IE7945 Master's Project

Literature Review Report

Predictive Analytics and Clustering of ESG Performance for Global Countries

Subtitle:
Combining Machine Learning Techniques to Forecast and Analyze ESG Trends

Author: Susu Ping
Instructor: Professor Srinivasan Radhakrishnan

Date: June 1, 2024

Section 1: Introduction to the Project

Title:
Predictive Analytics and Clustering of ESG Performance for Global Countries

Subtitle:
Combining Machine Learning Techniques to Forecast and Analyze ESG Trends

Introduction:

Environmental, Social, and Governance (ESG) criteria are pivotal in evaluating the sustainability and societal impact of companies and countries. These criteria help in assessing how an entity performs across various aspects crucial for sustainable growth and ethical operations. With the increasing importance of sustainable investment and corporate responsibility, analyzing and predicting ESG performance has become vital for stakeholders.

Objective:

The primary objective of this project is to develop a comprehensive analytical framework that leverages machine learning techniques to predict ESG scores and cluster countries based on their ESG performance. By doing so, the project aims to provide actionable insights into the future sustainability trends and categorize countries into distinct groups based on their ESG characteristics. This dual approach will help in understanding not only the future trajectory of ESG performance but also the underlying patterns that define different groups of countries.

Tasks:

1. **Predicting ESG Scores:**
    - **Objective:** Develop models to predict ESG scores for countries based on historical ESG data.
    - **Approach:** Utilize time-series forecasting techniques or regression models to analyze historical ESG indicators and forecast future scores.
2. **Clustering Countries Based on ESG Performance:**
    - **Objective:** Cluster countries into different groups based on their ESG performance.
    - **Approach:** Apply clustering algorithms like K-means or hierarchical clustering to categorize countries into distinct groups based on their ESG characteristics.

Significance:

Understanding and predicting ESG performance is crucial for policymakers, investors, and corporations aiming to foster sustainable development. Predictive analytics can offer foresight into future trends, enabling proactive decision-making. Clustering countries based on ESG performance can reveal hidden patterns and similarities, providing a nuanced understanding of global ESG dynamics.

Section 2: Annotated Bibliography

1. Gupta, A., Sharma, U., & Gupta, S. (2021). The Role of ESG in Sustainable Development: An Analysis Through the Lens of Machine Learning. *2021 IEEE International Humanitarian Technology Conference (IHTC)*, 1-5.

- **Key Insights:** This paper presents a methodology for creating a dataset that combines ESG and financial parameters of companies. It uses machine learning to analyze the importance of ESG parameters for investment decisions and their impact on financial performance.
- **Methods Used:** Linear regression, Random Forest regression.
- **Relevance:** Provides a framework for combining ESG and financial data, essential for building predictive models in ESG analysis.

2. Menten, C., Cekic, B., Atici, K. B., Camgoz, S. M., & Ulucan, A. (2024). ESG Performances of Energy Companies in OECD Countries: A Clustering Approach. In J. Thewissen, Ö. Arslan-Ayaydin, W. Westerman, & A. Dorsman (Eds.), *The ESG Framework and the Energy Industry* (pp. 87-111). Springer. https://doi.org/10.1007/978-3-031-48457-5_6

- **Key Insights:** The paper uses clustering methods to evaluate ESG performance in energy companies across OECD countries. It identifies patterns and evaluates high and low performance across different ESG pillars.
- **Methods Used:** K-means clustering.
- **Relevance:** Demonstrates the application of clustering in ESG analysis, providing insights into grouping companies based on ESG performance.

3. Jiao, J., Shuai, Y., & Li, J. (2024). Identifying ESG types of Chinese solid waste disposal companies based on machine learning methods. *Journal of Environmental Management, 360*, 121235. https://doi.org/10.1016/j.jenvman.2024.121235

- **Key Insights:** This paper explores the identification of ESG types for Chinese solid waste disposal companies using various machine learning methods. It demonstrates how machine learning can effectively categorize and analyze ESG performance.
- **Methods Used:** K-means clustering, random forest.
- **Relevance:** Provides a case study on the application of machine learning in ESG analysis, relevant for understanding how similar methods can be applied in this project for predicting ESG scores and clustering countries.

4. Rusu, Ș., Boloș, M. I., & Leordeanu, M. (2023). K-means and agglomerative hierarchical clustering analysis of ESG scores, yearly variations, and stock returns: Insights from the energy sector in Europe and the United States. *Journal of Financial Studies, 8*(Special-J), 166-180.

- **Key Insights:** This paper uses K-means and agglomerative hierarchical clustering to analyze ESG scores, yearly variations, and stock returns in the energy sector. It provides insights into clustering techniques and their application in ESG analysis.
- **Methods Used:** K-means clustering, agglomerative hierarchical clustering.

- **Relevance:** Aligns with the project's focus on clustering analysis using machine learning methods, offering practical insights and methodologies applicable to the project's goals.

5. Lee, O., Joo, H., Choi, H., & Cheon, M. (2022). Proposing an Integrated Approach to Analyzing ESG Data via Machine Learning and Deep Learning Algorithms. Sustainability, 14, 8745. https://doi.org/10.3390/su14148745

- **Key Insights:** This paper proposes an integrated approach for analyzing ESG data using machine learning and deep learning algorithms, demonstrating enhanced predictive accuracy and insightful pattern recognition in ESG datasets.
- **Methods Used:** Various machine learning and deep learning techniques including neural networks, SVM, and ensemble methods.
- **Relevance:** Directly aligns with the project's focus on using advanced machine learning techniques for ESG analysis, providing a comprehensive framework for predictive modeling and clustering.

6. Svanberg, J., Ardeshiri, T., Samsten, I., Öhman, P., Neidermeyer, P. E., Rana, T., Semenova, N., & Danielson, M. (2022). Corporate governance performance ratings with machine learning. *Intelligent Systems in Accounting, Finance and Management*, 29(1), 50–68. https://doi.org/10.1002/isaf.1505

- **Key Insights**: This paper explores the use of machine learning algorithms to evaluate corporate governance performance ratings. It demonstrates the effectiveness of machine learning in assessing corporate governance, a critical component of ESG.
- **Methods Used:** Various machine learning techniques including supervised learning algorithms.
- **Relevance:** Directly aligns with the project's focus on using machine learning to analyze ESG performance, providing valuable methodologies and insights.

Section 3: Dataset Overview and Repository Details

Dataset Overview

1. ESGData.csv:

- **Source:** Kaggle: Environment, Social, and Governance Data
- **Description:** Contains historical ESG scores and indicators for various countries.
- **Size:** 16,013 rows and 67 columns with multiple features representing different ESG indicators.
- **Attributes:**
  - **Country Name**
  - **Country Code**
  - **Indicator Name**
  - **Indicator Code**
  - **Yearly values (1960-2020)**
- **Availability:** Readily available, good quality with potential gaps in earlier years.

- **Challenges:** Handling missing values and normalizing data for consistent analysis.

2. ESGCountry.csv:

- **Source:** Kaggle: Environment, Social, and Governance Data
- **Description:** Contains additional country-specific metadata for enriched clustering analysis.
- **Size:** 239 rows and 31 columns.
- **Attributes:**
  - **Country Code**
  - **Region**
  - **Income Group**
  - **Special Notes**
- **Availability:** Readily available.
- **Challenges:** Ensuring metadata alignment with ESG scores for accurate clustering.

Repository Details

Document Repository: A local folder which contains the selected research papers and datasets is uploaded on Canvas.