**The Online Payment Fraud Detection System.**



A Minor Project Report

**Computer Science & Engineering**

By Team 117

Roll.No: 2103A2010 Name: CH. RISHIKA

Roll.No: 2103A2011 Name: CH. ALEKHYA

Roll.No: 2103A2018 Name: G. SANJANA

Roll.No: 2103A2090 Name: K. MADHU SRI

**Under the Guidance of**

**Dr. Balajee Maram,**

**Professor,**

**School of CS & AI**



**DEPARTMENT OF COMPUTER SCIENE AND ARTIFICIAL INTELLIGENCE**

**CERTIFICATE**

This is to certify that the Minor Project Report Titled “ONLINE PAYMENT FRAUD DETECTION” is a record of a bonafied work carried out by the student(s) Ch. Rishika, Ch. Alekhya, G. Sanjana, K. Madhu Sri bearing Roll No(s) 2103A52010, 2103A52011, 2103a52018, 2103A52090 during academic year 2023-24.

**Supervisor Head of the Department**

**Table of contents:**

|  |  |
| --- | --- |
| **CONTENTS** | **PAGE.NO** |
| 1.ABSTRACT | 1 |
| 2.INTRODUCTION | 1 |
| 3.OBJECTIVE | 2 |
| 4.EXISTING SOLUTION | 2 |
| 5.PROPOSED SOLUTION | 2 |
| 6.LITERATURE SURVEY | 3 |
| 7.PROBLEM STATEMENT | 4 |
| 8.PROPOSED WORK AND METHODOLOGY | 4 |
| 9.EXPERIMENTAL WORK | 5 |
| 10.IMPLEMENTATION | 7 |
| 11.RESULTS | 8 |
| 12.CONCLUSION | 11 |
| 13.FUTURE SCOPE | 11 |
| 14.REFERENCES | 12 |

**THE ONLINE PAYMENT FRAUD DETECTION**

1. **ABSTRACT:**

In this document, we delve into the essentials of crafting a robust online payment fraud detection system. We explore the critical elements like gathering data, employing machine learning models, real-time monitoring, and integrating external data sources. Throughout, we stress the importance of staying adaptable and innovative to counter evolving fraud techniques. By grasping these methodologies, businesses can fortify their defenses against fraud, protecting finances and nurturing customer confidence in online transactions. This guide offers practical wisdom and insights, empowering organizations to proactively detect and combat fraudulent activities. Through clear and concise language, we aim to facilitate a deeper understanding of fraud detection strategies, fostering secure and trustworthy online payment environments.

**2.INTRODUCTION:**  
  
As e-commerce continues to expand at a fast pace and digital transactions become more prevalent, online payment fraud has become an important issue for both businesses and consumers. The convenience brought by online shopping and banking has its own downsides as criminals always find new ways to exploit payment systems’ weaknesses. These include stolen credit card details and account takeover scams among others which are manifold and sophisticated whereby they continuously pose a threat for financial security as well as consumer trust.

For organizations to address this emerging problem, they must establish strong detection system that detect and prevent real-time fraudulent activities. To achieve this, these systems use a variety of advanced technologies such as machine learning algorithms, behavioral analysis techniques, and real-time monitoring functionalities. These mechanisms can separate authentic from fake transactions in order to lower the monetary losses and reputation ruin caused by online transaction frauds through an extensive examination of transaction information recorded together with customer’s behaviour patterns. The effectiveness of a fraud detection system depends on its ability to actively combat emerging threats and adapt to changing fraud strategies. Innovation and continuous improvement are important in order to be one step ahead of criminals who continuously device new tricks of avoiding security measures. Moreover, collaboration and sharing information among industry players is crucial in cementing collective defense against online payment fraud.

This paper examines the major constituents and approaches that can help develop an efficient system for identifying the frauds associated with online payments. Data collection, machine learning models, real-time monitoring, and integration with external data sources are just some components essential for improving the overall security within digital payment ecosystems. Implementation of these actions by firms ensures continuity of their operations while making sure that buyers have confidence hence a environment where e-commerce can thrive.……………………………………………….

**3.OBJECTIVE:**  
  
This document aims to offer an overall picture of what is needed in terms of the fundamental components and methodologies for a well-functioning online payment fraud detection system. This will explore themes such as data collection, machine learning models, real-time monitoring and data integration from external sources in order to equip businesses with the knowledge and tools they need to detect fraudulent activities and prevent them from happening in their digital transactions. Besides, it is meant to emphasize on the significance of always adjusting and innovating because frauds are dynamic and change tactics through time.

This guide seeks therefore concise explanation as well as practical insights that would help organizations strengthen their capabilities in detecting fraud so as to protect financial assets while retaining customer trust. Ultimately, this could assist businesses in identifying fraudulent transactions before they take place by outlining the best practices and strategies for fraud detection, hence minimizing the risks linked with these activities, thereby strengthening trustworthiness of online payment ecosystems.

**4.EXISTING SOLUTION:**

Online fraud detection mechanisms are designed to ensure that online payment services are safeguarded from fraudulent activities. Existing solutions for online payment fraud detection, employ a multifaceted approach combining machine learning algorithms, behavioral analytics, rules-based systems, biometric authentication, device fingerprinting, IP Geo-location, transaction velocity monitoring, collaborative filtering, and real-time monitoring. ML algorithms examine the data of transactions for identifying patterns that will suggest fraudulence while behavioral analytics observes user behavior for anomalies. Rules-based systems establish predetermined thresholds to identify suspicious transactions whereas biometric authentication verifies the identity of users. Beyond this device fingerprinting and IP Geo-location detect irregularities in device or location patterns while transaction velocity monitoring identifies abnormal spikes of transactions. Collaborative filtering and network analysis expose fraudulent networks while real-time monitoring triggers the alerts on possible frauds. Organizations have integrated these technologies into strategies to mitigate the risks associated with online payment processes and protect themselves against financial losses.

**5.PROPOSED SOLUTION:**

We propose a fraud detection solution for the data set "fraud\_detection\_dateset.csv" that utilizes advanced optimization techniques to enhance accuracy. Initially, we begin processing the data by doing such activities as cleansing of data, the creation of new attributes and normalization in preparation for analyses. Moving on, we design machine learning models using various algorithms which include logistic regression, random forest and gradient boosting for this data to enable us detect unusual behaviors. We optimize these models through hyper-parameter tuning, feature selection and ensemble learning among other strategies. Moreover, with regards to fraud detection datasets imbalanced due to class imbalance issues over sampling and under sampling are re-sampling methods used. All through our model development process, we evaluate performance metrics like precision recall and f1 score extensively so as to ensure effectiveness in detecting fraudulent activities while limiting false positives. Optimizing our fraud detection solution should improve its accuracy in spotting online payment fraud risk exposures faced by businesses and financial institutions at a large scale and also make it more robust during implementation.

**6.LITERATURE SURVEY OR RELATED WORK:**

Li et al. (2018) proposed that Enriching Online Payment Fraud Detection by Way of Combining Learning Approaches. This work seeks to investigate how effective ensemble learning techniques, including boosting, and bagging, are at improving the fraud detection accuracy on online payment systems. It also compares the performance of ensemble methods and stand-alone classifiers while explaining its prospects in real life. Yan et al. (2017) proposed that an Application of Machine Learning Techniques in Fraud Detection for Online Payment Systems.The writers examine the usage of the machine learning techniques in identifying online payment system frauds. They explore patterns that indicate potential risks in the transaction records with various machine-learning algorithms.

Zhang et al. (2016) Anomaly Detection in Online Payment Systems showed that this essay examines anomaly detection techniques in online payment systems. It details statistical, machine learning and behavioral approaches that can be exploited to identify deceitful transactions. Li et al. (2020) Behavioral Analysis for Fraud Detection in Online Payments proposed that it is about investigating the use of behavior analysis as a tool for preventing frauds associated with digital payments by Li and his colleagues. In relation to this, user behaviors are compared with transaction histories so as to enable the adoption of an approach that is based on behaviorally based analytical techniques which can be used to detect fraud through such indicators.

Chen et al. (2018) Data Mining Techniques for Online Payment Fraud Detection says that the paper’s topic revolves around data mining methods used to unveil electronic payment scams by Chen, Chen and Chen. In this situation, they employ clustering, association rule mining together with decision tree analysis among other strategies aimed at identifying suspicious patterns. Zhou et al. (2019) Fraud Detection in Blockchain-Based Payment Systems says that in this article, the authors address how blockchain technology could potentially help uncover instances of fraud within the payment systems built on its platform. As for transparency and accountability.

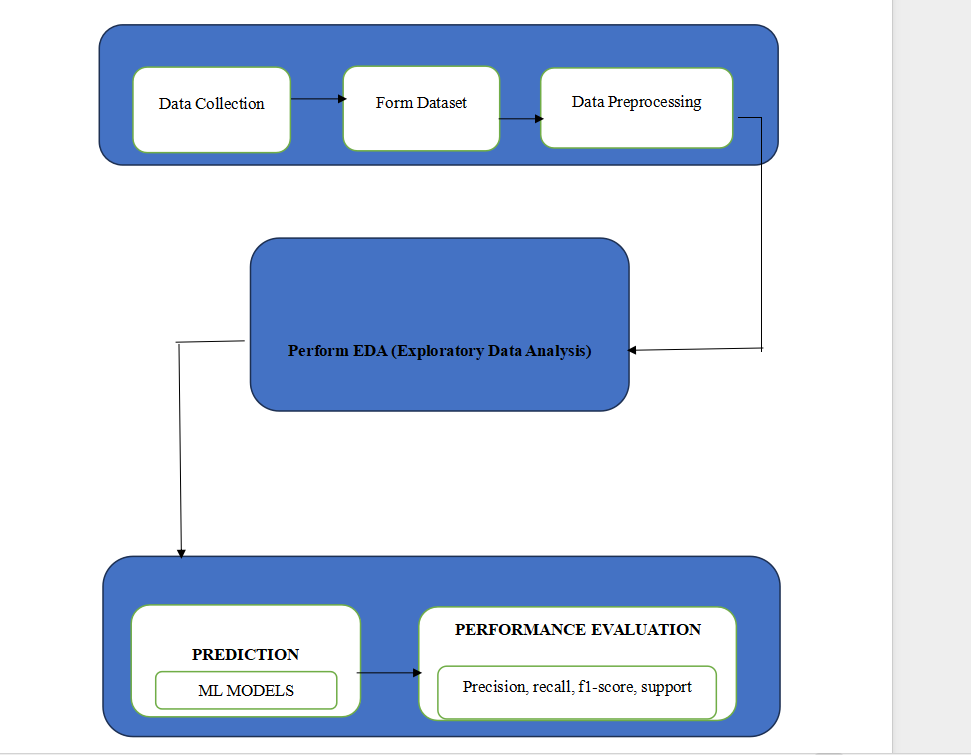
Liu et al. (2017) Real-Time Fraud Detection in Online Payment Systems proposed that a real-time fraud detection system for payment transactions is presented by Liu et al. They also look at some of the setbacks that come with working in real-time and share out their strategies on how to effectively track fraudulent activities in online payment systems. Yang et al. (2018) Risk Management Techniques for Online Payment Fraud Detection says that this paper investigates risk management techniques which can be employed to detect and counteract online payments frauds. There is a discussion on proactive risks assessment as well as detection of potential indicators of fraud from transaction data.

X et al. (2020) Hybrid Models for Online Payment Fraud Detection Proposed hybrid models, X et al. combine many fraudulent detecting methods so that they could be much more accurate and reliable. They also talk about combining rule-based systems, machine learning algorithms, and anomaly detection methods to bring in different viewpoints on integration.Wang et al. (2021) Deep Learning for Fraud Detection in Online Payments proposed that in this research paper, Wang et al. scrutinize methods such as convolutional neural networks, and recurrent neural networks, which uses deep learning that are used for detecting frauds on online payment transactions. The applicability of these approaches is shown on large volumes of transaction data with complex patterns.

**7.PROBLEM STATEMENT:**

The threat of online payment fraud to business and financial institutions is enormous, leading to financial loss, a tarnished image and loss of customer loyalty. The challenge lies in identifying precisely fraudulent transactions out of a vast number of legitimate ones especially with the changing ways through which fraudsters are operating. The goal of this project is to create an enhanced system for real time detection of fraud that will correctly identify illegal activities. It should be capable of analyzing different transaction attributes, user behavior patterns as well as contextual information in order to differentiate between genuine and fake transactions. In addition, the alternative may be scalable, adjustable in case there is a change on modus operandi by culprits, thus reducing false positives so that consumers can have a smooth experience while using it. With these issues properly addressed it will become possible for businesses and financial institutions to obtain effective anti-fraud measures which ensure operational efficiency and customer satisfaction even as they combat against internet-based payment system malpractice.

1. **PROPOSED WORK AND METHODOLOGY:**



**Steps:**

1. Firstly, we need to collect data.
2. Form a dateset from the collected data.
3. Then we do data preprocessing on the dateset formed.
4. Then we perform- EDA (Exploratory Data Analysis) on the dateset.
5. For the prediction process, we use different Machine Learning (ML) models.
6. Lastly, for evaluating the performance we use precision, Recall, F-Score, and Support.
7. **EXPERIMENTAL WORK:**

Here, we usually follow the architecture and perform the steps according to architecture.

We collected data using various parameters and formed a dateset.

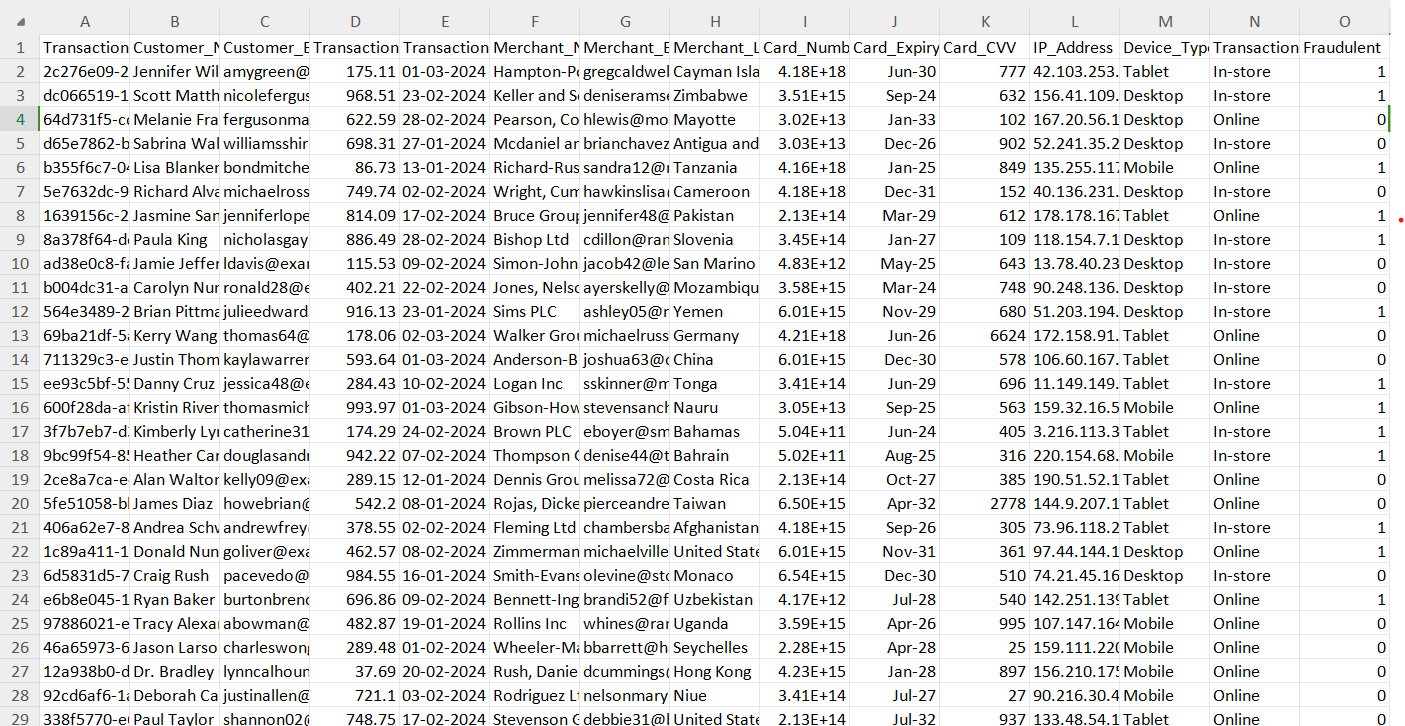
**Dataset Description:**

1. The dateset used is “fraud\_detection\_dateset.csv”.
2. It consists of 2000 rows and 15 columns.
3. Each entry includes a unique transaction ID, customer details such as name and email, transaction specifics like amount and date, merchant information, including name, email, and location, card details, including number, expiry, and CVV, IP address,the device type used for the transaction (desktop, mobile, or a tablet), transaction type (online or in-store), and a binary indicator for potential fraud. This dateset provides a diverse range of synthetic data reflecting transaction patterns, enabling the development and testing of fraud detection algorithms and models.
4. In this dataset- Fraudulent is the target column which consists of 0 or 1, where 0 implies Not Fraudulent and 1 implies Fraudulent.
5. Transaction\_ID, Customer\_Name, Customer\_Email, Transaction\_Amount, Transaction\_Date, Merchant\_Name, Merchant\_Email, Merchant\_Location, Card\_Number, Card\_Expiry, Card\_CVV, IP\_Address, Device\_Type, Transaction\_Type are the freatures.
6. Out of 2000 rows there are 1003 rows that are not Fraudulent and 997 rows that6 are Fraudulent.

Dataset link: <https://www.kaggle.com/datasets/smmmmmmmmmmmm/fraud-detection-financial-transactions/data>

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset Name** | **Number of rows** | **Number of columns** | **Features** | **Target** |
| fraud\_detection\_dateset.csv | 2000 | 15 (Features + Target) | Transaction\_ID, Customer\_Name,  Customer\_Email,  Transaction\_Amount,  Transaction\_Date,  Merchant\_Name, Merchant\_Email, Merchant\_Location, Card\_Number, Card\_Expiry, Card\_CVV, IP\_Address, Device\_Type, Transaction\_Type | Fraudulent  Values:  0-Not Fraudulent  1-Fraudulent |

**Dataset:**



1. **IMPLEMENTATION:**

**Algorithms used:**

1. **Random Forest:**

An approach to collective learning is Random Forest Regression. During training, several decision trees are constructed. The average (for regression) or mode (for classification) of each tree's contribution to the final prediction is used. It is renowned for its capacity in handling different types of data and reliability against over fitting.

1. **Logistic Regression:**

Logistic regression is a statistical technique used to model binary outcomes by estimating the likelihood of an event occurring based on one or more predictors. This logistic function maps output to a probability between 0 and 1 whereby coefficients that best fit the data are calculated in order to predict the probability of an event for new observations.

1. **Support Vector Machine (SVM):**

SVR is a regression method that extends Support Vector Machines to issues related to regression. It locates a hyperplane that minimizes the error margin and most accurately matches the data. SVR works well at identifying complex patterns in data and is especially helpful when working with high-dimensional data.

1. **K Nearest Neighbour(KNN):**

K-Nearest Neighbors (KNN) is a type of non-parametric supervised learning algorithm used for classification and regression tasks. It involves finding, among the training set, K closest data points to a new point, then assigning their majority class (classification) or averaging their values (regression) to the new point. The value of k i.e., the number closer neighbors considered affects performance of this model.

1. **Decision Tree:**

Decision tree can be described as is that it’s a type of superised learning algorithm most commonly applied within both regession and classification. This would divide feature space into sections on basis of input features’ values and create tree like structure having leaf nodes as well as decision nodes. Each node will make some decision concerning impurity minimization in regards to variance for classification/regression cases thus forming sequences involving binary decisions finally ending with estimated classes at leaves.

1. **XGBoost Classifier:**

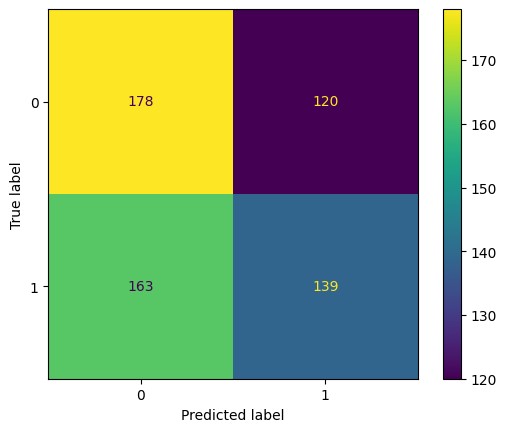
A advanced use of gradient boosting is called XGBoost (Extreme Gradient Boosting).It is appropriate for huge datasets because of its great efficiency and flexibility. To avoid overfitting, XGBoost employs normalized trees and performs well on regression problems.

1. **Artificial Neural Networks Regression:**

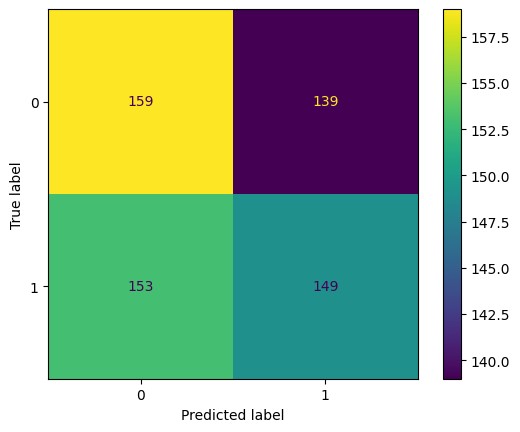
Artificial neural networks are used in neural networks regression to represent complicated connections in data. It is made up of layers of linked neurons that can recognize patterns in input data. Although very adaptable and capable of collecting complex patterns, neural networks may need careful tuning and considerable data.

1. **RESULTS:**

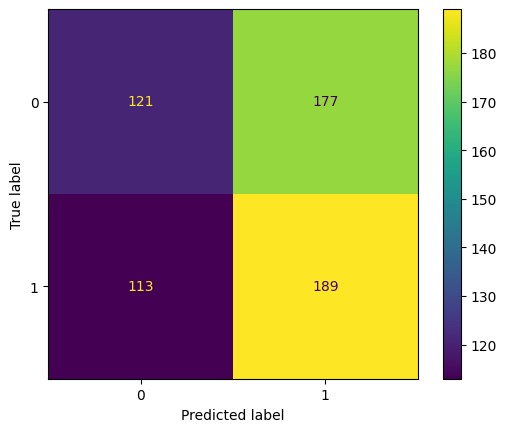
**Random Forest:**



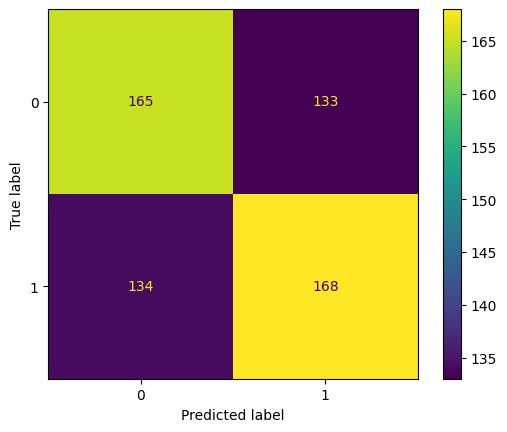
**Logistic Regression:**



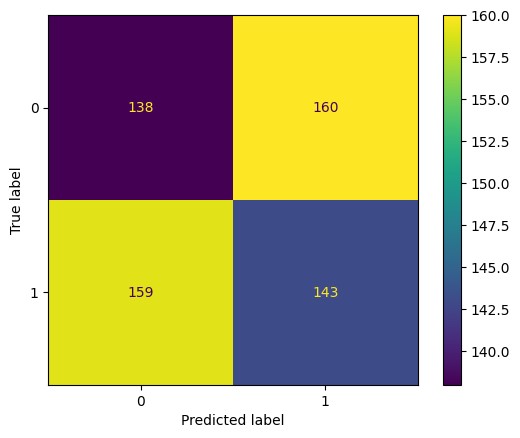
**Support Vector Machine:**



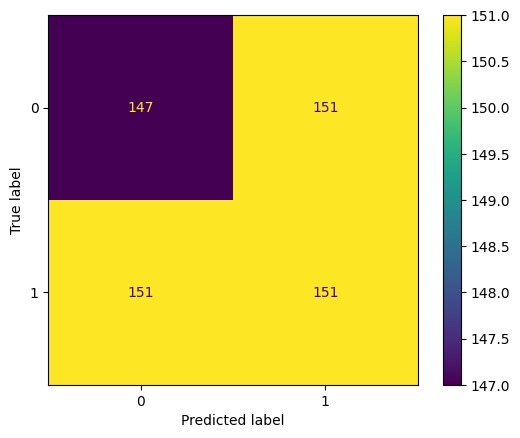
**K Nearest Neighbour:**



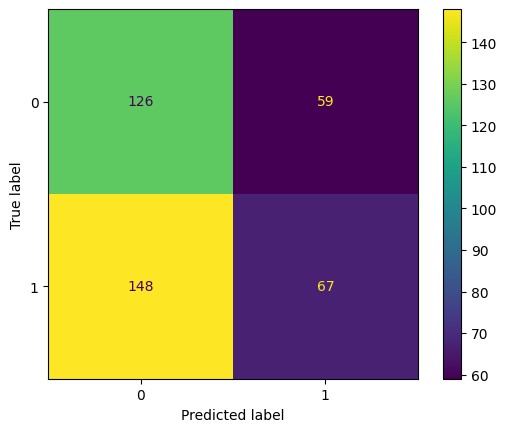
**Decision Tree:**



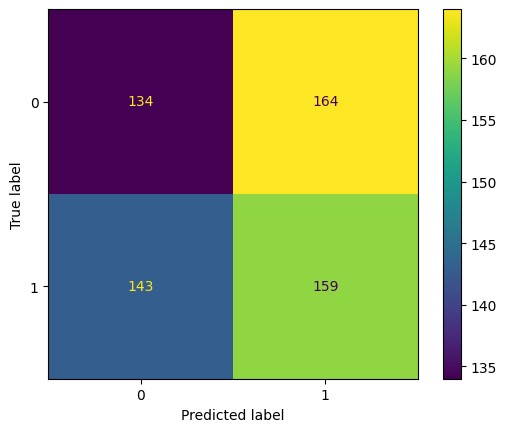
**XGB Classifier:**



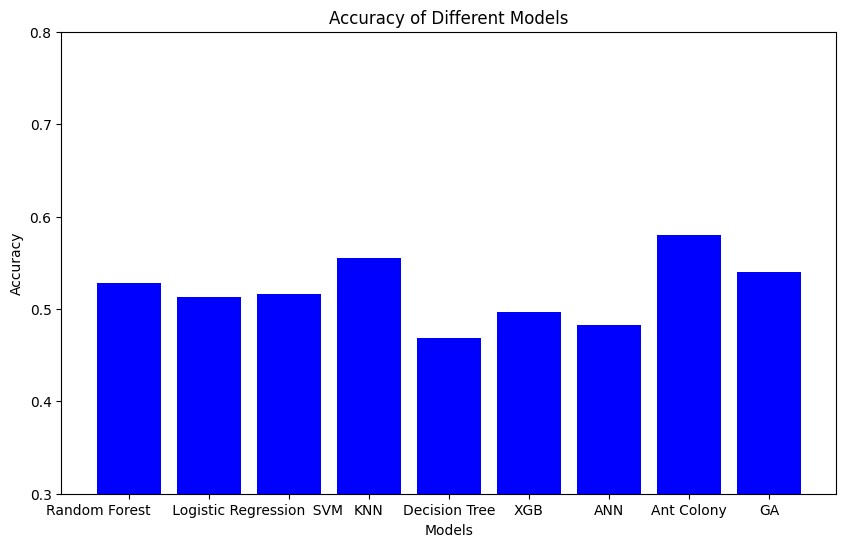
**ANN:**



**Ant Colony Optimization:**



**Accuracy of all the models :**



1. **CONCLUSION:**

Online payment fraud detection landscape is being shaped by rapid technological changes and cooperation. Technological innovation and concerted efforts are the driving forces behind the evolution of the online payment fraud detection landscape. The field of machine learning, behavioral analytics, and biometric authentication has witnessed various levels of progress hence it will be easier to detect any deception in transactions that are carried out. Additionally, block-chain technology integration along with real-time monitoring mechanisms will strengthen security measures while compliance with government regulations and educational programs will continue to play a major role in fraud prevention initiatives. By embracing these advancements and adopting a multi-layered approach to fraud detection, stakeholders can navigate the evolving threat landscape and safeguard the integrity of digital transactions, ensuring a secure and seamless online payment experience for consumers and businesses alike.

**14. FUTURE SCOPE:**

The future of online payment fraud detection is about to take leap forward thanks to high-tech solutions and partnership. Both machine learning and AI algorithms are going through constant improvement, which allows them to make analysis on large amounts of data in real time, detecting any fraudulent pattern that may occur. More so, it will be possible to regard behavioral analytics as an essential technique for assessing risk by analyzing user behavior and transaction patterns. In addition, security and authentication processes will be improved through the integration of block-chain technology as well as biometric authentication methods. What this means is that when they are combined together with collaborative fraud prevention networks, real-time monitoring and alerts can enable proactive identification and prevention of fraudulent transactions. At the same time, the best approach would be adaptive security measures that keep on adjusting dynamically to changing threats while also meeting regulatory standards and industry guidelines. Educating consumers and businesses on fraud schemes as well as the best practices will play a great role in reducing risks associated with such activities. Simply put, innovations in terms of techniques used in the online payment fraud detection process would stem from collaboration among different stakeholders who employ multiple strategies across various levels of defense for digital transaction protection.

1. **REFERENCES:**

Here are the references for the literature survey you provided in APA format:

[1] Lin, W., Sun, L., Zhong, Q., Liu, C., Feng, J., Ao, X., & Yang, H. (2021, August). Online credit payment fraud detection via structure-aware hierarchical recurrent neural network. In *IJCAI* (pp. 3670-3676).

[2] Cui, J., Yan, C., & Wang, C. (2021, January). Learning transaction cohesiveness for online payment fraud detection. In *The 2nd International Conference on Computing and Data Science* (pp. 1-5).

[3] Zhou, R., Zhang, Q., Zhang, P., Niu, L., & Lin, X. (2021). Anomaly detection in dynamic attributed networks. *Neural Computing and Applications*, *33*(6), 2125-2136.

[4] Xie, Y., Liu, G., Yan, C., Jiang, C., Zhou, M., & Li, M. (2022). Learning transactional behavioral representations for credit card fraud detection. *IEEE Transactions on Neural Networks and Learning Systems*.

[5] Zhou, X., Cheng, S., Zhu, M., Guo, C., Zhou, S., X, P., ... & Zhang, W. (2018). A state of the art survey of data mining-based fraud detection and credit scoring. In *MATEC Web of Conferences* (Vol. 189, p. 03002). EDP Sciences.

[6] Yang, M., Zhu, T., Liang, K., Zhou, W., & Deng, R. H. (2019). A blockchain-based location privacy-preserving crowdsensing system. *Future Generation Computer Systems*, *94*, 408-418.

[7] Cao, S., Yang, X., Chen, C., Zhou, J., Li, X., & Qi, Y. (2019). Titant: Online real-time transaction fraud detection in ant financial. *arXiv preprint arXiv:1906.07407*.

[8] Yang, C. H., & Lee, K. C. (2020). Developing a strategy map for forensic accounting with fraud risk management: An integrated balanced scorecard-based decision model. *Evaluation and program planning*, *80*, 101780.

[9] Wickramanayake, B., Geeganage, D. K., Ouyang, C., & Xu, Y. (2020). A survey of online card payment fraud detection using data mining-based methods. *arXiv preprint arXiv:2011.14024*.

[10] Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. *Information Sciences*, *557*, 302-316.