RESEARCH ARTICLE OPEN ACCESS

Evaluation of Supervised Machine Learning Algorithms' Performance in Identifying Credit Card Fraud

Ms. Shweta Anil Kanojia
Research Scholar,Department of Computer Science,Science College, Nagpur.
Dr. R.N. Jugele
Professor, Department of Computer Science, Science College, Nagpur.

ABSTRACT
As online transactions have grown in popularity, the use of credit card transactions have become the most common online method. As the use of credit card is day by day, therefore the issues are also increasing rapidly. The process of making transactions with a fake cards is known as credit card fraud. Because it's becoming more and more prevalent, fraud seems to be a significant problem for the credit card industry. Advancements in technology and greater internet transactions have led to huge financial losses due to fraud. Machine learning automatically detects credit card fraud, but it ignores behavioural problems or fraud that could cause alerts. In this paper, Researcher have studied various Supervised Machine Learning Algorithm to predict and identify fraudulent credit card transactions.
Keywords: Credit Card Fraud Detection, Supervised Machine learning Algorithms, Fraud Detection.

I. INTRODUCTION

In the current era, the use of credit card is notably rising, as it has ease the daily needs. Due to the increasing popularisation the usage of credit card is also increasing in e-service fields like e-commerce sector, e-finance and mobile payments [14]. After the pandemic, covid-19 the use of credit card is significantly increasing because people, now prefers to make online payments [3]. Credit card fraud is the process in which someone's uses another persons card to make the transactions [6]. Credit card fraud occurs in banks, automakers and appliance manufactures [13]. This frauds are of two types: Application fraud and Behavioral fraud. When an application is fraudulent it is refers as application fraud. After the card is approved and issued behavior fraud occurs [14]. A branch of artificial intelligence called machine learning that learns from the past experience and makes the predictions based on that data. Supervised, unsupervised, and reinforcement learning are three distinct types of machine learning.

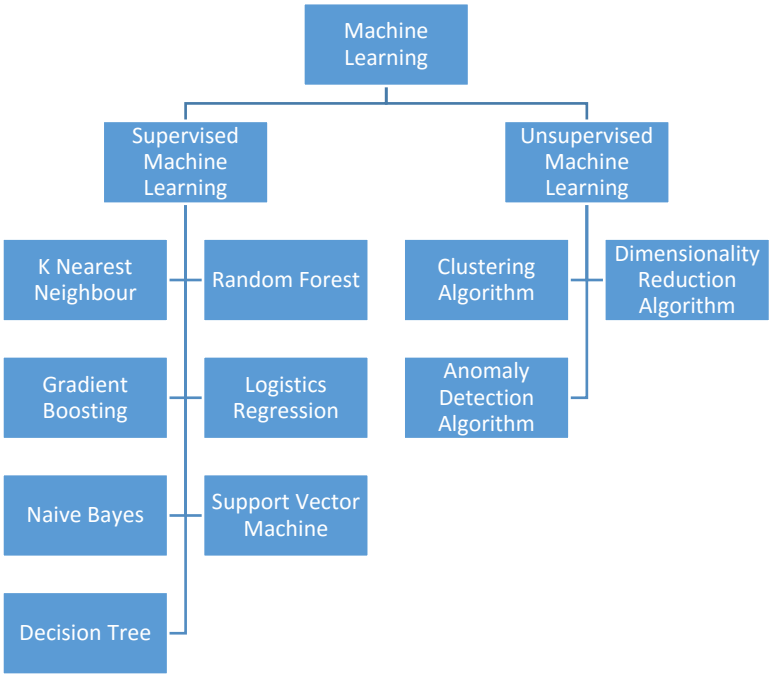

Fig 1: Classification of Machine Learning Algorithms.

This study employs seven supervised machine learning techniques to confirm which algorithm produces the best results and quickly identifies fraud in any circumstance. These techniques include the following:

a) K Nearest Neighbour
b) Random Forest
c) Gradient Boosting
d) Logistic Regression
e) Naïve Bayes
f) Support Vector Machine
g) Decision Tree

II. METHODOLOGY:

a) K-Nearest Neighbour:

A non-linear classifier called K Nearest Neighbour makes it easy to classify fresh data into an appropriate category, even if the data is not linearly separable [1]. It is the simplest way to classify the algorithm by looking at what's nearby. Another name for it is lazy learner algorithm, as it stores the information and then performs an action on it throughout the classification process, instead from the training set.

b) Random Forest:

A common machine learning method for real-world models and applications is random forest [1]. It is also called as the ensemble algorithm, which generates numerous decisions and uses major votes for classification and average for regression [2]. As the number of trees increases its accuracy increases [5]. It is Similar to decision tree, the random forest offers more choices for improving outcomes [4]. It is a blend of several classifiers [15]. In order to understand irregular patterns, deep trees are utilized [11]. Because of its efficiency and speed, an RF can be used to handle skewed datasets with hundreds of variables. To classify a class, each tree votes. The most class votes are given to object new when it is formed [14].

c) Gradient Boosting:

An algorithm, which is supervised is used to precisely predict target variables which combine estimates from a number of weaker and simpler models [12][3]. It is implemented by open source software called as XGBoost.

d) Logistic Regression:

This method is employed to find out the relationship between the variables [4]. It is an analysis approach of the prediction type. Another name for it is a classification algorithm [12]. One technique for predicting binary values (1/0, Yes/No, False/True) from independent variables is logistic regression [2]. Basically it is used for classification [13]. The model can estimate parameter coefficients using the sigmoid function and predict both binary and multinominal outcomes.

e) Naïve Bayes:

This algorithm uses Bayesian principle. It is used to find out the result and it is reliant on probability [10]. It determines if a transaction is fraudulent or not by taking the chance of an event (feature) and calculating the probability of another output [4].

f) Support Vector Machine:

SVM is an effective text method that works on linear model. It distinguishes between positive and negative examples with large margins. Compared to naïve bayes, the SVM produced better results for fraud detection [8]. This method creates a hyperplane and it divides a variety of qualities into groups in order to efficiently classify future data. The extreme points used while building a hyperplane using this method are known as support vectors [6].

g) Decision Tree:

To obtain the final outcome of those choices, it is utilized to divide the dataset according to attributes. The various nodes and branches that comprise this decision tree represent the dataset, functions, choices and output generated by this algorithm [1].

2.1 EVALUATION MEASURE:

Precision, Recall, F1 score and Accuracy are computed, and the confusion matrix is used to assess the outcome. The actual class and the expected class are its two classes. These characteristics serve as the foundation for the confusion metrics:

- True positive (TP)- It is a result in where both the predictive and actual values are positive.

- False positive (FP)- It is a result in which the predictive value is positive and actual value is negative.
- True negative (TN)- It is a result in which the predictive and actual values both are negative.
- False negative (FN)- It is a result in which the predictive value is negative and actual value is positive.

III. RESULTS and DISCUSSION :

A. Dataset Summary: The dataset consist of 550,000 European Card holders transactions which is taken from Kaggle, having 28 different attributes and it was saved as a CSV file.

B. Performance Measures: The algorithms performance is calculated by using the four following features-

i. Precision- It is defined as –

$$\text{Precision} = \frac{TP}{TP+FP}$$

ii. Recall - It is defined as-

$$\text{Recall} = \frac{TP}{TP+FN}$$

iii. F1 Score: The F1-score is defined as-

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

iv. Accuracy- It is defined as-

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Algorithms	Precision	Recall	F1 Score	Accuracy
K-Nearest Neighbour	0.9987	0.9999	0.9993	0.9993
Random Forest	1	1	1	0.9998
Gradient Boosting	0.9997	0.9995	0.9995	0.9996
Logistic Regression	1	1	1	0.9983
Naïve Bayes	1	0.9887	0.9943	0.9943
Support Vector Machine	1	1	1	0.9986
Decision Tree	0.9996	0.9996	0.9996	0.9996

Table 1: Represents the precision , recall, F1 score and accuracy

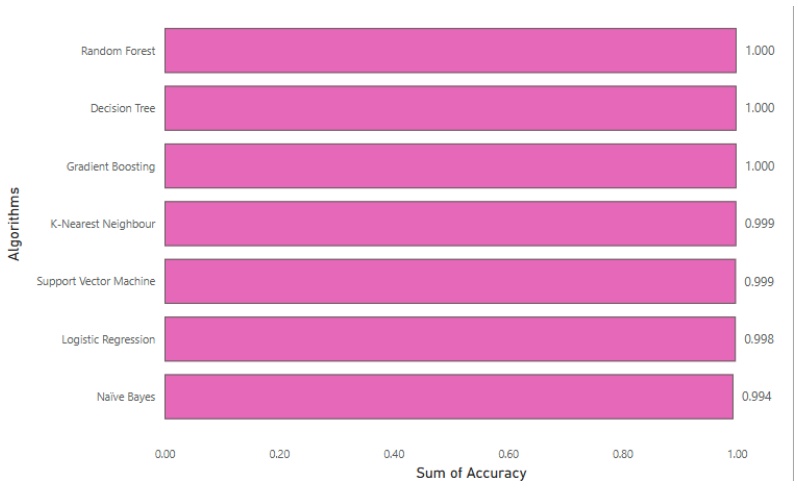

Fig 2: Accuracy of ML models.

Conclusion: In this research paper, Researcher have studied seven supervised machine learning techniques. Numerous fields can benefit from the application of machine learning such as Image Recognition, Speech Recognition, Medical or Healthcare sector, Fraud Detection and many more. Credit card Fraud detection is one of it. Fraud detection is much needed as day by day fraudster tries to attempt new method. Here, Researcher have studied several algorithms and it is found that the accuracy of Random Forest is comparatively high than other algorithm. Also, the execution time of KNN and SVM is more. Naïve Bayes and Logistics Regression execution speed is high and gives high result when feature scaling is used.

REFERENCE:

[1] Rupali Aggarwal, Pradeepta Kumar Sarangi, Ashok Kumar Sahoo, "Credit Card Fraud Detection: Analyzing the Performance of Four Machine Learning Models", 2023 International Conference on Disruptive Technologies (ICDT) | 979-8-3503-2388-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICDT57929.2023.10150782

[2] Vipul Jain, Kavitha H, Mohana Kumar S, "Credit Card Fraud Detection Web Application using Streamlit and Machine Learning", 2022 IEEE International Conference on Data Science and Information System (ICDSIS) | 978-1-6654-9801-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICDSIS55133.2022.9915901

[3] Suhas Jain G. M, N. Rakesh, Pranavi K, Lahari Bale, "A novel approach in credit card fraud detection system using machine learning techniques", 2021 International Conference on Forensics, Analytics, Big Data, Security (FABS) | 978-1-6654-2005-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/FABS52071.2021.97026721

[4] D. Tanouz, R Raja Subramanian, D. Eswar, G V Parameswara Reddy, A. Ranjith kumar, CH V N M praneeth, "Credit Card Fraud Detection Using Machine Learning", 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) | 978-1-6654-1272-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICICCS51141.2021.9432308

[5] Aditi Singh, Anoushka Singh, Anshul Aggarwal and Anamika Chauhan, "Design and Implementation of Different Machine Learning Algorithms for Credit Card Fraud Detection", 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME) | 978-1-6654-7095-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICECCME55909.2022.9988588

[6] Mr. P.Yogendra Prasad, A Sreni Chowdary, Cherapalli Bavitha, Earagaraju Mounisha, Chatna Reethika, "A Comparison Study of Fraud Detection in Usage of Credit Cards using Machine Learning", Proceedings of the 7th International Conference on Trends in Electronics and Informatics (ICOEI 2023) IEEE Xplore Part Number: CFP23J32-ART; ISBN: 979-8-3503-9728-4, DOI: 10.1109/ICOEI56765.2023.10125838

[7] Uqba Jabeen, Dr. Karan Singh, Satvik Vats, "Credit Card Fraud Detection Scheme using Machine Learning and Synthetic Minority Oversampling Technique (SMOTE)", 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) | 979-8-3503-2142-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICIRCA57980.2023.10220646

[8] Fawaz Khaled Alarfaj Muhammadramzan, Iqra Malik, Hikmat Ullah Khan , And muzamil Ahmed, Naif Almusallam, "Credit Card Fraud Detection Using State-of-the-Art Machine Learning and Deep Learning Algorithms", IEEE Access, April 12, 2022, DOI: 10.1109/ACCESS.2022.3166891

[9] Asifuddin Nasiruddin Ahmed and Ravinder Saini, "Detection of Credit Card Fraudulent Transactions Utilizing Machine Learning Algorithms" 2023 2nd International Conference for Innovation in Technology (INOCON) | 979-8-3503-2092-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/INOCON57975.2023.10101137

[10] Indrani Vejalla, Sai Preethi Battula, Kartheek Kalluri and Hemantha Kumar Kalluri, "redit Card Fraud Detection Using Machine Learning Techniques" , 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS) | 979-8-3503-1071-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/PCEMS58491.2023.10136040

[11] Prateeksha M.S , B. Naga Swetha and Manjula Patil , Credit Card Fraud Detection Using Machine-Learning , DOI: 10.21474/IJAR01/16824

[12] Nishank Jain, Alka Chaudhary and Anil Kumar," Credit Card Fraud Detection using Machine Learning Techniques ", 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART) | 978-1-6654-8734-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/SMART55829.2022.10047360

[13] Asha RB, Suresh Kumar KR, "Credit card fraud detection using artificial neural network," Global Transitions Proceedings , volume 2, Pages 35-41, June 2021, https://doi.org/10.1016/j.gltp.2021.01.006

[14] Xinwei Zhang, Yaoci Hana, Wei Xu, Qili Wang, "HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture", Information Sciences, vol. 2, no. 1, pp. 35-41, June 2021, Volume 557, Pages 302-316, May 2021, https://doi.org/10.1016/j.ins.2019.05.023