Credit Card Fraud Detection using Machine Learning

A Project Report (Project-II) submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

DECEMBER-2024

Supervised By

Dr. S.C. Gupta
(HoD, Dept. CSE)

Submitted By

Kunal Saini (2821172)
Sunidhi (2821122)
VIIth Sem

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

PANIPAT INSTITUTE OF ENGINEERING AND TECHNOLOGY
SAMALKHA, PANIPAT-132103

(Approved by AICTE and Affiliated to the Kurukshetra University, Kurukshetra)

DECLARATION

We certify that

i. The work presented in this project report is an authentic record of our own work under the guidance of our supervisor. It has not been submitted to any other Institute for the award of any other degree or diploma.

ii. Whenever we have used information (text, data, figure, photograph, chart, analysis, inference, etc.) from other sources, we have given due credit by citing it in the text of the report and providing its details in the references.

iii. We have followed the guidelines provided by the department for preparing the report.

_____

Name(s) of the student(s), Roll Number(s):

Kunal Saini (2821172)

Sunidhi (2821122)

Project report title: Credit Card Fraud Detection using Machine Learning

Semester: 7th Sem

Date:

APPROVAL FROM SUPERVISOR

This is to certify that the project entitles “Credit Card Fraud Detection using Machine Learning” presented by “Kunal Saini (2821172), Sunidhi (2821122)” under my supervision is an authentic work. To the best of my knowledge, the content of this report has not been submitted for the award of any previous degree to anyone else.

It is recommended that the report be accepted as fulfilling this part of the requirements for the award of the degree.

_____

Ms. Saba
Assistant Professor
Department of Computer Science & Engineering
Date:

_____

(Counter Signed by)
Dr. S.C. Gupta
(Prof. & Head Department of CSE)

CERTIFICATE

This is to certify that the work embodied in this report, entitled *"Credit Card Fraud Detection using Machine Learning"* carried out by *"Kunal Saini (2821172), Sunidhi (2821122)"* is approved for the degree of "*Bachelor of Technology (B.Tech.) in CSE*" at the department of "Computer Science & Engineering", Panipat Institute of Engineering and Technology, Samalkha.

_____

Internal Examiner

_____

External Examiner

Date:

Place: Panipat

ACKNOWLEDGEMENT

We would like to express our deepest gratitude to Supervisor ***Ms. Saba**, Assistant Professor, Department of CSE, Panipat Institute of Engineering and Technology*. Her never-ending patience, intellectual direction, persistent encouragement, constant and vigorous supervision, constructive criticism, helpful suggestions, and reading numerous poor versions and revising them at all stages allowed this project to be completed.

We would like to extend our sincere thanks to ***Dr. S.C. Gupta**, Head Department of Computer Science & Engineering, Panipat Institute of Engineering and Technology, Samalkha* for his full support and assistance during the development of the project.

We would also like to express our gratitude to everyone who has assisted us in making this project a success, whether directly or indirectly. In this unusual scenario, we would like to express our gratitude to the different staff members, both teaching and non-teaching, who have provided us with valuable assistance and assisted our project. Finally, we must express our gratitude for our parents' unwavering support and patience.

Kunal Sani (2821172)
Sunidhi (2821122)
Date:

ABSTRACT

Credit card fraud is a pressing issue in today's digital economy, where electronic payments and online transactions are the norm. Credit card fraud is a significant problem, with billions of dollars lost each year. Machine learning can be used to detect credit card fraud by identifying patterns that are indicative of fraudulent transactions. Credit card fraud refers to the physical loss of a credit card or the loss of sensitive credit card information. Many machine-learning algorithms can be used for detection. This project proposes to develop a machine-learning model to detect credit card fraud. The system identifies anomalies in credit card transactions by analyzing historical data to distinguish between legitimate and fraudulent activities. Machine learning algorithms are trained to recognize patterns indicative of fraud through processes like data preprocessing, feature selection, and model optimization.

To address the challenge of class imbalance often seen in fraud detection datasets, techniques like Oversampling, Undersampling, and Synthetic Minority Oversampling (SMOTE) are implemented. These methods ensure the model's ability to accurately detect fraudulent transactions despite the inherent skew in the data.

This project demonstrates the potential of machine learning in fraud detection, emphasizing the importance of data-driven methods in enhancing the security and trustworthiness of financial transactions. The outcomes provide a robust framework for future research and improvements in credit card fraud prevention systems.

Keywords: Credit Card Fraud Detection, Fraud Detection, Fraudulent Transactions, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, Decision Tree.

LIST OF FIGURES

FIGURES	PAGE NO.
FIG.1 Credit Card fraud detection image	1
FIG.2 System framework	3
FIG.3 Data-Driven decision making	5
Fig.4 System architecture	22
FIG.5 Use case Diagram	22
FIG.6 Data Flow Diagram	23
FIG.7 Logistic Regression function	24
FIG.8 XG Boost diagram	25
FIG.9 XG Boost Model Building	26
FIG.10 Graph plotting of XG Boost	27
FIG.11 Plotting of the mean test and train score	28
FIG. 12 Printing the Accuracy, Sensitivity, Specificity, F1-Score	28
FIG. 13 Model building of Decision Tree	29
FIG. 14 Creating a Decision Tree Classifier	30
FIG.15 Model building of Decision Tree after balancing data	31
FIG.16 Roc curve plotting	32
FIG.17 ROC curve of decision tree	32
FIG.18 Fraud and Not-Fraud Transactions	36
FIG.19 Density Time graph	37
FIG.20 Train case of XG Boost	37
FIG.21 Test case of XG Boost	37
FIG.22 Density Time graph	38
FIG.23 Result of LR	39
FIG.24 Result of Decision Tree	40
FIG.25 XG Boost Result	40
FIG.26 Result of LR after balancing	41
FIG.27 Result of Decision Tree after balancing	41

FIG.28 Result of XG Boost after balancing	42
FIG.29 Graph of Fraud vs Not fraud	43
FIG.30 ROC curve of XG Boost classifier	43
FIG.31 ROC Curve for XG Boost balanced dataset	44

LIST OF ABBREVIATIONS

S. No.	ABBREVIATIONS	Full Form
1	RF	Random Forest
2	DT	Decision Tree
3	KNN	K-Nearest Neighbors
4	ANN	Artificial Neural Network
5	SVM	Support Vector Machine
6	HMM	Hidden Markov Model
7	IC3	Internet Crime Complaint Centre
8	PCA	Principal Component Analysis
9	GA	Genetic Algorithm
10	KCGAN	K-Conditional Generative Adversarial Network
11	GAN	Generative Adversarial Network
12	SMOTE	Synthetic Minority Over-sampling Technique
13	XG Boost	eXtreme Gradient Boosting

TABLE OF CONTENTS

Declaration i

Approval from Supervisor ii

Certificate iii

Acknowledgement iv

Abstract v

List of Figures vi

List of Abbreviations viii

Chapter 1: Introduction

1.1 Introduction 1

1.2 Problem Statement 3

1.3 Objectives 5

1.4 Significance and Motivation of the Project Work 7

1.5 Organization of Project Report 9

Chapter 2: Literature Review

2.1 Overview of Relevant Literature 11

2.11 Credit Card Fraud Detection using ML Algorithms 12

2.12 Comparative Analysis of Techniques 14

2.13 Anomaly Detection in Credit Card Transactions 15

2.14 Key Gaps in the Literature 17

Chapter 3: System Development

3.1 Requirements and Analysis 20

3.11 Functional Requirements 21

3.12 Non-Functional Requirements 23

3.2 Project Design and Architecture 25

3.21 Data Collection 26

3.22 Data Preprocessing and Feature Designing 27

3.23 Model Development 28

3.24 Model Evaluation and Deployment 29

3.3 Key Challenges 32

Chapter 4: Testing

4.1 Testing Strategy 34

4.2 Test Cases and Outcomes 36

Chapter 5: Results and Evaluation

5.1 Results 38

5.11 Results for Unbalanced Dataset 39

5.12 Results for Balanced Dataset 41

5.2 Evaluation 43

Chapter 6: Conclusion and Future Scope

6.1 Conclusion 45

6.2 Future Scope 45

References

References 49