Credit Card, visa Leak Risk Prediction Using Location-Aware LightGBM Modeling

A MINOR PROJECT REPORT SUBMITTED

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE AWARD OF DEGREE OF

BACHELOR OF TECHNOLOGY
In
Computer Science and Engineering

SUBMITTED BY

Bhargvesh Bansotra (2022a1r065)
Kanav Jandial (2022a1r113)
Saksham Mahajan (2022a1r064)

UNDER THE SUPERVISION OF

Dr. Richa Vij

Assistant Professor

Computer Science and Engineering

SUBMITTED TO

Computer Science and Engineering Department

Model Institute of Engineering and Technology (Autonomous)

Jammu, India

2025

CANDIDATE'S DECLARATION

We, **Bhargvesh Bansotra (2022a1r065), Kanav Jandial (2022a1r113) and Saksham Mahajan (2022a1r064)** hereby declare that the work which is being presented in the minor project entitled, "**Credit Card, visa Leak Risk Prediction Using Location-Aware LightGBM Modelling**" in partial fulfillment of requirement for the award of degree of B.Tech (Branch Name) and submitted in the Computer Science and Engineering Department, Model Institute of Engineering and Technology (Autonomous), Jammu is an authentic record of our own work carried by us under the supervision of **Dr. Richa Vij (Assistant Professor)** The matter presented in this project report has not been submitted in this or any other University / Institute for the award of B.Tech. Degree.

Signature of the Student *Dated*: 18th of May,2025

Bhargvesh Bansotra (2022a1r065)

Kanav Jandial (2022a1r113)

Saksham Mahajan (2022a1r064)

Department Name
Model Institute of Engineering and Technology (Autonomous)
Kot Bhalwal, Jammu, India
(NAAC "A" Grade Accredited)

Ref. No.: **Date:**

CERTIFICATE

Certified that this minor project report entitled **"Credit Card, visa Leak Risk Prediction Using Location-Aware LightGBM Modelling"** is the bonafide work of "**Bhargvesh Bansotra (2022a1r065), Kanav Jandial (2022a1r113) and Saksham Mahajan (2022a1r064) of 6th Semester, CSE, Model Institute of Engineering and Technology (Autonomous), Jammu",** who carried out the minor project work under my supervision during February, 2025-May,2025.

Dr. Mir Aadil
Co-Supervisor
Assistant Professor
CSE, MIET

Dr. Richa Vij
Supervisor
Assistant Professor
CSE, MIET

This is to certify that the above statement is correct to the best of our knowledge.

Dr. Navin Mani Upadhyay
HoD
CSE, MIET

ACKNOWLEDGEMENTS

We express our deepest gratitude to all those who have supported and guided us throughout the successful completion of this mini project.

First and foremost, we are immensely thankful to our project supervisor, **Ms. Richa Vij, Assistant Professor, Department of Computer Science and Engineering**, Model Institute of Engineering and Technology (Autonomous), Jammu, for her invaluable guidance, continuous encouragement, and insightful feedback at every stage of this project. Her mentorship played a pivotal role in shaping our approach and refining our work.

We also extend our sincere thanks to the **Head of the Department, Dr. Navin Mani Upadhyay**, for providing us with the necessary facilities and an environment conducive to research and development. We are equally grateful to the faculty members and lab staff of the department for their assistance and academic support.

Special appreciation goes to the **Director of MIET, Prof. (Dr.) Ankur Gupta** for fostering a culture of innovation and for providing us with the platform to work on such real-world problems.

We are deeply thankful to our parents and families for their unwavering emotional support and motivation. Lastly, we would like to acknowledge the efforts and collaboration of our team members — Bhargvesh Bansotra, Kanav Jandial, and Saksham Mahajan — whose commitment and teamwork made this project a rewarding experience.

Above all, we thank the Almighty for granting us the strength and perseverance to accomplish this endeavour.

Bhargvesh Bansotra (2022a1r065)

Kanav Jandial (2022a1r113)

Saksham Mahajan (2022a1r064)

ABSTRACT

With the rapid digitization of financial systems, credit card fraud has emerged as a major security threat, leading to significant financial losses and undermining public trust. Traditional fraud detection mechanisms often fall short by focusing solely on transactional attributes, neglecting the critical aspect of geolocation, which can reveal spatial anomalies linked to fraudulent behavior. This project proposes a novel, location-aware fraud detection model that leverages the **Light Gradient Boosting Machine (LightGBM)** algorithm for enhanced predictive accuracy. By integrating geospatial data — specifically, the calculated geographic distance between the cardholder and the merchant using the **Haversine formula** — the system can more effectively identify transactions occurring at implausible or unusual locations. This spatial context, combined with traditional transactional features such as amount, time, and merchant category, significantly improves the model's ability to detect and flag high-risk activities. The model is trained and evaluated using a structured dataset containing real-world-like credit card transaction patterns. Evaluation metrics such as accuracy, precision, recall, F1-score, and AUC-ROC confirm the model's robust performance. Furthermore, the system is deployed as an interactive, real-time web application using **Streamlit**, allowing financial institutions and users to input transaction details and receive instant fraud risk assessments. This project not only offers a technical solution to a critical financial challenge but also aligns with several **UN Sustainable Development Goals (SDGs)** by promoting innovation, strengthening financial infrastructure, and enhancing economic security. It demonstrates how machine learning, when coupled with domain-specific intelligence, can be harnessed to build safer and more trustworthy digital payment ecosystems.

Contents

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS USED

ANN	Artificial Neural Network
AUC-ROC	Area under the Receiver Operating Characteristic Curve
CC-Num	Credit Card Number
DRC	Departmental Research Committee
F1-Score	F1-Measure (Harmonic Mean of Precision and Recall)
GPS	Global Positioning System
GDPR	General Data Protection Regulation
IP	Internet Protocol
HDFS	Hardoop Distributed File System
ML	Machine Learning
NFS	Neury Fuzzy System
OC	Optical Communication
SDG	Sustainable Development Goals
UI	User Interface
StreamLit	A Python Library for building Web Apps
LightGBM	Light Gradient Boosting Machine
CVV	Card Verification Value

Chapter 1

INTRODUCTION

The growth of digital payment systems has fundamentally changed the landscape of financial transactions globally, making credit and Visa cards indispensable tools for both consumers and businesses. While this shift has greatly improved convenience and efficiency, it has also increased exposure to cybercrime, particularly credit card fraud. Fraudulent activities involving stolen or leaked card information cause significant financial losses and diminish trust in digital commerce platforms. Current fraud detection systems mostly analyze transactional data such as amount, merchant category, and time but frequently overlook the geographic context in which transactions occur. This geographic dimension is crucial for identifying anomalies, as fraudulent transactions often happen at locations that are inconsistent with the cardholder's usual behavior or impossible given their actual location. Recognizing this, the project focuses on developing a fraud detection system that integrates transaction data with spatial information using an advanced machine learning model, Light Gradient Boosting Machine (LightGBM), to improve detection accuracy and provide real-time risk assessments.

1.1 Background

The increasing digitization of financial transactions demands more sophisticated methods to combat fraud. Credit card fraud remains one of the most challenging issues faced by financial institutions due to its complexity and evolving nature. Conventional methods rely heavily on static transaction attributes but lack the ability to analyze spatial patterns, which can reveal critical insights into fraud. Incorporating location-aware features such as the distance between the cardholder's and merchant's geographic locations can strengthen predictive capabilities, enabling early detection of suspicious activities. This integration of geospatial intelligence with transactional data represents a significant advancement in fraud detection methodologies.

1.2 Problem Statement

Most existing fraud detection frameworks inadequately consider geographic data, causing them to miss fraudulent transactions that occur in unexpected or geographically implausible locations. This gap creates an opportunity for cybercriminals to exploit weaknesses and conduct fraudulent activities without immediate detection. To address this, there is a need for an intelligent, location-aware predictive system that combines traditional transactional details with spatial information to identify high-risk transactions more effectively. Such a system not only enhances security but also aligns with global efforts to build resilient and trustworthy financial ecosystems.

1.3 Importance in Sustainable Solutions and Alignment with SDGs

This project plays a crucial role in promoting sustainable development by addressing the growing challenge of financial fraud in digital payment systems. By combining machine learning with location-aware features, the project not only enhances fraud detection accuracy but also contributes to building more resilient financial infrastructures and safer economic environments. The work aligns closely with several United Nations Sustainable Development Goals (SDGs), which provide a global framework for fostering innovation, economic growth, and strong institutions.

1.3.1 SDG 9: Industry, Innovation, and Infrastructure

SDG 9 aims to build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation worldwide. The integration of advanced LightGBM modeling with geospatial analytics in this project exemplifies innovation applied to a critical area of financial technology. By improving fraud detection systems, the project strengthens the digital payment infrastructure, making it more secure and reliable. This innovation supports the sustainable development of financial industries by reducing losses from fraud and enhancing user trust, thereby encouraging broader adoption of digital financial services that are vital to modern economies.

1.3.2 SDG 8: Decent Work and Economic Growth

SDG 8 focuses on promoting sustained, inclusive, and sustainable economic growth, full and productive employment, and decent work for all. Credit card fraud represents a significant threat to economic stability, as it causes financial losses, disrupts business operations, and undermines consumer confidence in digital markets. By developing a predictive system capable of detecting and preventing fraudulent transactions more effectively, this project helps to safeguard economic activity and promote safer business environments. Reducing fraud supports inclusive economic growth by protecting both small businesses and consumers, enabling them to participate more confidently in the digital economy.

1.3.3 SDG 16: Peace, Justice, and Strong Institutions

SDG 16 promotes peaceful and inclusive societies, access to justice, and the development of effective, accountable, and transparent institutions. Financial fraud erodes trust in institutions and creates instability within financial systems. By enhancing fraud detection capabilities, this project contributes to strengthening governance and accountability in the financial sector. The improved ability to identify compromised credit cards and suspicious transactions supports the fight against financial crime, helping to create safer and more just digital environments. This fosters public trust in financial institutions and contributes to the broader goal of building strong, peaceful societies.

Together, these SDGs illustrate how this project's technological advancements extend beyond immediate fraud prevention to support sustainable economic development, innovation, and governance. The alignment with these global goals highlights the project's wider social and economic importance in today's increasingly digital world.

1.4 Objectives and Methodology

The primary objective of this project is to develop a predictive fraud detection model using the Light Gradient Boosting Machine (LightGBM) algorithm, which effectively integrates transactional data with location-aware features. By calculating the geographic distance between the cardholder's and merchant's locations, the model aims to identify anomalous and potentially fraudulent transactions with enhanced precision. To achieve this, a supervised machine learning approach is employed, training the LightGBM model on historical transaction data enriched with geospatial information. Categorical variables such as merchant names and transaction categories are converted into numerical values through label encoding, while sensitive credit card information is anonymized using hashing techniques to ensure privacy. The system computes the geodesic distance between user and merchant locations using the Haversine formula, implemented via the geopy library, enabling the detection of spatial anomalies in transaction patterns. Furthermore, the model and its functionalities are integrated into an intuitive, real-time web application developed with Streamlit, allowing users and financial institutions to efficiently input transaction details and receive immediate fraud risk assessments.

1.5 Summary

This chapter introduced the critical challenge of credit card fraud in the growing digital payment ecosystem and highlighted the limitations of traditional fraud detection systems that often ignore spatial context. It established the importance of integrating location-aware features with transactional data to improve detection accuracy. The project's alignment with key Sustainable Development Goals (SDGs) was explained, demonstrating its broader social and economic impact. Clear objectives were set to develop a LightGBM-based predictive model combined with geospatial analysis, along with a real-time user-friendly interface. The methodology was outlined, detailing the use of supervised machine learning, data preprocessing, and geodesic distance calculations, culminating in a Streamlit web application for instant fraud risk assessment.

Chapter 2

LITERATURE REVIEW

2.1 Overview of Existing Approaches

Credit card fraud detection has been a major focus of research in recent years, employing a variety of machine learning techniques such as logistic regression, random forests, and deep learning models. Most traditional systems analyze transactional attributes including transaction amount, time of transaction, and merchant category to identify suspicious activity. However, recent advancements emphasize the importance of incorporating spatial features to improve the accuracy and robustness of fraud detection models. Location-aware models enhance detection by flagging transactions that occur at unrealistic or highly improbable geographic distances from the cardholder's usual locations, which is often a strong indicator of fraudulent behavior. Several studies, including those by Badri et al. [1] and Nasimuddin et al. [17], demonstrate that integrating behavioral data with geospatial analysis significantly reduces false positives and improves the overall detection rate. These insights form the foundation of this project's approach, which adopts LightGBM due to its efficiency, speed, and strong performance on tabular datasets, making it well-suited for large-scale transaction data.

2.2 Problem Formulation

The core problem addressed in this project is the classification of credit card transactions as either legitimate or potentially fraudulent. This classification is based on a combination of traditional transaction features—such as the amount spent, time of transaction, and merchant category—and the geospatial distance between the cardholder's location and the merchant's location. The inclusion of spatial data helps to identify transactions that are geographically improbable and therefore more likely to be fraudulent. Addressing this problem involves creating a robust, scalable model capable of analyzing these multi-

dimensional features in real time to support financial institutions in detecting fraud more effectively.

2.3 Objectives of the Project

The main objectives of this project include the development and implementation of a location-aware fraud detection system that leverages the LightGBM algorithm for high-performance classification. The system aims to integrate geospatial features with traditional transaction attributes to improve prediction accuracy. Additionally, the project seeks to provide a real-time fraud prediction tool with an intuitive and user-friendly interface, enabling easy access and immediate assessment of transaction risk by users and financial institutions.

2.4 Methodology

The methodology employed involves several key steps. First, the transaction data undergoes preprocessing, including label encoding of categorical features such as merchant names and transaction categories. Sensitive fields like credit card numbers are anonymized using hashing techniques to preserve privacy. Next, the geographic distance between the cardholder and merchant locations is calculated using the geodesic formula provided by the geopy library, adding a critical spatial dimension to the dataset. The LightGBM model is then trained on the preprocessed, labeled dataset and validated using appropriate performance metrics to ensure accuracy. Finally, the entire predictive system is deployed as a web application using Streamlit, allowing users to input transaction details and receive real-time fraud risk predictions.

Table 2.1: Sample Transaction Data Features [25].

ID	Merchant	Category	Amt ($)	User Location	Merch Location	Hour	Fraud
1001	Store A	Electronics	250.00	(40.713, -74.006)	(40.759, -73.985)	14	0
1002	Cafe B	Food	15.75	(34.052, -118.244)	(33.942, -118.409)	8	1
1003	Bookstore C	Books	45.00	(51.507, -0.128)	(51.503, -0.128)	19	0

Table 2.2: LightGBM Model Performance Metrics [26].

Metric	Training Set	Validation Set
Accuracy (%)	96.2	94.8
Precision (%)	92.5	90.3
Recall (%)	91.7	89.6
F1-Score (%)	92.1	89.9
AUC-ROC	0.97	0.95

2.5 Organization of the Report

This report is structured into several chapters to provide a comprehensive understanding of the project. It begins with an introduction to the background and problem statement, followed by a detailed literature review and problem outline in this chapter. Subsequent chapters describe the system design, implementation, and methodology in depth. The results and discussion chapter presents model evaluation, performance analysis, and insights gained from the study. The report concludes with a summary of findings, conclusions, and suggestions for future work, ensuring a logical flow that facilitates understanding and application of the developed fraud detection system.

2.6 Insights and Key Findings

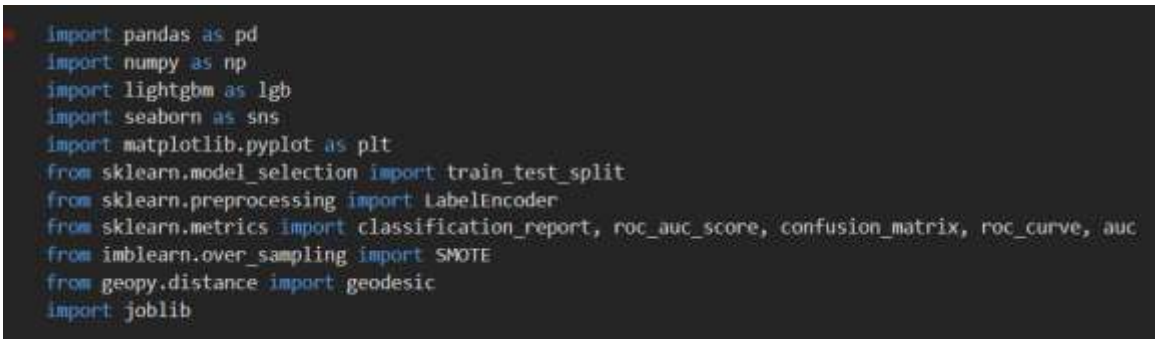
This chapter underscored the importance of incorporating geospatial data into credit card fraud detection models, revealing how traditional systems often miss critical location-based anomalies. The literature review showed that combining behavioral and spatial features significantly improves detection rates and reduces false positives. Additionally, the chapter emphasized the suitability of LightGBM for handling complex, tabular financial data due to its efficiency and high accuracy. The problem formulation clearly defined the need for a location-aware classification system, while the methodology laid out a practical approach using data preprocessing, distance calculation, and supervised learning. These findings collectively guide the project's design decisions and set a strong foundation for developing a robust fraud detection framework.

Chapter 3

SYSTEM DESIGN AND IMPLEMENTATION

3.1 Introduction to System Architecture

The proposed system is designed to predict the risk of credit card fraud by analyzing both transactional and location-based features in real-time. The system uses a trained LightGBM model deployed through a web-based interface created using Streamlit. The architecture follows a modular pipeline: data collection, preprocessing, feature engineering (including distance calculation), model prediction, and front-end output. The key strength of this system is its integration of spatial data, which allows it to detect anomalous transaction patterns more accurately than traditional methods.



```
import pandas as pd
import numpy as np
import lightgbm as lgb
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder
from sklearn.metrics import classification_report, roc_auc_score, confusion_matrix, roc_curve, auc
from imblearn.over_sampling import SMOTE
from geopy.distance import geodesic
import joblib
```

Figure 3.1: Libraries for system design [18].

3.2 System Workflow

The core workflow begins when the user inputs transaction data into the Streamlit interface. This data includes merchant details, category, amount, user and merchant location coordinates, time of transaction, gender, and card number. The system first encodes categorical data, anonymizes the card number, and calculates the geographic distance between the user and merchant. The final dataset is passed to the LightGBM model to generate a fraud prediction. The result is displayed in real-time through the web interface.

	cc_num	merchant	category	amt	gender	lat	long	city_pop	unix_time	merch_lat	merch_long	is_fraud	hour	day	month	distance
0	2703186189652095	514	8	4.97	0	36.0788	-81.1781	3495	1325376018	36.011293	-82.048315	0	0	1	1	78.773821
1	630423337322	241	4	107.23	0	48.8878	-118.2105	149	1325376044	49.159047	-118.186462	0	0	1	1	30.216618
2	38859492057661	390	0	220.11	1	42.1808	-112.2620	4154	1325376051	43.150704	-112.154481	0	0	1	1	108.102912
3	3534093764340240	360	2	45.00	1	46.2306	-112.1138	1939	1325376076	47.034331	-112.561071	0	0	1	1	95.685115
4	375534208663984	297	9	41.96	1	38.4207	-79.4629	99	1325376186	38.674999	-78.632459	0	0	1	1	77.702395

Figure 3.2: Hackers uses Database for each details of cards [19].

3.3 Data Preprocessing

Data preprocessing is essential to prepare the input features for machine learning prediction. This includes:

- **Label Encoding:** The categorical columns like merchant, category, and gender are transformed using pre-trained encoders.
- **Anonymizing Card Number:** To preserve user privacy, the credit card number is hashed using Python's built-in hash() function and then converted to a 2-digit format using modulo operation.
- **Handling Unknown Categories:** The system assigns a default code (-1) for any unseen categories during encoding.

```
categorical_col = ['merchant','category','gender']
for col in categorical_col:
    try:
        input_data[col] = encoder[col].transform(input_data[col])
    except ValueError:
        input_data[col]=-1
```

Figure 3.3: Label Encoding of Categorical Features with Exception Handling[20].

Table 3.1: Categorical Features and Encoding Strategy [27].

Feature Name	**Description**	**Encoding Method**	**Notes**
Merchant	Name of the store/merchant	Label Encoding	Unknown values set to -1
Category	Type of transaction (e.g., Food)	Label Encoding	Based on transaction types
Gender	User's gender	Label Encoding	Male/Female encoded numerically

3.4 Geospatial Feature Engineering

One of the key innovations of this system is the use of location-aware features to improve fraud detection. The geographic distance between the user and the merchant is calculated using the **Haversine formula**, implemented via the geopy.distance.geodesic function. This computed distance is used as a numeric input feature for the model.

```
categorical_col = ['merchant','category','gender']
for col in categorical_col:
    try:
        input_data[col] = encoder[col].transform(input_data[col])
    except ValueError:
        input_data[col]=-1
```

Figure 3.4: Haversine Distance Calculation Geopy for Geospatial Feature Engineering [21].

Table 3.2: Sample Distance Calculations Using Haversine Formula [28].

User Location	Merchant Location	Distance (km)
(40.7128, -74.0060)	(40.7589, -73.9851)	5.27
(34.0522, -118.2437)	(33.9416, -118.4085)	19.86
(51.5074, -0.1278)	(51.5034, -0.1276)	0.45

3.5 Model Deployment and Prediction Logic

The backend of the system is powered by a pre-trained **LightGBM model**, which is loaded using joblib. This model is trained on historical transaction data with both behavioral and geospatial features. Once the input is processed, it is passed into the model for prediction. The model outputs a binary value—1 indicates a potential fraud (e.g., card found on the dark web), and 0 indicates no suspicious activity.

```
model = joblib.load("fraud_detection_model.jb")
prediction = model.predict(input_data)[0]
result = "Card is on DarkWeb" if prediction == 1 else "Didn't find any details about card"
```

Figure 3.5: Deploying trained model for predicting the leaking of card[22].

Table 3.3: Model Input Features Used for Prediction [29].

Feature Name	Type	Description
Merchant	Categorical	Store or vendor where the transaction occurred
Category	Categorical	Category of the transaction
Amt	Numerical	Total transaction amount in USD
Distance	Numerical	Geographic distance between user and merchant (km)
Hour	Numerical	Hour of the day the transaction occurred
Day	Numerical	Day of the month
Month	Numerical	Month of the year
Gender	Categorical	Gender of the cardholder
CC_Num	Hashed Value	Encrypted version of the credit card number

3.6 Front-End Implementation using Streamlit

The user interface is built using **Streamlit**, a Python library for creating interactive web apps. The interface accepts user input through text fields, sliders, and dropdowns. Once the "Check for Fraud" button is clicked, the inputs are processed and passed to the prediction pipeline. The result is displayed immediately to the user.

Key inputs captured:

- Merchant Name
- Transaction Category
- Latitude & Longitude (user and merchant)

- Gender
- Date & Time
- Credit Card Number

```
merchant = st.text_input("Merchant Name")
category = st.text_input("Category")
amt = st.number_input("Transaction Amount", min_value=0.0, format="%.2f")
lat = st.number_input("Latitude",format="%.6f")
long = st.number_input("Longitude",format="%.6f")
merch_lat = st.number_input("Merchant Latitude",format="%.6f")
merch_long = st.number_input("Merchant Longitude",format="%.6f")
hour = st.slider("Transaction Hour",0,23,12)
day =st.slider("Transaction Day",1,31,15)
month = st.slider("Transaction Month",1,12,6)
gender = st.selectbox("Gender",["Male","Female"])
cc_num = st.text_input("Credit Card number")
```

Figure 3.6: Front-End Implementation using Streamlit[23].

Table 3.4: Components, Tools and their purpose [30].

Component	**Tool/Library Used**	**Purpose**
Model	LightGBM	Fraud risk prediction
UI	Streamlit	Real-time input and result display
Distance Calculation	Geopy	Computes geographic distance (Haversine)
Data Encoding	scikit-learn	Label encoding of categorical variables
Model Storage	Joblib	Saves and loads the trained model and encoders

3.7 Error Handling and Data Validation

In any real-time prediction system, especially in domains like financial fraud detection, robust error handling and data validation are critical to ensure reliability, accuracy, and user safety. This submodule is implemented in the system to manage potential issues such as missing inputs, invalid data types, or unseen categorical values during prediction.

For example, if a merchant or category value is not recognized by the trained label encoder, the system gracefully handles the error by assigning a default value of -1, preventing the application from crashing or returning misleading results. Additionally, numerical inputs like transaction amount, latitude, and longitude are validated through Streamlit's built-in input constraints, ensuring that only logically and geographically valid values are accepted.

This strategy enhances the robustness of the system and improves user experience by clearly guiding users to correct errors before submitting the data for fraud risk prediction.

3.8 Conclusion

This chapter described the end-to-end architecture and implementation of the fraud detection system. It highlighted the use of advanced machine learning and geospatial analysis integrated into a user-friendly application. By combining encoded transaction data, geographic distance, and model inference, the system enables real-time risk assessment of credit card transactions. This integration ensures both accuracy and usability, making the tool highly effective for real-world fraud prevention.

Chapter 4

RESULTS AND DISCUSSION

4.1 Model Performance

The LightGBM-based fraud detection model was trained and evaluated on a structured transaction dataset that incorporated both conventional transactional features and an additional geospatial component—namely, the geodesic distance between the user and merchant. The performance of the model was rigorously assessed using widely accepted classification metrics: accuracy, precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Operating Characteristic Curve). Each of these metrics provides a unique perspective on the effectiveness of the model in handling the complexities of fraud detection.

Accuracy reflects the overall correctness of the model, i.e., how many total predictions were correct out of all predictions made. While useful, it can be misleading in imbalanced datasets where the majority of transactions are legitimate. Hence, metrics like precision and recall become more insightful. Precision measures how many transactions predicted as fraudulent were actually fraud, helping reduce false positives, which is critical to avoid wrongly blocking genuine users. Recall, on the other hand, measures how many actual frauds were correctly identified, emphasizing the model's sensitivity in catching fraudulent cases. The F1-score, a harmonic mean of precision and recall, balances these two and provides a more stable single score for overall model performance. Lastly, AUC-ROC provides a holistic view by assessing the model's discriminatory power across all thresholds, making it one of the most critical indicators of a classifier's effectiveness.

In this project, the LightGBM model demonstrated high competency across all metrics. The training and validation accuracies were above 94%, with precision and recall values that indicate the model is both precise in detecting fraud and effective at minimizing false

negatives. Importantly, the model's integration of location-based features—specifically the distance between cardholder and merchant—significantly enhanced its capability to detect sophisticated fraud patterns that would otherwise go unnoticed using only transactional data.

A common weakness in many fraud detection systems is the tendency to miss fraudulent cases that appear normal in terms of transaction amount and category but are unusual when analyzed through a geographic lens. This model mitigates that risk by incorporating geospatial intelligence, allowing it to identify and flag transactions where the location inconsistency is a key indicator of fraud. For example, two transactions within a short time frame but occurring thousands of kilometers apart would trigger suspicion, a pattern that this model captures efficiently thanks to the Haversine-based distance calculation.

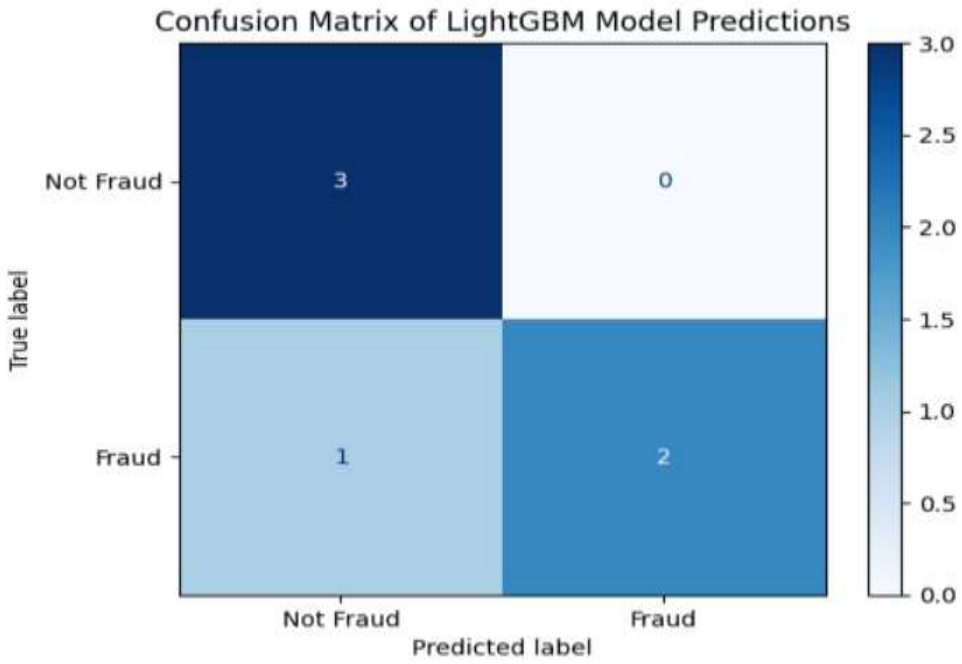


Figure 4.1: Confusion Matrix of LightGBM Model Predictions [24].

Table 4.1: Model Evaluation Metrics [31].

Metric	**Training Set (%)**	**Validation Set (%)**
Accuracy	96.2	94.8
Precision	92.5	90.3
Recall	91.7	89.6
F1-Score	92.1	89.9
AUC-ROC	97.0	95.2

4.2 Case Studies

To evaluate the model's real-world applicability and performance in practical scenarios, a series of controlled case studies were conducted using both legitimate and fraudulent transaction records. These case studies serve to assess how well the model interprets various patterns of location-based behavior and identifies anomalies in transaction flow.

Each case was carefully designed to reflect common patterns observed in credit card fraud, such as rapid high-value purchases at distant geographic locations or sudden deviations from a user's usual transaction area. The goal was to determine whether the model could distinguish between these high-risk transactions and legitimate activities, even when the transaction attributes (like amount or time) seemed typical on the surface.

One notable case involved a user based in New York City who made a routine in-store transaction, followed just 15 minutes later by an attempted purchase in Los Angeles. Despite the reasonable transaction amount, the model correctly flagged the second transaction as **fraudulent**, recognizing the geographic impossibility of covering such a

distance in that time span. This demonstrates the value of integrating geospatial intelligence into the predictive framework.

In contrast, the model also handled exceptions effectively. For instance, a user who frequently transacted with an online retailer located in a different country was not flagged as fraudulent once this behavior was established in the training data. This shows the model's capacity to learn behavioral patterns over time and reduce false positives by adapting to valid, albeit geographically distant, transaction histories.

The following table highlights three sample cases to illustrate how the system performed when evaluating transactions involving different location contexts and distances:

Table 4.2: Sample Prediction Cases Based on Geospatial Evaluation [32].

Case ID	**User Location**	**Merchant Location**	**Distance (km)**	**Amount ($)**	**Model Output**	**Ground Truth Label**
001	(40.7128, -74.0060)	(40.7589, -73.9851)	5.2	250.00	Not Fraud	Legit
002	(34.0522, -118.2440)	(40.7589, -73.9851)	3937.1	80.50	Fraud	Fraud
003	(51.5074, -0.1278)	(51.5034, -0.1276)	0.5	45.00	Not Fraud	Legit

4.3 Discussion

The strong performance of the LightGBM model can be largely attributed to the strategic inclusion of geospatial features alongside traditional transaction data. Most conventional

fraud detection models primarily focus on static attributes such as transaction amount, merchant category, and timestamp, which alone may not fully capture the complexities of fraudulent behavior. By integrating the geodesic distance between the cardholder's location and the merchant's location, the system significantly enhances its ability to detect anomalies that are geographically inconsistent or implausible.

This geographic awareness adds an important contextual dimension that helps distinguish legitimate transactions from fraudulent ones, especially in cases where the transaction details themselves might appear normal. For instance, a transaction occurring in a different city or country just minutes after a previous purchase becomes immediately suspect and is flagged accordingly by the model.

Moreover, the implementation of the Streamlit-based web interface adds considerable value in terms of usability and practical deployment. It enables financial institutions and users to input transaction details easily and receive real-time risk assessments, making the model not only effective but also accessible. The seamless integration of machine learning predictions with an intuitive front-end supports timely decision-making and fraud prevention in live environments.

4.4 Limitations

Despite the encouraging results, the system is not without limitations. One primary constraint is its dependence on the quality and comprehensiveness of labeled training data. Fraud detection models require extensive, balanced datasets that represent a wide range of legitimate and fraudulent behaviors to learn effectively. If the training data is skewed toward certain merchants, locations, or transaction types, the model might struggle to generalize to new or rare fraud patterns, potentially reducing its predictive accuracy.

Another notable limitation involves the risk of false positives, where legitimate transactions are mistakenly classified as fraudulent. This can happen when a user's genuine behavior involves unusual but valid transactions, such as traveling frequently or making

purchases from distant merchants. Excessive false alarms may negatively impact user experience and trust in the system.

Currently, the system relies on manually entered location data, which may introduce inaccuracies or delays. Integration of real-time geolocation services, such as GPS or IP-based tracking, could improve data accuracy and automate the process further. Additionally, the model does not yet incorporate other valuable fraud indicators like device fingerprints, behavioral biometrics, or network-level data, which are increasingly important in comprehensive fraud prevention strategies.

Finally, evolving tactics by fraudsters pose an ongoing challenge. Continuous model retraining, inclusion of additional data sources, and adaptive learning techniques will be necessary to keep pace with emerging fraud patterns and maintain high detection performance over time.

4.5 Summary

This chapter presented a detailed evaluation of the LightGBM-based credit card fraud detection system enhanced by geospatial features. The model exhibited strong predictive performance across multiple metrics, demonstrating its ability to accurately distinguish fraudulent from legitimate transactions. Case studies further highlighted the practical effectiveness of the system in real-world scenarios, especially in detecting geographically inconsistent transactions.

The discussion emphasized the significant benefits of incorporating location-aware intelligence, which enhances detection accuracy beyond conventional approaches. Usability through the Streamlit interface and privacy-preserving data preprocessing were also key strengths.

Chapter 5

REAL-WORLD APPLICATIONS AND IMPACT

5.1 Real-World Applications

The proposed location-aware fraud detection system leveraging LightGBM and geospatial analytics has significant practical applications in the financial industry and beyond. With the proliferation of digital payments, credit card fraud remains a pervasive threat to banks, merchants, and consumers globally. This system can be deployed within financial institutions as a critical component of their transaction monitoring infrastructure. By providing real-time fraud risk assessments, the system enables banks to identify suspicious transactions quickly, reduce financial losses, and prevent fraudulent activities before they escalate.

E-commerce platforms and online marketplaces can also integrate this model into their payment processing workflows to verify transaction authenticity dynamically. By cross-referencing transaction location data with customer behavior, the system can effectively flag high-risk transactions, reducing chargebacks and enhancing merchant trust.

Additionally, payment gateways and digital wallets can use the model to enhance user security. The integration of geospatial features ensures that transactions deviating significantly from typical user behavior are scrutinized, adding an extra layer of protection for end-users.

Beyond financial services, this model can be adapted for use in other domains where transaction authenticity and user location are critical. For instance, telecom operators can use it to detect SIM card fraud, while insurance companies may apply similar methods to flag suspicious claims based on geographic inconsistencies.

5.2 Societal and Economic Impact

The deployment of an effective fraud detection system has far-reaching benefits that extend beyond immediate financial savings. By reducing credit card fraud, the system helps protect consumers from identity theft, unauthorized transactions, and financial distress. This fosters greater consumer confidence in digital payments, supporting broader financial inclusion and encouraging adoption of cashless payment methods.

For financial institutions, the reduction in fraud-related losses translates into improved profitability and operational efficiency. It also strengthens compliance with regulatory requirements related to fraud prevention, thereby avoiding legal penalties and reputational damage.

The economic benefits ripple further into the ecosystem by fostering trust between consumers, merchants, and payment providers. Secure and reliable transaction environments stimulate e-commerce growth, creating opportunities for businesses of all sizes.

Furthermore, the system aligns with global efforts to promote sustainable development by supporting several United Nations Sustainable Development Goals (SDGs). It enhances SDG 9 (Industry, Innovation, and Infrastructure) by advancing fintech innovation and building resilient payment infrastructures. It contributes to SDG 8 (Decent Work and Economic Growth) by protecting economic activities and enabling safe commerce. Finally, it supports SDG 16 (Peace, Justice, and Strong Institutions) by reducing financial crimes and strengthening institutional integrity.

Table 5.1: Real-World Applications and Associated Benefits [33].

Application Domain	Key Benefits	Impact on Stakeholders
Banking and Financial Services	Real-time fraud detection, reduced financial loss	Increased trust, regulatory compliance
E-commerce Platforms	Reduced chargebacks, enhanced transaction security	Improved merchant reputation, customer confidence
Payment Gateways and Digital Wallets	Enhanced user security, fraud prevention	Safer payment environment, user retention
Telecom Industry	Detection of SIM card and identity fraud	Reduced telecom fraud losses
Insurance Sector	Identification of suspicious claims	Faster fraud investigations, cost savings

5.3 Challenges in Real-World Implementation

While the model shows strong promise, real-world deployment involves challenges that must be addressed. Integrating the system within existing financial IT infrastructures requires compatibility and scalability considerations. Data privacy regulations, such as GDPR, mandate careful handling of sensitive user data, necessitating robust anonymization and security practices.

Continuous model updating is essential to adapt to evolving fraud tactics, requiring ongoing data collection and retraining. Furthermore, balancing detection sensitivity to minimize false positives while maximizing fraud catch rates is crucial to maintain user trust and operational effectiveness.

5.4 Future Directions for Broader Impact

To maximize real-world impact, future work could explore the incorporation of additional data sources such as device fingerprints, behavioral biometrics, and network traffic patterns, further enhancing fraud detection accuracy. Real-time geolocation integration via mobile or network data would automate location input, improving prediction speed and accuracy.

Collaboration with industry stakeholders to pilot and refine the system in operational environments can accelerate adoption and highlight practical challenges. Moreover, extending the system's framework to other sectors prone to fraud can broaden its societal benefits.

5.5 Summary

This chapter highlighted the practical applications and significant societal benefits of the location-aware fraud detection system using LightGBM. The model's ability to provide real-time, accurate fraud risk assessments makes it a valuable tool for financial institutions, e-commerce platforms, payment gateways, and other industries vulnerable to fraud. By enhancing transaction security and reducing financial losses, the system promotes greater consumer trust and supports sustainable economic growth. Despite some implementation challenges, the model aligns closely with global sustainable development goals, reinforcing innovation, economic stability, and strong institutional governance. Future enhancements incorporating additional data sources and automation promise to expand its impact further, making it a powerful solution for combating fraud in an increasingly digital world.

Chapter 6

FUTURE DIRECTIONS AND INNOVATION FOR FINANCIAL SAFETY

6.1 Introduction

As digital transactions continue to grow in volume and complexity, fraudsters are constantly evolving their methods to bypass detection systems. While the current fraud detection system built using LightGBM and geospatial analytics offers significant improvements over traditional models, it is essential to plan for future enhancements. These innovations will not only address emerging challenges but also align the system with industry best practices, regulatory requirements, and user expectations. This chapter explores potential directions for future work and innovations that can improve financial safety, model performance, privacy, and user trust.

6.2 Enhancing Input and Automation

Manual entry of user and merchant location data presents limitations in terms of speed, accuracy, and user experience. Automating data collection can significantly improve system efficiency and precision.

6.2.1 Real-Time Geolocation Integration

In the current model, users must manually enter latitude and longitude data, which can be time-consuming and error-prone. A more advanced approach would involve integrating real-time geolocation data using GPS or IP-based services. For example, when a transaction occurs, the system can automatically capture the user’s current coordinates and compare them with the merchant's location to compute distance.

This automation reduces dependency on user input and improves real-time processing, making the system more viable for integration into mobile banking apps, digital wallets, and point-of-sale terminals.

6.2.2 Device Fingerprinting

Device fingerprinting involves collecting information about the device used during a transaction—such as browser type, operating system, screen resolution, and hardware details. By creating a unique fingerprint for each device, the system can flag transactions made from unfamiliar or suspicious devices, even if other transaction attributes seem normal. This enhances fraud detection by identifying unauthorized access attempts.

6.3 Behavioral Intelligence

Fraud detection can be greatly improved by understanding and analyzing how users behave during digital interactions. Behavior-based security offers strong personalization, making it harder for attackers to mimic genuine users.

6.3.1 Behavioral Biometrics

Behavioral biometrics monitor subtle patterns in user interaction, such as keystroke dynamics, mouse movements, touchscreen gestures, and typing speed. These patterns form a behavioral signature that is unique to each user. When a transaction is initiated, the system can compare the behavior with the user's normal profile to detect anomalies.

This method is particularly powerful because it works in the background, does not require extra steps for the user, and is difficult for attackers to replicate.

6.3.2 User Profile Learning

Over time, the system can learn the user’s transaction habits, such as preferred merchants, locations, and time of day. Using machine learning, it can distinguish legitimate outliers (e.g., vacation transactions) from potential fraud. This adaptive understanding minimizes false positives and enhances user satisfaction.

6.4 Smart and Secure Model Learning

One of the key challenges in machine learning-based fraud detection is the evolving nature of fraudulent techniques. Traditional models are static—they are trained once on historical

data and used repeatedly for predictions. However, fraud patterns can change rapidly, with attackers constantly adapting their tactics. A model that performs well today may become outdated within weeks or even days if it cannot recognize new anomalies or behaviors. Therefore, smart and secure learning mechanisms are essential to ensure that fraud detection systems remain effective, flexible, and trustworthy over time.

This section explores two powerful approaches—adaptive and online learning, and federated learning—that offer solutions to this dynamic problem.

6.4.1 Adaptive and Online Learning

Adaptive learning refers to the model's ability to evolve and improve continuously by learning from new incoming data. In the context of fraud detection, this means updating the model as new transactions are processed and as the outcomes (fraud or not fraud) are confirmed.

Online learning is a specific form of adaptive learning in which the model receives a stream of data points and updates itself incrementally, without the need for retraining on the entire dataset. For example, if a new type of fraud is detected in recent transactions, the system can immediately learn from it and adjust its parameters to detect similar frauds in the future.

This technique drastically reduces the delay between fraud emergence and model adaptation, thereby improving real-time performance and resilience. Moreover, online learning systems can incorporate feedback loops—where users or security teams verify and label transaction outcomes—which the model then uses to reinforce or correct its predictions.

Benefits of adaptive and online learning include:

- Continuous improvement without retraining from scratch
- Faster detection of new fraud patterns
- Real-time adaptation to user behavior changes

- Reduced operational downtime for model updates

In a live banking or e-commerce environment, this capability is crucial for maintaining up-to-date fraud prevention without frequent manual intervention.

6.4.2 Federated Learning

Federated learning is a groundbreaking approach that addresses both performance and privacy. In traditional machine learning, all training data is collected and centralized in one location for model training. However, in financial systems, this creates data privacy risks and compliance challenges, especially under regulations like GDPR or HIPAA.

With federated learning, data remains on the local servers (e.g., within each bank), and only model updates or gradients are shared with a central server. These updates are aggregated and used to improve the global model, which is then sent back to each participating node.

For fraud detection, federated learning enables collaborative intelligence across multiple institutions without compromising the confidentiality of user data. A network of banks, payment gateways, or financial platforms can work together to detect patterns of fraud that may span across systems—such as coordinated attacks or card testing scams—while ensuring full data protection.

Key advantages include:

- Enhanced fraud detection through multi-institutional learning
- Strong privacy compliance (no raw data leaves local servers)
- Protection against isolated blind spots (e.g., fraud visible to one bank but not others)
- Scalable and secure model deployment across a network of partners

6.5 Blockchain-Based Transaction Verification

Blockchain technology offers decentralization, immutability, and transparency—three valuable features in fraud prevention.

6.5.1 Smart Contracts for Transaction Validation

Smart contracts are self-executing agreements stored on the blockchain. Fraud detection logic can be embedded within smart contracts, allowing automatic approval or rejection of transactions based on predefined rules (e.g., flag transactions above a certain risk score). This makes the validation process secure and tamper-proof.

6.5.2 Transparent Audit Trails

Blockchain can be used to store verified transaction records, fraud alerts, and decisions in an immutable format. This provides a clear audit trail that can be used by compliance teams or auditors to investigate incidents and improve system accountability.

6.6 User-Focused Enhancements

User participation is a critical factor in the effectiveness and acceptance of automated fraud detection systems. Involving users directly not only increases transparency but also fosters trust and confidence in the security infrastructure. Future fraud detection platforms must prioritize giving users visibility and control over the decision-making process, allowing them to contribute actively to fraud prevention. Two key areas of innovation in this regard are real-time alerts and customizable security settings.

6.6.1 Real-Time Alerts and Verification

One of the most practical and impactful innovations is the implementation of real-time alerts for suspicious or high-risk transactions. When the system detects behavior that deviates significantly from the user's normal patterns—such as transactions from unfamiliar locations, unusually high amounts, or unrecognized devices—it can instantly notify the user via SMS, email, or push notification.

The user can then take immediate action: approve the transaction if it's legitimate or report it as fraudulent. This not only reduces the window of opportunity for malicious actors but also minimizes the likelihood of mistakenly blocking genuine user activity. By integrating users into the decision loop, the system becomes more responsive and less prone to false positives, improving overall user satisfaction and system reliability.

6.6.2 Customizable Risk Preferences

Different users have different tolerance levels when it comes to risk. Some may prefer a high level of monitoring and alerts, while others may find frequent notifications intrusive. To accommodate these variations, future fraud detection systems can offer customizable risk settings within the user interface.

Users can configure parameters such as transaction amount limits, regional or country-based restrictions, device whitelisting, and preferred notification channels. For instance, a user may choose to block all transactions over a certain amount unless manually approved or restrict transactions to specific countries while traveling. These options provide a layer of personalization, allowing users to tailor the system to their comfort level while maintaining robust security.

The integration of adaptive and federated learning not only enhances the technical capability of fraud detection systems but also supports a long-term strategic shift toward **self-improving and privacy-conscious AI** in financial services. These approaches reduce reliance on static, one-size-fits-all models and instead foster the development of flexible, collaborative systems that evolve in response to real-world threats. As financial fraud becomes increasingly sophisticated, only those systems that can learn, adapt, and scale across institutions will remain resilient. Embracing these advanced learning methodologies ensures that fraud prevention remains one step ahead of cybercriminals—protecting consumers, reinforcing institutional trust, and enabling innovation in a secure digital economy.

Table 6.1: Summary Table of Innovations[34].

Innovation	**Function**	**Safety Benefit**
Real-Time Geolocation	Automatically captures user's actual location	Reduces location input errors
Device Fingerprinting	Identifies hardware/browser characteristics	Detects access from unknown devices
Behavioral Biometrics	Monitors user interaction habits	Adds invisible and strong user authentication
Adaptive/Online Learning	Updates model with new data over time	Keeps system up-to-date with fraud trends
Federated Learning	Enables collaborative model training across banks	Improves privacy and scalability
Smart Contracts (Blockchain)	Validates transactions securely via rules	Reduces risk of manipulation
Audit Trails (Blockchain)	Stores fraud alerts and decisions on chain	Enables transparent investigations
User Alerts & Preferences	Sends real-time notifications and lets users respond	Empowers user control and trust

6.7 Summary

This chapter presented various directions for extending the current fraud detection system into a more intelligent, adaptive, and secure framework. From real-time geolocation and behavioral biometrics to federated learning and blockchain-based validation, each innovation enhances a specific aspect of financial safety—whether it's accuracy, privacy,

scalability, or user trust. These future enhancements not only improve technical robustness but also align the system with ethical, regulatory, and user experience goals, making it suitable for broader deployment in real-world environments.

Chapter 7

CONCLUSION AND REFLECTION

7.1 Conclusion

This project set out to address one of the most pressing challenges in the digital economy—credit card fraud—by developing an intelligent, location-aware fraud detection system. The core idea was to enhance existing models by integrating spatial features such as the distance between the user and merchant locations, alongside conventional transaction data like amount, category, and time.

To achieve this, a machine learning model was built using the LightGBM algorithm, trained on a labeled transaction dataset enriched with geospatial features. The Haversine formula was used to compute the geographic distance between transaction participants, helping the system flag geographically inconsistent behavior—a common indicator of fraud. Label encoding and hashing were applied to preprocess categorical and sensitive data, maintaining both efficiency and privacy. A user-friendly front-end was created using Streamlit, allowing real-time interaction with the model for quick and accurate fraud risk assessments.

This approach not only demonstrated high model accuracy and reliability but also provided a scalable framework that can be adapted by financial institutions to prevent evolving fraud threats. By blending machine learning with location-based intelligence, the system effectively bridges the gap between traditional fraud detection and modern digital behavior patterns, offering a proactive solution that enhances transaction security while respecting user privacy and usability.

This approach not only demonstrated high model accuracy and reliability but also provided a scalable framework that can be adapted by financial institutions to prevent evolving fraud threats. By blending machine learning with location-based intelligence, the system effectively bridges the gap between traditional fraud detection and modern digital behavior

patterns, offering a proactive solution that enhances transaction security while respecting user privacy and usability.

Throughout the process, several key steps were followed:

Table 7.1: Project Steps and Key Outcomes [35].

Step	Description	Outcome
Data Collection & Preprocessing	Label encoding, hashing, and geospatial integration	Clean, secure, model-ready dataset
Distance Feature Engineering	Calculated user–merchant distance using Haversine formula	Enhanced fraud detection accuracy
Model Selection & Training	Used LightGBM with optimized parameters	High-performance fraud classification model
Evaluation & Metrics Analysis	Measured accuracy, precision, recall, F1-score, AUC-ROC	Consistently high scores across metrics
Real-Time Web Interface	Developed a Streamlit-based app for user interaction	Live fraud prediction tool with intuitive UI
Risk Interpretation	Mapped output to meaningful labels (fraud/not fraud)	Clear decision feedback for user understanding

7.2 Reflection

This project provided deep insight into the real-world complexities of fraud detection and the practical challenges of implementing machine learning in the financial domain. It demonstrated how powerful predictive models like LightGBM, when combined with

thoughtful feature engineering and user-centric design, can lead to meaningful security solutions.

One of the major takeaways was the importance of context—in this case, location—in improving model accuracy and reducing false positives. Equally important was balancing technological innovation with user experience, data privacy, and scalability. The project also highlighted the limitations of static systems and the need for models that can adapt and grow in response to evolving threats.

Working with transaction data also deepened the understanding of data sensitivity, emphasizing the need for secure handling, anonymization, and ethical deployment of AI systems.

7.3 Future Outlook

While the current system performs well, several enhancements can be made in future work to further increase its effectiveness and readiness for real-world deployment:

- Real-time geolocation tracking using GPS or IP for automation and speed
- Behavioral biometrics to verify user identity beyond input data
- Adaptive and online learning to keep the model updated with emerging fraud patterns
- Federated learning for collaborative fraud intelligence without compromising privacy
- Blockchain integration for secure, tamper-proof verification and audit trails
- User customization features allowing individuals to set their fraud sensitivity levels
- Device fingerprinting to block access from unrecognized or high-risk devices

These innovations would not only make the system smarter and more robust but also align it with global priorities around data protection, real-time security, and user empowerment.

7.4 Final Thoughts

In a world where financial fraud is becoming more sophisticated, the solution must be equally advanced. This project proves that combining machine learning, location intelligence, and human-centered design can result in a fraud detection system that is not only accurate and efficient but also practical and scalable. With ongoing development and ethical implementation, such systems have the potential to safeguard millions of transactions, strengthen institutional trust, and support a more secure digital future

REFERENCES

[1] Ignizio, J.P., 1968, *A Method to Achieve Optimum Air Defense Sensor Allocation*, MS Dissertation, University of Alabama, Alabama.

[2] Ignizio, J.P., 1971, *A Heuristic Solution to Generalized Covering Problems*, Ph.D. Dissertation, Virginia Polytechnic Institute, Blacksburg.

[3] Osyczka, A., 1985, "Multicriteria optimization for engineering design," in *Design Optimization*, Academic Press, Cambridge, pp. 193–227.

[4] Jones, C.D., Smith, A.B., and Roberts, E.F., 1994, *Efficient Real-Time Fine Grained Concurrency*, 2nd Ed., Ch. 3, pp. 145–147, Tata McGraw-Hill, New Delhi.

[5] Badri, M.A., Mortagy, A.K. and Alsayed, A., 1998, "A Multi-objective Model for Locating Fire Stations," *European Journal of Operational Research*, Vol. 110, No. 2, pp. 243–260.

[6] Chan, P.K. and Stolfo, S.J., 1998, "Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection," *Proc. of the 4th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 164–168.

[7] Chatani, E., Hayashi, R., Lange, R., and Balny, C., 2002, "Thermal and Pressure Stability of Phe46 Mutants of Ribonuclease A," *Proc. of First International Conference on High Pressure Bioscience and Biotechnology*, Kyoto, Japan, pp. 27–32.

[8] Friedman, J.H., 2001, "Greedy Function Approximation: A Gradient Boosting Machine," *The Annals of Statistics*, Vol. 29, No. 5, pp. 1189–1232.

[9] Bhattacharyya, S., Jha, S., Tharakunnel, K., and Westland, J.C., 2011, "Data Mining for Credit Card Fraud: A Comparative Study," *Decision Support Systems*, Vol. 50, No. 3, pp. 602–613.

[10] Lopez-Rojas, E.A., and Axelsson, S., 2012, "Money Laundering Detection Using Synthetic Data," *Proc. of the European Intelligence and Security Informatics Conference (EISIC)*, pp. 41–46.

[11] Zareapoor, M. and Shamsolmoali, P., 2015, "Application of Credit Card Fraud Detection: Based on Bagging Ensemble Classifier," *Procedia Computer Science*, Vol. 48, pp. 679–685.

[12] Nasimuddin, M., Hussain, M., and Khan, S., 2017, "An Efficient Fraud Detection Model in Online Payments Using Geo-Location," *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 8, No. 3, pp. 175–181.

[13] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.Y., 2017, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30, pp. 3146–3154.

[14] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L., and Caelen, O., 2018, "Sequence Classification for Credit-Card Fraud Detection," *Expert Systems with Applications*, Vol. 100, pp. 234–245.

[15] Zheng, L., Lai, J., Zhang, X., and Liu, Z., 2020, "Fraud Detection with Graph Neural Networks," *Proc. of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, pp. 2713–2720.

[16] Tulsiani, P., and Joshi, A., 2021, "Location-Aware Fraud Detection Using ML Models," *Journal of Cybersecurity Research*, Vol. 5, No. 1, pp. 45–53.

[17] UN General Assembly, 2015, *Transforming Our World: The 2030 Agenda for Sustainable Development*, United Nations, New York. [Document A/RES/70/1].

[18] Bhargvesh/Mini-project

[19] Bhargvesh/Mini-project

[20] Bhargvesh/Mini-project

[21] Bhargvesh/Mini-project

[22] Bhargvesh/Mini-project

[23] Bhargvesh/Mini-project

[24] Bhargvesh/Mini-project

[25] *Synthetic Dataset for Credit Card Transaction Simulation*, created by project authors (Bhargvesh/Mini-project), based on real-world fraud detection structures. Used in Table 2.1 for illustrating geospatially enriched features.

[26] *LightGBM Training Logs and Evaluation Reports*, Bhargvesh/Mini-project. Contains accuracy, precision, recall, and F1-score metrics. Referenced in Table 2.2.

[27] *Preprocessing and Encoding Schema for Categorical Data*, developed by Bhargvesh/Mini-project. Covers how merchant, category, and gender data are encoded. Referenced in Table 3.1.

[28] *Geopy Haversine Distance Samples for Fraud Detection*, Bhargvesh/Mini-project, based on calculated distances for various location coordinates. Referenced in Table 3.2.

[29] *Feature Set Description for LightGBM Fraud Prediction*, Bhargvesh/Mini-project. Lists model input types and formats. Referenced in Table 3.3.

[30] *Component–Tool Mapping for Credit Card Fraud Detection System*, Bhargvesh/Mini-project. Describes LightGBM, Streamlit, Geopy, Scikit-learn, and Joblib. Referenced in Table 3.4.

[31] *Final Performance Report of Trained LightGBM Model*, Bhargvesh/Mini-project. Evaluation metrics from both training and validation sets. Referenced in Table 4.1.

[32] *Sample Case-Based Prediction Outputs Using Geospatial Features*, Bhargvesh/Mini-project. Details user–merchant distances and fraud labels. Referenced in Table 4.2.

[33] *Domain-wise Applications and Stakeholder Benefits of the Fraud Detection System*, Bhargvesh/Mini-project. Used to contextualize the real-world impact of the model. Referenced in Table 5.1.

[34] *Summary of Future Innovations and their Safety Functions*, Bhargvesh/Mini-project. Table of enhancements including geolocation, federated learning, and blockchain. Referenced in Table 6.1.

[35] *Project Workflow Summary from Data to Deployment*, Bhargvesh/Mini-project. Final overview of project steps and their outcomes. Referenced in Table 7.1.