**ORIGINAL ARTICLE**

# Region-income-based prioritisation of Sustainable Development Goals by Gradient Boosting Machine

Atie Asadikia[1] · Abbas Rajabifard[1] · Mohsen Kalantari[1]

## Abstract

The Sustainable Development Goals (SDGs) seek to address complex global challenges and cover aspects of social development, environmental protection, and economic growth. However, the holistic and complicated nature of the goals has made their attainment difficult. Achieving all goals by 2030 given countries' limited budgets with the economic and social disruption that the COVID-19 pandemic has caused is over-optimistic. To have the most profound impact on the SDGs achievement, prioritising and improving co-beneficial goals is an effective solution. This study confirms that countries' geographic location and income level have a significant relationship with overall SDGs achievement. This article applies the Gradient Boosting Machine (GBM) algorithm to identify the top five SDGs that drive the overall SDG score. The results show that the influential SDGs vary for countries with a specific income level located in different regions. In Europe and Central Asia, SDG10 is among the most influential goals for high-income countries, SDG9 for upper-middle-income, SDG3 in low and lower-middle-income countries of Sub-Saharan Africa, and SDG5 in Latin America and the Caribbean upper-middle-income countries. This systematic and exploratory data-driven study generates new insights that confirm the uniqueness, and non-linearity of the relationship between goals and overall SDGs achievement.

**Keywords** Co-beneficial SDGs · SDG prioritisation · Machine learning · Gradient boosting machine · Data science · Sustainable development goals

## Introduction

Sustainable Development Goals (SDGs) as a global framework tool try to balance three overarching objectives: economic, social, and environmental. However, achieving one specific aspect of sustainability, such as preserving natural resources, can directly or indirectly be related to social or economic aspects. This relationship becomes more complicated when one goal's progress can negatively impact upon another (Mans Nilsson et al. 2017). For example, improving SDG4 (Quality education) impacts positively on SDG8 (Jeong et al. 2020) (Economic growth). Nonetheless,

progress towards SDG13 (Climate action) can compromise this by its negative effect on SDG8. Such relationships among or within goals make the SDGs complex or challenging to achieve. The complicated nature of SDGs, limited resources, and differences in countries' national priorities undermine goals' achievability.

A study using a scenario analysis discusses that even in the 'best-case' situation for a high-income developed country, all SDGs cannot be achieved by 2030 (Allen et al. 2019). The progress towards realising all SDGs was already behind schedule when the unexpected COVID-19 pandemic erupted in late 2019. This unwelcome disruption has consumed enormous resources and budgets and requires dedication for recovery. The world is facing a serious crisis that prevents countries from fulfilling SDGs, and "reversing decades of progress" on some goals (United Nations 2021). The question arises "if attaining all goals by 2030 is not possible, can countries maximise what they can achieve by prioritising some SDGs?".

This study aims to identify: (1) the significance of countries characteristics with their overall SDGs achievement;

✉ Atie Asadikia
a.asadikia@student.unimelb.edu.au

1 SDGs Research Group, Centre for Spatial Data Infrastructures and Land Administration, Faculty of Engineering and IT, The University of Melbourne, Parkville, VIC 3010, Australia

(2) the magnitude of the positive or negative influence of each goal on the overall SDGs achievement of countries; (3) the type of relationship of each goal with the overall SDGs achievement (linear or non-linear); and (4) co-beneficial or influential goals that emerge based on countries characteristics.

In this analysis the Gradient Boosting Machine (GBM), a powerful supervised learning algorithm famous for identifying variables' importance is applied. Using this method of machine learning, we explore the influence of each SDG on the overall SDGs achievement.
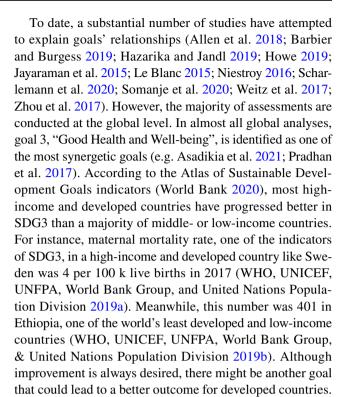
## Evidence-based prioritisation of SDGs is a necessity

Countries cherry-pick SDGs over others based on their economic or national political interests (Horn and Grugel 2018). This unhealthy prioritisation is against the intention of the 2030 Agenda and jeopardises the coherence of the integrated and indivisible goals (UN 2018). In other words, prioritising SDGs in isolation without including the potential interactions between them, can act unfavourably towards overall goal achievement. Although the ultimate objective of Agenda 2030 is fulfilling all goals, with the damage wrought by the COVID-19 pandemic and the unexpected resource allocation to control the virus, attaining all goals by the intended year (2030) is far from the reality. As opposed to bias or judgmental selection of SDGs, evidence-based prioritisation, which maximises synergies (positive impacts) and minimises trade-offs (negative impacts), can be one of the solutions to rapid acceleration of the progress towards goals (Asadikia et al. 2021; Weitz et al. 2017). This solution which offers decision-makers to focus on some SDGs instead of all (or nothing), opens a window of hope to not stop the progress because attaining all goals does not seem practical.

This type of evidence-based prioritisation supports developing effective and coherent policies or strategies throughout different jurisdictions. It also helps to define the governance structures to capitalise on co-beneficial goals and accelerate progress towards the 2030 targets. However, a holistic approach is required to identify those synergetic goals and to ensure that attaining those benefits overall SDG achievement.

## The role of "scale" and "context" in prioritisation of SDGs

Literature has long argued that the characteristics of a country can guide SDGs relationships (e.g. Måns Nilsson et al. 2016; Weitz et al. 2015). In this regard, several studies have shown that SDGs interact differently when the scale or context changes (Allen et al. 2018; Le Blanc 2015; Niestroy 2016; Weitz et al. 2017).

To date, a substantial number of studies have attempted to explain goals' relationships (Allen et al. 2018; Barbier and Burgess 2019; Hazarika and Jandl 2019; Howe 2019; Jayaraman et al. 2015; Le Blanc 2015; Niestroy 2016; Scharlemann et al. 2020; Somanje et al. 2020; Weitz et al. 2017; Zhou et al. 2017). However, the majority of assessments are conducted at the global level. In almost all global analyses, goal 3, "Good Health and Well-being", is identified as one of the most synergetic goals (e.g. Asadikia et al. 2021; Pradhan et al. 2017). According to the Atlas of Sustainable Development Goals indicators (World Bank 2020), most high-income and developed countries have progressed better in SDG3 than a majority of middle- or low-income countries. For instance, maternal mortality rate, one of the indicators of SDG3, in a high-income and developed country like Sweden was 4 per 100 k live births in 2017 (WHO, UNICEF, UNFPA, World Bank Group, and United Nations Population Division 2019a). Meanwhile, this number was 401 in Ethiopia, one of the world's least developed and low-income countries (WHO, UNICEF, UNFPA, World Bank Group, & United Nations Population Division 2019b). Although improvement is always desired, there might be another goal that could lead to a better outcome for developed countries.

There are a limited number of studies that assess the SDGs relationship at the national (e.g. Allen et al. 2018; Baffoe et al. 2021; Lusseau and Mancini 2019; Weitz et al. 2017) and local scales (e.g. Bandari et al. 2021). Although all prior studies provide potent insights for SDG interactions, most research incorporates a subjective selection of criteria or judgmental scoring in their analyses (Alcamo et al. 2020). To ensure this prioritisation is not included any subjective opinions or judgmental criteria, robust and detailed data analysis can be beneficial.

During the past decade, a growing body of literature has emphasised the context-sensitivity of SDG interactions (Måns Nilsson et al. 2018; Pradhan et al. 2017; Scherer et al. 2018; Singh et al. 2018; Weitz et al. 2017). To identify co-beneficial SDGs and tailor the prioritisation for countries, it is crucial to consider the contexts. However, the diversity of natural resource endowments between countries, levels of economic development, political, socio-cultural, and many other factors add further complexity to the contextual prioritisation of goals.

According to the literature, economic factors can impact the relationships of goals (Kurniawan and Managi 2018; Lusseau and Mancini 2019; Marcondes dos Santos and Perrella Balestieri 2018). In fact, the level of income is described as one of the major descriptors of the macroeconomic context of countries (Galor and Zeira 1993), which has a direct and indirect impact on sustainable development (Lusseau and Mancini 2019). One of the other factors that has been mentioned in literature as an important feature that impacts SDGs relationships is the geographical location of

a country (Allen et al. 2018; Bali Swain and Ranganathan 2021; Pedercini et al. 2019; Pradhan et al. 2017). The geographical context of a country provides an estimation of its natural resources. The countries region is an overarching factor which covers several contributing sub-factors such as environmental and political aspects, and the culture of neighbouring countries that are crucial to be considered for sustainable development (Scott and Rajabifard 2017).

The population is also one of the major factors directly and indirectly connected with the SDGs. Several recent studies have highlighted the importance of population for social equity (Chen et al. 2019), sustainable development (La Torre et al. 2019), and identifying the relationships among SDGs (Warchold et al. 2021). The population also influences the investment capital among countries. There are also other factors such as socio-political (Mukherjee and Chakraborty 2013), and governance (Stojanović et al. 2016) which might impact the relationships of SDGs.

## A need for a robust and unified quantitative approach to identify co-beneficial SDGs

All prior research has built a foundation for interpreting the SDGs' interactions from a global to a national perspective. However, most studies use qualitative methodology, including judgmental scoring or subjective selection of criteria. Several purely quantitative and data-driven statistical approaches have investigated how the goals are related (e.g. Bali Swain and Ranganathan 2021; Kroll et al. 2019; Lusseau and Mancini 2019; Pradhan et al. 2017; Stafford-Smith et al. 2017; Warchold et al. 2021). These studies help to assess the current state of the challenge; however, they mainly used pairwise correlation analysis. The main drawback of this approach is that the second, third or more levels of interactions among variables are ignored. For example, there might be a high correlation between a pair of goals in isolation but improving other SDGs can also affect that correlation positively or negatively. In other words, pairwise correlation analyses cannot reveal the complex associations between SDGs by ignoring the impact of more than one level of interactions. In fact, pairwise analysis is unable to identify indirect effects among goals.

Moreover, SDG-related data are multi-dimensional, heterogeneous, and highly correlated, with information from various sources of which some may be missing. Traditional statistical methods present challenges for computing relationships between goals. Some studies have used an explorative, data-driven approach with system dynamics or a machine learning technique to priorities the SDGs (Asadikia et al. 2021; Spaiser et al. 2017). However, these studies are based on global data, and their limitation is that the outcome might not apply to all countries.

This study offers a machine learning approach to overcome the limitations of traditional statistical methods so that SDGs' relationships can be identified.

The second section describes the data and method of analysis, while the third section elaborates on these results. Then, the fourth section covers the discussion, policy implications and limitations. Finally, the fifth section presents the conclusion.

## Methods and data

In this research, we use quantitative methods to identify co-beneficial SDGs. This section explains data resources, parameters, features and methods.

### Research methodology

This article describes four phases with six steps to identify co-beneficial SDGs. Figure 1 depicts the research methodology followed in this study.

In the first phase, input data are collected and prepared. In this phase, we begin with two steps. Data are collected from two different data sources in the first step (see "Source of data" section). In the second step of Phase 1, data are cleaned and merged as the primary dataset. The output of Phase 1 is passed to Phase 2 as an input. Then to set the context in Phase 2, predictors are identified (see "Context-sensitive SDGs prioritisation" section). In step 3, the primary dataset is converted to five data frames based on predictors (step 3, see Fig. 2). In the next phase (Phase 3), GBM models are fitted by tuned hyperparameters (step 4, see "Method of analysis" section for the method selection), and in step 5, models are validated (Fig. 7. Finally, in Phase 4, the result (co-beneficial SDGs) is illustrated (step 6).

### Context-sensitive SDGs prioritisation

SDGs try to balance sustainability's economic, social, and environmental objectives. This study examines one factor from each of those sustainability pillars based on what has been frequently cited in the literature.

From the economic perspective, one of the significant descriptors of the macroeconomic context of countries is income level (Galor and Zeira 1993) which also impacts sustainable development (Lusseau and Mancini 2019).

There are economic factors such as Gross Domestic Product (GDP) per capita, Gross National Income (GNI) per capita, and level of economic development that define a countries economic situation. GNI measures a country's overall economic conditions, including its foreign investments, when GDP includes only the income earned within a country. Since many countries earn much of their income from
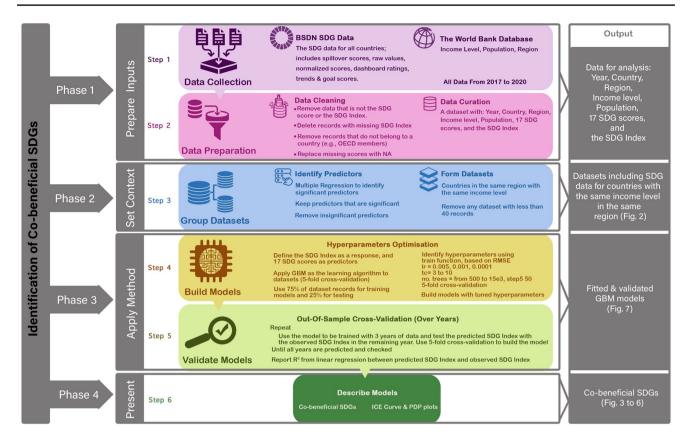
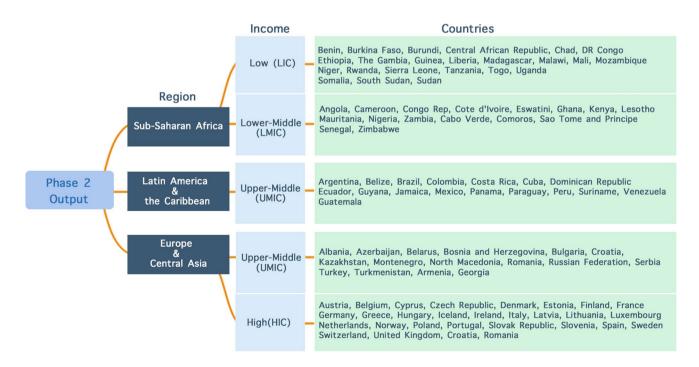**Fig. 1** Summary of the research methodology to identify co-beneficial SDGs



**Fig. 2** Data observations are grouped based on the region and income level. The last column includes a list of countries that their data are used for training the GBM models

abroad, which is included in GNI, we used the World Bank income classification of countries (Atlas method) based on the GNI per capita (The World Bank Group 2021). The income level classifications are high-income (HIC), upper-middle-income (UMIC), lower-middle-income (LMIC), and low-income (LIC). Since there is a high correlation between income level and development level and it may cause multi-collinearity, level of development is excluded.

A recent study suggests that region-specific identification of interactions among SDGs is crucial to facilitate contextual prioritisation (Bali Swain and Ranganathan 2021). To include an overarching perspective of natural resources, socio-political culture, territory, and environmental aspects, we added countries' geographic location as a contextual factor to be assessed in our analysis. To identify the significance of geographic location, this study uses the World Bank's seven region classifications (The World Bank Group 2021). The regions collated by the World Bank are East Asia and Pacific, Europe and Central Asia, Latin America and the Caribbean, Middle East and North Africa, North America, South Asia and Sub-Saharan Africa.

From the social perspective, the country's population is another factor that plays a mediating role in a wide range of public policy areas such as health, education, poverty, sustainable development, urban and environmental planning (Aguirre 2002; La Torre et al. 2019; Profiroiu et al. 2020).

The population data (indicator code SP.POP.TOTL) are obtained from the World Bank database (The World Bank Group 2021).

For this study, we first investigated the significance of the relationship between region, income level, and population with the SDG Index based on 640 observations in our dataset. To do this, we used Multiple Linear Regression (MLR), known as multiple regression with a confidence interval of 95%. The SDG Index is a dependent variable for this regression, while region, income level, and population are independent explanatory variables. We used dummy coding to convert each region to dichotomous variables and the income level to ordinal variables (1 to 4). The issue of multicollinearity has been tested by the Variance Inflation Factor (VIF). The VIF for all factors is less than five, indicating there is no multicollinearity among the independent variables. Regarding the residual analysis, the residual plot of the regression model is examined, and the plot has been found to be 'healthy' for the following reasons. First, the plot has random scatter around zero. The line is relatively straight (Linearity assumption). Second, the variances of the errors are about equal (Homoscedasticity assumption), and third the error terms do not appear to be related to adjacent error terms (Independence of error). In addition, the normality assumption of residuals is assessed, while the *p-value* of the Shapiro–Wilk test (0.280) suggests that residuals have no significant departure from normality. We also tested the

assumption of a linear relationship between the dependent variable and each independent variable. Since the population does not have a linear relationship with the SDG Index, it is removed from the multiple regression analysis.

This statistical analysis confirms there is a significant relationship between region and income level with the SDG Index. To do a custom-tailor prioritisation, we used both factors to identify co-beneficial SDGs for countries in the same region with the same income level.

The output of phase 2 leads to five datasets for which our models are trained and tested based on those observations. Figure 2 lists the countries that their SDG data is used to fit the machine learning models in this study. It is evident that some countries' income levels have changed from 2017 to 2020. In Fig. 2, we list those countries based on most of their observations (e.g. Sudan is listed as LIC from 2017 to 2019, but LMIC in 2020). Where there is an equal number of observations (e.g. Senegal was a LIC country in 2017 and 2018, but LMIC in 2019, and 2020), we listed them based on the most recent income level.

## Method of analysis

A regression model emerges as a suitable approach to examine the magnitude of the enhanced SDG Index by improving individual SDGs in which the relationship between a dependent (the SDG Index) and independent variable (17 goals) is explored. However, there are many reasons that traditional statistical methods cannot be suitable for the purpose of this study. The first reason is that there are 17 independent variables in our analysis that add complications to interpret the outputs because of the multicollinearity problem. Since there is a high correlation among some SDGs (Pradhan et al. 2017) (independent variables) a method capable of dealing with multicollinearity is required. The second reason why traditional statistical methods are not suitable for this study is that there is a non-linear relationship among many goals. This makes formulating or interpreting the higher-order interactions and indirect effects almost impossible.

Moreover, in multiple regression, it is recommended to remove independent variables that do not significantly affect the dependent variable to strengthen the model fit. However, since the aim is to explore the impact of all SDGs on the SDG Index, we want to include all goals as predictors. There are also missing scores in our dataset for some goals but we do not want to delete them. Consequently, a method which provides high predictive power and ability to cope with multicollinearity, handles uninformative independent variables, deals with missing data, and includes higher-order interactions can fit the purpose of this study.

Machine learning models already popular in the development fields (Khediri et al. 2021; Rahimi et al. 2021) have garnered more attention to overcome the SDG
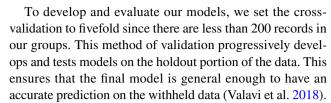
framework's complexity (Holloway and Mengersen 2018) from predicting poverty (SDG1) (Jean et al. 2016) to linking social programme supports (Gonzales Martinez 2019; Pincet et al. 2019). Machine learning techniques can measure SDGs indicators by analysing earth observations, satellite imagery and remote sensing data (e.g. Andries et al. 2019; Efremova et al. 2019; Holloway et al. 2018). In contrast to statistical approaches that start with some assumptions, machine learning uses an algorithm to understand the relationship between the response and its predictors by finding dominant patterns (Breiman 2001).

GBM is an ensemble machine learning algorithm that combines boosting and regression trees. GBM is a popular and powerful supervised learning algorithm that can effectively capture complex non-linear function dependencies (Friedman 2001). This method is widely used in practice to explore the complex relationships between variables. GBM models are flexible, can accommodate missing data and do not require outliers to be removed (Breiman et al. 2017). In these models, influential independent variables will be selected without being affected by multicollinearity (Buston and Elith 2011). To fit models with GBM, interaction depth can be set up to 10 levels by the Tree-Complexity ($tc$). This parameter is used to consider interactions among dependent variables by setting the interaction depth; when $tc$ of 1 means no interaction, and 10 means high interaction (ten levels of interaction).

Model selection in GBM provides a coherent and robust alternative to traditional approaches such as stepwise variable selection (Elith et al. 2008). Growing trees stagewise and slowly causes the GBM models to function better than a single decision tree model (Elith et al. 2008). In this method, the contribution of each tree shrinks according to the learning rate ($lr$). The fitted values are calculated based on the total of all trees multiplied by the $lr$. In general, a smaller $lr$ and larger number of trees are preferable (Elith et al. 2008). The feature of controlling $tc$ and $lr$ by the data scientist makes GBM models superior to similar methods like Random Forest by providing robust predictive performance and explanatory power (Elith et al. 2008). Since this study is interpreting the behaviour of SDGs (descriptive), GBM is preferred over Random Forest and other single decision tree models.

In the gradient boosting algorithm, the variable's "relative influence" is calculated as the average improvement made by each variable across all the trees that the variable is used. This improvement in the split-criterion is calculated based on the Mean Squared Error (MSE) in regression.

For fitting models, we used gbm.step function which is part of the dismo package (Hijmans et al. 2017) in R 3.6.3 (R Core Team 2021). The selection of hyperparameters is explained in detail in Appendix A.

To develop and evaluate our models, we set the cross-validation to fivefold since there are less than 200 records in our groups. This method of validation progressively develops and tests models on the holdout portion of the data. This ensures that the final model is general enough to have an accurate prediction on the withheld data (Valavi et al. 2018).

Apart from fivefold cv, out-of-sample cv helps evaluate the model, and its generalisation. We trained the algorithm (Appendix B) with three years of data for this validation, and then used the model to predict the fourth year SDG Index. This process is executed four times to predict all years using the trained model in the algorithm. Then $R^2$ is calculated from the linear regression between the predicted index and the SDG Index in the testing data set.

After identifying the highest influential variables, we used Partial Dependence Plots (PDs) (Friedman 2001), and Individual Conditional Expectation (ICE) plots (Goldstein et al. 2015) to demonstrate how the response variable changes based on the most relevant variables. These plots are visual, model-agnostic techniques that can help to explain insights from "black-box" machine learning algorithms (Goldstein et al. 2015).

## Method limitations

The main limitation in the GBM method is the minimum number of records required to fit models (almost 40). North America, East Asia and Pacific, South Asia, Middle East and North Africa records did not meet this criterion; hence they are omitted from our study. Another limitation is that although GBM can deal with missing scores of SDGs, the quality of data can impact learning curves.

## Source of data

There is an expanding list of 231 unique indicators to assess progress towards the SDGs (UN Statistical Commission 2020). The UN has classified these indicators into three different tiers (UN 2017). This categorisation is based on the availability of data and method of measurement. A variety of data resources can provide the required SDG data for Tier 1 and some Tier 2 indicators (Sachs et al. 2020; United Nations 2020; World Bank 2020).

It is evident that SDGs scores can be affected if inconsistent indicators or methods are used for calculation purpose (Miola and Schiltz 2019). Bertelsmann Stiftung and Sustainable Development Solutions Network (BSDN) gathered data for 111 indicators, of which 86 apply to all countries plus 35 indicators for OECD countries to measure the score of each goal. Those indicators are available for at least 80% of the UN member states whose populations are greater than 1 million. These data have been described as "the most comprehensive picture of national progress on the SDGs and

offers a useful synthesis of what has been achieved so far" (Nature Sustainability Editorial 2018).

Various methods are proposed to measure the SDGs progress, such as measuring the distance to targets (OECD 2019), or progress (Eurostat 2018), based on the best and worst achiever (Lafortune et al. 2018), and a method that captures the interaction between goals by rewarding synergies and penalising trade-offs (Biggeri et al. 2019). To assess where each country stands, BSDN calculates the arithmetic mean of SDG scores and this is called the SDG Index. The SDG Index is a single composite measure to quantify the progress of each country towards achieving SDGs (Sachs et al. 2016) and can be used for "identifying priority areas for action" (Biggeri et al. 2019). For this study, since we needed a single composite measure that aggregated equally weighted goals, we used the SDG Index as the overall SDGs achievement that countries made. The method and data used to calculate SDG scores and the SDG Index are explained in the BSDN methodology report in detail (Lafortune et al. 2018).

To improve models' performance with more observations, we used 17 SDG scores and the SDG Index of countries in the BSDN dataset from 2017 to 2020 (Sachs et al. 2020). Records of those countries without the SDG Index are deleted. In total, 640 observations (157, 156, 162, and 165 in 2017, 2018, 2019, and 2020, respectively) describe the SDGs behaviour in this analysis. The parameter of year and country names are not included in the analysis; however, 'year' is used to cross-validate results (refer to Method of Analysis). Table 1 lists SDGs icons used to represent each goal in the result section with their name.

## Results

The score of 17 SDGs (independent variables) along with the SDG Index (dependent variable) of countries are used to train our GBM models for regression. Tuning hyperparameters with the training function returned the *lr* of 0.001 for all models and the *tc* of 3 for all groups except Europe and Central Asia HIC. For this group, the *tc* of 5 fits the model better.

Table 2 in Appendix C lists how many trees the gbm.step function was used to fit each model along with the fivefold cross-validation result. The result of this cross-validation indicates how well our models are fitted. Considering the range of independent variables from 0 to 100, calculated RMSE and Mean Absolute Error (MAE) are perfect.

To evaluate whether this result is because the models are over-fitted or not, we have conducted out-of-sample

**Table 1** SDGs with their icons used in this study. Credit (icon): The United Nations



| Goal | | Goal | |
|---|---|---|---|
| 1 | No poverty | 10 | Reduced inequalities |
| 2 | Zero hunger | 11 | Sustainable cities and communities |
| 3 | Good health and well-being | 12 | Responsible consumption and production |
| 4 | Quality education | 13 | Climate action |
| 5 | Gender equality | 14 | Life below water |
| 6 | Clean water and sanitation | 15 | Life on land |
| 7 | Affordable and clean energy | 16 | Peace, justice and strong institutions |
| 8 | Decent work and economic growth | 17 | Partnerships for the goals |
| 9 | Industry, innovation and infrastructure | **Total** | **17** |

accuracy validation. For this validation, data for three years are used to fit the model, and the trained model is used to predict the fourth year for testing. Then, the predicted and observed SDG Index are compared. Fig. 7 in Appendix D displays the error rate for out-of-sample validation over the years. It can be seen that the observed and predicted SDG Index falls well on the straight line. When the model is fitted, the coefficient of determination ($R^2$) between the predicted values and actual data is $\geq 0.7$. This means more than 70% of the SDG Index variance is explained by the variability of the SDGs.

## Region-income-based SDGs prioritisation

In all models, improving each goal has a positive influence on the SDG Index; however, the magnitude of their influence varies. For countries in the same region with the same income level, some goals resulted in a higher SDG Index than their neighbours.

The top five co-beneficial SDGs, which made a difference to having a higher SDG Index in Europe and Central Asia, HIC and UMIC countries are depicted in Fig. 3.

Reducing inequality within and among countries (SDG10) and building resilient infrastructure (SDG9) are the most influential SDGs in the HIC and UMIC countries of Europe and Central Asia, respectively. After SDG10, strengthening the means of implementation besides revitalising the global partnership for sustainable development (SDG17), decent work and economic growth (SDG8) and ensuring sustainable consumption and production patterns (SDG12) are SDGs which are among the top co-beneficial goals in Europe and Central Asia's HIC countries. Notably SDG17, and SDG12 have been identified as influential SDGs only in HIC countries for Europe and Central Asia.

Promoting sustainable use of terrestrial ecosystems (SDG15) and taking urgent action to combat climate change (SDG13) are among the most influential SDGs for UMIC countries in Europe and Central Asia. These two goals have not been identified in any other models as the top 5 influential goals.

In Sub-Saharan Africa (Fig. 4) for both LMIC and LIC countries good health and well-being (SDG3) constitutes the most co-beneficial goal of all. The second- and third-ranked SDGs with more than 10% relative influence on the SDG Index in LMIC countries in Sub-Saharan Africa are ensuring inclusive and equitable quality education (SDG4), and promoting sustained, inclusive and sustainable economic growth (SDG8).

It is notable that building resilient infrastructure (SDG9), which has the second most importance in LIC countries of Sub-Saharan Africa, is also one of the top 5 influential SDGs in LMIC countries of this region. Even though LIC countries score lower than LMIC countries in SDG1 (Sachs

et al. 2020), this goal is one of the main influential goals in Sub-Saharan Africa UMIC countries but not in LIC countries. In those countries making cities human inclusive, safe, resilient, and sustainable (SDG11), achieving food security, improving nutrition, promoting sustainable agriculture (SDG2), and achieving gender equality and empowering women and girls (SDG5) are among the five top influencer goals. The results for the UMIC countries of Latin America and the Caribbean (Fig. 5) reveal that achieving gender equality and empowering women and girls (SDG5) is the top co-beneficial goal which improves the SDG Index by almost 30%. SDG3 and SDG9 each wield more than 10% influence on the SDG Index of UMIC countries in this region. Sustainable use of the oceans, seas and marine resources (SDG14) is the influential SDG which uniquely has been identified as co-beneficial only in Latin America and the Caribbean, UMIC countries.

## PD and ICE plots

To depict how the GBM identified the most relative variables, we plotted the top three influential variables, which accumulate 50% of the relative influence on the SDG Index (Fig. 6). ICE plots refine the partial dependence plot by graphing the functional relationship between the predicted response, and the feature for individual observations. These plots enable data scientists to drill much deeper to explore individual differences and identify subgroups and interactions between model inputs. It can be seen from the ICE plots that the average of observations is fairly well aligned with individual inputs. It is also apparent that an improvement in each goal exerts a unique impact on improving the SDG Index. For many goals, it is a steplike positive trend in the predicted SDG Index.

However, some goals show a different pattern. SDG9 and SDG13 in UMIC countries of Europe and Central Asia have shown a different pattern compared to the goals, even in other models. In those countries, SDG9 has a substantial impact on the SDG Index when the goals score is about 28%, but then after that, its impact is relatively steady, even when the goal score is double. SDG13 has shown no improvement in the predicted SDG Index when the goal score is less than 68%. However, it can be seen from the plot that when the score of this goal is above 68%, the improvement starts with a sharp rise, followed by a minor steplike enhancement. Interestingly, a minor trade-off (negative influence) can be seen in SDG15 of UMIC countries of Europe and Central Asia when the goal score is between 50 and 60%. Some declines are noted in all top three influential goals of UMIC countries of Latin America and the Caribbean (SDG5, SDG3, and SDG9). However, even with those declines, improvement in this goal positively influences the predicted SDG Index. In
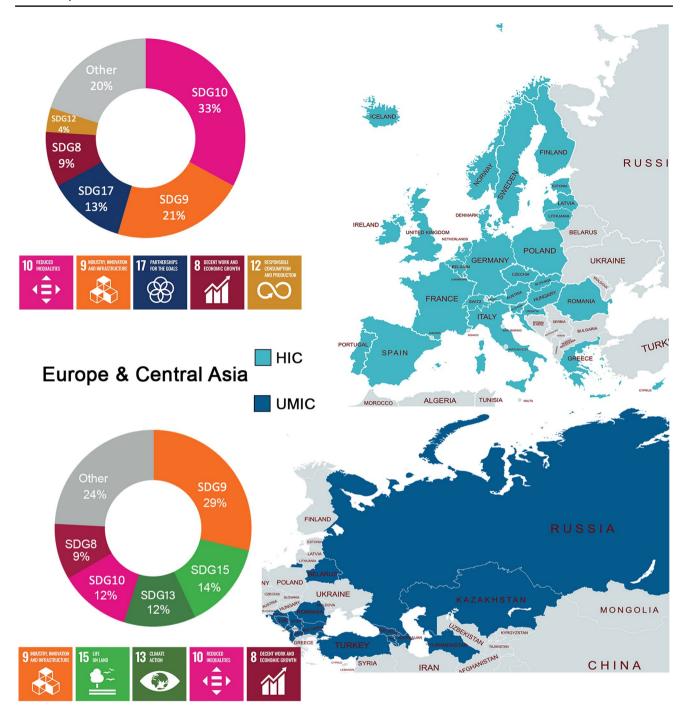
**Fig. 3** Top five high influential SDGs of HIC (light blue) and UMIC (dark blue) countries of Europe and Central Asia. Credit (maps): mapchart. net, Credit (SDG icons): United Nations

LMIC countries of Sub-Saharan Africa, SDG3 depicts a sharp increase when the score exceeds 52%. Meanwhile it remains steady for scores greater than 60%. Apart from HIC countries in Europe and Central Asia, SDG9 shows a minimum impact on the predicted SDG Index of other models after a certain score.

## Discussion

This study used a machine learning approach to overcome the limitations of traditional statistical methods to identify co-beneficial SDGs among 17 goals. The results of this
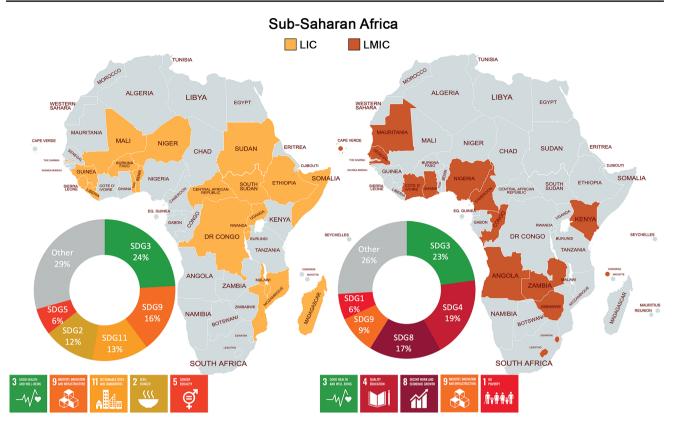
**Fig. 4** Top five high influential SDGs of LIC (light orange) and LMIC (dark orange) countries of Sub-Saharan Africa. Credit(maps): mapchart. net, Credit (SDG icons): United Nations

analysis confirm region and income level have a statistically significant relationship with the SDG Index of countries. Fitting GBM models for countries in the same region with the same income level supports us to identify the influential SDGs in those nations. Findings demonstrate that in general, improvement in all goals (for countries included in our analysis) has a positive influence on the SDG Index.

However, the degree of this influence changes depending on two things: the region, and income level of countries.

Promoting some goals significantly affects the SDG Index, while in others it can be minor. Although some SDGs are identified as co-beneficial in all models (e.g. SDG9), or some (e.g. SDG8 and SDG3), the majority are not acting the same, even in countries with the same level of income located in diverse regions (e.g. UMIC countries of Europe and Central Asia versus Latin America and the Caribbean). It is shown (see Fig. 6) that the pattern in improving co-beneficial goals can vary in the same region, but at different income levels (e.g. Europe and Central Asia HIC versus LIC countries). This strongly suggests how important region and income level are in terms of influence on the SDGs' behaviour.

## Co-beneficial SDGs

The result has shown that SDG10 is the most influential goal for HIC countries in Europe and Central Asia. Income inequality directly impacts population health (Wilkinson and Pickett 2006), mortality, obesity, teenage birth rates, mental illness, low levels of trust, hostility, racism, poor education of children, greater imprisonment rates, drug overdose mortality, etc. (Wilkinson and Pickett 2007). This goal directly or indirectly reshapes many other SDGs in HIC countries (Lusseau and Mancini 2019). Although improving SDG10 promotes all goals in these countries (Kroll et al. 2019; Lusseau and Mancini 2019), a study on HIC nations from 1993 to 2013 suggests that one widely recognised social, political, moral, and macroeconomic problem in HIC countries is rising income inequalities (Tomaskovic-Devey et al. 2020). Norway, Sweden, and Slovenia are countries that scored highest in the SDG10; these countries are among the top 10 nations in terms of their progress towards achieving SDGs (Sachs et al. 2020). This study suggests that reducing inequality and improving SDG10 in HIC countries of Europe and Central Asia can significantly enhance other SDGs and, consequently, lead to the SDGs being attained more by those countries.
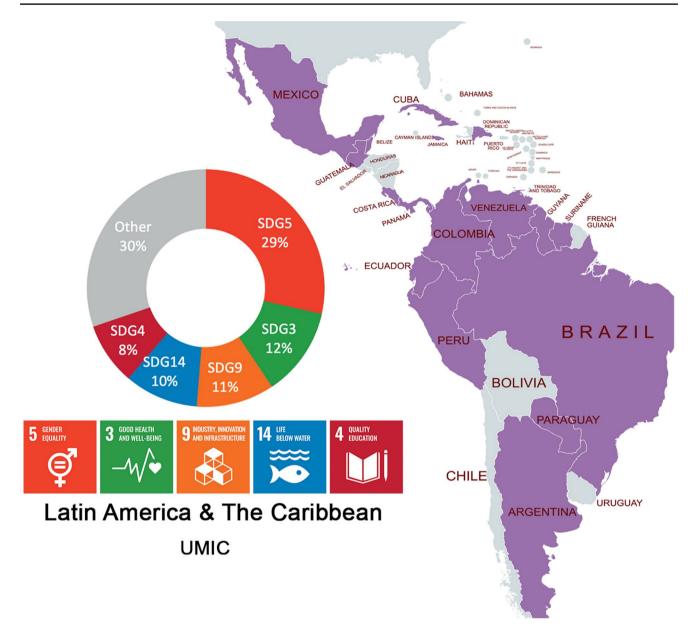
**Fig. 5** Top five high influential SDGs of UMIC countries of Latin America & the Caribbean. Credit (maps): mapchart.net, Credit (SDG icons): United Nations

Another influential goal evident in every model is building resilient infrastructure and promoting sustainable industrialisation (SDG9). This goal covers several related aspects, namely technology adoption, science-based development, and sustainable transportation systems. Despite the title given to SDG9, which seems to refer more to economic characteristics, this goal's targets can improve environmental and social SDGs (Mantlana and Maoela 2019). This goal directly affects SDG8, promoting sustainable economic growth which is also among the top co-beneficial goals in both Europe and Central Asia and

the LMIC countries of Sub-Saharan Africa. This positive impact on other goals through achieving SDG9 can lead to a significant improvement of those countries' overall SDGs attainment.

Identifying SDG5 as the top influencer goal for Latin America and the Caribbean indicates the importance of gender equality for countries in that region. SDG5 aims to "eliminate all forms of discrimination, and violence against women in public and private spheres and to undertake reforms to give women equal rights to economic resources and access to ownership of property".
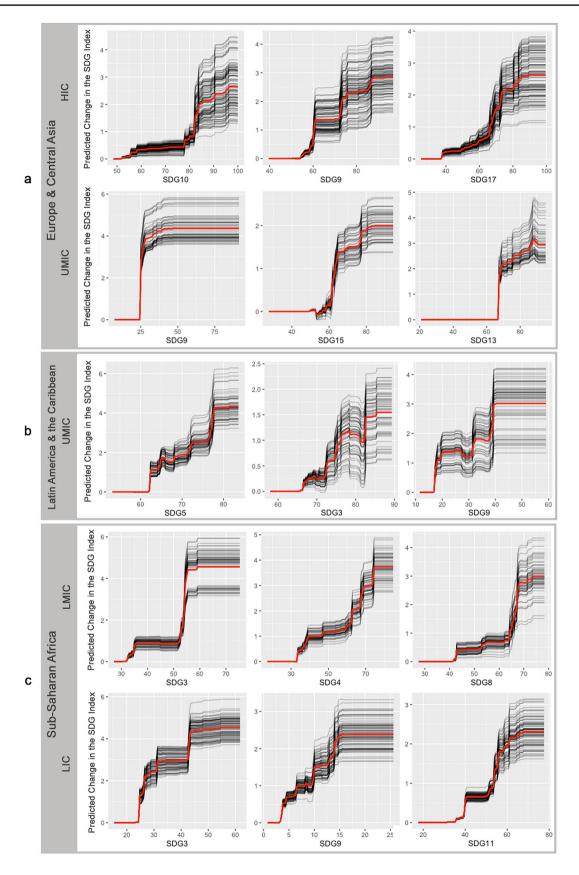
**Fig. 6** ICE plots for top three influential SDGs of each group. The red line displays the average of individual/black lines

In Latin America and the Caribbean region, one in six girls annually enters either a formal or informal relationship before the actual legal age of adulthood, i.e. 18 (OECD 2020). In this region, women are still vulnerable to many forms of harassment in public areas, workplaces, and education institutions. Furthermore, the extent of inequitable distribution of benefits to women is so immense that in nine Latin American and Caribbean countries women's access to some professions is restricted if not denied outright (OECD 2020). A recent report of the gender snapshot on progress on the SDGs (The UN Women 2020) shows both the importance of SDG5 to achieve sustainability, and its relationship with other SDGs. With this knowledge that SDG5 is one of the top influential goals in Latin America and the Caribbean, it suggests that the COVID-19 pandemic has hugely compromised this goal (The UN Women 2020) and may cause the SDG Index of those countries to drop significantly.

In one study, SDG1 "No Poverty" is identified as one of the most synergetic goals for LIC countries (Lusseau and Mancini 2019). However, it only has been identified as the fifth influential goal for LMIC countries in the Sub-Saharan Africa region in our analysis. For LIC and LMIC countries of Sub-Saharan Africa, SDG3 has been identified as the most influential goal that can significantly shape their SDG Index. Although eliminating poverty is an important objective, without good health, it is unlikely that poverty will actually be eliminated. People's lack of access to health care entrenches poverty. Almost 100 million people were pushed into extreme poverty worldwide because of health expenditures (Daepp and Arcaya 2017). Good health is also identified as one of the essential factors that can make economic growth possible in Sub-Saharan Africa (Ogundari and Awokuse 2018). Promoting national health insurance and strengthening the health system has emerged as a solution "to get rid of the poverty trap" in the Sub-Saharan region (Wang et al. 2021).

SDG4 is the second influencer in LMIC countries of Sub-Saharan Africa. Education can promote economic growth (Jeong et al. 2020), and economic growth contributes to eliminating extreme poverty by providing opportunities to find employment. Quality education is also identified as a critical factor in "addressing environmental and sustainability issues and ensuring human well-being" (Council of Science 2015; WHO 2018). Unfortunately, this goal is another SDG that has been greatly impacted by the COVID-19 pandemic in LIC and LMIC countries.

SDG2 is another influential goal for LIC countries of Sub-Saharan Africa. The COVID-19 pandemic may disrupt food systems, resulting in hunger, malnutrition, and food insecurity in Sub-Saharan Africa (Nchanji et al. 2021).

Armed with this knowledge that such a pandemic not only impacts on the SDG3, but also SDG1, SDG2, and SDG4 which are among the top influential goals in Sub-Saharan Africa, this might significantly help to reduce the SDG Index of those countries.

The PD and ICE curve plots provide a local explanation of how the "black-box" of the GBM algorithm calculated the most influential SDGs. These visualisations made the unique nature of SDGs' behaviour obvious. Each goal follows a unique pattern to improve the SDG Index even in the same region with varying income levels (e.g. SDG9 in Europe and Central Asia HIC versus UMIC countries). This clearly indicates that interpreting relationships between goals without considering influencing factors cannot accurately interpret SDGs behaviour. The non-linearity of the relationships between the SDGs and the SDG Index has been made evident by plotting individual goals and their impact on the predicted SDG Index.

## Policy implications

Although SDGs are global goals, nations and their governments are responsible for implementing the Agenda 2030. All UN members have agreed to adopt the SDGs framework and achieve goals but these do not seem to be an urgent priority. There are scarce resources and governments have only limited budgets for fulfilling "*to do*" lists. Given the current damage wrought globally by the COVID-19 pandemic, perhaps not all 17 goals can be attained by the due date, however, a well-focused investment strategy that improves co-beneficial goals can accelerate SDGs achievement. Scientific evidence assists government, industry, and policy-makers to identify goals for investment to improve them, in order to maximise the SDGs' overall attainment.

The positive influence of building resilient infrastructure, promoting sustainable industrialisation and fostering innovation on other SDGs confirms the importance of this goal to meet the 2030 Agenda. Research shows that middle-income countries at an early stage of industrialisation have more opportunities to enjoy fast economic growth and introduce policies focusing on structural transformation (Kynčlová et al. 2020). Moreover, adopting technology in various industries helps environmental sustainability and improves the long-term viability of society and the economy (Bai et al. 2020). Improvement in introducing and promoting new technologies, facilitating international trade and enabling the efficient use of resources can open new horizons for countries included in this study, specifically UMIC countries of Europe and Central Asia.

The importance of achieving gender equality and empowering all women and girls in Latin America and the Caribbean has been discussed above. Currently, there is no

protective policy in several Caribbean countries to delink women's marital status from their citizenship rights (OECD 2020). Moreover, there is no flawless legal framework that has been mandated to protect women from various sorts of violence, such as sexual assault, rape, and domestic violence (OECD 2020). The development of legal frameworks, especially about child marriage, violence against women, workplace rights and political voice, can be the starting point to improve those countries' SDGs overall. This also requires both advancements being made in the public sector and the legal system, and awareness of stakeholders and the public.

In terms of the Sub-Saharan Africa region, a recent study has found that "to get rid of the poverty trap in this region, the government needs to promote the national health insurance and strengthen the health system as a whole plan" (Wang et al. 2021). According to Ifeagwu et al. (2021), 27 out of 48 countries in Sub-Saharan Africa are affected by direct out-of-pocket payments for healthcare services. In 2018 the average spending on the health system in this region was recorded at about 5% of GDP, with some countries investing less than 3% of their GDP (World Bank 2018). During this time, the average expenditure for most countries virtually doubled this amount. Policies that support government subsidies of healthcare and reduce childhood early deaths help Sub-Saharan African countries improve goal 3 which is linked to many other objectives such as SDG1.

It is worth stating that to exploit synergies among SDGs, policy coherence and relevant, effective actions are critical. Identification of co-beneficial SDGs might be meaningless without coherent development of policies to improve them.

### Limitations

This article investigates the impact of SDGs on the SDG Index based on countries' geographic location and income levels. Since the analysis is data-driven, it is essential to note that SDG scores are calculated based on indicators, of which the majority belong to the Tier 1 classification. Including Tier 2 and Tier 3 indicators in goal scores might affect the result.

Moreover, in the BSDN SDG data, some indicators measure progress towards goals that are not the same as UN indicators. Apart from some differences in those indicators, many Tier 1 indicators have not been included in the score calculation due to data limitations. Hence, alternation in indicators or adding more indicators can change SDG scores. Furthermore, although machine learning algorithms are less reliant on judgemental or biased criteria, they depend on data quality. More analysis is required when data become available for missing indicators. The findings of this research are based on cross-sectional data, and subsequently,

causality cannot be inferred. It is also important to mention this study has been conducted with one factor from each aspect of sustainability for contextual prioritisation. There might be other factors that impact on SDGs relationship which are not included in this study. Furthermore, this investigation is for countries with the same income level in the same region. Hence, it does not provide a specific and detailed prioritisation for an individual country.

### Conclusion

The holistic and complicated nature of the 17 SDGs objectives coupled with the high level of economic destruction brought on by the COVID-19 pandemic has severely impaired the progress towards attaining such goals. With less than ten years left to meet the goals of the Agenda 2030, optimising the budget and resource allocation is necessary. To optimise this investment, prioritising co-beneficial SDGs to boost the overall SDG achievement is essential. However, this prioritisation requires excluding any judgmental or subjective criteria, based on maximising overall achievement. Identifying co-beneficial SDGs are critical to inform the policy maker for prioritising strategies and plans for an integrated approach.

In this article, we employed a machine learning method (GBM) that is capable of identifying the relative co-beneficial goals systematically when more than two levels of interactions are included.

This study sheds light on how SDGs behave uniquely in various contexts. The result confirms that even though the SDG Index is the arithmetic means of SDG scores, the relationship with goals is not linear. Promoting some goals (depending on the context) can significantly improve the SDG Index when improvement in others might have a minor effect. Even the same goal acts differently in relation to boosting the SDG Index. It is important not to generalise how goals are behaving from a global perspective or only based on the income level of countries. The results of this study constitute supporting evidence for why prioritisation of SDGs should be based on a country's features rather than a global-scale assessment. In all models in this study, combinations of the social, economic, and environmental SDGs are identified as the top five influential goals and no dimension of sustainability has been "left behind". This emphasises the importance of having a holistic approach when prioritising SDGs. Science and technology can help to ensure that economic progress is sustainable when the planet's natural resources and environment are protected and managed responsibly, considering human rights.

## Future studies

There is a need for additional research investigating the relationships of countries features with the SDGs to identify what other factors have a significant relationship with the SDG Index. Further investigation is also required at the national level for each country in which other factors such as their natural resources (such as land or water availability), cultural contexts, human resources, governance, and institutional capacity are included.

It is noteworthy that analyses conducted in this study are at the goal level; it is worth exploring the relationships of targets or indicators that have the most significant influence on co-beneficial goals.

## Appendix A

### Hyperparameter setting

In the gbm.step function, apart from dependent and independent variables, we set the method (family) of regression as gaussian, bag.fraction to 0.75 (recommended for small training samples). According to Greg Ridgeway (2007) "shrinkage$=0.001$ will almost certainly result in a model with better out-of-sample predictive performance than setting shrinkage$=0.01$". However, lower $lr$ (shrinkage) will impact negatively on the computation time. Since our dataset is not large, we ignored the computing time (the difference is less than 1 min) and memory usage. Instead, we focused on setting hyperparameters based on predictive performance with fewer error rates without over-fitting our models. This assists us to establish a more precise interpretation of SDGs relationships.

In GBM the $tc$ and the $lr$ determine the required number of trees for optimal prediction (Elith et al. 2008). To tune those hyperparameters ($tc,lr$) we used the train function which is part of the Caret package (Kuhn 2021) with cross-validation (cv) method using R 3.6.3 (R Core Team 2021). The method is set as "gbm", and the metric is based on Root Mean Square Error (RMSE). The range of 3–10 for $tc$, and three options for $lr$ (0.005, 0.001, 0.0001), and the number of trees set as minimum 500, maximum 15e3 with steps 50.

## Appendix B

Algorithm: Out-Of-Sample cross-validation over years by authors

| | Inputs | $Y$:set of years, $Y=$ {2017, 2018, 2019, 2020} |
|---|---|---|
| | | $y$: time, $y \in Y$ |
| | | $i_y$: SDG Index of year $y$ |
| | | $i'_y$: Predicted SDG Index of year $y$ |
| | | $g_{ky}$: Goal number $k$ in year $y$, $k \in$ {1,2,...,17} |
| | Output | : $R^2$ |
| 1 | INITIALISE | |
| 2 | SET $tc \leftarrow$ tuned $tc$ | |
| 3 | SET $lr \leftarrow$ tuned $lr$ | |
| 4 | SET n.fold $\leftarrow 5$ | |
| 5 | FOR $y$ in $Y$ DO: | |
| 6 | UPDATE test_set $\leftarrow i_y, g_{ky} \forall_k$ | |
| 7 | UPDATE train_set $\leftarrow i_{/\{y\}}, g_{k/\{y\}} \forall_k$ | |
| 8 | fitted_model $\leftarrow$ RUN BRT (train_set, $tc$, $lr$, n.fold) | |
| 9 | fitted_model $\leftarrow$ RUN BRT (train_set, $tc$, $lr$, n.fold) | |
| 10 | $i'_y \leftarrow$ Fitted_Model ( $g_{ky} \forall_k$) | |
| 11 | $R^2 \leftarrow linearRegression (i_y, i'_y)$ | |
| 12 | RETURN $R^2$ | |

## Appendix C

See Table 2.

## Appendix D

See Fig. 7.

**Table 2** Number of trees, lr, and tc used to fit each model using gbm.step function along with the result of RMSE and MAE calculated from fivefold cross-validation

| Region-income groups | $lr$ | $tc$ | No. trees | RMSE | MAE |
|---|---|---|---|---|---|
| Europe and Central Asia-HIC | 0.001 | 5 | 12,150 | 0.33 | 0.13 |
| Europe and Central Asia-UMIC | 0.001 | 3 | 9500 | 0.85 | 0.56 |
| Sub-Saharan Africa-LMIC | 0.001 | 3 | 15,000 | 0.31 | 0.21 |
| Sub-Saharan Africa-LIC | 0.001 | 3 | 13,850 | 0.23 | 0.15 |
| Latin America and the Caribbean-UMIC | 0.001 | 3 | 10,950 | 0.56 | 0.35 |

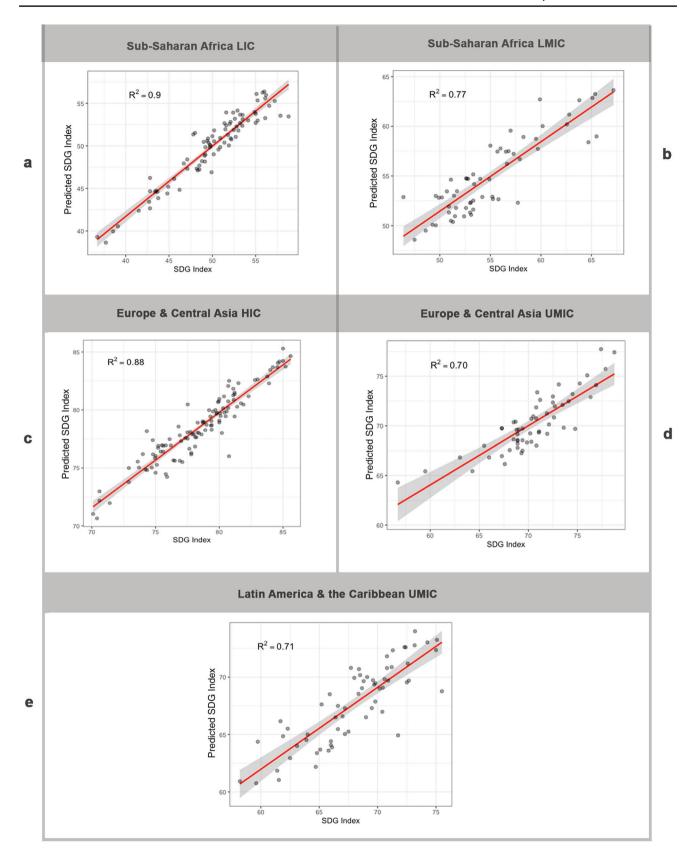**Fig. 7** The results of model validation. The x-axis is the actual SDG Index, and the y-axis is the predicted SDG Index

# References

Aguirre MS (2002) Sustainable development: why the focus on population? Int J Soc Econ 29:923–945

Alcamo J, Thompson J, Alexander A, Antoniades A, Delabre I, Dolley J, Scharlemann JPW (2020) Analysing interactions among the sustainable development goals: findings and emerging issues from local and global studies. Sustain Sci 15(6):1561–1572. https://doi.org/10.1007/s11625-020-00875-x

Allen C, Metternicht G, Wiedmann T (2018) Prioritising SDG targets: assessing baselines, gaps and interlinkages. Sustain Sci 14(2):421–438. https://doi.org/10.1007/s11625-018-0596-8

Allen C, Metternicht G, Wiedmann T, Pedercini M (2019) Greater gains for Australia by tackling all SDGs but the last steps will be the most challenging. Nat Sustain 2(November):1041–1050. https://doi.org/10.1038/s41893-019-0409-9

Andries A, Morse S, Murphy RJ, Lynch J, Woolliams ER (2019) Seeing sustainability from space: using Earth observation data to populate the UN sustainable development goal indicators. Sustainability 11(18):5062

Asadikia A, Rajabifard A, Kalantari M (2021) Systematic prioritisation of SDGs: machine learning approach. World Dev 140:105269. https://doi.org/10.1016/j.worlddev.2020.105269

Baffoe G, Zhou X, Moinuddin M, Somanje AN, Kuriyama A, Mohan G, Takeuchi K (2021) Urban–rural linkages: effective solutions for achieving sustainable development in Ghana from an SDG interlinkage perspective. Sustain Sci (Tacoli 2004). https://doi.org/10.1007/s11625-021-00929-8

Bai C, Dallasega P, Orzes G, Sarkis J (2020) Industry 4.0 technologies assessment: a sustainability perspective. Int J Prod Econ 229:107776. https://doi.org/10.1016/j.ijpe.2020.107776

Bali Swain R, Ranganathan S (2021) Modeling interlinkages between sustainable development goals using network analysis. World Dev 138:105136. https://doi.org/10.1016/j.worlddev.2020.105136

Bandari R, Moallemi EA, Lester RE, Downie D, Bryan BA (2021) Prioritising sustainable development goals, characterising interactions, and identifying solutions for local sustainability. Environ Sci Policy. https://doi.org/10.1016/j.envsci.2021.09.016

Barbier EB, Burgess JC (2019) Sustainable development goal indicators: analyzing trade-offs and complementarities. World Dev 122:295–305. https://doi.org/10.1016/j.worlddev.2019.05.026

Biggeri M, Clark DA, Ferrannini A, Mauro V (2019) Tracking the SDGs in an 'integrated' manner: a proposal for a new index to capture synergies and trade-offs between and within goals. World Dev 122:628–647. https://doi.org/10.1016/j.worlddev.2019.05.022

Breiman L (2001) Statistical modeling: the two cultures. Stat Sci 16(3):199–215. https://doi.org/10.1214/ss/1009213726

Breiman L, Friedman JH, Olshen RA, Stone CJ (2017) Classification and regression trees. pp 1–358. https://doi.org/10.1201/9781315139470

Buston PM, Elith J (2011) Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. J Anim Ecol 80(3):528–538. https://doi.org/10.1111/j.1365-2656.2011.01803.x

Chen Y, Bouferguene A, Shen Y, Al-Hussein M (2019) Assessing accessibility-based service effectiveness (ABSEV) and social equity for urban bus transit: a sustainability perspective. Sustain Cities Soc 44:499–510

Council of Science (2015) Review of the SDGs: the science perspective. ICSU, Paris

Daepp MIG, Arcaya MC (2017) The effect of health on socioeconomic status: using instrumental variables to revisit a successful randomized controlled trial. Econ Hum Biol 27:305–314. https://doi.org/10.1016/j.ehb.2017.09.002

Efremova N, West D, Zausaev D (2019) AI-based evaluation of the SDGs: the case of crop detection with earth observation data. ArXiv Preprint ArXiv:1907.02813

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77(4):802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Eurostat (2018) Sustainable development in the European Union: monitoring report on progress towards the SDGs in an EU context. Publications Office of the European Union, Luxembourg. https://doi.org/10.2785/221211

Friedman J (2001) Greedy function approximation: a gradient boosting machine. Author(s): Jerome H. Friedman source. Ann Stat 29(5) (2001), 1189–1232. Institute of Mathematical Statistics Stable. Ann Stat 29(5):1189–1232

Galor O, Zeira J (1993) Income distribution and macroeconomics. Rev Econ Stud 60(1):35–52. https://doi.org/10.2307/2297811

Goldstein A, Kapelner A, Bleich J, Pitkin E (2015) Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. J Comput Graph Stat 24(1):44–65. https://doi.org/10.1080/10618600.2014.907095

Gonzales Martinez R (2019) Which social program supports sustainable grass-root finance? Machine-learning evidence. Int J Sust Dev World 00(00):1–7. https://doi.org/10.1080/13504509.2019.1706059

Hazarika R, Jandl R (2019) The nexus between the Austrian forestry sector and the sustainable development goals: a review of the interlinkages. Forests 10(3):205. https://doi.org/10.3390/f10030205

Hijmans RJ, Phillips S, Leathwick J, Elith J (2017) Dismo: species distribution modeling

Holloway J, Mengersen K (2018) Statistical machine learning methods and remote sensing for sustainable development goals: a review. Remote Sens. https://doi.org/10.3390/rs10091365

Holloway J, Mengersen K, Helmstedt K (2018) Spatial and machine learning methods of satellite imagery analysis for sustainable development goals. In: Proceedings of the 16th conference of international association for official statistics (IAOS). International Association for Official Statistics (IAOS), pp 1–14

Horn P, Grugel J (2018) The SDGs in middle-income countries: setting or serving domestic development agendas? Evidence from Ecuador. World Dev 109:73–84. https://doi.org/10.1016/j.worlddev.2018.04.005

Howe P (2019) The triple nexus: a potential approach to supporting the achievement of the sustainable development goals? World Dev 124:104629. https://doi.org/10.1016/j.worlddev.2019.104629

Ifeagwu SC, Yang JC, Parkes-Ratanshi R, Brayne C (2021) Health financing for universal health coverage in Sub-Saharan Africa: a systematic review. Glob Health Res Policy. https://doi.org/10.1186/s41256-021-00190-7

Jayaraman R, Colapinto C, La Torre D, Malik T (2015) Multi-criteria model for sustainable development using goal programming applied to the United Arab Emirates. Energy Policy 87:447–454. https://doi.org/10.1016/j.enpol.2015.09.027

Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. Science 353(6301):790–794. https://doi.org/10.1126/science.aaf7894

Jeong S, Lee Y, Kang SH (2020) Government spending and sustainable economic growth: based on first- and second-level COFOG data. Public Money Manage 40(2):140

Khediri A, Laouar MR, Eom SB (2021) Improving intelligent decision making in urban planning: using machine learning algorithms. Int J Bus Anal 8(3):40–58. https://doi.org/10.4018/IJBAN.20210 70104

Kroll C, Warchold A, Pradhan P (2019) Sustainable development goals (SDGs): are we successful in turning trade-offs into synergies? Palgrave Commun 5(1):1–11. https://doi.org/10.1057/ s41599-019-0335-5

Kuhn M (2021) caret: Classification and regression training. R package

Kurniawan R, Managi S (2018) Measuring long-term sustainability with shared socioeconomic pathways using an inclusive wealth framework. Sustain Dev 26(6):596–605. https://doi.org/10. 1002/sd.1722

Kynčlová P, Upadhyaya S, Nice T (2020) Composite index as a measure on achieving sustainable development goal 9 (SDG-9) industry-related targets: the SDG-9 index. Appl Energy 265:114755

Lafortune G, Fuller G, Moreno J, Schmidt-Traub G, Kroll C (2018) SDG index and dashboards 2018. Global responsabilities. Implementing the goals, September, pp 1–476

La Torre D, Liuzzi D, Marsiglio S (2019) Population and geography do matter for sustainable development. Environ Dev Econ 24(2):201–223

Le Blanc D (2015) Towards integration at last? The sustainable development goals as a network of targets. Department of Economic and Social Affairs 1(141):1–17

Lusseau D, Mancini F (2019) Income-based variation in sustainable development goal interaction networks. Nat Sustain 2(March). https://doi.org/10.1038/s41893-019-0231-4

Mantlana KB, Maoela MA (2019) Mapping the interlinkages between sustainable development goal 9 and other sustainable development goals: a preliminary exploration. Bus Strat Dev 3(3):344–355. https://doi.org/10.1002/bsd2.100

Marcondes dos Santos HT, PerrellaBalestieri JA (2018) Spatial analysis of sustainable development goals: a correlation between socioeconomic variables and electricity use. Renew Sustain Energy Rev 97(July):367–376. https://doi.org/10.1016/j.rser. 2018.08.037

Miola A, Schiltz F (2019) Measuring sustainable development goals performance: how to monitor policy action in the 2030 Agenda implementation? Ecol Econ 164(June):106373. https://doi.org/10. 1016/j.ecolecon.2019.106373

Mukherjee S, Chakraborty D (2013) Is environmental sustainability influenced by socioeconomic and sociopolitical factors? Cross-country empirical evidence. Sustain Dev 21(6):353–371

Nature Sustainability Editorial (2018) Tracking progress on the SDGs. Nat Sustain 1(8):377. https://doi.org/10.1038/s41893-018-0131-z

Nchanji E, Lutomia CK, Karanja D (2021) Immediate impacts of COVID-19 measures on bean production, distribution, and food security in eastern Africa. J Agric Food Syst Commun Dev 10(2):259–263. https://doi.org/10.5304/jafscd.2021.102.007

Niestroy I (2016) How are we getting ready? The 2030 agenda for sustainable development in the EU and its member states: analysis and action so far. German Development Institute. Retrieved from https://www.die-gdi.de/discussion-paper/article/how-are-we-getti ng-ready-the-2030-agenda-for-sustainable-development-in-the-eu-and-its-member-states-analysis-and-action-so-far/

Nilsson M, Griggs D, Visbeck M (2016) Policy: map the interactions between sustainable development goals. Nature 534(7607):320–322

Nilsson M, Griggs D, Visbeck M, Ringler C, McCollum D (2017) A framework for understanding development success. Development

success: statecraft in the South, pp 1–29. https://doi.org/10.1057/ 9780230223073

Nilsson M, Chisholm E, Griggs D, Howden-Chapman P, McCollum D, Messerli P, Stafford-Smith M (2018) Mapping interactions between the sustainable development goals: lessons learned and ways forward. Sustain Sci 13(6):1489–1503. https://doi.org/10. 1007/s11625-018-0604-z

OECD (2019) Reader's guide. In: Measuring distance to the SDG targets 2019. OECD, pp 13–13. https://doi.org/10.1787/0afed8fb-en

OECD (2020) SIGI 2020 regional report for Latin America and the Caribbean. https://doi.org/10.1787/cb7d45d1-en

Ogundari K, Awokuse T (2018) Human capital contribution to economic growth in Sub-Saharan Africa: does health status matter more than education? Econ Anal Policy 58:131–140. https://doi. org/10.1016/j.eap.2018.02.001

Pedercini M, Arquitt S, Collste D, Herren H, Dc W (2019) Harvesting synergies from sustainable development goal interactions. PNAS 116(46):23021–23028. https://doi.org/10.1073/pnas. 1817276116

Pincet A, Okabe S, Pawelczyk M (2019) Linking aid to sustainable development goals: a machine learning approach. OECD Development Co-operation Working Papers. OECD Publishing, Paris

Pradhan P, Costa L, Rybski D, Lucht W, Kropp JP (2017) A systematic study of sustainable development goal (SDG) interactions. Earth's Future 5(11):1169–1179. https://doi.org/10.1002/2017EF000632

Profiroiu CM, Bodislav DA, Burlacu S, Rădulescu CV (2020) Challenges of sustainable urban development in the context of population Growth. Eur J Sustain Dev 9(3):51

R Core Team (2021) R: a language and environment for statistical computing. Retrieved from https://www.r-project.org/

Rahimi M, Moosavi SM, Smit B, Hatton TA (2021) Toward smart carbon capture with machine learning. Cell Rep Phys Sci. https:// doi.org/10.1016/j.xcrp.2021.100396

Ridgeway G (2007) Generalized boosted models: a guide to the gbm package. Compute 1(4):1–12

Sachs J, Schmidt-Traub G, Kroll C, Lafortune G, Fuller G (2016) SDG index and dashboards—global report. Bertelsmann Stiftung and Sustainable Development Solutions Network (SDSN), New York. Retrieved from https://www.sdgindex.org/reports/sdg-index-and-dashboards-2016/

Sachs J, Schmidt-Traub G, Kroll C, Lafortune G, Fuller G, Woelm F (2020) Sustainable development report 2020. Cambridge

Scharlemann JPW, Brock RC, Balfour N, Brown C, Burgess ND, Guth MK, Kapos V (2020) Towards understanding interactions between sustainable development goals: the role of environment–human linkages. Sustain Sci 15(6):1573–1584. https://doi.org/10.1007/ s11625-020-00799-6

Scherer L, Behrens P, de Koning A, Heijungs R, Sprecher B, Tukker A (2018) Trade-offs between social and environmental sustainable development goals. Environ Sci Policy 90(October):65–72. https://doi.org/10.1016/j.envsci.2018.10.002

Scott G, Rajabifard A (2017) Sustainable development and geospatial information: a strategic framework for integrating a global policy agenda into national geospatial capabilities. Geo-Spatial Inf Sci 20(2):59–76. https://doi.org/10.1080/10095020.2017.1325594

Singh GG, Cisneros-Montemayor AM, Swartz W, Cheung W, Guy JA, Kenny T-A, Ota Y (2018) A rapid assessment of co-benefits and trade-offs among sustainable development goals. Mar Policy 93:223–231

Somanje AN, Mohan G, Lopes J, Mensah A, Gordon C, Zhou X, Takeuchi K (2020) Challenges and potential solutions for sustainable urban-rural linkages in a Ghanaian context. Sustainability (Switzerland) 12(2):1–19. https://doi.org/10.3390/su12020507

Spaiser V, Ranganathan S, Swain RB, Sumpter DJT (2017) The sustainable development oxymoron: quantifying and modelling the incompatibility of sustainable development goals. Int J Sust Dev

World 24(6):457–470. https://doi.org/10.1080/13504509.2016.1235624

Stafford-Smith M, Griggs D, Gaffney O, Ullah F, Reyers B, Kanie N, O'Connell D (2017) Integration: the key to implementing the sustainable development goals. Sustain Sci 12(6):911–919. https://doi.org/10.1007/s11625-016-0383-3

Stojanović I, Ateljević J, Stević RS (2016) Good governance as a tool of sustainable development. Eur J Sustain Dev 5(4):558

The UN Women (2020) Progress on the sustainable development goals: The gender snapshot 2020, pp 1–24. Retrieved from https://www.unwomen.org/-/media/headquarters/attachments/sections/library/publications/2019/progress-on-the-sdgs-the-gender-snapshot-2019-single-pages-en.pdf?la=en&vs=5813

The World Bank Group (2021) World bank country and lending groups. Retrieved January 6, 2021, from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

Tomaskovic-Devey D, Rainey A, Avent-Holt D, Bandelj N, Boza I, Cort D, Tufail Z (2020) Erratum: rising between-workplace inequalities in high-income countries. Proc Natl Acad Sci USA (2020)117:9277–9283. https://doi.org/10.1073/pnas.1918249117). Proc Natl Acad Sci USA 117(24):13847. https://doi.org/10.1073/pnas.2009782117

UN (2017) Work of the statistical commission pertaining to the 2030 agenda for sustainable development (A/RES/71/313), annex (6 July 2017), 11371 (July), 25

UN (2018) The sustainable development goals report 2018. Retrieved from https://unstats.un.org/sdgs/files/report/2018/TheSustainableDevelopmentGoalsReport2018.pdf

United Nations (2020) SDG indicators, United Nations Global SDG Database. Retrieved from https://unstats.un.org/sdgs/indicators/database/

United Nations (2021) The sustainable development goals report 2016. The sustainable development goals report 2021. https://doi.org/10.29171/azu_acku_pamphlet_k3240_s878_2016

UN Statistical Commission (2020) Global indicator framework for the sustainable development goals and targets of the 2030 agenda for sustainable development. Work of the statistical commission pertaining to the 2030 agenda for sustainable development, pp 1–21

Valavi R, Elith J, Lahoz-Monfort JJ, Guillera-Arroita G (2018) blockCV: an r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. BioRxiv, 357798

Wang Q-S, Hua Y-F, Tao R, Moldovan N-C (2021) Can health human capital help the sub-saharan Africa out of the poverty trap? An ARDL model approach. Front Public Health 9:642. https://doi.org/10.3389/fpubh.2021.697826

Warchold A, Pradhan P, Kropp JP (2021) Variations in sustainable development goal interactions: population, regional, and income disaggregation. Sustain Dev 29(2):285–299. https://doi.org/10.1002/sd.2145

Weitz N, Persson Å, Nilsson M, Tenggren S (2015) Sustainable development goals for Sweden: insights on setting a national agenda. Stockholm Environment Institute, Stockholm

Weitz N, Carlsen H, Nilsson M, Skånberg K (2017) Towards systemic and contextual priority setting for implementing the 2030 agenda. Sustain Sci 13(2):531–548. https://doi.org/10.1007/s11625-017-0470-0

WHO (2018) Towards a global action plan for healthy lives and well-being for all: uniting to accelerate progress towards the health-related SDGs. World Health Organization, Geneva PP, Geneva. Retrieved from https://apps.who.int/iris/handle/10665/311667

WHO, UNICEF, UNFPA, World Bank Group, & United Nations Population Division (2019a) Maternal mortality 2000–2017: Sweden, pp 1–8

WHO, Unicef, UNFPA, World Bank Group, & United Nations Population Division (2019b) Maternal mortality 2000 to 2017: ethiopis. World Health Organization, Geneva, p 104

Wilkinson RG, Pickett KE (2006) Income inequality and population health: a review and explanation of the evidence. Soc Sci Med 62(7):1768–1784. https://doi.org/10.1016/j.socscimed.2005.08.036

Wilkinson RG, Pickett KE (2007) The problems of relative deprivation: why some societies do better than others. Soc Sci Med 65(9):1965–1978. https://doi.org/10.1016/j.socscimed.2007.05.041

World Bank (2018) World development indicators. Retrieved from http://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS

World Bank (2020) SDG atlas 2020. Retrieved from https://databank.worldbank.org/source/sustainable-development-goals-(sdgs)

Zhou X, Moinuddin M, Xu M (2017) Sustainable development goals interlinkages and network analysis: a practical tool for sdg integration and policy coherence

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.