Modeling and prediction of the achievement level with related goals for SDG 11: Sustainable Cities and Communities

Baskaran, Venkatesh
Govrnment College of Engineering

Velkennedy, R
Thiagarajar College of Engineering

Murugan, S
Thiagarajar College of Engineering

Theerumalai, Ga
College of Engineering, Guindy

https://doi.org/10.5109/5909057

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 8, pp.19-24, 2022-10-20. Interdisciplinary Graduate School of Engineering Sciences, Kyushu University
バージョン：

Modeling and prediction of the achievement level with related goals for SDG 11: Sustainable Cities and Communities

Venkatesh Baskaran[1*], R Velkennedy[2], S Murugan[2], Ga Theerumalai[3]
[1]Govrnment College of Engineering, Tirunelveli, India
[2]Thiagarajar College of Engineering, Madurai, India
[3]College of Engineering, Guindy, Chennai, India
*Corresponding author email: venkatesh@gcetly.ac.in

Abstract: *United Nations has announced the current decade as a 'decade of action' to deliver the 2030 agenda successfully. The 17 Sustainable Development Goals (SDG) achievement level of all the nations varies. Almost 70% of the world's population is expected to settle in the urban region by 2050. So SDG 11, the goal of developing a resilient and sustainable city, was focused on in the study. Since the targets of each goal are interrelated to others, the progress in each goal will also reflect on the achievement scores of different goals. An attempt was made in this study to develop a model to predict the achievement status of SDG 11 based on the performance of the country's related goals. The Random Forest machine learning algorithm was used in developing the prediction model with the KNIME analytics platform, which attained overall prediction accuracy up to 83.67%.*

Keywords: SDG 11; sustainable city; random forest; machine learning; prediction

1. INTRODUCTION

The most ambitious global development goals in history were introduced when the 193 member states of the United Nations General Assembly unanimously adopted the 2030 Agenda for Sustainable Development in 2015. Member states are committed to the seventeen Sustainable Development Goals (SDGs) (2015–2030) to end poverty, move the world toward sustainability, and promote greater inclusion [1]. SDG targets are prioritized through target trade-offs which also depend on the current status at the local or any administration level [2]. By 2050, 68 percent of the world's population, up from 55 percent today, is anticipated to reside in urban regions. According to projections, an additional 2.5 billion people could live in the urban areas by 2050 due to urbanization, which is the steady human population migration from rural to urban zones [3]. The complete goals for sustainable development by U.N are shown in figure 1 below.

The rapid rise in urbanization was brought on due to limited resources, rising pollution, and compelled urban economic transformation [4]. These pave enormous challenges to urban sustainability, making the United Nations think of focusing on cities in one of the proposed 17 SDGs titled 'Goal 11: Make cities inclusive, safe, resilient, and sustainable' [5].

SDG 11 created a set of quantifiable targets for future cities, such as ensuring universal access to essential public services, power, housing, and transportation. Moreover, Cities will determine whether we succeed or fail in attaining our goals of poverty eradication, equality, climate change mitigation, and ensuring healthy lives. Cities will decide whether or not we achieve inclusive economic growth or increase inequality. People will seek higher education and employment prospects in cities. And cities will determine whether we continue to deplete the world's resources at an alarming rate or whether we can navigate a more sustainable route [6].

The specific target sets for SDG 11 from the 2030 agenda are

- To ensure access to housing with adequate, safe, and affordable essential services.
- To provide safe, inclusive, sustainable, and economical transportation systems, focusing on public transportation.
- To increase affordable and sustainable urbanization.
- To protect and preserve the world's natural heritage and culture.
- To achieve sustainable resource management and efficient resource utilization.
- To provide public spaces that are accessible, inclusive, and safe for all people.

Fig. 1. Sustainable Development Goals to transform our world by 2030

The SDGs are significantly more integrated, comprehensive, and complicated than the millennium development goals (MDGs), with significant coverage of sustainable development's economic, social, and environmental components spread among 169 targets and 232 indicators. Because the SDG targets are integrated, progress toward one target is linked to other targets via complex feedback mechanisms [7]. Previously, an author presented the interlink matrix based on the interaction strength between goals 1 to 16 as per the International Council for Science (ICS) report [8]. In that network study, the matrix has zero entries on the main diagonal because self-reinforcing links (i.e., links from a Goal to itself) were not permitted. In these databases, 162 of the remaining 240 were non-zero, including 152 positive and ten negative entries.

Nowadays, machine learning is prevalent in estimating the dependency and relevancy of an outcome from big data. Around 80% of the Sustainable Development Goals could be achieved with artificial intelligence's (A.I) help. A.I. has the potential to play a significant role in enabling a rotary economy and creating resource-efficient smart cities. Machine learning is a branch of artificial intelligence (A.I.) that, among other things, aids in more effectively designing, carrying out, advising, and planning the future of our planet and its sustainability [9,10]. Machine learning (KNN, Naive Bayes, Decision Trees, SVM) can also be used to comprehend sustainable development priorities.

This article attempted to model the achievement level of SDG 6 with the status of the related target's goal as per their interlinkages released in the metadata by the U.N.. This modeling was performed with one of the supervised classification algorithms called random forest. Random forest is a versatile, user-friendly machine learning algorithm that often produces excellent results without hyper-parameter tuning. Because of its simplicity and diversity, it is also one of the most widely used algorithms and can be applied to classification and regression tasks [11].

2. METHODOLOGY

The workflow starts with collecting the current achievement status of SDG 11 and its related goals from the Sustainable Development Report 2022 and an interactive dashboard [12]. It contains target-wise status from the 17 goals for about 194 countries. All these countries were grouped into eight regions based on their geographical position. Data were formerly reported for nations in "developed" and "developing" areas, which were then further divided into geographical subregions as follows, and the same is shown in figure 2 below.

1. Sub-Saharan Africa
2. Northern Africa and Western Asia
3. Central and Southern Asia
4. Eastern and South-Eastern Asia
5. Latin America and the Caribbean
6. Oceania
7. Australasia
8. Europe and Northern America

2.1 Interlinkages between SDG 11 and other targets

Rising cities necessitate facilitating safe, regular, and responsible migration and mobility of people (goal 10.7), as growing cities result in increased waste production and emissions. As a result, changes in production and consumption must be made (targets 12.3, 12.4, and 12.5) to limit and even prevent cities from becoming vulnerable to climate change and natural catastrophes. It is crucial for cities with improved infrastructure (targets 9.1, 9.2, 9.3, and 9.4), such as sustainable transportation systems (target 11.2), to invest in technology research and innovation. Increasing security and safety, improving access to adequate sanitation and clean drinking water (targets 6.1 and 6.2) [13], and reducing the impact of communicable diseases and maternal and child mortality (targets 3.2 and 3.3) can all be achieved by ensuring access to safe and affordable housing and basic services (target 11.1) [14]. Similarly, Table 1 below shows the interlinkages between self targets (SDG 11) and the targets of other goals [15].

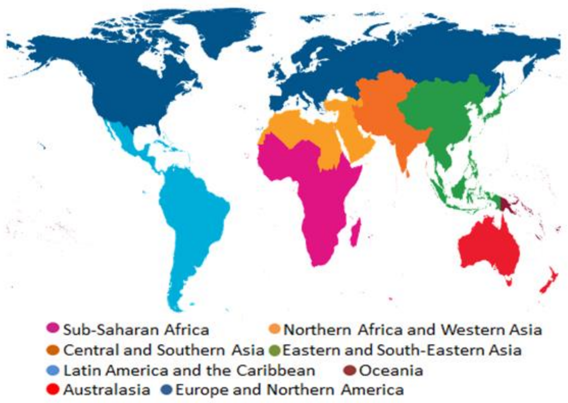

Fig. 2. Regional grouping for SDG indicators reporting

One cannot directly monitor and assess the goal like whether it is in progress. Each goal may have targets varied from 6 to 9. Further, each target is measured only with the help of indicators with the appropriate unit for measurement, such as the population proportion, land area, and so on. These indicators' score varies from 0 to 100. The maximum range of value represents the accomplishment of targets, and the minimum side represents many challenges with the initiation of making the target in progress [16].

In this study, the authors assumed that the indicator score of related targets from other goals directly reflects the achievement level of the corresponding gaol. So hereafter, SDG 11 is related only to goals 1, 3, 5, 6, 9, 10, 12, and 16, not with targets.

The official website displays the achievement level of each goal in the dashboard in five different colors. The class color code and its status inference are listed below

- Green - Goal Achievement
- Yellow - Challenges Remain
- Orange - Significant Challenges Remain
- Red - Major Challenges Remain
- Grey - Data Insufficient

The datasets for the study are prepared by modifying the original datasets, which contain all the target's indicator values and goal-wise status and track direction of all countries. To complete the initial data preparation, we organized only the score of eight goals, in addition to SDG 11 status, with a population of countries without any missing values.

Table 1. Interlinkage of SDG 11 with other goals

Goal	Relevancy (targets)
SDG 1: No Poverty	Upgrade slums and access to transport systems, and reduce the number of deaths by disasters with a focus on the poor (11.1, 11.2, 11.5)
SDG 3: Good Health and Well-being	Reduce deaths and injuries from traffic accidents, reduce illness from air pollution, and access to safe transportation (3.6, 3.9, 11.2)
SDG 5: Gender Equality	Access to public transport and public space with particular attention to women (11.2, 11.7)
SDG 6: Clean Water and Sanitation	Assure access to drinking water (6.1)
SDG 9: Industry Innovation and Infrastructure	Quality, reliable, sustainable, and resilient infrastructure (9.1)
SDG 10: Reduced Inequalities	Universal access to public spaces, access to adequate, safe, and affordable housing, and basic services (11.7, 11.1)
SDG 12: Responsible Consumption and Production	Sustainable public transport reduces the adverse environmental impact on cities, and improves air quality, municipal and other waste management (11.2, 11.6)
SDG 16: Peace, Justice, and Strong Institutions	Inclusive urbanization, capacities for participatory human settlement planning and management (11.3)

2.2 Random Forest algorithm with KNIME

KNIME is a platform designed for robust analytics with a graphical user interface. This Analytics Platform is a top visual programming package capable of implementing ML applications. Users can use these deep learning extensions to read, construct, modify, train, and execute deep neural networks [17]. This paper uses the Random Forest machine learning algorithm, one of the most trusted ML algorithms for modeling and predicting datasets with classification. It is based on ensemble learning, a method of combining various classifiers to address complex issues and enhance model performance. Random Forest, as its name suggests, is a classifier that uses a number of decision trees on different subsets of the given dataset and averages the results to increase the dataset's predictive accuracy. Instead of relying on a single decision tree, the random forest uses predictions from each tree and predicts the result based on the votes of the majority of predictions. The sequence process in this R.F. predictive algorithm is shown in figure 3 below.

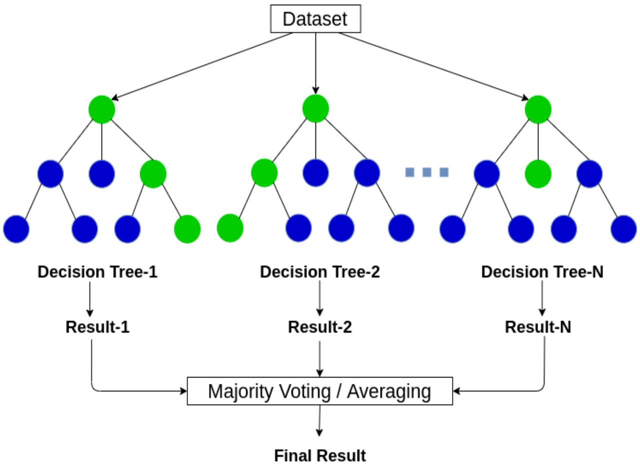

Fig. 3. The basic principle of the Random Forest algorithm

The workflow for executing the prediction starts with the node, which reads the dataset from excel. Nodes in the KNIME Analytics Platform represent individual tasks. Each node is represented by a colored box with input and output ports. Nodes can conduct various functions, such as reading and writing files, manipulating data, training models, and creating visualizations. The complete workflow for predicting the achievement class of SDG 11 using other related targets with RF-ML is shown in figure 4 below.

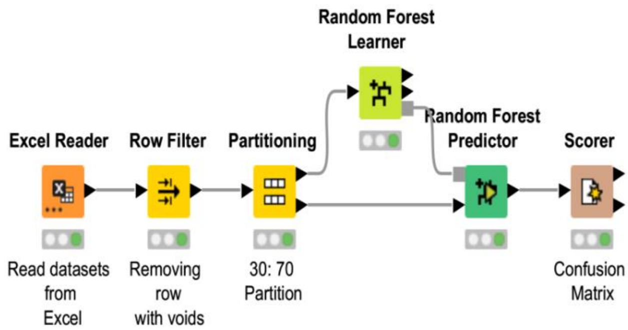

Fig. 4. Workflow of Random Forest Prediction model for SDG 11: Level of achievement

These base datasets had total entries for 194 countries. But after applying the row filter to remove rows with missing values, the row count decreased to 162 counties. The workflow continues with the partition node, which splits the filtered input table into row-wise two partitions. These two train and test data will be available at the two outer ports. Here we have chosen the 70:30 partition with the 'draw random' option. This split was popularly adopted by various authors using the R.F. prediction model in their research [18–20]. The country population and status of related goals (SDG 1, 3, 5, 6, 9, 10, 12, and 16) were included as an attribute for predicting SDG 11 as the target column in the R.F. learner node. We have set the decision tree split criteria with an option called Gini Index. This Gini Index, commonly called Gini impurity, determines the likelihood that a specific feature would be erroneously classified when chosen randomly. It can be

considered pure if every element is connected to a single class [21]. The fifth node R.F predictor will run the so far created learner model using the data set for testing. While executing this predictor node, KNIME creates a new column, 'Prediction_SDG 11'.
The scorer is the final node which helps with calculating the confusion matrix. For any machine learning model, the accuracy assessment directly expresses the level of reliability. This confusion matrix is vital in choosing an appropriate classifier prediction model with overall accuracy [22]. It evaluates how well classification models perform when they make predictions based on test data and indicates how effective our classification model is. It identifies not only the classification error but also the specific sort of error, such as type-I or type-II error.

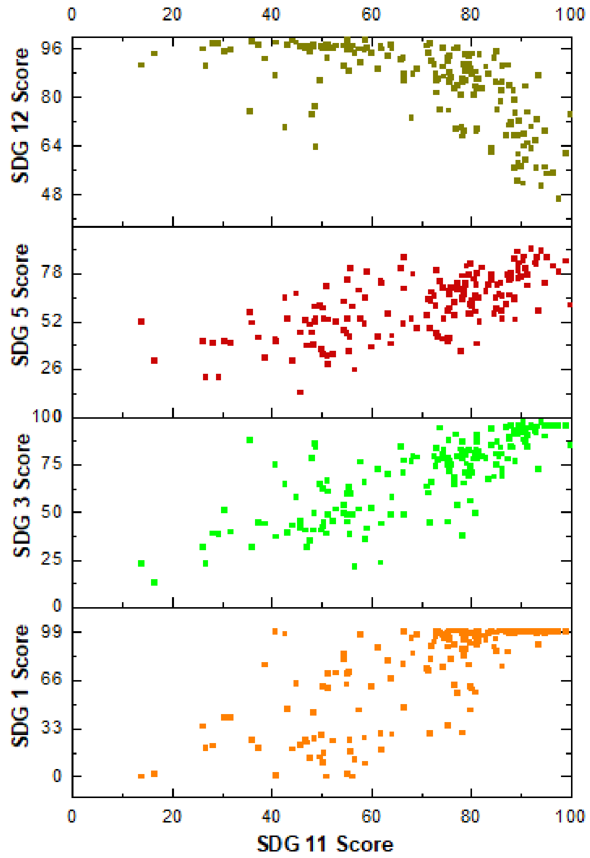

Fig. 5. Relation between the score of SDG 11 and explicitly linked goals

3. RESULT AND DISCUSSION

The interlinkage of SDG 11 with the related goal was classified into explicitly linked and substantially linked [23]. The nexus approach helps identify both links for the desired goal by detecting synergies and trade-offs [24]. SDG 1: No Poverty, SDG 3: Good Health and Well-being, SDG 5: Gender Equality, and SDG 12: Responsible Consumption and Production are explicitly linked with our desired goal. In this link, except for SDG 12 **(-0.61)**, all three have a moderate positive correlation **(SDG 1: 0.76, SDG 3: 0.79, and SDG 5: 0.67)** with SDG 11. The stacked plot below in figure 5 depicts the relation between explicitly linked goals with SDG 11 in terms of achievement score.

SDG 6: Clean Water and Sanitation, SDG 9: Industry Innovation and Infrastructure, SDG 10: Reduced Inequalities, and SDG 16: Peace Justice Strong Institutions are substantially linked with our desired goal. In this link, SDG 10 has a weak positive correlation **(0.32)**, while the other three have a moderate positive correlation **(SDG 6: 0.73, SDG 9: 0.68, and SDG 16: 0.72)** with SDG 11. The stacked plot below in figure 6 depicts the relation between substantially linked goals with SDG 11 in terms of achievement score.

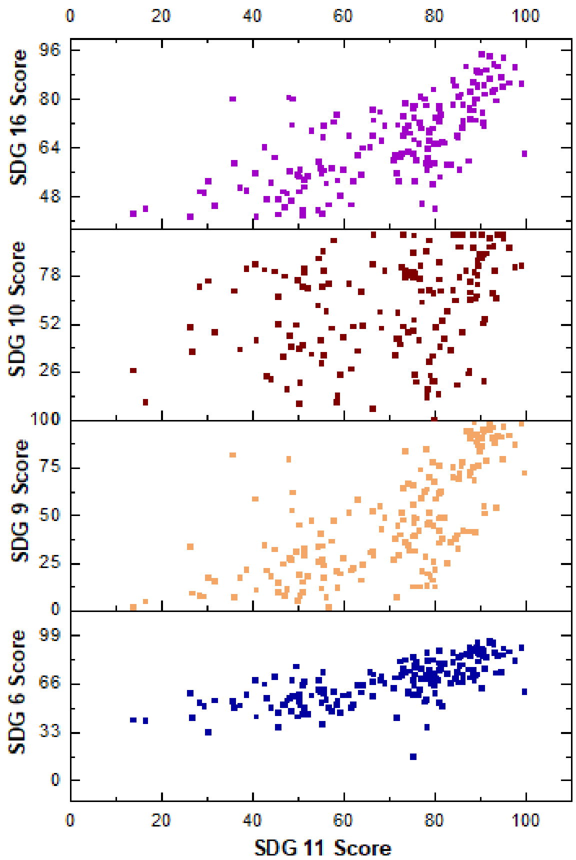

Fig. 6. Relation between the score of SDG 11 and substantially linked goals

Initially, we considered the scores of eight SDGs, population, and subregions as attributes in the designed R.F model for predicting SDG 11 achievement status. This resulted in 81.633% overall accuracy for the prediction model. Since this dataset does not have enough grey and green in SDG 11 status column, the test data also has nil grey and green, which is also reflected in the confusion matrix. Out of 49 test data, this model has correctly classified 40 entries and wrongly classified the remaining nine entries, as shown in figure 7.

Since the rough visualization plot between subregions with SDG 11 status does not make any significant relation, we have modified the model by removing subregions from the attribute list. While running this revised model, the overall accuracy improved to 83.673%. The confusion matrix for this modified prediction model is shown in figure 8. The country's population has been removed from the decider attribute list in the subsequent modification. But the performance of the model has not changed. The overall accuracy remains the same at 83.673%. It is a shred of clear evidence that the population is not significantly related to a country's achievement status of SDG 11.

Goal 11 Da...	yellow	orange	red	grey	green
yellow	6	2	0	0	0
orange	3	17	4	0	0
red	0	0	17	0	0
grey	0	0	0	0	0
green	0	0	0	0	0

Correct classified: 40 Wrong classified: 9

Accuracy: 81.633% Error: 18.367%

Cohen's kappa (κ): 0.709%

Fig. 7. Confusion matrix for the test data in the model (with subregion)

Goal 11 Da...	yellow	orange	red	grey	green
yellow	6	2	0	0	0
orange	2	18	4	0	0
red	0	0	17	0	0
grey	0	0	0	0	0
green	0	0	0	0	0

Correct classified: 41 Wrong classified: 8

Accuracy: 83.673% Error: 16.327%

Cohen's kappa (κ): 0.739%

Fig. 8. Confusion matrix for the test data in the model (without subregion)

The prediction for each entity is based on the results of R.F.'s 100 tree models [25]. So each predicted status of SDG 11 from this test set has its confidence value. In the yellow class, five predictions have higher confidence between 0.90 and 1.0 which are correctly classified [26]. The relation between the predicted class and its confidence value is illustrated in figure 9, as shown below.

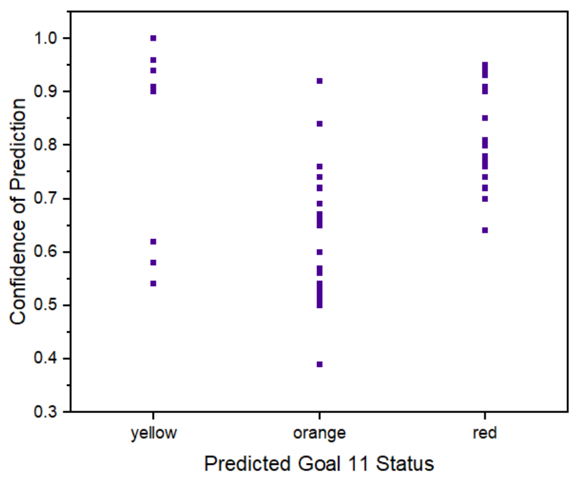

Fig. 9. Confidence of entries from each class

But in the orange class, out of 24 actual entries, only 16 were correctly classified; the same has been reflected in the confidence plot. Similarly, all the 17 red classes were correctly classified because of their confidence distribution in higher regions.

4. CONCLUSION

Monitoring and controlling is a much-needed process in achieving all 17 SDGs. Substantial financial assistance is required for each country to deliver this 2030 agenda. A few developing and underdeveloped countries were not even on track to achieving these goals. Every country has its priority in focusing on the goals and corresponding targets. A few goals may need less financial commitment, but their progress is reflected positively in the costlier goal due to their interrelated targets. Considering a vast urban settlement expected in the mid- 22nd century, we have chosen urban-related goal SDG 11 to develop a machine learning prediction model to assess its relation with interlinked goals.

Random Forest is a popular classification prediction algorithm utilized on the KNIME analytics platform with proper workflow. We created a model to predict the achievement status of SDG 11 using the achievement score of related goals SDG 1, 3, 5, 6, 9, 10, 12, and 16 and the country's population and subregions as a decider attribute. The optimized model performed with an overall accuracy of 83.673%. In addition to the model, the achievement score of the related goal of all countries was plotted against SDG 11. Except for SDG 12, all the remaining seven goals were positively correlated with SDG 11. However, their positive significance varied from weak to moderate only.

REFERENCE

[1] V. Herrera, Reconciling global aspirations and local realities: Challenges facing the Sustainable Development Goals for water and sanitation, World Dev. 118 (2019) 106–117.

[2] B. Venkatesh, R. Velkennedy, A systematic review on the role of geographical information systems in monitoring and achieving sustainable development goal 6 : Clean water and sanitation, Sustain. Dev. (2022) 1–9.

[3] Z. Patel, S. Greyling, D. Simon, H. Arfvidsson, N. Moodley, N. Primo, C. Wright, Local responses to global sustainability agendas: learning from experimenting with the urban sustainable development goal in Cape Town, Sustain. Sci. 12 (2017) 785–797.

[4] X. Xu, Z. Zhang, T. Long, S. Sun, J. Gao, Mega-city region sustainability assessment and obstacles identification with GIS–entropy–TOPSIS model: A case in Yangtze River Delta urban agglomeration, China, J. Clean. Prod. 294 (2021) 126-147.

[5] The 2030 Agenda for Sustainable Development, United Nations Dep. Econ. Soc. Aff. (2015) 1–41.

[6] V. Rozhenkova, S. Allmang, S. Ly, D. Franken, J. Heymann, The role of comparative city policy data in assessing progress toward the urban SDG targets, Cities. 95 (2019) 102-128.

[7] C. Allen, G. Metternicht, T. Wiedmann, Prioritising SDG targets: assessing baselines, gaps and interlinkages, Sustain. Sci. 14 (2019) 421–438.

[8] J.H.P. Dawes, J.H.P. Correspondence, Are the Sustainable Development Goals self-consistent and mutually achievable?, Sustain. Dev. 28 (2020) 101–117.

[9] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S.D. Langhans, M. Tegmark, F. Fuso Nerini, The role of artificial intelligence in achieving the Sustainable Development Goals, Nat. Commun. 11 (2020) 1–10.
[10] A. Hajikhani, A. Suominen, Mapping the sustainable development goals (SDGs) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection, Springer International Publishing, 2022.
[11] M. Schonlau, R.Y. Zou, The random forest algorithm for statistical learning:, Stata J. Promot. Commun. Stat. Stata. 20 (2020) 3–29.
[12] J.D. Sachs, G. Lafortune, C. Kroll, G. Fuller, F. Woelm, From Crisis to Sustainable Development: the SDGs as Roadmap to 2030 and Beyond, Sustain. Dev. Rep. 2022. (2022).
[13] M. Bilal, M. Sultan, T. Miyazaki, Spatiotemporal Investigation of Atmospheric Water Harvesting Potential Using Response Surface Methodology for Multan (Pakistan) and Fukuoka (Japan)., in: Int. Exch. Innov. Conf. Eng. Sci., 2020: pp. 128–133.
[14] M. Salem, N. Tsurusaki, A. Eissa, T. Osman, Detection of Slums from Very High-Resolution Satellite Images Using Machine Learning Algorithms: A Case Study of Fustat Area in Cairo, Egypt., in: Int. Exch. Innov. Conf. Eng. Sci., 2020: pp. 219–224.
[15] T. Devisscher, C. Konijnendijk, L. Nesbitt, J. Lenhart, F. Salbitano, SDG 11 : Sustainable Cities and Communities – Impacts on Forests and Forest-Based Livelihoods, in: Sustain. Dev. Goals Their Impacts For. People, Cambridge University Press, 2021: pp. 349–385.
[16] L. Niemann, T. Hoppe, How to sustain sustainability monitoring in cities: Lessons from 49 community indicator initiatives across 10 Latin American countries, Sustain. 13 (2021).
[17] F. Villarroel Ordenes, R. Silipo, Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications, J. Bus. Res. 137 (2021) 393–410.
[18] W. Chen, Y. Li, W. Xue, H. Shahabi, S. Li, H. Hong, X. Wang, H. Bian, S. Zhang, B. Pradhan, B. Bin Ahmad, Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods, Sci. Total Environ. 701 (2020) 134-159.
[19] T. Amestoy, A. Mercat, W. Hamidouche, D. Menard, C. Bergeron, Tunable VVC frame partitioning based on lightweight machine learning, IEEE Trans. Image Process. 29 (2020) 1313–1328.
[20] N. Arora, P.D. Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment, Appl. Soft Comput. 86 (2020) 105-126.
[21] L.E. Raileanu, K. Stoffel, Theoretical Comparison between the Gini Index and Information Gain Criteria, Ann. Math. Artif. Intell. 2004 411. 41 (2004) 77–93.
[22] B. Venkatesh, M.A.M. Mannar Thippu Sulthan, An Automated Geoprocessing Model for Accuracy Assessment in Various Interpolation Methods for Groundwater Quality, in: P.K. Paul, A. Choudhury, A. Biswas, B.K. Singh (Eds.), Environ. Informatics, 1st ed., Springer Nature Singapore, Singapore, 2022: pp. 285–299.
[23] S. Kumar, P. Das, B. Maity, S. Rudra, Understanding future urban growth, urban resilience and sustainable development of small cities using prediction-adaptation-resilience (PAR) approach, Sustain. Cities Soc. 74 (2021) 103-126.
[24] J. Liu, V. Hull, H.C.J. Godfray, D. Tilman, P. Gleick, H. Hoff, C. Pahl-Wostl, Z. Xu, M.G. Chung, J. Sun, S. Li, Nexus approaches to global sustainable development, Nat. Sustain. 2018 19. 1 (2018) 466–476.
[25] T. Lan, H. Hu, C. Jiang, G. Yang, Z. Zhao, A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification, Adv. Sp. Res. 65 (2020) 2052–2061.
[26] X. Deng, Q. Liu, Y. Deng, S. Mahadevan, An improved method to construct basic probability assignment based on the confusion matrix for classification problem, Inf. Sci. (Ny). 340–341 (2016) 250–261.