Machine Learning for Sustainable Development

Sunday TUNMIBI & David OKHAKHU
Department of Information Management,
Lead City University, Ibadan, Oyo State

Abstract

This paper presents machine learning as a tool for sustainable development. In 2015, the sustainable development goals (SDGs) were adopted by the United Nations as a global call to eliminate poverty, keep the planet safe, and ensure peace and prosperity for all people by 2030. Machine learning basically focuses on design of algorithms that enable machines to learn from data and make predictions based on the attributes of the data. Hence, one of the major approaches in machine learning is the adoption of artificial neural network (ANN) learning algorithm. With this in mind, we discussed the machine learning process and also reviewed scholarly work on how machine learning has been adopted in sustainable development. This paper shows that machine learning, together with appropriate data mining techniques, can help to gather and summarize data for decision makers. It can also help decision makers at every sector to make predictions and projections into what is expected in the future.

Keywords: Machine Learning, Artificial Neural Network, Sustainable Development Goals, United Nation, Data Mining.

Introduction

Machine learning has its root in artificial intelligence. It basically refers to the design of algorithms which aids machines to learn and predict attributes of data. Hence, machine learning helps in solving issues associated with traditional programming approach, in which decisions are based on input samples. According to Levin, Pomares and Alvarez (2016) "Machine learning procedures use statistical tools to find patterns in the data that reveal new and relevant information that may prove useful for performing an action or task".

In 2015, leaders from 193 countries of the world came together to look at existing situations and plan for the future. They realized that their own countries, together with other countries around the world, needed to navigate a path out of famines, drought, wars, plagues and poverty. So they inaugurated some goals. The goals, which are seventeen in number, focus on "ending poverty and

hunger; good health and well-being; quality education; gender equality; clean water and sanitation; affordable and clean energy; decent work and economic growth; industry, innovation and infrastructure; reduced inequalities; sustainable cities and communities; responsible consumption and production; climate action; life below water; life on land; peace, justice and strong institutions; and partnerships for the goals" (Abata-Ebiri, Adebowale and Ojokuku, 2018). These 17 goals encapsulate five key areas which are "people, planet, prosperity, peace and partnership".

Achieving the sustainable development goals require a coordinated approach from all development organizations and institutions (Maepa and Marumo, 2016; Mojapelo, 2018). Without doubt, thorough analysis of data generated at every sector is important for the actualization of the United Nations 2030 Agenda for Sustainable Development. Hence, this paper focuses on presenting machine learning as a tool for sustainable development. Specifically, we also examined the origin of machine learning, discussed the concept of artificial neural network, and examined the machine learning process.

Background on Machine Learning

Machines are not designed to be intelligent. The question of intelligent machine started with Alan Turing in 1955, this birthed the artificial intelligence field. Alan Turing proposed a test that starts with three people: a man, a woman and an interrogator with the aim of answering the question: can machine think? The test summarily involved a machine fooling an interrogator into believing that it was a woman. However, some scholars are of the opinion that the question "can machine think?" be replaced with "can machine do what humans can do?"

Machine learning basically focuses on design of algorithms that enable machines to learn from data and make predictions based on the attributes of the data. Hence, one of the major approaches in machine learning is the adoption of artificial neural network (ANN) learning algorithm. ANN is based on modelling human brain with the use of artificial neurons for processing information. ANN has been widely accepted and used basically because of these two reasons: ability to learn and ability to generalize. The different approaches adopted in ANN learning algorithms are

"supervised learning, semi-supervised learning, unsupervised learning and reinforcement learning" (Mohammed, Khan and Bashier, 2017).

Supervised learning is guided learning by training dataset (supervisor) and it is similar to the learning experience of students in school. This learning approach uses labeled (classified) data. The objective of this approach is to make predictions from the labeled data.

In semi-supervised learning, the training dataset provided is a mixture of labeled (classified) and unlabeled (unclassified) data. This is similar to the learning experience of a child. A child is provided with unlabeled data from the environment as well as labeled data from parent or tutor in school. The objective of this approach is to allow the machine to learn a model that will predict data better than that from the labeled data alone.

There is no training dataset or supervisor in unsupervised learning. Hence, the machine learns from unlabeled (unclassified) data or input data. The system uses the data to learn about the patterns without any previous knowledge about it. This is similar to the learning experience of employee or employer in an organization. The objective of this approach is to discover patterns or features in the unlabeled data with no help from a supervisor (training dataset).

In reinforcement learning, data is usually not given but it is generated by interactions with the environment. The objective is to use observations gathered from the interaction of the machine with the environment to make decisions that maximum reward or minimize risk.

The common types of machine learning algorithm include Naïve Bayes; K-means clustering; Random forest; Q-learning; and Neural network training algorithms. Naïve Bayes makes use of classifying algorithm (Holloway and Mengerson, 2018; and Aissaoui et al., 2019). The target of classification is to group data with similar features into groups. Although these classifying algorithms can be extended and adapted for semi-supervised learning, they are mostly used for supervised learning. K-means clustering, on the hand, is widely used in solving clustering problems in both semi-supervised and unsupervised learning algorithms. The focus is to classify a given dataset through a number of clusters. A random forest algorithm is a classifier with a number

of decision trees. It can also be used for unsupervised learning. Q-learning algorithm adopts a stochastic approximation method and it is used in reinforcement learning.

The different neural network training algorithms are multilayer perceptron, training multilayer perceptron and back propagation algorithm. The multilayer perceptron is used for unsupervised learning while the back propagation algorithm is widely used for supervised learning.

Machine Learning and Sustainable Development

Porciello et al (2020) used machine learning approach for network of seventy-seven researchers from 23 countries in gathering and summarizing data (evidence syntheses) towards achieving United Nations Sustainable Development Goal 2. Goal 2 is focused on "ending hunger, achieving food security and improved nutrition, and promoting sustainable agriculture".

Also, Holloway and Mengerson (2018) were able to show that earth science and statistical domains have well defined interface, as machine learning was applied to remote sensing data to derive key metrics for agriculture and the environment. They focus particularly on "applications related to the United Nations World Bank Sustainable Development Goals, including agriculture (food security), forests (life on land), and water (water quality), providing a review of useful statistical machine learning methods and how they work in a remote sensing context".

In another study on machine learning for sustainable development, Al-Abadi and Alsamaani (2020) focuses on using machine learning technique to model groundwater artesian condition in an area of Iraq. They used "an inventory map of flowing and non-flowing groundwater wells along with six explanatory factors (distance to faults, faults density, lineament density, aquifer saturated thickness, well depths, and ground surface elevation)". Their approach could help in successfully drilling artesian wells with little cost and efforts.

Lee and Tae (2020) also developed a "decision support model based on machine learning for applying greenhouse gas reduction". Their study also suggested an effective method of controlling greenhouse gas emissions based on a comparison of predictive power. Their study identified optimal greenhouse gas reduction technologies and also supported efficient greenhouse gas

emissions control when considering the energy consumption patterns and environment of a building.

Machine learning has also been applied to tax reform for sustainable development. Zheng, Zheng and Ye (2016) discussed how machine learning could be used in designing tax instrument for controlling environmental pollution. Their study also addressed issues related to environmental taxes, environmental policy and economic development. They used quantitative methods in machine learning to evaluate Environmental Fee to Tax reform effect in a province in China. They "quantitatively assesses the impact of environmental tax reform on SO2 emissions. By creating a synthetic control city for every pilot city, their study simulates the pollutant trend under a scenario with no EFT reform for each target city, which allows the authors to assess the actual impact of pollution intensity".

In addressing sustainable education, Abidi et al (2018) Incorporated substantial and sustainable development issues into teaching and learning by developing a machine learning tool called an Intelligent Tutoring System. The machine learning tool could assist teachers in identifying groups of students who were confused attempting homework exercise. This tool also highlights which skills need more attention for further practice. Furthermore, teachers can also govern learning behaviors and student performances during various mastery skills, allowing them to focus only problematic skills in the next day of the class, which will save a lot of time and effort for both tutors and students.

In a related study, Aissaoui et al. (2019) focused on developing a method which automatically detect learning styles using machine learning algorithms. According to them, "Learning styles refer to the preferred way in which an individual learns best". Their approach was based on the following two steps: "in the first step the learners' sequences were extracted from the log file then transformed to an input of the K means algorithm. The k means algorithm was used to group students into sixteen clusters based on FSLSM (Felder and Silverman learning style model), where each cluster was labeled with a learning style combination. The second step consists in performing an unsupervised algorithm (Naïve Bayes) to predict the learning style for a new sequence".

Conclusion and Recommendations

The reviewed literatures have provided evidence to the application of machine learning to agriculture, education, greenhouse gas reduction and environmental tax reform. Hence, synthesized information gathered through data mining and machine learning could guide decision makers in designing appropriate framework and making right policies in agriculture, education and human capital development, environment, health and nutrition and management of natural resource. This implies that there would be food security, preserved environment and sustainable development.

The United Nations' Sustainable Development Goals (SDGs) are guideposts for providing better life for all people, irrespective of their race. This paper has shown that machine learning, together with appropriate data mining techniques, can help to gather and summarize data for decision makers. It can also help decision makers at every sector to make predictions and projections into what is expected in the future. It is our submission that machine learning can thereby provide objective recommendations to stakeholders towards achieving the SDGs.

References

Abata-Ebiri, B. D., Adebowale, J. A. and Ojokuku, B. Y. (2018). Achieving Sustainable Development Goals (SDGS) in Nigeria: The Roles of Libraries. International Journal of Applied Technologies in Library and Information Management, 4(2), 89-95

Abidi, S. M. R., Hussain, M., Xu, Y. and Zhang, W. (2019). Prediction of confusion attempting algebra homework in an intelligent tutoring system through machine learning techniques for educational sustainable development. Sustainability, 11, 105. doi:10.3390/su11010105

Aissaoui, O., Alami, Y. E., Oughdir, L. and Allioui, Y. (2019). A hybrid machine learning approach to predict learning styles in adaptive e-learning system. Springer Nature Switzerland, 772-786. https://doi.org/10.1007/978-3-030-11928-7_70

Al-Abadi, A. M. and Alsamaani, J. J. (2020). Spatial analysis of groundwater flowing condition using machine learning techniques. Elsevier. https://doi.org/10.1016/j.gsd.2020.100418

Holloway, J. and Mengersen, K. (2018). Statistical machine learning methods and remote sensing for sustainable development goals: A review. Remote Sensing, 10, 1365. doi:10.3390/rs10091365

Lee, S. and Tae, S. (2020). Development of a decision support model based on machine learning for applying Greenhouse Gas reduction technology. Sustainability, 12, 3582. doi:10.3390/su12093582

Levin, I., Pomares, J. and Alvarez, R. M. 2016. Using machine learning algorithms to detect election fraud. Cambridge University Press, 266-294. https://doi.org/10.1017/CBO9781316257340.012

Maepa, M. E. and Marumo K. (2016). The contribution of the National Library of South Africa towards the achievement of Sustainable Development Goals (SDGs). IFLA, Columbus.

Mohammed, M., Khan, M. B. and Bashier, E. B. M. 2017. Machine learning: algorithms and applications. CRC Press, Taylor & Francis Group. DOI: 10.1201/9781315371658

Porciello, J., Ivanina, M., Islam, M., Einarson, S. and Hirsh H. (2020). Accelerating evidence-informed decision-making for the sustainable development goals using machine learning. Nature Machine Intelligence, 2, 559-565. https://doi.org/10.1038/s42256-020-00235-5

Zheng, Y., Zheng, H. and Ye, X. (2016). Using machine learning in environmental tax reform assessment for sustainable development: A case study of Hubei Province, China. Sustainability, 8, 1124. doi: 10.3390/su8111124