**RV UNIVERSITY**

**School of Computer Science and Engineering**

**Bengaluru – 560059**

****

**FUNDAMENTALS OF DATA SCIENCE**

**COURSE CODE: CS3236**

**V Semester B.Sc/B.Tech (HONS.)**

**Project Report**

|  |  |
| --- | --- |
| **Name** | **Nishita Bankapur** |
| **USN** | **1RVU22BSC063** |
| **Academic Year** | **2024 - 2025** |
| **Project Title** | **Credit Card Fraud Detection** |

**2. Abstract**

**Objective**: The project aims to detect fraudulent credit card transactions using multiple machine learning models. It addresses the problem of fraud detection in financial transactions, which is critical for minimizing losses for businesses and enhancing consumer security.

**Key Findings**: Initial results indicate that Random Forest classifier achieved high accuracy and precision, with Gradient Boosting yielding the best performance in terms of overall accuracy. Logistic Regression offers interpretable insights that can be valuable for understanding model decisions.

**Outcome**: The project highlights the potential for deploying these models in real-world scenarios, with Gradient Boosting showing promise as an optimal model for fraud detection.

## 3. Introduction

Credit card fraud is a prevalent issue in the financial industry, leading to substantial financial losses for both consumers and financial institutions. Detecting fraudulent transactions is crucial to maintaining the integrity of financial systems and protecting users from unauthorized activities. This report outlines the implementation of machine learning models for credit card transactions fraud detection using the provided dataset.

## 4. Dataset Overview

The dataset used for this analysis comprises credit card transaction data, including various features such as transaction amount, gender, and location information. The target variable is "is\_fraud," indicating whether a transaction is fraudulent (1) or not (0). The dataset is split into training and testing sets for model development and evaluation.

Dataset - <https://www.kaggle.com/datasets/kartik2112/fraud-detection>

## 4. Data Preprocessing

### 4.1. Data Cleaning

* Duplicates: Duplicate records were identified and removed to ensure data integrity.
* Null Values: Null values were handled by dropping rows with missing data.

### 4.2. Feature Selection

* Irrelevant Columns: Columns such as 'Unnamed: 0' and personally identifiable information were dropped as they do not contribute to fraud detection.

### 4.3. Data Transformation

* Categorical Encoding: The 'gender' column was encoded into numerical values (0 and 1) for model compatibility.
* Standard Scaling: Numerical features were standardized using StandardScaler to bring them to a common scale.

## 5. Exploratory Data Analysis

* Visualizations: Data distribution was visualized using count plots to understand the distribution of fraudulent and non-fraudulent transactions.

## 6. Model Implementation and Evaluation

### 6.1. Logistic Regression

* A logistic regression model was trained and evaluated.
* Confusion Matrix: The confusion matrix was used to assess the model's performance.

### 6.2. Random Forest

* A Random Forest classifier was employed for fraud detection.
* Confusion Matrix: The confusion matrix was analyzed to evaluate the model's effectiveness.

### 6.3. Gradient Boosting Classifier

* A Gradient Boosting Classifier was trained and assessed.
* Confusion Matrix: Model performance was analyzed using a confusion matrix.

## 7. Model Performance

* Accuracy Scores: The accuracy of each model was calculated.
* High Accuracy: All models achieved high accuracy, indicating their ability to correctly classify transactions.

## 8. Considerations and Recommendations

* **Imbalanced Data:** The report emphasizes the need to consider class imbalance and recommends exploring additional metrics such as precision, recall, and F1-score.
* **Hyperparameter Tuning:** Further improvement may be achieved by fine-tuning hyperparameters for each model.
* **Real-world Deployment:** Before deployment, models should be rigorously tested on new, unseen data to ensure generalizability.

## 9. Conclusion

This report outlines the implementation of machine learning models for credit card transactions fraud detection. While the models demonstrate high accuracy, additional metrics and considerations are essential for a comprehensive evaluation. Continuous monitoring, periodic model updates, and collaboration with domain experts are crucial for maintaining the effectiveness of fraud detection systems.