values (close to zero), such as a constant, Price-to-Book ratio (P/B ratio), and Earnings Per Share - Return on Equity Interaction variable suggest that these features are statistically significant in relation to your target variable (market return or Log return).

Features with intermediate p-values (between 0 and 0.5), like a Price-to-Earnings ratio (P/E ratio), FCF Margin, ESG score, and Debt-to-Equity - Debt-to-Asset Interaction variable indicate a bit weaker evidence against the null hypothesis. These features may not have a strong statistical impact on a target variable.

Note: a significant constant suggests that there is a meaningful baseline effect on the target variable even when all other predictors are absent or have zero effect.

Debt-to-Equity - Debt-to-Asset Interaction is a derived variable obtained by multiplying the values of the Debt-to-Equity (D/E) ratio and the Debt-to-Asset (D/A) ratio. And Earnings Per Share - Return on Equity Interaction by multiplying the values of the EPS and the ROE.

7.4.3 Lasso test

Based on the Lasso test results, it is observed that only the coefficient for the Sales-Per-Share has been reduced to zero. This suggests that this variable doesn't significantly contribute to the model's predictive capability for the target variable. Therefore, we can conclude that this variable may be excluded from further analysis or model building processes due to their lack of impact on the target variable. ESG score has not appeared.

7.4.4 Decision Tree, Random Forest and XGBoost

In this section we compare the feature importance results from the Decision Tree Regressor, Random Forest Regressor (RF), and XGBoost Regressor:

Feature	Importance (RF)	Importance (DT)	Importance (XGBoost)
Price-to-Book ratio	0.46	0.38	0.39
ESG score	**0.11**	**0.20**	**0.21**
Price-to-Earnings	0.10	0.21	0.17
D/E-D/A interaction	0.08	0.04	0.07
Book value	0.11	0.13	0.06
Return-on-total-Capital	0.03	0.003	0.03
Price-to-Sales	0.01	0.01	0.03
Price to Free Cash Flow	0.07	0.03	0.03

All three models prioritize Price-to-Book ratio, Price-to-Earnings ratio, ESG score, and Book Value as important features for predicting market return. Debt-to-Equity - Debt-to-Asset Interaction also consistently appears among the top features across all models.

The consistent agreement among various tree-based models regarding essential features highlights the reliability of these financial metrics in forecasting market returns. Notably, the ESG score emerged as a highly significant feature, ranking second in importance for predicting market trends. In summary, the collective analysis underscores the significance of the ESG score in anticipating market returns.

7.5 Models creation and making predictions

Firstly, we have divided our dataset into training, validation, and test sets. The training and validation sets cover the years 2015 to 2019, while the test set includes data from 2020. We established a time series cross-validation (CV) framework with specified parameters like the number of splits and training years to ensure robust evaluation across all time periods.

We defined a set of regression models along with their hyperparameters for tuning using GridSearchCV. Employing TimeSeriesSplit for cross-validation, we segmented the data into year groups. Each model underwent training using cross-validation on various combinations of training years, optimizing hyperparameters based on negative root mean squared error (RMSE).

The regression models used for training and validation are as follows:

- Linear Regression
- Random Forest Regressor
- Support Vector Regression (SVR)
- XGBoost Regressor
- Ridge Regression
- Lasso Regression

These models were fine-tuned with specific parameter grids to enhance their predictive performance. Through this structured approach, we aim to identify the most suitable model for predicting market returns effectively.

Those hyperparameters were considered for each model:

model	params
LinearRegression	{}
RandomForestRegressor	{'n_estimators': [50, 100, 200], 'max_depth': [None, 10, 20]}
SVR	{'kernel': ['linear', 'rbf'], 'C': [0.1, 1, 10]}
XGBRegressor	{'n_estimators': [50, 100, 150], 'max_depth': [3, 5, 7], 'learning_rate': [0.1, 0.01]}
Ridge	{'alpha': [0.1, 1, 10, 50]}
Lasso	{'alpha': [0.1, 1, 10, 50]}

We identified the best-performing model by evaluating the average Root Mean Squared Error (RMSE) across all folds and combinations of training years.

Following this, we created **two distinct datasets**: one with the ESG score feature included and another without it. Both datasets underwent separate training processes, maintaining the methodology described earlier.

The optimal model was trained using data from 2015 to 2019 and its performance was assessed on the 2020 test dataset to compute the RMSE.

Ultimately, we identified the best models for each dataset using performance metrics from the validation sets. These best models were then used to predict outcomes on the respective test datasets for comparison. The comparison aimed to evaluate the impact of including or excluding the ESG score feature on model predictions, providing valuable insights into the predictive power of ESG scores within our models.

Final Model Training and Testing:

After rigorous evaluation, the model with the lowest RMSE on the validation set for dataset **included ESG score** was determined to be the **Support Vector Regressor (SVR)**:

- Best Model: SVR
- Best Model Parameters: {'C': 0.1, 'kernel': 'linear'}
- Average RMSE of Best Model is **0.1895**
- Final RMSE on Test Set (year 2020) is **0.2673**

The model with the lowest RMSE on the validation set for dataset **excluded ESG score** was determined to be the **Ridge**:

- Best Model: Ridge
- Best Model Parameters: {'alpha': 10}
- Average RMSE of Best Model is **0.1857**
- Final RMSE on Test Set (year 2020) is **0.2113**

Final Prediction:

The SVR and Ridge models were employed to predict log returns on the 2020 features, and symbols were ranked accordingly based on these predicted log returns for each dataset.

This structured approach ensured that our model selection was rigorous and that the final predictions were based on the most accurate model identified during the validation process.

8 PORTFOLIOS CREATION AND RESULT

After completing the final predictions and organizing companies by their predicted log_returns, we determined the top symbols exhibiting the highest positive predicted returns under two distinct scenarios: one incorporating ESG scores and the other without ESG scores.

Following this, we arranged the companies based on their Log_return and selected only the top-performing companies to establish two portfolios: one (portfolio1) constructed using a model that incorporates ESG score information, and another (portfolio2) relying solely on financial data without ESG scores. The number of top companies selected was based on positive ESG scores; specifically, we included the first N companies with positive ESG scores. The process ceased once a negative ESG score was encountered. The same number of companies were selected for portfolio2.

Next, we obtained historical prices for these portfolios and the benchmark S&P 500 index (^GSPC) using Yahoo Finance (yfinance) for the period from 01.01.2021 to 01.01.2024 to analyze their real-world performance.

Daily returns were calculated for both portfolios and the benchmark, which were then used to derive cumulative monthly returns for each portfolio and the benchmark.

Figure 1 illustrates the cumulative monthly returns of the portfolios against each other and the benchmark S&P 500, providing a visual comparison of their performance over time.

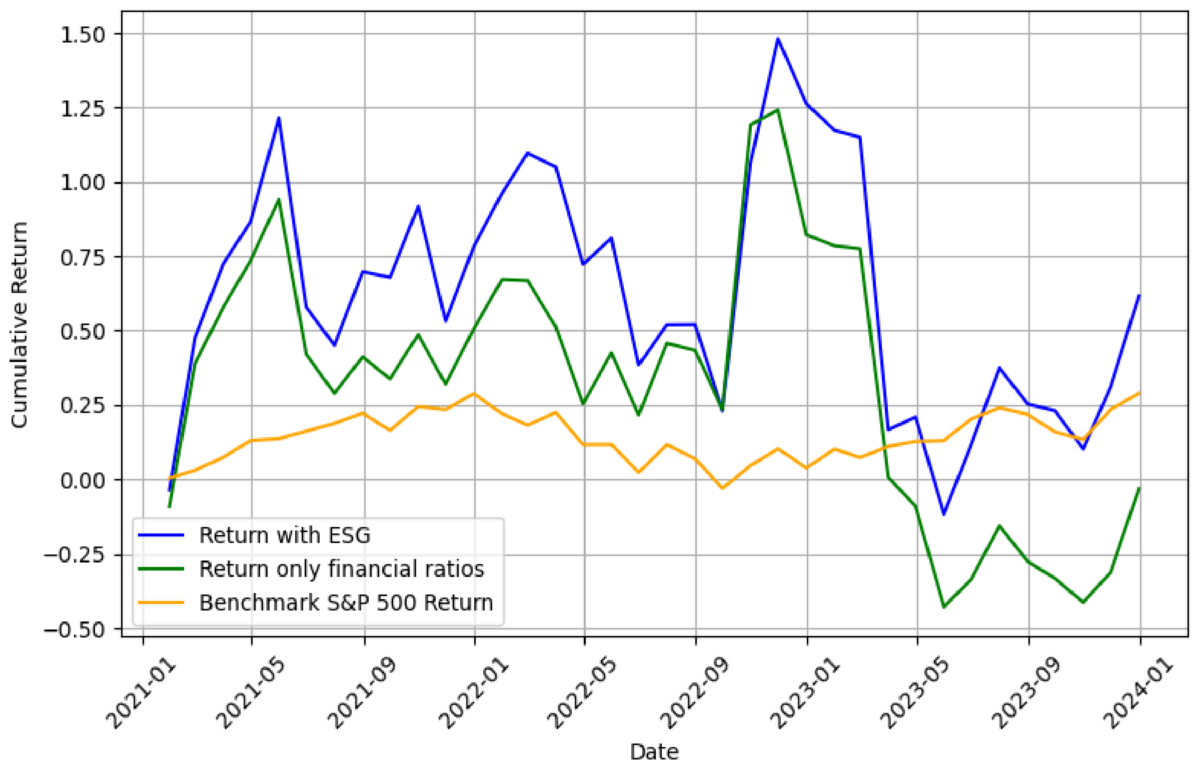

Figure 1: Cumulative Monthly Return of Portfolios vs Benchmark (S&P 500)

In Figure 1, we can observe that the cumulative return of the portfolio constructed using ESG knowledge tends to be slightly higher than that created without ESG scores in the model over the long run. An intriguing discovery is that during the initial two years, both portfolios significantly outperformed the benchmark S&P 500. However, in the last year, the cumulative return for the portfolio 2 without ESG score considered dropped below that of the S&P 500.

9 CONCLUSION

Efforts and policies aimed at ESG objectives may not be captured by conventional financial metrics, yet they significantly influence a company's business growth and stock performance. Leveraging ESG components effectively can yield substantial benefits in investments. ESG data serves as a reliable reflection of sustainable business practices and provides a meaningful gauge of a company's ESG premium.

In this study, we leverage ESG data to create alpha signals using practical machine learning techniques. Our ESG alpha strategy outperforms benchmark indexes and baseline portfolios constructed solely on traditional financial indicators in terms of cumulative returns over a long-term investment horizon (3 years). These findings underscore the potential of Sanctify's ESG data to expand traditional investment strategies and enhance profitability by deriving ESG alpha insights from the data.

Moving forward, we aim to enrich the data feature space and extend the training periods, possibly on a monthly or quarterly basis. Additionally, exploring sectors beyond banking and financial institutions could provide comprehensive insights. With more extensive data, implementing Deep Recurrent networks can lead to improved predictions and overall performance enhancements.

Furthermore, our analysis extends beyond conventional methods by initially filtering companies with a beta coefficient equal to 1 from the CAPM model. This approach allows us to draw a crucial conclusion: our portfolio's outperformance against the benchmark S&P 500 is not solely attributable to companies with beta coefficients exceeding 1. Instead, it underscores a distinct factor—our meticulous consideration and integration of ESG scores. It is this incorporation of ESG data that has contributed to our portfolio's alpha, representing an excess return derived from ESG-related insights.

BIBLIOGRAPHY

[1] Ahmed, U., Mahmood, A., Tunio, M. A., Hafeez, G., Khan, A. R., & Razzaq, S. (2024). Investigating boosting techniques' efficacy in feature selection: A comparative analysis. *Energy Reports, 11*, 3521-3532. https://doi.org/10.1016/j.egyr.2024.03.020

[2] Awad, Mariette and Khanna, Rahul. 2015. Support vector regression. In Efficient Learning Machines. Springer, 67–80.

[3] Bajic, Stevan and Yurtoglu, Burcin. Which aspects of csr predict firm market value? Journal of Capital Markets Studies, 2018.

[4] Basel Committee on Banking Supervision (2022), Principles for the effective management and supervision of climate-related financial risks. Retrieved 2 May, Available at: https://www.bis.org/bcbs/publ/d532.pdf)

[5] Breiman, Leo. Random Forests. Machine Learning, 45(1):5–32, October 2001.

[6] Breiman, Leo. Bagging predictors. Machine Learning, 24(2):123–140, August 1996.

[7] Chen, Q., & Liu, X.-Y. (2020). Quantifying ESG alpha using scholar big data: an automated machine learning approach. In Proceedings of the First ACM International Conference on AI in Finance (ICAIF '20) (pp. 1–8). https://doi.org/10.1145/3383455.3422529

[8] Dixon, Matthew; Halperin, Igor and Bilokon, Paul. Machine Learning in Finance: From Theory to Practice. Springer International Publishing, Cham, 2020.

[9] Fang, Y; Liu, X and Yang, H. 2019. Practical Machine Learning Approach to Capture the Scholar Data Driven Alpha in AI Industry. In 2019 IEEE International Conference on Big Data (Big Data). IEEE, 2230–2239.

[10] Fletcher, Tristan. Support Vector Machines Explained.

[11] Friedman, Jerome; Hastie, Trevor and Tibshirani, Robert. 2001. The elements of statistical learning. Vol. 1. Springer series in statistics New York.

[12] Friedman, Milton. The Social Responsibility of Business Is to Increase Its Profits. In Walther Ch Zimmerli, Markus Holzinger, and Klaus Richter, editors, Corporate Ethics and Corporate Governance, pages 173–178. Springer, Berlin, Heidelberg, 2007.

[13] Galetta, S., Goodell, J.W., Mazzù, S., & Paltrinieri, A. (2022). Bank reputation and operational risk: The impact of ESG. Finance Research Letters, Vol 51.

[14] Gerakos, Joseph J and Robert Gramacy. 2013. Regression-based earnings forecasts. Chicago Booth Research Paper 12-26 (2013).

[15] Guo, T.; Jamet, N.; Betrix, V.; Piquet, L.A.; Hauptmann, E. ESG2Risk: A deep learning framework from ESG News to stock volatility prediction. arXiv 2020, arXiv:2005.02527.

[16] Gupta, A.; Sharma, U.; Gupta, S.K. The role of ESG in sustainable development: An analysis through the lens of machine learning. In Proceedings of the 2021 IEEE International Humanitarian Technology Conference (IHTC), Virtual, 2–4 December 2021.

[17] Howard-Grenville, Jeniffer. Esg impact is hard to measure-but it's not impossible. 2021

[18] Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. Journal of Finance, 23(2), 389–416.

[19] Jordan, M and Mitchell, T. Machine learning: Trends, perspectives, and prospects. Science, 349(6245):255–260, July 2015.

[20] Koller, Tim, Nuttall, Robin, and Henisz, Witold. Five ways that esg creates value. The McKinsey Quarterly, 2019.

[21] KPMG International (2022), Survey of Sustainability Reporting 2022 Retrieved 2 May, Available at:
https://assets.kpmg.com/content/dam/kpmg/se/pdf/komm/2022/Global-Survey-of-SustainabilityReporting-2022.pd

[22] Kubat, Miroslaw. An Introduction to Machine Learning. Springer International Publishing, Cham, 2021.

[23] Liaw, Andy; Wiener, Matthew, et al. 2002. Classification and regression by random Forest. R news 2, 3 (2002), 18–22.

[24] Lintner, J. (1965a). "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." The Review of Economics and Statistics 47, 13–37.

[25] Lintner, J. (1965b). "Securities Prices, Risk, and Maximal Gains from Diversification." The Journal of Finance 20(4), 587–615

[26] Lintner (1965a), p. 25, and Mossin (1966), pp. 775–776.

[27] Loh, W-Y. Classification and regression trees. WIREs Data Mining and Knowledge Discovery, 1(1):14–23, 2011.

[28] Margot, V.; Geissler, C.; De Franco, C.; Monnier, B. ESG Investments: Filtering Versus Machine Learning Approaches. Appl. Econ. Finance. 2021, 8, 1.
[29] McKinsey's (2022), Does ESG really matter—and why? Retrieved 2 May, Available at:
https://www.mckinsey.com/capabilities/sustainability/our-insights/does-esg-really-matter-and-why

[30] Nilsson, A and Robinson, David. What Is the Business of Business? Innovation Policy and the Economy, 18:79–106, January 2018. Publisher: The University of Chicago Press.

[31] Piotrowski, Joseph D. 2000. Value investing: The use of historical financial statement information to separate winners from losers. Journal of Accounting Research (2000), 1–41.

[32] Pugliese, Raffaele; Regondi, Stefano and Marini, Riccardo. Machine learning based approach: Global trends, research directions, and regulatory standpoints. Data Science and Management, 2021

[33] Sharma, U.; Gupta, A.; Gupta, S.K. The pertinence of incorporating ESG ratings to make investment decisions: A quantitative analysis using machine learning. J. Sustain. Finance. Investig. 2022, 2021, 2013151.

[34] Sharpe, W. F. (1964). "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." The Journal of Finance 19(3), 425–442.

[35] Yang,Y; Liu, X and Wu, Q. 2018. A practical machine learning approach for dynamic stock recommendation. In 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). IEEE, 1693–1697

[36] Yu, G.; Liu, Y.; Cheng, W.; Lee, C.-T. Data analysis of ESG stocks in the Chinese Stock Market based on machine learning. In Proceedings of the 2022 2nd International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, 14–16 January 2022.

[37] Zabarankin, M., Pavlikov, K., & Uryasev, S. (2014). Capital Asset Pricing Model (CAPM) with drawdown measure. *European Journal of Operational Research*, 234(2), 508-517. ISSN 0377-2217. https://doi.org/10.1016/j.ejor.2013.03.024.