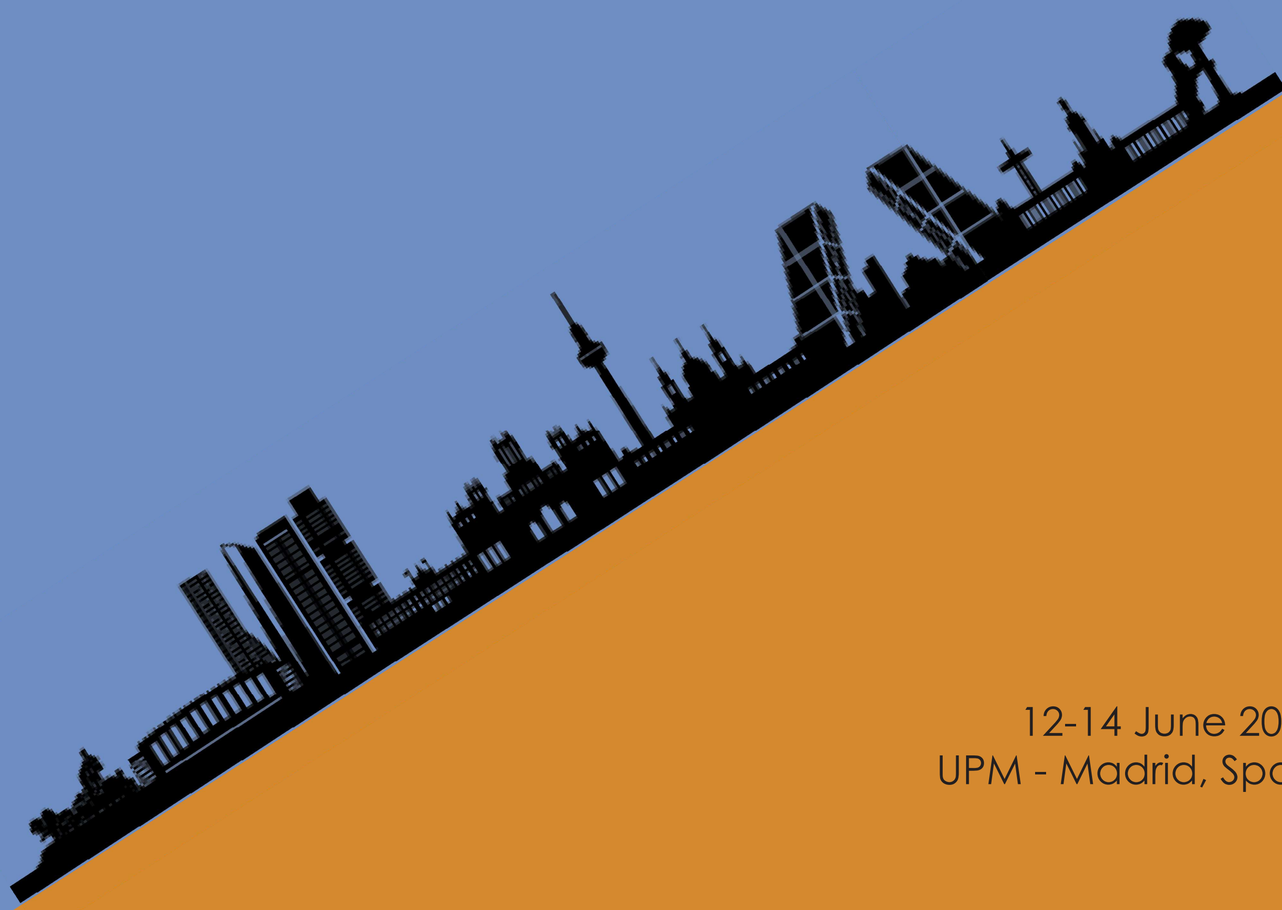
IFKAD
International Forum on Knowledge Asset Dynamics

19th International Forum on Knowledge Asset Dynamics

PROCEEDINGS

Translating Knowledge into Innovation Dynamics

IFKAD 2024



12-14 June 2024
UPM - Madrid, Spain

19th International Forum on Knowledge Asset Dynamics

PROCEEDINGS
IFKAD 2024

12-14 June 2024
Madrid - Spain

Translating Knowledge into Innovation Dynamics

Proceedings IFKAD: Translating Knowledge into Innovation Dynamics
Distribution: IFKAD 2024 – Madrid, Spain 12-14 June 2024
Polytechnic University of Madrid
Arts for Business Institute
LUM University

ISBN 978-88-96687-17-8
ISSN 2280-787X

Edited by: Gustavo Morales-Alonso, Giovanni Schiuma
Published by: Institute of Knowledge Asset Management (IKAM)
Realized by: Gabriela Jaroš

ISBN 978-88-96687-17-8 || ISSN 2280-787X

Unlocking Agenda 2030 Progress: Artificial Intelligence for SDGs' Indicator Prioritization in OECD Nations

Armando Calabrese

Department of Enterprise Engineering
University of Rome "Tor Vergata"
Rome, Italy

Roberta Costa

Department of Enterprise Engineering
University of Rome "Tor Vergata"
Rome, Italy

Nathan Levialdi Ghiron

Department of Enterprise Engineering
University of Rome "Tor Vergata"
Rome, Italy

Juan Carlos Martínez Pérez*

Department of Enterprise Engineering
University of Rome "Tor Vergata"
Rome, Italy

Luigi Tiburzi

Department of Enterprise Engineering
University of Rome "Tor Vergata"
Rome, Italy

** Corresponding author*

Abstract

As the imminent deadline of the 2030 Agenda approaches, a substantial number of its 17 Sustainable Development Goals (SDGs) remain unattained, necessitating accelerated efforts to make significant progress within the remaining six years. Since the large number of indicators, associated to the SDGs, are intertwined in complex non-linear ways, considering their actual synergies and trade-offs to prioritize them plays a key role in their successful implementation. This prioritization is essential for taking impactful and actionable steps and efficiently utilizing available resources. In particular, indicator-level analysis is of special importance as it offers a more granular perspective enabling the formulation of clear governance recommendations for stakeholders. Efforts in this field

ISBN 978-88-96687-17-8 || ISSN 2280-787X

have encounter challenges, such as the inadequacy of traditional regression methods due to the inherent correlations and complex interdependencies among the SDGs. Additionally, the scarcity of data from many countries' SDG indicators has limited the scope of previously performed analyses. This article performs an Artificial Intelligence method, a Random Forest, to address these challenges due to its ability to deal with correlated and non-linear data. Furthermore, it focuses on high-quality data from OECD countries to reduce the necessity for extensive data manipulation that might introduce uncertainties and diminish the validity of the analysis. The obtained results pinpoint income inequality as the most significant factor impacting OECD countries. This underscores income inequality's profound influence on high-income countries and its impact on other SDGs. Further key indicators include SO_2 emissions embodied in imported goods and services (kg/capita), proportion of urban population living in slums and population's access to clean fuels and cooking technology. These results showcase the multifaceted nature of sustainability challenges in OECD nations and provide new insights into key drivers for OECD members that significantly impact the attainment of the 2030 Agenda. Furthermore, the prioritized indicators found by this study will support stakeholders in developing effective resource allocation strategies.

Keywords – SDG indicator, SDG prioritization, Artificial Intelligence, OCED countries, Agenda 2030.

Paper type – Academic Research Paper

1 Introduction

The 2030 Agenda for Sustainable Development, adopted by all United Nations Member States in 2015, serves as a universal call to action to end poverty, protect the planet, and ensure that all people enjoy peace and prosperity by 2030 (UN, 2015). The Agenda is driven by 17 interconnected objectives, known as Sustainable Development Goals (SDGs). Each SDGs is accompanied by specific targets totalling 169 and each target is delineated by indicators that serve as metrics to measure progress into their achievement (UN, 2015). With just over five years remaining until the targeted year for accomplishing the goals arrives, the urgency to meet the set objectives intensifies even for advanced economies (Allen et al., 2019a). Especially when considering that the achievement of the SDGs has been negatively impacted by global crises, such as the COVID-19 pandemic, international armed conflicts, supply chain disruptions, inflation, and environmental problems. (Leal Filho et al., 2023). All countries are falling short on fulfilling the goals, raising significant concern on the agenda's attainment (Sachs et al. 2023). Because of this, research aimed at supporting stakeholders in implementing the SDGs and overcoming these challenges holds special importance. (Warchold et al., 2022). Nevertheless, important hurdles exist in this endeavour. For instance, measuring indicators across nations can be complex, resulting in poor-quality data (Sachs et al., 2023; Warchold et al., 2022). Additionally, the intricate interrelationships among the goals exhibit non-linear

ISBN 978-88-96687-17-8 || ISSN 2280-787X

behaviour, making traditional analytic methods inadequate for their study (Asadikia et al., 2021). Moreover, analysis of potential interventions is necessary at an indicator level to propose specific governance priorities (Pradhan, 2019). Lastly, having hundreds of indicators and limited resources available, prioritization to efficiently allocate them becomes a critical challenge (Allen et al., 2019b). To overcome these setbacks this research proposes employing an Artificial Intelligence method, leveraging high-quality OCED data, to extract key indicators that have the greatest impact on progress toward achieving the 2030 Agenda. Furthermore, as the analysis is performed at the indicator level, the results of this study will offer clearer insights into the interventions policymakers need to develop for creating effective strategies.

2 Theoretical Background

A vast amount of evidence-based research has been conducted on the SDGs with data assuming a pivotal role (Warchold et al., 2022). Researchers primarily rely on databases that provide quantifiable metrics for each SDG indicator (Warchold et al., 2022), allowing to perform research analysis on the SDGs in a quantitative way (Sachs et al., 2023). This is of great importance, especially considering that the components of the SDGs conform a system with complex interactions (Lusseau et al.,2019). Such interactions are usually classified as synergies, progress in one favour progress in another, and trade-offs, progress in one hamper progress in another (Pradhan et al., 2017; Allen et al., 2019b). Interactions are key for properly understanding the SDGs as a system and allow stockholders perform effective action (Pradhan et al., 2017). Scholars have investigated the interactions between SDGs though different analytical methods. Early studies attempted to measure these interactions through correlation coefficients (Pradhan et al., 2017). Further research has complemented these interaction calculations with diverse network analysis (Huan et al., 2022; Lusseau et al., 2019). As well, machine learning approaches have been recently developed that try to investigate the synergies among SDGs (Asadikia et al., 2021). However, many of these studies have often overlooked the non-linear complex relationships between the SDGs. For example, by considering only the neighbouring targets in the network analysis and ignoring second-order influences (Huan et al., 2022) or by not addressing at all the interaction between other goals on the studied correlation pairs (Pradhan et al., 2017). Furthermore, the achievement of SDGs involves significant costs and requirements, posing a challenge in determining effective and swift strategies (Lozano et al., 2018). Because of this, prioritization is essential for taking impactful and actionable steps and efficiently utilizing available resources to achieve the urgent and vital transformation the planet needs (Allen et al., 2019b; Biermann et al., 2022). Unfortunately, studies that address interactions do not always focus on prioritization (Asadikia et al., 2023). Nevertheless, different endeavours have been

ISBN 978-88-96687-17-8 || ISSN 2280-787X

made to prioritize SDGs. For instance, early efforts focus on using thresholds for pairwise correlation coefficients to categorize interactions as either synergic or trade-offs (Pradhan et al., 2017). Allen et al., 2019b applied Multi-Criteria Decision Analysis (MCDA) in order to prioritize SDGS based on their *"urgency level, policy gaps and systematic impact"*. Later research has tried to identify the more synergetic SDGs by means of network analysis (Lusseau et al., 2019; Huan et al., 2022). Some efforts have also aimed to study the prioritization of goals across different scenarios that either involve no interaction or low interaction between the SDGs by means of Boosting Regression Trees (BRT). (Asadikia et al., 2021; Asadikia et al., 2022). However, specificity and measurability to achieve the goals is often disregarded given their broad and englobing nature (Biermann et al., 2022). Because of that, most research that has maintained its focus at the level of the 17 SDGs has lost important detailed indicator-level information, hampering the ability to provide specific advice or action plans (Pradhan et al., 2019; Asadikia, 2021). Alternatively, studies addressing prioritization at a target or indicator level have restricted their focus to a specific SDG or to a city/nation context (Cheng et al., 2024; Al-Saidi, 2022). In this way, hindering the development of global applicable strategies since solutions and priorities identified may not be necessarily applicable in an international context (Pradhan et al., 2019). Artificial Intelligence (AI) models become a valuable tool for studying goal, target or indicator prioritization considering the complex nonlinearity interdependences of the SDGs (Asadikia et al., 2021). AI has been extensively utilized to enhance sustainability development efforts (Nishant et al., 2020). However, these studies mainly focus on the application of AI methods in specific industries or the resolution of very specific social or environmental issues (Gue et al., 2020; Gupta et al., 2021). In the prioritization field, only few studies have been performed. Mainly, Asadikia et al., 2021 that used Boosted Regression Trees (BRT) to study which goals are the better enablers of other goal's progress. For this, they considered scenarios with no interactions, pair-wise interactions, and up to five levels of interactions between the goals. However, from an extensive literature review, and to the best of our knowledge, no studies have yet addressed the prioritization at indicator level using AI methods.

3 Research Method

3.1 Data Source

For the present study, we utilized the SDG indicator data provided by the Bertelsmann Stiftung and Sustainable Development Solutions Network (BE-SDSN) (Sachs et al., 2023). The BE-SDSN database was preferred over other alternatives for multiple reasons. Firstly, the database upholds a clear and transparent methodology (Lafortune et al., 2018). Moreover, a normalization process is performed on the indicators that provides consistency across all of them and

ISBN 978-88-96687-17-8 || ISSN 2280-787X

ensures that they are comparable among each other (Lafortune et al., 2018; Asadikia et al., 2021). Furthermore, the BE-SDSN database introduces the *Score Index,* which assesses the percentage towards achieving optimal SDG performance (Sachs et al., 2023). This score not only enables us to track the progress of the overall 2030 Agenda but also provides a reliable target variable to be used in AI supervised learning methods. For the 2023 edition, the BE-SDSN 2023 database provided data on 98 global indicators for all 193 UN member countries spanning the years 2000 to 2023 (Sachs et al., 2023). Indicators are aligned with or closely correspond to SDG targets and are primarily sourced from the UN Statistical Commission and International organizations' databanks. In c-ases where data was insufficient, alternative reliable sources were utilized as complement. To be included in the BE-SDSN database, indicators must fulfil five criteria: global relevance and applicability, represent valid and reliable measures, being up-to-date, availability for at least 80% of UN Member States, and measurability to enable an optimal performance assessment (Sachs et al., 2023). This study has chosen to utilize data from the OCED countries. Primarily because these nations have more reliable data collection and reporting systems (OECD, 2018). A key aspect, considering the pivotal role data quality has in analytical models and the limitations posed by data availability in previous studies (Warchold et al., 2022). Additionally, it reduces the necessity for extensive data manipulation that might introduce uncertainties and diminish the validity of the analysis (Pradhan et al., 2017; Huan et al., 2022). Simultaneously, it maintains some global perspective by encompassing both developing and developed countries that are still facing substantial challenges in achieving the SDGs (Allen et al., 2019a; Sachs et al., 2023).

3.2 Analytical Method

In this study, the AI regression tree ensemble method Random Forest (RF) was adopted to understand which indicators have a larger contribution in predicting the SDG Index. The use of RF for this end is motivated by many reasons. Firstly, since beyond the aim of prediction, the construction of a predictive model through regression methods has the purpose of understanding the structural relationships between the dependent and independent variables (Breiman et al., 1984). Additionally, previous research has demonstrated correlations among the SDGs (Allen et al., 2019b, Warchold et al., 2022), making tree-ensemble regression methods more suitable than traditional regression methods that assume no correlation between their variables (Breiman et al., 1984). Furthermore, tree methods prove valuable when predictors exhibit non-linear associations, as they do not assume any specific underlying relationships between predictor variables and the response (James et al., 2023; Lunetta et al., 2004). Moreover, an essential aspect of RF is their capability to rank variables based on their importance for prediction (Behnamian et al., 2017; Boulesteix et al., 2012). Variable importance

ISBN 978-88-96687-17-8 || ISSN 2280-787X

measures in RF have been proved to be able to capture interactions between the predictors (James et al., 2023; Lunetta et al., 2004), imperative for our prioritization objective. In addition, unlike other ensemble methods, RF considers a pool of all predictor variables during the splits, uncovering interaction effects that might have been overlooked otherwise (Breiman, 2001; Strobl et al., 2008). To achieve this, RF averages a large number of decorrelated decision trees on bootstrapped training samples (Breiman et al., 1984). It constructs the trees by randomly selecting the predictive variables each time a split is considered (Breiman, 2001). In this way, RF is able to utilize all the involved variables while allowing the trees to be fairly deep than other ensemble tree methods (James et al., 2023). Further in-depth information on the RF method can be found in the original publication (Breiman, 2001).

3.3 Implementation

3.3.1 Data Preparation

The utilized data corresponds to a merge of the normalized indicator scores present in the 2030 Backdated SDG Index and the SDR 2023 Data tables of the BE-SDSN database. The former table contains the historical SDG Index Score as well as the goal and indicator scores for each nation from the previous years of the report. The latter contains the indicators normalized scores for the latest edition, i.e. the 2023 edition. Mounting to an overall of 304 observations, one for each of the 38 OECD countries for every year since the launch of the BE-SDSN SDG Index. To avoid unnecessary data manipulation, indicators with more than 20% missing data were excluded, adhering to the missing data criteria outlined by the BE-SDSN database guidelines (Sachs et al., 2023). An exception to this rule was applied to indicators belonging to the SDG 14 given that six OECD countries are landlock and this SDG addresses the conservation and sustainable use of the oceans, seas, and marine resources. For these indicators the 20% threshold for removal of missing data continued to apply, provided that the missing values did not originate from landlocked countries. Following these guidelines, three indicators were removed out of the initial 98, resulting in a final set of 95 indicators. All indicator's names and descriptions can be found in the Appendix A. The indicators removed were:

- 'n_sdg3_hiv': New HIV infections (per 1,000 uninfected population).
- 'n_sdg4_literacy': Literacy rate (% of population aged 15 to 24).
- 'n_sdg17_govrev': Government revenue excluding grants (% of GDP).

Within the remaining dataset, 2.05% of the values were still missing. These missing values were handled by replacing them with the median of the non-missing values of the indicator. This median-imputation strategy, recommended by the creator of the Random Forest, Leo Breiman, is only employed when dealing with small amounts of missing data (Breiman, 2003).

ISBN 978-88-96687-17-8 || ISSN 2280-787X

3.3.2 Method Deployment

The Random Forest (RF) algorithm was implemented using the RandomForestRegressor package from the scikit-learn library in Python (Pedregosa et al., 2011). To fit the RF to our data, four main parameters were fine-tuned. Firstly, max_feat that controls the maximum number of indicators considered at each split in the decision trees. Then max_depth and min_samples_split, where the former controls the maximum depth of the trees and the latter the minimum number of samples required to split an internal node. These last two parameters allow to prevent overfitting, by controlling the modulation of tree depth. Finally, n_estimators that selects for the number of estimators, or trees, that are in the forest. To determine the optimal parameter values, a grid search was performed on a training set together with a 5-fold cross-validation process to avoid overfitting. The value of 5 in the k-fold cross-validation was selected as it is the recommended value in the literature to provide a good bias-variance trade-off (James et al., 2023). Moreover, the model was evaluated on the test dataset by comparing the predicted values generated by the selected best-performing model against the actual data through a coefficient of determination R2. To conclude and obtain the indicator's predictor importance, a single stable forest comprising 10,000 trees was constructed with the optimal parameters to ensure variable importance stability (Behnamian et al., 2017).

4 Results

The obtained optimal parameter values though the grid analysis for the Random Forest model can be found in Table 1. The maximum number of indicators considered at each split was 31. Agreeing with the recommended value for RF regressions of three times less than the total number of predictors (Boulesteix et al., 2012). The maximum possible depth of the considered trees was 30 provided that there were more than 5 samples available for splitting in the node. Additionally, it was found that increasing the number of trees beyond 2,000 did not result in a significant improvement in performance.

Table 1. Optimal hyper-parameter values for the Random Forest Model

Parameter	Value
n_estimator	2000
max_depth	30
max_features	31
min_samples_split	5

The optimized model was evaluated by creating a prediction from the test dataset and comparing it with the real data. The resulting coefficient of determination was R^2 = 0.98, with an MSE score of 0.216. These values indicate that the model generalizes well to unseen data. Furthermore, that the model

ISBN 978-88-96687-17-8 || ISSN 2280-787X

explains a high percentage of the variance in the test data. A scatter plot showing the predicted against the actual data can be found in Appendix B. A variable importance analysis was conducted to determine the indicators that contribute most to predicting the SDG Index. To ensure result stability, a Random Forest model comprising 10,000 trees was fitted to the full data using the optimal model configuration (Behnamian et al., 2017). The top ten indicators that provide the highest relative influence on predicting the SDG Index are shown in Figure 1. Only nine indicators exhibited an influence exceeding 3% in predicting the SDG Index, collectively accounting for over 58% of the importance in shaping the SDG Index. We can observe that the indicator labelled 'n_sdg10_gini', that refers to the Gini Coefficient (Appendix A), has the largest relative influence on the SDG Index with about 15.8%. Furthermore, after this dominant indicator is considered, there are three salient indicators that had a similar importance on predicting the SDG Index. In particular: 'n_sdg12_so2import', which pertains to SO_2 emissions embodied in imports; 'n_sdg11_slums', reflecting the proportion of urban population living in slums (%); and 'n_sdg7_cleanfuel', indicating the percentage of the population with access to clean fuels and technology for cooking. The SDG index is influenced over 22% by these three indicators and together with the 'n_sdg10_gini' indicator account for over 38% on the contribution to predicting the SDG Index. Moreover, an additional five indicators play a noteworthy role, individually contributing over 3% each to the explanation of the SDG Index. Among these, 'n_sdg3_u5mort', representing the under-5 mortality rate, holds the fifth position with a 5% contribution. Following closely are indicators 'n_sdg5_familypl', 'n_sdg15_cpfa', 'n_sdg10_palma', and 'n_sdg6_wastewat'. Notably, the cumulative influence of these five indicators amounts to 19.7%. Additionally, twelve indicators demonstrate a low contribution, each exerting between 3% and 1% influence on the SDG Index. These indicators being: n_sdg15_redlist, n_sdg9_articles, n_sdg8_accounts, n_sdg8_rights, n_sdg17_oda, n_sdg7_renewcon, n_sdg14_biomar, n_sdg3_neonat, n_sdg1_wpc, n_sdg15_forchg, n_sdg2_stunting and n_sdg16_cpi. Collectively, the importance of these indicators on predicting the SDG Index amount to 20.1%. Conversely, the 74 remaining indicators exhibit a negligible impact, individually contributing less than 1% to the SDG Index.

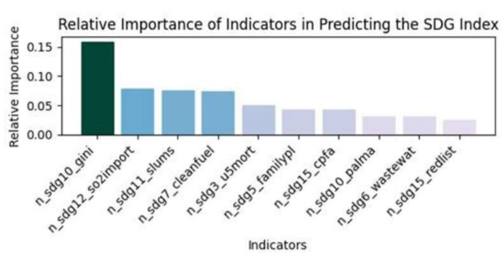


Figure 1. Top ten indicators with the highest relative importance in predicting the SDG Index via Random Forest.

ISBN 978-88-96687-17-8 || ISSN 2280-787X

5 Discussions

This study employed RF regression, an AI method, to identify the most significant indicators for predicting the SDG Index. This approach was taken to overcome limitations of classical regression methods when dealing with data containing correlated variables and non-linear relationships between predictors and the target variable (Breiman et al., 1984; Breiman, 2001). The fine-tuned model achieved high performance on unseen data from the test set, demonstrating its generalizability and its ability to explain a large proportion of the variance within the test data. Our model identified the indicator representing the Gini coefficient, denoted as 'n_sdg10_gini', as the most influential factor in predicting the SDG Index, accounting for a substantial 15.8%. The Gini coefficient is an aggregate measure of income inequality within a nation. This metric aligns perfectly with SDG 10's objective of reducing inequality both within and among countries. In particular, the Gini coefficient can indirectly track progress toward target 10.1, which seeks to achieve and sustain income growth for the bottom 40% of the population (Sachs et al., 2023). Previous studies have shown the high influence income inequality has on a wide range of social issues including health problems, violence, educational performance, obesity, and substance abuse (Wilkinson et al. 2007). Furthermore, previous research has shown that, for high-income economies, the reduction of inequalities has a great impact on other SDGs (Lusseau et al., 2019; Asadikia et al., 2022). This contrasts with other findings that have identified SDG 3 (Good Health and Well-being) as the most impactful goal when less developed regions were included in the analysis (Pradhan et al., 2017, Asadikia et al., 2021). This can be explained as more egalitarian societies, like the OECD nations, tend to be healthier (Wilkinson et al., 2007). Therefore, countries that are not currently facing critical challenges related to SDG 3, unlike many less developed nations (Sachs et al., 2023), will have other indicators as their main priorities. The high influence of income inequality on predicting the SDG Index highlights its importance within OECD nations. This is evidenced by the fact that some OECD countries, despite their boasting advanced economies, still suffer of substantial income disparities (OECD, 2018). However, further research is needed to fully understand the precise impact the Gini coefficient has on the SDG Index. The next indicators in explanation power after the Gini coefficient are 'n_sdg12_so2import', 'n_sdg11_slums' and 'n_sdg7_cleanfuel'. As can be seen from Figure 1, these three indicators contribute almost equally to the prediction of the SDG Index with a bit over 7% each. The indicator representing the SO_2 emissions embodied in imported goods and services (kg/capita), denoted as 'n_sdg12_so2import', emerged as the second-most influential factor, explaining 7.8% of the variation in the SDG Index. SO_2 pollution ranks among the foremost environmental and human health issues, leading to severe health consequences (Lum et al., 2023). This indicator is closely aligned with target 9.4, which aims to assess the sustainability of industries through the adoption of clean

ISBN 978-88-96687-17-8 || ISSN 2280-787X

and environmental-friendly technologies and industrial processes (Sachs et al., 2023). Notably, this indicator also aligns with SDG 12, which focuses on ensuring sustainable consumption and production patterns (Sachs et al., 2023). Furthermore, previous research has suggested that SDG 12 has only been influential for high-income countries (Asadikia et al., 2022). Air pollution from SO_2 represents a challenge for the OECD nations, as they need to reduce their comparably larger environmental footprint with respect to other less developed countries (Pradhan et al., 2017). The indicator measuring the proportion of urban population living in slums, denoted as 'n_sdg11_slums', arrived third with 7.5% influence over the SDG Index. This indicator is associated with the target 11.1 which aims at "ensure access for all to adequate, safe, and affordable housing and basic services, and upgrade slum" (UN, 2015). While most of the OECD countries have attained the goal of reducing to zero the urban population living in slum households, some OECD countries like Türkiye, Mexico and Colombia still face significant challenges on this front (Sachs et al., 2023). The rapid urbanization observed in these countries helps us understand why achieving target 11.1 is a priority in these regions (Laumann et al., 2022). Research suggests resolving slum conditions is intricately linked to other SDGs, as fundamental infrastructure services such as waste management, access to clean water, and reliable energy are indispensable for both public health and economic advancement (Teferi et al., 2017). It can be observed that each of these top three indicators corresponds to one of the three pillars of sustainability. Specifically, the indicator 'n_sdg10_gini' addresses the economic pillar, 'n_sdg12_so2import' the environmental pillar, and 'n_sdg11_slums' the social pillar. Hence, no dimension of sustainability has been left behind by the output of our model. Furthermore, our individuated primary contributors to the SDG Index provide a panoramic perspective of the multifaceted and intertwined challenges associated with sustainable development. The indicator measuring the population's access to clean fuels and cooking technology, labelled as 'n_sdg7_cleanfuel', ranked fourth with 7.4% impact on predicting the SDG Index. This indicator is closely aligned with the target 7.1 that aims at ensuring universal access to affordable, reliable, and modern energy services by improving the percentage of population with primary reliance on clean fuels and technology (UN, 2015). While OECD countries may have achieved universal energy access, many are lagging behind in two crucial areas: renewable energy adoption and energy efficiency (OECD, 2018). Advancing the adoption of clean energy for cooking is a precondition to attaining sustainable development, as dirty fuels have a negative impact on both, human and environmental health (Sreeja et al., 2023). Furthermore, it plays a crucial role in advancing various SDGs by improving public health, gender equality, and environmental conservation (Aberilla et al., 2020). However, despite improved access to cleaner cooking and energy consumption in most developed countries, many still resort to fuel stacking practices (Sreeja et al., 2023), presenting a challenge to sustainable energy transition efforts. This highlights the need for

ISBN 978-88-96687-17-8 || ISSN 2280-787X

further actions and policies to promote energy practices within OECD nations. The subsequent five indicators in the relative importance ranking exert influence ranging from 3% to 5% on the SDG Index. These indicators have a medium influence on the SDG Index and portrait a 19.7% explanatory power as a whole. These indicators are: 'n_sdg3_u5mort' representing the mortality rate for kids under five years of age, 'n_sdg5_familypl' indicating the demand for family planning satisfied by modern methods (percentage of females aged 15 to 49), 'n_sdg15_cpfa' measuring the proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, 'n_sdg10_palma' denoting the Palma ratio, and 'n_sdg6_wastewat' representing the percentage of anthropogenic wastewater that receives treatment. Again these indicators demonstrate a balanced representation across the social, environmental, and economic dimensions of sustainable development. Where 'n_sdg3_u5mort' and 'n_sdg5_familypl' pertain to the social dimension, 'n_sdg15_cpfa' and 'n_sdg6_wastewat' to the environmental dimension and 'n_sdg10_palma' to the economic dimension. It is worth noting the apparition of the Palma ratio, akin to the Gini coefficient, as a measure of income inequality. This ratio measures the income inequality by comparing the income share of the top 10% to that of the bottom 40% of earners in a population (Palma, 2011). The arrival of this indicator in the eighth position within the importance ranking further strengthens the significance of addressing income inequalities within OECD nations. Also, reinforces the importance highlighted by the Gini coefficient in explaining the SDG Index. The nine indicators discussed thus far collectively account for over 58% of the explanatory power of the SDG Index. The remaining percentage is attributed to the remaining 86 indicators. These additional indicators exert minimal influence on the SDG Index, each contributing less than 3%. Consequently, their impact on understanding the SDG Index is negligible.

6 Conclusions

The fulfilment of the 2030 Agenda presents a complex challenge especially considering setback from global crisis. The high costs and intricate interaction between SDGs make prioritizing efforts crucial for taking impactful and actionable policy to efficiently utilize available resources. In particular, prioritizing at the indicator level can provide specific and detailed information for creating applicable strategies that are lost at goal level. However, significant challenges in dealing with SDG data, such as multicollinearity and non-linear relations among variables, pose obstacles to properly understanding their interactions. In this study we propose an AI method to address these challenges. In particular, a Random Forest, an ensemble tree method, was utilized due to its ability to account for interactions among all variables in the model, while effectively addressing the challenges associated with traditional regression methods mentioned earlier. Moreover, it allows us to rank variables based on their

ISBN 978-88-96687-17-8 || ISSN 2280-787X

importance for prediction. The data used for modeling were indicators from OECD nations sourced from the BE-SDSN database. This dataset provided high-quality data for the nations involved and enabled the utilization of the SDG Index as the target variable for the AI model. The trained model achieved a high performance on unseen data proving its generalizability and its capacity to understand national progress towards the SDGs. Our analysis identified the Gini coefficient, represented by 'n_sdg10_gini', as the most influential factor in predicting the SDG Index. This result emphasis the substantial influence income inequality has on high-income countries and the impact it has on other SDGs. Furthermore, it highlights the need for targeted interventions to address this issue. Additionally, high influential predictors where: 'n_sdg12_so2import', 'n_sdg11_slums', and 'n_sdg7_cleanfuel'. These indicators reflect key aspects of environmental sustainability, social equity, and access to clean energy, respectively, underscoring the multifaceted nature of sustainable development challenges in OCED nations. Addressing issues such as SO_2 air pollution, a significant environmental and public health concern, is particularly emphasized. Furthermore, managing rapid urbanization and providing essential infrastructure services, especially in the least developed OECD countries, are imperative actions. These results signify progress in identifying the most influential indicators shaping the SDG Index. Through the modeling process, multicollinearity and non-linear relationships between the variables were effectively addressed allowing full interaction among indicators. Nevertheless, while the results highlight influential factors, they do not explicitly elucidate whether this influence represents synergy or trade-offs. Therefore, further research must be conducted to properly understand these relationships.

7 Limitations

The primary limitations of this study are attributed to the data utilized. While our analysis leveraged from high-quality OECD data, the obtained results may not directly be applied to non-member nations. Additionally, the used supervised learning model in identifying the main indicators that impact the SDG Index provides only a picture of the current situation with the utilized set of indicators. Potential result variations may arise from new indicator inclusion or updated data. Furthermore, the method developed in this study offers a solution to previously known problems with SDG prioritization, such as multicollinearity and non-linear relationships between indicators. However, further research is needed to fully understand how these key indicators interact with each other and how exactly they influence the SDG Index.

ISBN 978-88-96687-17-8 || ISSN 2280-787X

References

Aberilla J.M., Gallego-Schmid A., Stamford L. and Azapagic, A., (2020) "Environmental sustainability of cooking fuels in remote communities: Life cycle and local impacts", Science of The Total Environment, Vol. 713, 136445. https://doi.org/10.1016/j.scitotenv.2019.136445.

Allen, C., Metternicht, G., Wiedmann, T. and Pedercini, M., (2019a) "Greater gains for Australia by tackling all SDGs but the last steps will be the most challenging", Nature Sustainability, Vol. 2, No.11, pp. 1041–1050. https://doi.org/10.1038/s41893-019-0409-9.

Allen, C., Metternicht, G., and Wiedmann, T., (2019b) "Prioritising SDG targets: Assessing baselines, gaps and interlinkages", Sustainability Science, Vol. 14, No. 2, pp. 421–438. https://doi.org/10.1007/s11625-018-0596-8.

Al-Saidi, M., (2022) "Disentangling the SDGs agenda in the GCC region: Priority targets and core areas for environmental action", Frontiers in Environmental Science, Environmental Economics and Management Section, Vol. 10. https://doi.org/10.3389/fenvs.2022.1025337.

Asadikia, A., Rajabifard, A. and Kalantari, M., (2021) "Systematic prioritisation of sdgs: Machine learning approach", World Development, Vol. 140. https://doi.org/10.1016/j.worlddev.2020.105269.

Asadikia, A., Rajabifard, A. and Kalantari, M., (2022) "Region-income-based prioritisation of Sustainable Development Goals by Gradient Boosting Machine", Sustainability Science, Vol. 17, pp. 1939–1957. https://doi.org/10.1007/s11625-022-01120-3.

Asadikia, A., Rajabifard, A. and Kalantari, M., (2023) "A systems perspective on national prioritisation of sustainable development goals: Insights from Australia", Geography and Sustainability, Vol. 4, No. 3, pp. 255-267. https://doi.org/10.1016/j.geosus.2023.06.003.

Behnamian A., Millard K., Banks S.N., White L., Richardson M. and Pasher J., (2017) "A Systematic Approach for Variable Selection With Random Forests: Achieving Stable Variable Importance Values", IEEE Geoscience and Remote Sensing Letters, Vol. 14, No. 11, pp. 1988-1992. https://doi.org/10.1109/LGRS.2017.2745049.

Biermann, F., Hickmann, T., Sénit, C.A., Beisheim, M., Bernstein, S., Chasek P., Grob L., Kim R.E., Kotzé L., Nilsson M., Ordóñez Llanos A., Okereke C., Pradhan P., Raven R., Sun Y., Vijge M., van Vuuren D., and Wicke B., (2022) "Scientific evidence on the political impact of the Sustainable Development Goals", Nature Sustainability, Vol. 5, pp. 795–800. https://doi.org/10.1038/s41893-022-00909-5.

Boulesteix A.L., Janitza S., Kruppa J., Inke R. and Konig I.R., (2012) "Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 2, pp. 493–507. https://doi.org/10.1002/widm.1072.

Breiman L., (2001) "Random forests", Machine Learning, Vol. 45, No.1, pp. 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman L., (2003) Setting up, using, and understanding random forests V4.0, url: https://www.stat.berkeley.edu/~breiman/Using_random_forests_v4.0.pdf .

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J., (1984). Classification and regression trees, ed. Chapman and Hall, New York. https://doi.org/10.1201/9781315139470.

Cheng, Q., Zhang, C., Zou, Y., Pu. X. and Jin, H., (2024) "Unraveling interactions and priorities under sustainable development goals in less-developed mountainous areas: case study on the National Innovation Demonstration Zone for the 2030 Agenda for Sustainable Development, China", Environmental Science and Pollution Research, Vol. 31, pp. 5254–5274. https://doi.org/10.1007/s11356-023-31478-5.

Gue, I.H.V., Ubando, A.T., Tseng, ML. and Tan R.R., (2020) "Artificial neural networks for sustainable development: a critical review", Clean Technologies and Environmental Policy, Vol. 22, pp. 1449–1465. https://doi.org/10.1007/s10098-020-01883-2.

Gupta, S., Langhans, S.D., Domisch, S., Fuso-Nerini, F., Felländer, A., Battaglini, M., Tegmark, M. and Vinuesa R., (2021) "Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level", Transportation Engineering, Vol. 4. https://doi.org/10.1016/j.treng.2021.100064.

Huan, Y., Wang, L., Burgman, M., Li, H., Yu, Y., Zhang, J. and Liang, T., (2022) "A multi-perspective composite assessment framework for prioritizing targets of sustainable development goals", Sustainable Development. Vol. 30, No. 5, pp. 833–847. https://doi.org/10.1002/sd.2283.

James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J., (2023). An introduction to statistical learning, ed. Springer https://doi.org/10.1007/978-3-031-38747-0.

Lafortune, G., Fuller, G., Moreno, J., Schmidt-Traub, G., and Kroll, C., (2018) SDG Index and Dashboards Detailed Methodological Paper, Sustainable Development Solutions Network.
https://raw.githubusercontent.com/sdsna/2018GlobalIndex/master/2018GlobalIndex Methodology.pdf.

Laumann F., von Kügelgen J., Kanashiro Uehara T.H. and Barahona M., (2022) "Complex interlinkages, key objectives, and nexuses among the Sustainable Development Goals and climate change: a network analysis", The Lancet Planetary Health, Vol. 6, No. 5, pp. 422-430. https://doi.org/10.1016/S2542-5196(22)00070-5.

Leal Filho, W., Viera Trevisan, L., Simon Rampasso, I., Anholon, R., Pimenta Dinis, M.A., Londero Brandli, L., Sierra, J., Lange Salvia, A., Pretorius, R., Nicolau, M., Pries Eustachio, J.H. and Mazutti, J., (2023) "When the alarm bells ring: why the UN Sustainable Development Goals may not be achieved by 2030", Journal of Cleaner Production, Vol. 407. https://doi.org/10.1016/j.jclepro.2023.137108.

Lozano, R., Fullman, N., Abate, D., Abay, S. M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., AbdelRahman, O., Abdi, A., Abdollahpour, I., Abdulkader, R. S., Abebe, N. D., Abebe, Z., Abejie, A. N., Abera, S. F., Abil, O. Z., ... Murray, C. J. L., (2018) "Measuring progress from 1990 to 2017 and projecting attainment to 2030 of the health-related sustainable development goals for 195 countries and territories: A systematic analysis for the global burden of disease study 2017", The Lancet, Vol. 392, pp. 2091–2138. https://doi.org/10.1016/S0140-6736(18)32281-5.

Lunetta, K.L., Hayward, L.B., Segal, J. and Van Eerdewegh, P., (2004) "Screening large-scale association study data: exploiting interactions using random forests", BMC Genetics, Vol. 5. https://doi.org/10.1186/1471-2156-5-32.

Lum, M.M.X., Kim, H.N., Sin, Y.L., Mohamed A.R., Alsultan A.G., Taufiq-Yap Y.H., Koh M.K, Mohamed M.A., Vo, D.V.N., Subramaniam, M., Mulya K.S. and Imanuella, N., (2023) "Sulfur dioxide catalytic reduction for environmental sustainability and circular economy: A review", Process Safety and Environmental Protection, Vol. 176, pp. 580-604. https://doi.org/10.1016/j.psep.2023.06.035.

ISBN 978-88-96687-17-8 || ISSN 2280-787X

Lusseau, D., and Mancini, F., (2019) "Income-based variation in sustainable development goal interaction networks", Nature Sustainability, Vol. 2, pp. 242–247. https://doi.org/10.1038/s41893-019-0231-4.

Nishant, R., Kennedy, M. and Corbett, J., (2020) "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda" International Journal of Information Management, Vol. 53. https://doi.org/10.1016/j.ijinfomgt.2020.102104.

OECD (2018), Measuring Distance to the SDG Targets 2017: An Assessment of Where OECD Countries Stand, ed. OECD Publishing, Paris. https://doi.org/10.1787/9789264308183-en.

Palma, J.G., (2011) "Homogeneous Middles vs. Heterogeneous Tails, and the End of the 'Inverted-U': It's All About the Share of the Rich", Development and change, Vol. 42, No. 1, pp. 87-153. https://doi.org/10.1111/j.1467-7660.2011.01694.x.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... and Duchesnay, É., (2011) Scikit-learn: Machine learning in Python, Journal of Machine Learning Research, Vol. 12, pp. 2825-2830. https://doi.org/10.48550/arXiv.1201.0490.

Pradhan, P., Costa, L., Rybski, D., Lucht, W., and Kropp, J. P., (2017) "A Systematic Study of Sustainable Development Goal (SDG) Interactions", Earth's Future, Vol. 5, No.11, pp. 1169–1179. https://doi.org/10.1002/2017EF000632.

Pradhan, P., (2019) "Antagonists to meeting the 2030 agenda", Nature Sustainability, Vol. 2, pp. 171–172. http://doi.org/10.1038/s41893-019-0248-8.

Sachs, J.D., Lafortune, G., Fuller, G. and Drumm, E., (2023) Implementing the SDG Stimulus. Sustainable Development Report 2023, ed. Dublin University Press. https://doi.org/10.25546/102924.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T. and Zeileis, A., (2008) "Conditional variable importance for random forests", BMC Bioinformatics, Vol. 9, No.1. https://doi.org/10.1186/1471-2105-9-307.

Sreeja, A.T., Dhengle, A., Kumar, D. and Pradhan, A.K., (2023) "Does access to clean cooking fuels reduce environmental degradation? Evidence from BRICS nations", Environmental Science and Pollution Research, Vol. 30, pp. 78948–78958. https://doi.org/10.1007/s11356-023-27619-5.

Teferi, Z.A. and Newman, P., (2017) "Slum Regeneration and Sustainability: Applying the Extended Metabolism Model and the SDGs", Sustainability, Vol. 9, No. 12. https://doi.org/10.3390/su9122273.

UN, (2015) Transforming our world: the 2030 Agenda for Sustainable Development, General Assembley 70 Session. https://unstats.un.org/sdgs/iaeg-sdgs.

Warchold, A., Pradhan, P., Thapa, P., Putra, M.P.I.F. and Kropp, J.P., (2022) "Building a unified sustainable development goal database: Why does sustainable development goal data selection matter?", Sustainable Development, Vol. 30, No. 5, pp. 1278-1293. https://doi.org/10.1002/sd.2316.

Wilkinson, R.G and Pickett, K.E., (2007) "The problems of relative deprivation: why some societies do better than others", Social Science & Medicine, Vol. 65, No.9, pp. 1965–1978. https://doi.org/10.1016/j.socscimed.2007.05.041.

ISBN 978-88-96687-17-8 || ISSN 2280-787X

Appendix A

Table A. Indicators descriptions and association to goals and targets as provided by the BE-SDSN database (Sachs et al. 2023).

Indicator Name	Associated SDG	Indicator Description	SDG Target: Match or Closely Aligned?	Associated SDG Target
n_sdg1_wpc	1	Poverty headcount ratio at $2.15/day (2017 PPP, %)	Match	1.1.1
n_sdg1_lmicpov	1	Poverty headcount ratio at $3.65/day (2017 PPP, %)	Match	1.1.1
n_sdg2_undernsh	2	Prevalence of undernourishment (%)	Match	2.1.1
n_sdg2_stunting	2	Prevalence of stunting in children under 5 years of age (%)	Match	2.2.1
n_sdg2_wasting	2	Prevalence of wasting in children under 5 years of age (%)	Match	2.2.2
n_sdg2_obesity	2	Prevalence of obesity, BMI ≥ 30 (% of adult population)	Closely aligned	2.2
n_sdg2_trophic	2	Human Trophic Level (best 2-3 worst)	Not in UNSTATS	-
n_sdg2_crlyld	2	Cereal yield (tonnes per hectare of harvested land)	Closely aligned	2.3 & 2.4
n_sdg2_snmi	2	Sustainable Nitrogen Management Index (best 0-1.41 worst)	Closely aligned	2.4
n_sdg2_pestexp	2	Exports of hazardous pesticides (tonnes per million population)	Closely aligned	3.9
n_sdg3_matmort	3	Maternal mortality rate (per 100,000 live births)	Match	3.1.1
n_sdg3_neonat	3	Neonatal mortality rate (per 1,000 live births)	Match	3.2.2
n_sdg3_u5mort	3	Mortality rate, under-5 (per 1,000 live births)	Match	3.2.1
n_sdg3_tb	3	Incidence of tuberculosis (per 100,000 population)	Match	3.3.2
n_sdg3_hiv	3	New HIV infections (per 1,000 uninfected population)	Match	3.3.1
n_sdg3_ncds	3	Age-standardized death rate due to cardiovascular disease, cancer, diabetes, or chronic respiratory disease in adults aged 30–70 years (%)	Match	3.4.1
n_sdg3_pollmort	3	Age-standardized death rate attributable to household air pollution and ambient air pollution (per 100,000 population)	Match	3.9.1
n_sdg3_traffic	3	Traffic deaths (per 100,000 population)	Match	3.6.1
n_sdg3_lifee	3	Life expectancy at birth (years)	Closely aligned	3.1 : 3.9
n_sdg3_fertility	3	Adolescent fertility rate (births per 1,000 females aged 15 to 19)	Match	3.7.2
n_sdg3_births	3	Births attended by skilled health personnel (%)	Match	3.1.2
n_sdg3_vac	3	Surviving infants who received 2 WHO-recommended vaccines (%)	Closely aligned	3.b.1
n_sdg3_uhc	3	Universal health coverage (UHC) index of service coverage (worst 0-100 best)	Match	3.8.1
n_sdg3_swb	3	Subjective well-being (average ladder score, worst 0-10 best)	Closely aligned	3.4
n_sdg4_earlyedu	4	Participation rate in pre-primary organized learning (% of children aged 4 to 6)	Closely aligned	4.2.2
n_sdg4_primary	4	Net primary enrollment rate (%)	Closely aligned	4.1.2
n_sdg4_second	4	Lower secondary completion rate (%)	Match	4.1.2

ISBN 978-88-96687-17-8 || ISSN 2280-787X

n_sdg4_literacy	4	Literacy rate (% of population aged 15 to 24)	Match	4.6.1
n_sdg5_familypl	5	Demand for family planning satisfied by modern methods (% of females aged 15 to 49)	Match	3.7.1
n_sdg5_edat	5	Ratio of female-to-male mean years of education received (%)	Closely aligned	4.5.1
n_sdg5_lfpr	5	Ratio of female-to-male labor force participation rate (%)	Closely aligned	5.5
n_sdg5_parl	5	Seats held by women in national parliament (%)	Match	5.5.1
n_sdg6_water	6	Population using at least basic drinking water services (%)	Closely aligned	6.1.1
n_sdg6_sanita	6	Population using at least basic sanitation services (%)	Closely aligned	6.2.1
n_sdg6_freshwat	6	Freshwater withdrawal (% of available freshwater resources)	Match	6.4.2
n_sdg6_wastewat	6	Anthropogenic wastewater that receives treatment (%)	Match	6.3.1
n_sdg6_scarcew	6	Scarce water consumption embodied in imports (m3 H2O eq/capita)	Closely aligned	6.4
n_sdg7_elecac	7	Population with access to electricity (%)	Match	7.1.1
n_sdg7_cleanfuel	7	Population with access to clean fuels and technology for cooking (%)	Closely aligned	7.1.2
n_sdg7_co2twh	7	CO_2 emissions from fuel combustion per total electricity output ($MtCO_2$/TWh)	Closely aligned	7.2
n_sdg7_renewcon	7	Renewable energy share in total final energy consumption (%)	Match	7.2.1
n_sdg8_adjgrowth	8	Adjusted GDP growth (%)	Closely aligned	8.1.1
n_sdg8_slavery	8	Victims of modern slavery (per 1,000 population)	Closely aligned	8.7
n_sdg8_accounts	8	Adults with an account at a bank or other financial institution or with a mobile-money-service provider (% of population aged 15 or over)	Match	8.10.2
n_sdg8_unemp	8	Unemployment rate (% of total labor force, ages 15+)	Match	8.5.2
n_sdg8_rights	8	Fundamental labor rights are effectively guaranteed (worst 0–1 best)	Match	8.8.2
n_sdg8_impacc	8	Fatal work-related accidents embodied in imports (per 100,000 population)	Closely aligned	8.8.1
n_sdg8_impslav	8	Victims of modern slavery embodied in imports (per 100,000 population)	Closely aligned	8.7
n_sdg9_roads	9	Rural population with access to all-season roads (%)	Match	9.1.1
n_sdg9_intuse	9	Population using the internet (%)	Match	17.8.1
n_sdg9_mobuse	9	Mobile broadband subscriptions (per 100 population)	Closely aligned	9.c.1 & 17.6.1
n_sdg9_lpi	9	Logistics Performance Index: Quality of trade and transport-related infrastructure (worst 1-5 best)	Closely aligned	9.1
n_sdg9_uni	9	The Times Higher Education Universities Ranking: Average score of top 3 universities (worst 0-100 best)	Not in UNSTATS	-
n_sdg9_articles	9	Articles published in academic journals (per 1,000 population)	Closely aligned	9.5
n_sdg9_rdex	9	Expenditure on research and development (% of GDP)	Match	9.5.1
n_sdg10_gini	10	Gini coefficient	Closely aligned	10.1
n_sdg10_palma	10	Palma ratio	Closely aligned	10.1
n_sdg11_slums	11	Proportion of urban population living in	Match	11.1.1

ISBN 978-88-96687-17-8 || ISSN 2280-787X

		slums (%)		
n_sdg11_pm25	11	Annual mean concentration of particulate matter of less than 2.5 microns in diameter (PM2.5) ($\mu g/m^3$)	Match	11.6.2
n_sdg11_pipedwat	11	Access to improved water source, piped (% of urban population)	Closely aligned	11.1
n_sdg11_transport	11	Satisfaction with public transport (%)	Closely aligned	11.2.1
n_sdg12_msw	12	Municipal solid waste (kg/capita/day)	Closely aligned	12.5
n_sdg12_ewaste	12	Electronic waste (kg/capita)	Match	12.4.2
n_sdg12_so2prod	12	Production-based SO_2 emissions (kg/capita)	Closely aligned	9.4
n_sdg12_so2import	12	SO_2 emissions embodied in imports (kg/capita)	Closely aligned	9.4
n_sdg12_nprod	12	Production-based nitrogen emissions (kg/capita)	Closely aligned	9.4
n_sdg12_nimport	12	Nitrogen emissions embodied in imports (kg/capita)	Closely aligned	9.4
n_sdg12_explastic	12	Exports of plastic waste (kg/capita)	Closely aligned	12.4
n_sdg13_co2gcp	13	CO_2 emissions from fossil fuel combustion and cement production (tCO2/capita)	Closely aligned	13.2.2
n_sdg13_co2import	13	CO_2 emissions embodied in imports (tCO_2/capita)	Closely aligned	13.2
n_sdg13_co2export	13	CO_2 emissions embodied in fossil fuel exports (kg/capita)	Closely aligned	13.2
n_sdg14_cpma	14	Mean area that is protected in marine sites important to biodiversity (%)	Closely aligned	14.5.1
n_sdg14_cleanwat	14	Ocean Health Index: Clean Waters score (worst 0-100 best)	Closely aligned	14.1.1
n_sdg14_fishstocks	14	Fish caught from overexploited or collapsed stocks (% of total catch)	Closely aligned	14.4.1
n_sdg14_trawl	14	Fish caught by trawling or dredging (%)	Closely aligned	14.4
n_sdg14_discard	14	Fish caught that are then discarded (%)	Closely aligned	14.4
n_sdg14_biomar	14	Marine biodiversity threats embodied in imports (per million population)	Closely aligned	14.4
n_sdg15_cpta	15	Mean area that is protected in terrestrial sites important to biodiversity (%)	Match	15.1.2
n_sdg15_cpfa	15	Mean area that is protected in freshwater sites important to biodiversity (%)	Match	15.1.2
n_sdg15_redlist	15	Red List Index of species survival (worst 0-1 best)	Match	15.5.1
n_sdg15_forchg	15	Permanent deforestation (% of forest area, 3-year average)	Closely aligned	15.2
n_sdg15_biofrwter	15	Terrestrial and freshwater biodiversity threats embodied in imports (per million population)	Closely aligned	15.5
n_sdg16_homicides	16	Homicides (per 100,000 population)	Match	16.1.1
n_sdg16_detain	16	Unsentenced detainees (% of prison population)	Match	16.3.2
n_sdg16_safe	16	Population who feel safe walking alone at night in the city or area where they live (%)	Match	16.1.4
n_sdg16_u5reg	16	Birth registrations with civil authority (% of children under age 5)	Match	16.9.1
n_sdg16_cpi	16	Corruption Perceptions Index (worst 0-100 best)	Closely aligned	16.5.1 & 16.5.2
n_sdg16_clabor	16	Children involved in child labor (% of population aged 5 to 14)	Closely aligned	8.7.1
n_sdg16_weaponsexp	16	Exports of major conventional weapons (TIV constant million USD per 100,000 population)	Closely aligned	16.1
n_sdg16_rsf	16	Press Freedom Index (worst 0-100 best)	Closely	16.1

ISBN 978-88-96687-17-8 || ISSN 2280-787X

			aligned	
n_sdg16_justice	16	Access to and affordability of justice (worst 0–1 best)	Closely aligned	16.3.1 & 16.3.3
n_sdg16_admin	16	Timeliness of administrative proceedings (worst 0 - 1 best)	Closely aligned	16.6
n_sdg16_exprop	16	Expropriations are lawful and adequately compensated (worst 0 - 1 best)	Closely aligned	16.6
n_sdg17_govex	17	Government spending on health and education (% of GDP)	Closely aligned	1.a.1
n_sdg17_oda	17	For high-income and all OECD DAC countries: International concessional public finance, including official development assistance (% of GNI)	Closely aligned	17.2.1
n_sdg17_govrev	17	Other countries: Government revenue excluding grants (% of GDP)	Match	17.1.1
n_sdg17_cohaven	17	Corporate Tax Haven Score (best 0-100 worst)	Not in UNSTATS	-
n_sdg17_statperf	17	Statistical Performance Index (worst 0-100 best)	Closely aligned	17.18.1 : 17.19.2

Appendix B

Figure A. Scatter plot showing the SDG Index prediction values against the real SDG Index values for the test data with its corresponding coefficient of determination.