Swanirman Sunirmit Publications of Research Volume 4, Issue 2-Jun 2024 | [2024-25]
ISSN [Online]: 2583-2654

Explainable Machine Learning: Research, Application, and Future Perspectives

Ruchi Kawatra[1], Vanshika Jain[2], Upasna Singla[3], Vanshita Kakkar[4]
Department of Computer Science & Engineering, [1]SRM University Haryana, [2,3,4]Chitkara University, Punjab

Abstract: Machine learning (ML) which has recently experienced a substantial improvement is currently being used by large applications to construct automated or semi-automated processes with the creation of extremely trustworthy models over the past few decades, explainable machine learning has increased dramatically. Explainable ML models are habitually implied as black-box models since they license a predefined number of invalid limits, or center points, to be consigned values by AI estimations. Its algorithms are now able to learn from data and make predictions based on past patterns. This paper briefly reviews various types of explanatory machine learning models that are used for different purposes proposed by different researchers. It also sheds light on various fields where machine learning has greatly been impacted such as medicine, finance, education, marketing, etc. The futurescope of explainable machine learning will be tied to how well it can perform tasks such as predicting human behavior, diagnosing diseases, and understanding language. The devices designed by applying various explanatory machine learning algorithms can easily go for a patent as hardly any patent has been registered for such research.

Keywords- Explainable Machine Learning, reinforcementlearning, Multilayer perceptron, Scikit-learn package

I. INTRODUCTION

Recent advances in computing (AI) have caused its substantial business adoption, with the development of 'thinking' computer systems demonstrating superhuman overall performance on a huge variety of tasks. Machine learning is a subfield of computer science and artificial intelligence (AI) revolving around empirical computation, as opposed to rule-based programming. Machine learning is a branch of artificial intelligence that allows computer systems to learn from data without being explicitly programmed. Machine Learning is the next step in the evolution of AI. It is a type of predictive analytics that can continually learn from new information and use what was learned in previous experience to make predictions. This form of machine intelligence can take insights from past data and use them to find patterns in current data, then extrapolate what it believes might happen in the future. The main idea behind machine learning is that instead of writing out every logical step in detail, you give the computer many samples of what you expect it to do when given this type of input and let it figure out how to do that on its own. Instead of having a single program that needs to understand every detail of what it is supposed to do, you give it samples and let the machine figure out what it should do next.

In the Next Generation, a computer program is given basic instructions to follow and the program generalizes from those instructions to complete the task. Generally, machine learning can be classified into three categories of algorithms: such as supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is when you give a sample of desired output and the program learns on its own how to get closer to that output. In reinforcement learning, you reward it for doing well in certain situations as opposed to an error for doing poorly. Unsupervised learning is when you give it a collection of input data without an output goal or reward system. Unsupervised learning is a more difficult problem to solve since you cannot necessarily give it specific commands with which to classify input data. You have to manually categorize each piece of data, which can

ISSN [Online]: 2583-2654

be very hard even for humans. Machine learning is also useful outside of the world of computers, such as in medicine and biology where researchers have used machine learning applications in treating diseases and performing other biological manipulation tasks. It's also used in language translation and speech recognition. The main goal of this process is for the computer (or the machine) to learn from a set of training data and improve on the predictions that are made from it.

Explainable ML models are habitually implied as black-box models since they license a predefined number of invalid limits, or center points, to be consigned values by AI estimations. Even more absolutely, the backpropagation step is responsible for reviving the heaps according to its mix-up work. To fully understand the concept, one must know the concept, the processes involved, and how the processes occur, i.e. complete knowledge. For this reason, Explanatory ML focuses on the same thing, i.e. how to get results that can be understood by humans. It is not just a spell or a miracle that was previously envisioned for any process to happen, but we have full knowledge of how it happens[1]. Explainable ML is a subset of machine learning that aims to explain how a machine makes decisions, and more importantly, why a decision was made. It is based on the idea that a computer can be trained to understand and process information without being explicitly programmed. The use of explainable machine learning in today's world has increased tremendously. This algorithm can also be used for data mining, predictive analytics, and more. Machine learning has also been seen as one of the most promising ways to create more accurate self-driving cars and robots. Explainable machine learning is important because it helps data scientists produce more reliable models by giving them insight into the inner workings of their algorithms. The main reason to use was to make the model transparent and hence easy to understand.

The paper is further divided into five sections wherein the next section gives the difference between two commonly confused terms. Section 3 presents some papers studied by the author. Section 4 lists some of the applications and section 5 gives an overview of the technologies associated with explainable ML. In the end, the conclusion and futuredirections are there.

II. INTERPRETABILITY V/S EXPLAINABILITY

These phrases had been regularly taken into consideration considering that the method has an identical goal but there exists a minute distinction b/w the two. The first one bills for something we're guessing which may or may not provide us correct choice while the latter is going deep within the idea and displaying every issue minutely and subsequently giving us the correct result[20].

Explainable AI helps to make the ML model understand better. The most important aspect of model explainability are:

1 Transparency
2 Ability to Question
3 Ease of Understanding

To go briefly into the concept, explaining each observation seen at different points and how every feature individually affects the result can be explained using the local approach. A wider view of the model and how features collectively affect the result can be explained using the global approach. This paper reviews applications, algorithms, and techniques associated with explainable machine learning.

III. LITERATURE REVIEW

The utilitarian conclusions drawn out by different researchers in their respective fields are implemented to solve real-time problems. An explanatory machine learning model based on ML algorithms is vital to approach such day-to-day scenarios. This section briefly reviews various types of explanatory machine-learning models that are used for different purposes proposed by different researchers [18]. Table 1 gives an insight into their concerned fields, the methodologies they have used, and the outcomes they have achieved are observed and mentioned in their respective research papers.

ISSN [Online]: 2583-2654

Table 1. Summary of Tools and Technologies, Objectives and Pros & Cons of Explainable Machine Learning.

Ref	Tools and Techniques	Objective	Pros	Cons
[2]	• Extreme gradient boosting (XGBoost) model • Minimal Spanning Tree • EMST Dual-Tree Boruvka algorithm • Shapley approach using SHAP Lundberg and Lee computationalframework	To estimate the creditat the time of credit borrowing and employ peer-to-peerlending platforms.	• Effectively detectcredit risks. • Credit Risk supervising andprevention.	• The absence of humongous datasets may result in less accuracy. • The linear combination of features cannot be extracted by it.
[3]	• Shapley Additive exPlanations algorithm • Standard Python packages • Light Gradient Boosting Machine	Designed machine explainable machine prognostication algorithm that assesses the probability of in- hospital amputation among victims suffering from DFU.	• Higher accuracy • Little memory use	• Lack of clinical evidence and external validation cohorts. • Light GBM split the tree along the leaves, which cancause overfitting because it results in more complex trees.
[4]	• XGBoost model • Shapley Additive exPlanations • Numpy,pandas and scikit-learn.	Intend to enhance the accuracy in predictions of high-performance concrete compressivestrength HPCCS.	• Highly accurate as compared to other models • It May help in future research to build novel high-performance concrete mixes with advanced functions.	• The black box nature is likelythe biggest restriction. • TreeSHAP may generateillogical feature attributions.
[5]	• Logistic regression • Support vector machine • Adaptive boosting • Multilayer perceptron • XGBoost • Bootstrap method • Grid search method • Shapley Additive explanations • Scikit-learn package	Newly designed machine learning models that help in predicting the malnutrition status of children found with Congenital HeartDisease.	• Reliable andTransparent • Helps in findingnutritional treatment • Helps in determining follow-up strategiesfor children	• The model requires a large number of patients to verify its repeatability and robustness. • The period for follow-up is very short. • Not able to include information of racial origin despite the use of a large amount of data.
[6]	• Shapley values • XG Boost • Regression techniques • Machine learning approaches	An explainable AI model is applied for risk management in Fintech. All the credits borrowed have been reviewed at every peer-to-peer platform	• Gradient tree-boosting modelswere applied. • Prevents loss ofgenerality	• A quick clarification of the determinant of each individual's financial sound. • For sparse solutions with few features in a sample, it is not appropriate. Every feature is necessary.
[7]	• XGboost model • Logistic regression, • Machine learning • XGboost	To develop a model that helps in the prediction of whether victims diagnosed with AKI stage 1/2 would reach AKI stage 3.	• Performance is excellent for predictingAKI progression. • As opposed to Decision Trees, SVM models are challenging for humans to comprehend and interpret.	• Retrospective studies haveseveral chances ofunavoidable bias • Algorithms used to balance the sample have a bad effect on model generalization and reliability.
[8]	• Intraocular pressure measurement • fundus photography support vector machine • random forest • XGboost	To plan an explainable machine learning demonstration for theforecast of determination of glaucoma.	• Very less chances of misdiagnosis. • Using the Kernel Trick, SVM can effectively handle non-linear data.	• The sensitivity of the proposed model is 0.941 so it can also give a false negative ratio. • As opposed to Decision Trees, SVM models are challenging for humans to comprehend and interpret.

ISSN [Online]: 2583-2654

IV. APPLICATIONS

4.1 Medical

Explainable machine learning has enormously impacted the field of medical services. Numerous models have been created by researchers in medical care. It can be utilized to anticipate the length of stay of ICU disease-based hospitalization specifically cellular breakdown in the lungs of patients [9]. The use of XGBoost for Intense Myocardial Dead tissue from 12-lead ECG information presents a newer way to deal with the checking of Cardiovascular diseases (CVDs) [10]. Robotized forecast of readmission risk can save a great many dollars in medical care costs and can work on persistent consideration. AI models consider different aspects of patients, like socioeconomics, comorbidities, and slightness boundaries, to assess precisely their chance of being readmitted in 30 days.

4.2 Finance

Through explainable ML, a tonne of work has been done in the area of finance. Models are developed to predict house value credit risk for individuals using a real-world dataset, and they were shown how to interpret the results to make them more accessible to end users [11]. Framework innovations are the transformational calculations used in an innovative technique to generate plausible connection networks. The strategy widens the information space needed to prepare a decent ML program to identify the most fundamental characteristics in lattices, which results in the overall display of competing approaches to portfolio development. Reasonableness of an AI Conceding Scoring Model in Shared Lending. Credit risk AI approaches beat measurable methodologies, like calculated relapse, regarding order execution as well as logic [17].

4.3 Industry

Today, artificial intelligence (AI) is being utilized to conquer day-to-day life, industrial, and various business problems. In particular, it forms the pedestal to mount Industry 4.0. This task proposes an approach that defines the "importance of features" in the problem of anomaly detection in industrial scenarios [16]. Condition-based monitoring, predictive maintenance, quality control, and many tasks can be easily done using various models. The deployment of new standards was supported by the latest Industrial Internet of Things (IIoT) platform. It presents methods and software frameworks that let business players who aren't familiar with AI pick and choose the algorithms that are most suited to their requirements [14]. Most businesses rely on automated machine learning for predictive maintenance.

4.4 Marketing

Comprehensive marketing strategies are the driving force behind a company's profitability. Artificial intelligence (AI) and machine learning (ML) have become integral parts of a successful modern approach [15]. Companies can predict the magnitude of current and later stock price fluctuations. An advanced machine learning approach can support real-life problems in the financial and banking sectors where one needs to predict the likelihood of becoming a customer. In one of the studies, the outcome of a telemarketing campaign for a Portuguese bank was forecast using a tree-based model. It advocates employing an untested boosting algorithm to forecast campaign response and an explainable AI (XAI) technique to assess the model's effectiveness [12].

4.5 Education

Explainable machine learning is also being introduced into the field of education where researchers have proposed an explanatory, example-by-example approach using Bayesian teaching, where they aimed to study a small subset of data which helps learners to drive to conclusions [13]. There is a predictive model of student achievement in secondary education using five classification algorithms in which data is collected from school reports and surveys by training the Solver model by preferring a

ISSN [Online]: 2583-2654

locally interpretable knowledge model (LIME) that is interpretable for all classifiers. A novel approach that resorts to explanatory machine learning combined with learning analytics techniques provides actionable intelligent and automated feedback that supports the self-regulation of student learning in a data-drivenmanner.

4.6 Agriculture

The agriculture sector is another prominent domain to have been reaping the benefits of human advances in the field of AI. Explainable ML can identify and classify stress on various leaves of soybeans using a high-resolution top K feature map that separates the visual symptoms used for amazing accuracy and prediction. Explainable ML can be used for the prediction of the ripeness of fruit referred to as a black container set of rules which predicts the ripeness of peaches with the use of electric impedance.

Table 2. Different Algorithms And Techniques Are Used For Explainable Machine Learning

Algorithmsand Technique s	XGBoost	SVM	LR	SHAP	LightGBM
Full Form	Extreme Gradient Boosting	Support Vector Machine	Logistic Regression	Shapley AdditiveExplanation	Light Gradient Boosting Machine
Working	• Decision trees understoodas successive order. • Loads are designated to free figures thatare taken asinput into decision tree which predict results.	• Generalized between two different classes when a set of data labeled with atraining set of algorithms is provided. • Finds a hyperplane that can distinguish between the two classes.	• Multivariate analysis used once the variable quantity is categorical (e.g., yes/no). • Indispensable lose the faith estimates the likelihood of an occasion happening.	• It combines ideal credit attributes with explicit explanations using classic Shapley valuesfrom entertainment hypothesis and its associated extensions. • SHAP appreciation can emphasizeach contribution toincrease the show yield from basic appreciation	• Slope boosted tree-based outfit calculation that create forecast by averaging expansive number of person decisiontrees forecasts. • Decision trees are built consecutivelywith the objective to decrease the mistake of the past demonstrate at each emphasis.
Meaning	• Productive and adaptable execution of slope	• Supervised machine learning technique thatis employed	• This kind of quantifiable model is utilized for social affairs.	• SHAP is a mathematical method to explainpredictions of machine learning	• The strategy of machine learning can give an effective and exact apparatus for

V. ALGORITHMS AND TECHNIQUES ASSOCIATED WITH EXPLAINABLE MACHINE LEARNING

There are various algorithms designed by different scientists and researchers, which are summarized in Table 2 below, to make Explainable Machine Learning a success. XGBoost gives an equivalent tree supporting (generally called GBDT, GBM) that handles numerous data science projects [19]. A comparative code runs on major scattered environments (Hadoop, SGE, MPI) and can deal with issues past billions of models. SHAP evaluation has great potential to rationalize the expectations set by complex ML models. LGBT too centers on perceptions that are troublesome to foresee. Slope-boosted tree-based outfit calculation that creates forecasts by averaging an expansive number of person choice tree forecasts. LR is a regression model that is feasible, adaptable, and easily operated therefore has applications in many fields. Indispensable use in the faith estimates, the

ISSN [Online]: 2583-2654

likelihood of an occasion happening, for example, projected a surveying structure or didn't project a surveying structure, taking into account a given dataset of free factors. SVM is a machine learning model that can be generalized between two different classes when a set of data labeled with a training set of algorithms is provided. The main function of SVM is to find a hyperplane that can distinguish between the two classes.

VI. CONCLUSION AND FUTURE DIRECTIONS

Machine learning has made a lot of progress in the last few years. It is now possible to explain the logic behind the prediction models. This is done by using a visual representation and descriptive language to show how the modelarrived at its prediction. Machine learning is a type of AI that has been around for decades. It has been used in various industries to power algorithms and make predictions. There are two types of machine learning: non-explanatory and explanatory. Non-explanatory machine learning is what we have seen so far, where the prediction algorithm is not explainable to a human, but it can be explained to a computer. Explanatory machine learning will make the explanation understandable to humans, but this type of AI is still in its infancy stage. Undoubtedly, the past few years have witnessed developments around human-mimicking technologies like machine learning, but the future scope of explainable machine learning is still in its infancy. There are many potential applications, but we need to be mindful of the ethical implications. We have seen the rise of explainable AI in the last few years. This trend will continue to rise in the future. Now, owing to the digital advancements through the power of machine learning, businesses are a click away from gauging the demands of their target customers. Search engines like Google produce numerous options for their users looking around for a particular product. There is a possibility that we will see an increase in demand for explainable machine-learning tools and technologies.

VII. CURRENT & FUTURE DEVELOPMENTS

Explainable AI is a subset of artificial intelligence that deals with making machine learning more understandable to humans. This includes both generating explanations for decisions and improving transparency by providing datasets and models.

The future scope of explainable machine learning will be tied to how well it can perform tasks such as predicting human behavior, diagnosing diseases, and understanding language. Businesses are utilizing a wide range of software applications to improve their customer experience and, as a result, grow their revenues by leaps and bounds as the application of machine learning expands. The Gartner 2022 quadrant predicts that selecting the right AI technique provides global leaders a greater chance to increase productivity and workforce. Machine learning is a process of getting computers to do tasks that they were not programmed for. Explainable ML can be used for various purposes in the future such as medical diagnosis, legal decisions, and financial investments.

REFERENCES

[1] S. Arora, Kawatra Ruchi, and D. R. M. Agarwal, "PSE assessment-based e-learning: novel approach towards enhancing educationist performance," New Paradigm. eLearning Technol. Aris. Due To Covid- 19 Cris., p. 11, 2020.

[2] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable Machine Learning in CreditRisk Management," Comput. Econ., vol. 57, no. 1, pp.203–216, Jan. 2021, doi: 10.1007/s10614-020-10042- 0.

[3] P. Xie et al., "An explainable machine learning modelfor predicting in-hospital amputation rate of patients with diabetic foot ulcer," Int. Wound J., vol. 19, no. 4,pp. 910–918, May 2022, doi: 10.1111/iwj.13691.

[4] D. Chakraborty, I. Awolusi, and L. Gutierrez, "Anexplainable machine learning model to predict and elucidate the compressive behavior of high-performance concrete," Results Eng., vol. 11, Sep. 2021, doi: 10.1016/j.rineng.2021.100245.

[5] H. Shi et al., "Explainable machine learning model forpredicting the occurrence of postoperative malnutrition in children with congenital heart disease," Clin. Nutr., vol. 41, no. 1, pp. 202–210, Jan. 2022, doi: 10.1016/j.clnu.2021.11.006.

[6] C. Ntakolia, C. Kokkotis, P. Karlsson, and S. Moustakidis, "An explainable machine learning model for material backorder prediction in inventory management," Sensors, vol. 21, no. 23, Dec. 2021, doi: 10.3390/s21237926.

[7] C. Wei, L. Zhang, Y. Feng, A. Ma, and Y. Kang, "Machine learning model for predicting acute kidney injury progression in critically ill patients,"

BMC Med.Inform. Decis. Mak., vol. 22, no. 1, Dec. 2022, doi: 10.1186/s12911-021-01740-2.

[8] S. Oh, Y. Park, K. J. Cho, and S. J. Kim, "Diagnostics

[9] Explainable Machine Learning Model for Glaucoma Diagnosis and Its Interpretation," 2021, doi: 10.3390/diagnostics.

[10] B. Alsinglawi et al., "An explainable machine learning framework for lung cancer hospital length of stay prediction," Sci. Rep., vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-021-04608-7.

[11] L. Ibrahim, M. Mesinovic, K. W. Yang, and M. A. Eid, "Explainable Prediction of Acute Myocardial Infarction using Machine Learning and Shapley Values," IEEE Access, 2020, doi: 10.1109/ACCESS.2020.3040166.

[12] S. Ghosal, D. Blystone, A. K. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, "An explainable deep machine vision framework for plant stress phenotyping," Proc. Natl. Acad. Sci. U. S. A., vol. 115, no. 18, pp. 4613–4618, May 2018, doi: 10.1073/pnas.1716999115.

[13] Mesenchymal stem cells View project Market Risk Modelling View project." [Online]. Available: https://www.researchgate.net/publication/341914621.

[14] Kawatra Ruchi, V. Kumar, and S. Madan, "Hiding Information Along Fractal in a Digital Cover to Improve Capacity," in Lecture Notes in Electrical Engineering, 2020, vol. 605, pp. 1055–1070, doi: 10.1007/978-3-030-30577-2_93.

[15] Kawatra Ruchi, Saini Arora Sapna, "An Effective Approach towards Encryption of Limited Data", IITM Journal of Management and IT, vol. 7, Issue : 1, 32-36, 2016.

[16] Sharma, D., Kawatra, Ruchi (2023). "Security Techniques Implementation on Big Data Using Steganography and Cryptography". In: Fong, S., Dey, N., Joshi, A. (eds) ICT Analysis and Applications. Lecture Notes in Networks and Systems, vol 517. Springer, Singapore. https://doi.org/10.1007/978-981-19-5224-1_30

[17] Lipovetsky, S.; Conklin, M. Analysis of Regression in Game Theory Approach. Appl. Stoch. Models Bus. Ind. 2001, 17, 319–3

[18] V. Rachapudi, and G.L. Devi, "Feature Selection for Histopathological Image Classification using Improved Salp Swarm Optimizer", Recent Patents Comput. Sci., Vol. 12, No. 4, pp. 329-337, 2019. https://doi.org/10.2174/2213275912666181210165129

[19] R. Krishnamurthi, N. Aggrawal, L. Sharma, D. Srivastava, and S. Sharma, "Importance of Feature Selection and Data Visualization towards Prediction of Breast Cancer", Recent Patents Comput. Sci., Vol. 12, No. 4, pp. 317-328,2019. https://doi.org/10.2174/2213275912666190101121058.

[20] S. Arora, M. Agarwal, and R. Kawatra, "Prediction of educationist's performance using regression model," in 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), 2020, pp. 88–93.