iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** XII **Month of publication:** December 2022

DOI: https://doi.org/10.22214/ijraset.2022.48432

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XII Dec 2022- Available at www.ijraset.com

How Machine Learning Can Be Used To Improve Predictive Analytics

Monisha Gottam

Abstract: ***Machine learning algorithms are being used to improve predictive analytics. These algorithms are getting better at predicting future data. This is a great thing. But there are a few things you need to know before you begin to use these algorithms. Machine learning is an important part of predictive analytics and can increase the speed of data analysis and processing. It also allows predictive analytics algorithms to learn on larger data sets and conduct deep analysis on multiple variables.***

As a result, machine learning has become an important part of many businesses. Though there is still some controversy regarding its use, many industries are successfully implementing it. This technique has a long history of use in the financial sector, particularly in banking and investing, E-commerce, Customer service, Medical Diagnosis, Sales and Marketing, Financial Services, Cybersecurity and can help organizations forecast asset values. Additionally, it can help users understand the relationships between variables.

The latest advances in predictive analytics use neural networks, which simulate the functions of the human brain. Machine learning algorithms can be used by businesses for several applications, including healthcare. For example, predictive analytics can help organizations save money by optimizing staff schedules, identifying patients at risk of readmission, and improving pharmaceutical and supply management. The first step is finding datasets for the machine learning model. This can be done through various methods, including data discovery and data augmentation. Data discovery involves using existing datasets or searching for new ones. Data augmentation refers to using crowd-sourced datasets to complete machine-learning tasks. In this article we'll cover: automated machine learning and data-driven predictive analytics.

Keywords: Machine Learning, Predictive Analytics, Algorithm, XGBoost, Naive Bayes, K-fold, Variables, Time series, Scalability, Automated, Cybersecurity, Decision making.

I. INTRODUCTION

A machine-learning model can be trained by using a training dataset and a test dataset. The trained model can then predict the dependent variable for each instance of data. The dependent variable can be categorical. The difference between the predicted and actual category is called the generalization error. The underlying machine learning model can then be evaluated by comparing the values of the predicted and test data. Machine learning algorithms are often the basis for predictive analytics. These algorithms allow companies to make informed decisions based on data. In today's business environment, predictive analytics is increasingly becoming a vital component of a business' success. By understanding the data underlying customer behavior, predictive analytics algorithms can help companies make better decisions based on those insights. Machine learning models are often used to solve classification problems, but they can also be applied to project analytics. Unlike traditional statistical methods, machine learning models can identify significant features despite the collinearity and distribution of data. However, current studies of machine learning in project analytics have been relatively limited. Most have focused on project cost estimation and risk prediction, although some have used machine learning to improve the optimisation of existing processes. Machine learning models can be used to learn about data by training them with unsupervised or supervised learning methods. The former requires labels on training data, while the latter requires unlabeled data. Research has shown that using unlabeled data can improve learning accuracy.

II. LITERATURE REVIEW

In the last 20 years, there have been countless scientific papers published. It can be difficult to keep track of all of them, especially if you are an academic researcher, data science enthusiast, or even an industry expert. The good news is that there are now techniques for automated literature review. A lot of studies have focused on LA and EDM, few have addressed how to use machine learning to improve predictive analytics. While model interpretability is crucial for building trust and meeting compliance requirements, the key to developing effective predictive models is feature engineering. Currently, a variety of features are being used predictive analytics, which are effective and how they interact in digital world (Buchanan & Fan, 2017).

A. Machine Learning

Machine learning is a powerful analytical tool that uses data and mathematical models to predict the likely future behavior of things or people. This can help in a variety of areas, from predicting future sales to predicting the maintenance needs of machines. In addition, predictive analytics can help detect network intrusions. It can also help businesses better understand the behaviors of their customers (Ongsulee, *et al.*, 2018).

B. Predictive Analytics

Machine learning can help to improve predictive analytics in many ways. First, it can help to build an architecture for the application of predictive analytics. Secondly, it is essential to have quality data to learn from. This data should be centralized, unified, and in a consistent format. Additionally, it is essential to know what problems you want to solve with the data, since this will help you determine the best model to use. Finally, predictive analytics can help to improve efficiency across all stages of a business (Peterson, 2019).

C. Automated Machine Learning

Automated machine learning is a type of machine learning that allows non-scientists to gather data and make informed decisions. Its goal is to remove human error from the process, but it is not a substitute for a data scientist (Leung, *et al.*, 2020).

D. Data-driven Predictive Models

Data-driven predictive models are becoming increasingly powerful for a variety of uses. They help healthcare organizations predict patient prognosis and response to treatments. They can use data from a variety of sources, including a patient's medical history, environment, social risk factors, genetics, and more. They also help businesses understand what their customers want and how to meet their needs. This type of analysis can help businesses meet their targets and achieve higher levels of service (Nichol, Batten & Cho, 2021).

E. Customer Segmentation

One way to improve customer segmentation is to use machine learning algorithms. These algorithms cluster datasets based on certain patterns and are capable of dividing customers into highly specific groups. They also require little or no human supervision. Hence, they can be used for predictive analytics without hiring a data scientist (Zhang, Chen & Hong, 2022).

F. Cost-benefit Analysis

The adoption of predictive analytics in healthcare has the potential to transform the healthcare industry by delivering highly personalized patient care. Increasing pressure on health care providers to improve coordination of care, improve patient outcomes, and drive positive business results has led to the development of machine learning algorithms to process real-time and historical data (Henrys, 2021). These algorithms can help physicians to detect warning signs and potentially life-threatening conditions before they occur.

G. E-Commerce

Machine learning algorithms can learn patterns and trends from data through pattern recognition. They can also apply those patterns and trends to new data. Supervised machine learning involves labeled data to make predictions; unsupervised machine learning uses unlabeled data to find patterns (Antoniou, *et al.*, 2022).

H. Medical Diagnosis

Machine Learning algorithms can be programmed to use past patient data to predict future outcomes. This can help healthcare professionals make more informed decisions about a patient's risk for serious complications and mortality. These techniques can be used to improve patient outcomes, as well as improve the efficiency of the healthcare system (Buchanan & Fan, 2017).

I. Sales and Marketing

Machine learning is an emerging technology that uses algorithms to improve predictive analytics. This technology has multiple benefits for businesses. It helps organizations detect fraud, measure market risk, and identify opportunities. It also helps in improving KPIs (Ongsulee, *et al.*, 2018).

For example, machine Learning models have been developed and evaluated for predictive analytics. Typically, they are used for classification problems but can also be applied in project analytics. In these cases, it is necessary to convert non-categorical objective variables into categorical variables. For example, a project cost may be categorized into two possible values. Then, a machine learning model can be developed that can make correct predictions (Peterson, 2019).

J. Cybersecurity

Predictive analytics can be used to help cybersecurity practitioners detect and prevent threats. These analytics can make cybersecurity easier, less expensive, and more effective. But to make these tools effective, they need a complete picture of data. Machine learning algorithms require rich data from everywhere and must be able to represent as many potential outcomes as possible. Machine learning algorithms are becoming increasingly useful in cybersecurity, because they can help cybersecurity teams identify threats and respond to security incidents more quickly and effectively. These algorithms are already built into many security tools, and they are gradually replacing traditional methods of inference. A popular machine learning algorithm is regression, which seeks to identify correlations between datasets (Leung, *et al.*, 2020). For example, a regression algorithm can be used to predict the next system call of an operating system process, and then compare the predicted call to the actual call. Using machine learning techniques, a cybersecurity expert can create a model of network behavior and predict potential attacks. Then, he can determine which areas of the network are more vulnerable to attacks. This type of analysis can improve cybersecurity by identifying which areas are most susceptible to attacks and assessing the likelihood of each attack occurring. Machine learning algorithms can improve email spam detection, malware detection, and social engineering detection. They can also characterize email threats by reading the body and header data and detecting common patterns (Nichol, Batten & Cho, 2021).

III. PREDICTIVE ANALYTICS METHODOLOGY

Machine Learning is a powerful tool for predictive analytics. It can identify patterns in data and predict future behavior. Predict iQ is an advanced machine learning solution that uses data mining to create detailed images of customer behavior. For example, it can predict when a customer will leave a company, which can allow a company to repair the relationship before it breaks down. Its machine learning features are optional, so users can choose which datasets to analyze (Zhang, Chen & Hong, 2022).

1) *Unsupervised Learning:* Unsupervised learning is a powerful technique that enables computer algorithms to learn from a large set of data. This technique is useful in various fields, including gene sequence analysis, market research, and object recognition. For example, cell phone companies use machine learning algorithms to predict the number of clusters of people who rely on cell towers to receive service. Since cell phones can only communicate with one tower at a time, they need to identify clusters to optimize cell tower placement. The principle of unsupervised learning is to build a neural network from a set of input data. The training data is fed to the first layer, which trains on the data. The output of the first layer is then given to the second layer. This process is repeated until the desired number of layers is reached. The final layer then uses the representations it has learned from the previous layers to perform different tasks (Henrys, 2021).
2) *Automated Machine Learning:* Automated machine learning algorithms are used to improve predictive analytics and customer service. These algorithms use complex data sets to create a 360-degree view of a prospective customer. This type of analysis allows for improved lead conversion rates, customer segmentation, and marketing strategies. For example, machine learning-based predictive analytics can identify customers who are on the verge of leaving a company. This insight can then be used to better design packages to attract new customers and retain existing ones. Automated machine learning tools can process large data sets quickly and make accurate predictions based on that data. These tools can help make predictive analytics accessible to non-technical users, such as business analysts and marketers. In addition, automation makes ML technologies scalable, benefiting both large and small organizations (Antoniou, *et al.*, 2022)
3) *Variables in Machine Learning:* In machine learning, a variable represents a feature of a dataset. It is a feature that represents an attribute that an analyst wants to predict. In supervised machine learning, the model learns the relationship between the dataset and the target variable. The target variable may vary depending on the objective and the data available. Without a labeled target, a machine learning algorithm cannot map the data to the desired outcomes (Ongsulee, *et al.*, 2018)
4) *Time Series Modeling:* According to Peterson, (2019) another type of predictive analytics is called time series modeling, which captures data points as they change over time. Time is one of the most common independent variables used in predictive analytics. With this type of model, organizations can explore and forecast multiple scenarios.

5) *Scalability of Machine Learning:* Scalability is critical when deploying machine learning for predictive analytics. Machine learning applications must scale as users increase, and the number of models increases. Otherwise, the processing requirements may outgrow the capacity of the model. The LinkedIn article provides an example of how this can result in the incorrect recommendation of products or services (Leung, *et al.,* 2020). Scalability is a critical issue for machine learning because a system must be scalable in both space and time. Even though some simple systems can operate on a single computer, more complex models require a distributed architecture. The data must be available on dozens or hundreds of servers.
6) *Impact on Decision Makers:* Machine learning has several benefits for decision makers. It can make better predictions and reduce the need for human intervention. It can also speed up the decision making process. In many cases, it can make decisions that were once considered impossible. This technology can be a powerful tool for businesses of all sizes and sectors (Nichol, Batten & Cho, 2021).
7) *Predictive Analytics is a Growing Field*: Predictive analytics is a growing field that is taking advantage of machine learning and AI algorithms. These algorithms analyze massive data sets to identify trends and patterns. They can predict things like riskiness of mortgage applicants, interest rates and loan size, and can even detect fraud (Antoniou, *et al.,* 2022).

IV. RESULTS AND DISCUSSIONS

The use of machine learning algorithms continues to advance, resulting in better and more accurate predictive analytics. This article discusses the XGBoost algorithm, Convolutional neural networks, and K-fold cross-validation. You can also learn more about Naive Bayes models.

1) *XGBoost algorithm:* The XGBoost algorithm is a machine learning method that can solve predictive analytics problems. It works by creating decision trees in a sequential manner. In the first decision tree, weights are assigned to independent variables. In the second decision tree, these weights are increased, and individual classifiers combine to create a stronger model. The algorithm can work for classification, regression, ranking, and user-defined prediction problems (Ongsulee, *et al.,* 2018).
2) *Naive Bayes:* Machine learning is an important component of predictive analytics and provides a way for organisations to transform data into insights. It can help identify fraud, measure market risk, and identify opportunities. With the right amount of data storage, machine learning algorithms can help organisations make better business decisions (Peterson, 2019).
3) *Convolutional neural networks:* CNNs use Convolutional Neural Networks to recognize objects in images. It can even be used to find new drugs. The CNN algorithm works by capturing the way we view and operate within the world (Leung, *et al.,* 2020).
4) *K-fold cross-validation:* K-fold cross-validation is a statistical method used to test the predictive accuracy of a model. It involves dividing data into sets of near-equal size. One set is considered the test set, while the other k-1 sets are used to train the model. After fitting the model to the test data, the test error rate is calculated. Then, the process is repeated for all k sets, resulting in a mean error estimate. The k-value is chosen carefully to ensure that the test data set is representative of the entire dataset (Nichol, Batten & Cho, 2021).
5) *Image recognition:* Machine learning algorithms have a variety of applications in predictive analytics. They can predict consumer behavior, improve marketing strategies, improve financial analysis, detect network intrusions, and more. These algorithms draw on historical data and current statistics to provide insights and support business decisions (Zhang, Chen & Hong, 2022).
6) *Applications:* Machine learning is a powerful tool that helps companies predict and optimize business processes. Its applications span a wide range of fields. For example, predictive analytics can help businesses understand what kind of products and services their customers are looking for. It can also help companies understand their markets and develop more effective sales and marketing strategies (Henrys, 2021; Antoniou, *et al.,* 2022).

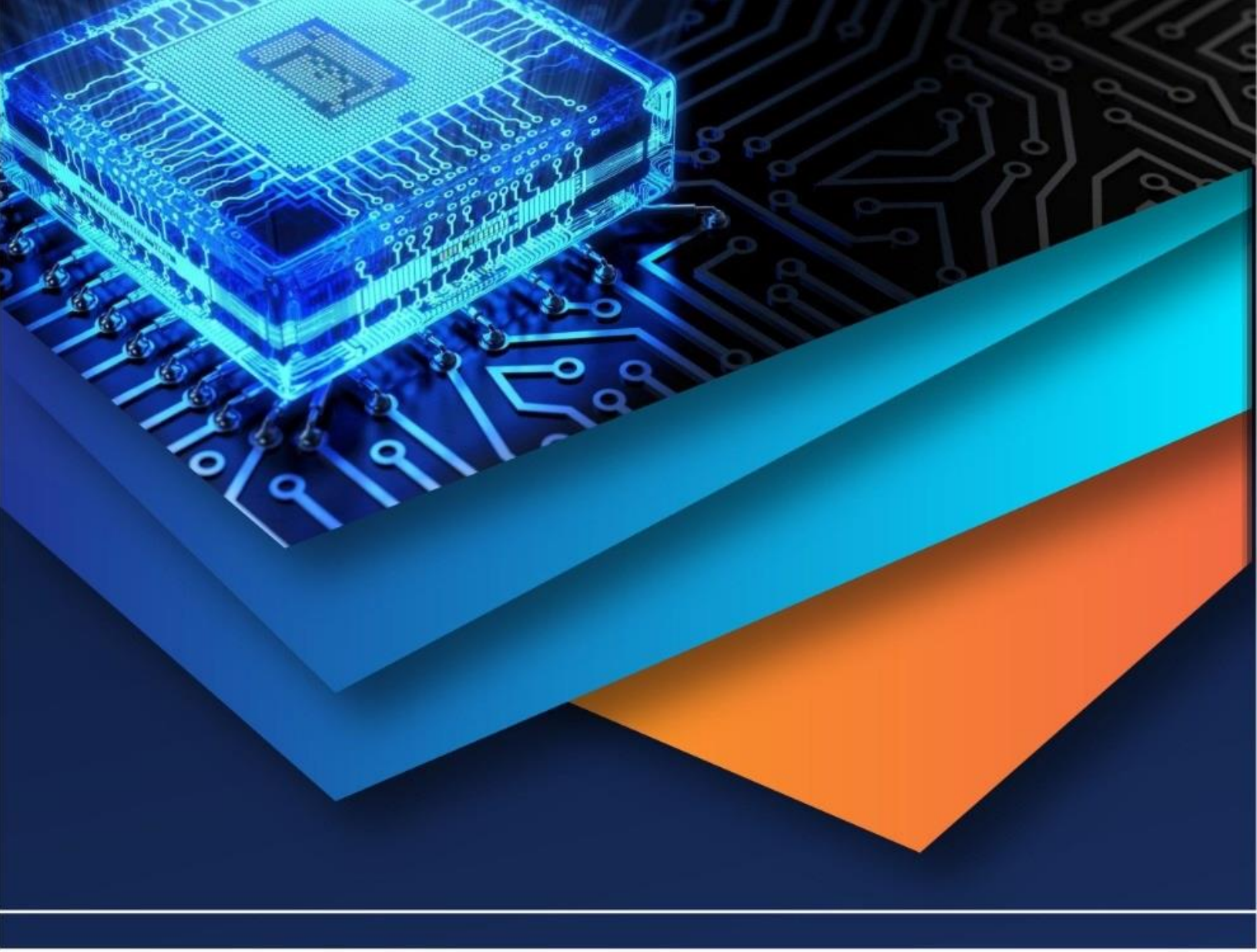
V. CONCLUSION

Machine Learning is an advanced analytics method that uses algorithms to generate and interpret data. It has many uses in businesses and is increasingly popular for commercial use. It is a powerful tool for organizations to gain hidden insights from historical data. It can identify and predict trends in consumer behavior. Predictive analytics is also used in the health industry to better identify and respond to high-risk customers. It can also be used in insurance claim collection and fraud detection. It can help optimize marketing campaigns and promotions, as well as evaluate customer behavior and identify up-sell and cross-sell opportunities. The education sector, for example, collects mountains of data. Data sampling can help organizations build better predictive models using existing student records. For instance, schools can use records from past students to create algorithms that predict new students' performance.

These insights can help improve student outcomes and improve the operation of institutions. Machine Learning is an artificial intelligence (AI) technology that uses historical data to predict future values. The most popular application of machine learning is in recommendation engines, but it is also used in spam filtering, fraud detection, malware detection, and business process automation. “Finally, machine learning is an important tool for predictive analytics. Its algorithms can learn from vast data troves, making predictive models more accurate and adaptive over time. With the right machine learning model, companies can improve their business operations and generate more revenue, increase competitive advantage, and reduce costs.”

REFERENCES

[1] Henrys, K. (2021). Role of predictive analytics in business. Available at SSRN 3829621.
[2] Leung, C. K., Elias, J. D., Minuk, S. M., de Jesus, A. R. R., & Cuzzocrea, A. (2020, July). An innovative fuzzy logic-based machine learning algorithm for supporting predictive analytics on big transportation data. In 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.
[3] Nichol, A. A., Batten, J. N., Halley, M. C., Axelrod, J. K., Sankar, P. L., & Cho, M. K. (2021). A Typology of Existing Machine Learning-Based Predictive Analytic Tools Focused on Reducing Costs and Improving Quality in Health Care: Systematic Search and Content Analysis. Journal of medical Internet research, 23(6), e26391.
[4] Ongsulee, P., Chotchaung, V., Bamrungsi, E., & Rodcheewit, T. (2018, November). Big data, predictive analytics and machine learning. In 2018 16th international conference on ICT and knowledge engineering (ICT&KE) (pp. 1-6). IEEE.
[5] Oyeleye, M., Chen, T., Titarenko, S., & Antoniou, G. (2022). A Predictive Analysis of Heart Rates Using Machine Learning Techniques. International Journal of Environmental Research and Public Health, 19(4), 2417.
[6] Peterson, E. D. (2019). Machine learning, predictive analytics, and clinical practice: can the past inform the present?. Jama, 322(23), 2283-2284.
[7] Uwagbole, S. O., Buchanan, W. J., & Fan, L. (2017, May). Applied machine learning predictive analytics to SQL injection attack detection and prevention. In 2017 IFIP/IEEE Symposium on Integrated Network and Service Management (IM) (pp. 1087-1090). IEEE.
[8] Zhang, Z., Chen, L., Xu, P., & Hong, Y. (2022). Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. Laparoscopic, Endoscopic and Robotic Surgery, 5(1), 25-34.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY