Deliverable 2 Problem understanding and work plan

Reza Nosrati, and Prabhakar Karna

Master of Science (Business Analytics) / University of North Florida

CEN6940 Computing Practicum / School of Computing

Date 10/06/2024

Table of Contents

Index Terms—Data mining, machine learning, Prediction of price, ESG, Sustainability market, ROE, Regression, LSTM,

Problem Statement

The project addresses the uncertainty surrounding the impact of Environmental, Social, and Governance (ESG) factors on corporate profitability. Despite the common reporting of ESG metrics, the direct correlation between high ESG performance and financial returns is not well understood. This ambiguity persists despite growing investor demand for transparency, complicating investment decisions related to sustainability.

Sustainable investing is a burgeoning field as investors increasingly prefer companies with robust ESG metrics. However, the lack of concrete data linking sustainable practices to positive financial returns causes hesitation. This project is crucial for providing clear insights into the profitability of ESG practices, fostering more informed decision-making among investors, corporations, and regulators.

This issue is complex due to the need to integrate diverse financial data with varied sustainability metrics across different industries and regions. The development of predictive models to accurately link ESG scores with financial performance involves sophisticated data analytics and machine learning techniques to effectively capture and analyze these trends.

It is recognized that ESG factors are increasingly significant to investors, with numerous studies indicating a positive correlation with long-term financial performance. However, the specific nature and extent of this correlation, particularly from a forward-looking perspective, remain elusive. Our project seeks to clarify these uncertainties by developing predictive models that offer actionable insights on how ESG scores impact financial metrics.

The project revolves around ESG scores—indicators of a company's adherence to environmental, social, and governance norms—and profitability metrics like return on equity and earnings per share. We propose to construct machine learning-based predictive models to explore how ESG performance affects these financial outcomes and to develop an interactive visualization tool for real-time analysis.

Our motivation derives from a desire to reconcile sustainability with profitability, catering to the growing demand for accurate, forward-looking financial impacts of ESG practices. The project will significantly contribute to the academic and practical understanding of sustainable investing, providing a predictive analysis that enhances decision-making for various stakeholders.

Work to be Performed

The project aims to establish a predictive framework evaluating the relationship between ESG performance and corporate profitability, encompassing data collection, model building, and visualization tool development.

Objectives

1. Establish data pipelines for efficient ESG and financial data collection and integration.
2. Develop machine learning models to predict financial impacts based on ESG performance.

3. Construct interactive visualization tools to display insights derived from the models.

Activities

1. **Data Collection and Integration**
   - Collect data from multiple reputable sources, ensuring accuracy and comprehensiveness.
   - Implement real-time data pipelines for continuous data updates.
2. **Model Development**
   - Select relevant ESG features and financial metrics.
   - Construct and fine-tune machine learning models for robust financial performance predictions.
3. **Visualization Development**
   - Create dynamic dashboards to visualize the correlation between ESG scores and financial outcomes, allowing customizable analyses through filters.
4. **Validation and Testing**
   - Employ cross-validation techniques to ensure model reliability and validity.
   - Optimize visualizations for intuitive, actionable investor insights.

Implementation Timeline

Goals are designed to guide the development of predictive models and tools that will enable investors and financial analysts to make data-driven decisions about the financial impacts of ESG practices.

Sr No	Activity ID	Activity	Start Week	Team Member(s)	Start Date	End Date	Outcome	8/23/2024	8/30/2024	9/6/2024	9/13/2024	9/20/2024	10/4/2024	10/18/2024	10/25/2024	11/1/2024	11/8/2024	11/15/2024	11/22/2024	11/29/2024	12/6/2024
1	PP1	Project Pitches & Team Formation & Research on Sustainability websites	3	Reza/Prabhakar	8/23/2024	8/29/2024	Agreement on the research Project														
2	D1	Data collection from multiple sources	4	Reza/Prabhakar	8/29/2024	9/5/2024	Raw ESG and financial data ready for preprocessing														
3	D2	Data cleaning and preprocessing	5	Reza	9/6/2024	9/12/2024	Cleaned dataset with normalized variables														
4	F1	Feature selection for machine learning	6	Prabhakar	9/13/2024	9/20/2024	Identified ESG features that significantly impact financial outcomes														
5	M1	Model development and training	7-8	Reza/Prabhakar	9/20/2024	10/4/2024	Initial predictive models built using Python (scikit-learn, TensorFlow)														
6	M2	Model tuning and optimization	9-10	Reza	10/4/2024	10/18/2024	Hyperparameters optimized for better model performance														
7	M3	Model validation and cross-validation	11	Prabhakar	10/18/2024	10/25/2024	Validated models with cross-validation; metrics like MSE and R-squared recorded														
8	V1	Visualization development in Tableau	12	Reza	10/25/2024	11/2/2024	Dashboards created to display ESG and financial performance relationships														
8	T1	Testing and refinement of models/tools	13-15	Reza/Prabhakar	11/2/2024	11/21/2024	Predictive models and visualization tools refined based on feedback from initial tests														
9	FRP1	Final review and presentation prep	14-16	Reza/Prabhakar	11/21/2024	12/10/2024	Final version of models and visualizations, ready for presentation to stakeholders														

List of Required Resources.

Software Tools

1. **Python**: Primary programming language for data processing, analysis, and model building. Python libraries such as Pandas for data manipulation, NumPy for numerical data, scikit-learn for machine learning, TensorFlow and Keras for advanced predictive modeling, and Matplotlib and Seaborn for data visualization.

2. **Tableau**: For creating interactive and shareable dashboards that visualize the relationship between ESG metrics and financial performance.

Hardware

1. **High-Performance Laptops**: With at least an Intel i7 processor or equivalent, 16GB RAM, and substantial SSD storage to handle large datasets and intensive computations.

APIs

1. **Financial Data APIs**: Yahoo website and S&P website and its API's for accessing real-time financial data, historical market data, and financial statements. The dataset is for 10-year stock market details for selected 100 companies. The data includes details such as "Open High Low Close Volume Dividends Stock Splits country revenue"

2. **ESG Data APIs**: From providers like MSCI or Sustainalytics to retrieve ESG scores and reports. Similar to financial data, 10-year Economic, Sustainability & Governance score, and ESG overall score, for Selected 100 companies.

Programming Platforms

1. **Jupyter Notebook**: For writing and testing code in an interactive environment which supports Python, and other programming languages. Below is proposed list of python function/library that will be used during data analysis and predictive modelling

Library/Function	Description
NumPy	Provides support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.
Matplotlib (pyplot)	Used for creating static, interactive, and animated visualizations in Python.
Seaborn	An extension to Matplotlib that makes it easier to generate certain types of plots and integrates well with pandas' data structures.
RandomForestRegressor (scikit-learn)	A regression tool that uses multiple decision trees to improve predictive accuracy and control over-fitting.
SimpleImputer (scikit-learn)	Provides basic strategies for imputing missing values in a dataset, such as using the mean, median, or most frequent value.
LinearRegression (scikit-learn)	Implements linear regression for fitting a linear model with coefficients to minimize the residual sum of squares between the observed targets in the dataset and the targets predicted by the linear approximation.

For visualization Tableau will be used to support below visualization. This will be used in conjunction with data procession and modeling.

Data Visualization Activity	Details	Type of Visualization
Correlation Matrix	Visualize correlations between different ESG factors and financial outcomes to identify significant relationships.	Correlation Matrix
Scatter Plot	Display the relationship between individual ESG metrics (e.g., carbon emissions) and financial outcomes (e.g., share price) to analyze potential patterns.	Scatter Plot
Time Series Analysis	Visualize how ESG scores and financial performance evolve over time to detect trends and patterns.	Line Chart, Time Series Plot
Bar Charts for Sector Comparison	Compare the ESG scores and financial performance across different sectors to see which industries have the highest or lowest ESG impact.	Bar Chart
Heatmap for Risk Assessment	Use a heatmap to highlight companies with higher ESG risks and their associated financial outcomes (e.g., low returns, high volatility).	Heatmap
Bubble Chart for Multivariate Analysis	Analyze how multiple factors (e.g., ESG score, market cap, financial risk) interact by representing them on a bubble chart with varying sizes and colors.	Bubble Chart

2. **Cloud Resources**- N/A

Datasets- the dataset will be complied in .csv format.

1. **ESG Scores**: From providers like MSCI, Sustainalytics, or Refinitiv, which offer comprehensive data on company performance in environmental, social, and governance aspects. Approx. data size is 1000c X 6r
2. **Financial Data**: Historical and current financial metrics from databases like Bloomberg, Yahoo Finance, or Google Finance, including stock prices, earnings per share, revenue data, etc. Approx. data size 1000c X 10r

   The final data after joining will look like below.

	Rank	Name	Symbol	revenue_ttm	price (USD)	country	Market Capital	ESG Score(100)
0	1	Walmart	WMT	622021000000	155.53	United States	418.81	84.0
1	2	Saudi Aramco	2222.SR	588889174131	8.59	Saudi Arabia	2078.00	NaN
2	3	Amazon	AMZN	524897000000	125.49	United States	1287.00	81.0
3	4	Sinopec	600028.SS	484144248616	0.91	China	100.61	66.0
4	5	PetroChina	601857.SS	462943680870	1.05	China	184.91	62.0

3. **Publicly Available Datasets**: Such as the datasets from the World Bank, IMF, or OECD, which could be used for macroeconomic indicators or to enrich the analysis (if required during analysis).

Miscellaneous Tools- if any will be listed in due course.

This detailed resource list ensures that all technical requirements are met for successfully completing the project, from data collection and processing to analysis, modeling, and visualization. Each tool and resource have been chosen to optimize the project's workflow and achieve the best possible outcomes in exploring the relationship between ESG performance and profitability.

This project aims to deliver predictive models and analytical tools that provide investors and companies with critical insights into the financial impact of Environmental, Social, and Governance (ESG) factors. By leveraging machine learning, data analytics, and real-time monitoring, this project will address the challenges of assessing and predicting how ESG performance influences financial outcomes. The project's outcomes will have a tangible and lasting impact on stakeholders, as outlined below.

Informed Decision-Making for Investors: Predictive models will provide forward-looking insights into how ESG initiatives impact profitability, stock performance, and risk, helping investors make data-driven decisions and optimize portfolios for better returns.

Simplified and Actionable ESG Insights for Investors: Visualization tools will offer clear, customizable ESG data, allowing investors to filter by sectors, regions, or criteria to make more informed investment choices and enhance profitability.

Risk Mitigation and Early Warning Systems: Real-time AI-powered tools will give investors and companies up-to-date ESG insights, helping them proactively address risks like regulatory changes and environmental disasters, improving financial stability and sustainability.

Long-Term Financial Stability and Growth for Companies: Predictive models will help companies identify areas for ESG improvement, leading to cost savings, operational efficiency, and revenue growth, ultimately boosting brand equity and shareholder value.

Below ML scheme is for indicative purposes only.

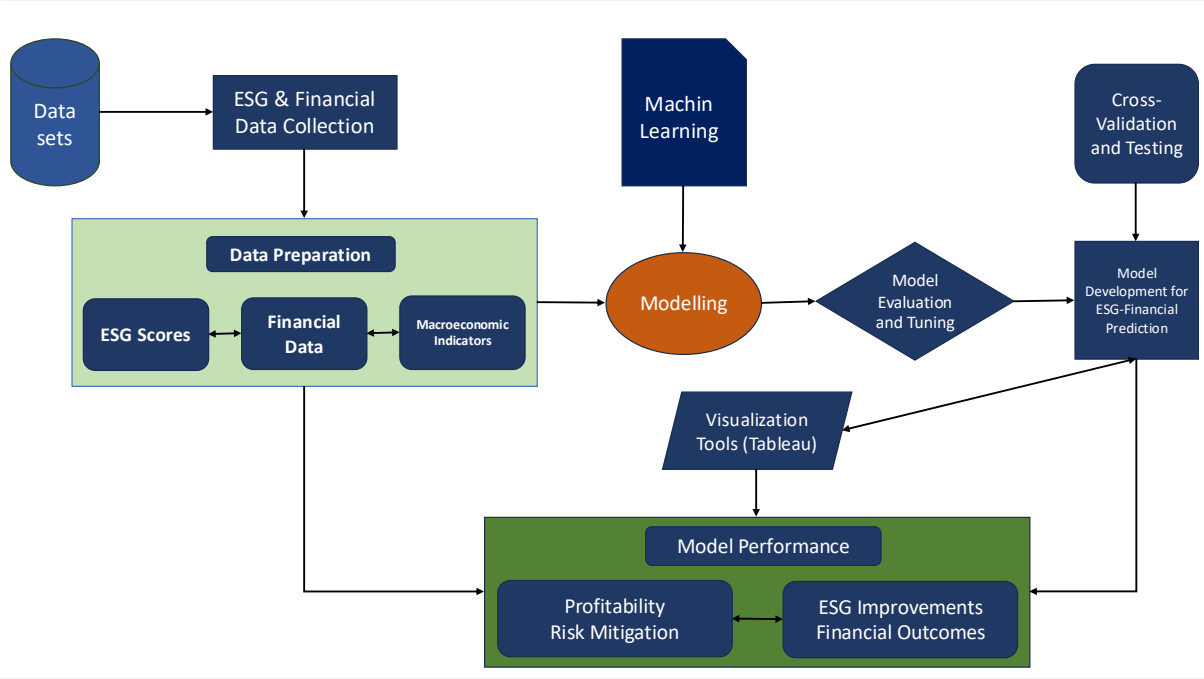

Figure-2 Predictive Framework for ESG Impact on Corporate Profitability

Bibliography

Eccles, R. G., Ioannou, I., & Serafeim, G. (2014). The impact of corporate sustainability on organizational processes and performance. Management Science, 60(11), 2835-2857. https://doi.org/10.1287/mnsc.2014.1984

Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment, 5(4), 210-233. https://doi.org/10.1080/20430795.2015.1118917

Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. The Accounting Review, 91(6), 1697-1724. https://doi.org/10.2308/accr-51383

Sustainable finance: Exploring the ESG framework. Luff, J. P., & Shimkus, D. C. (2021). Greenleaf Publishing.