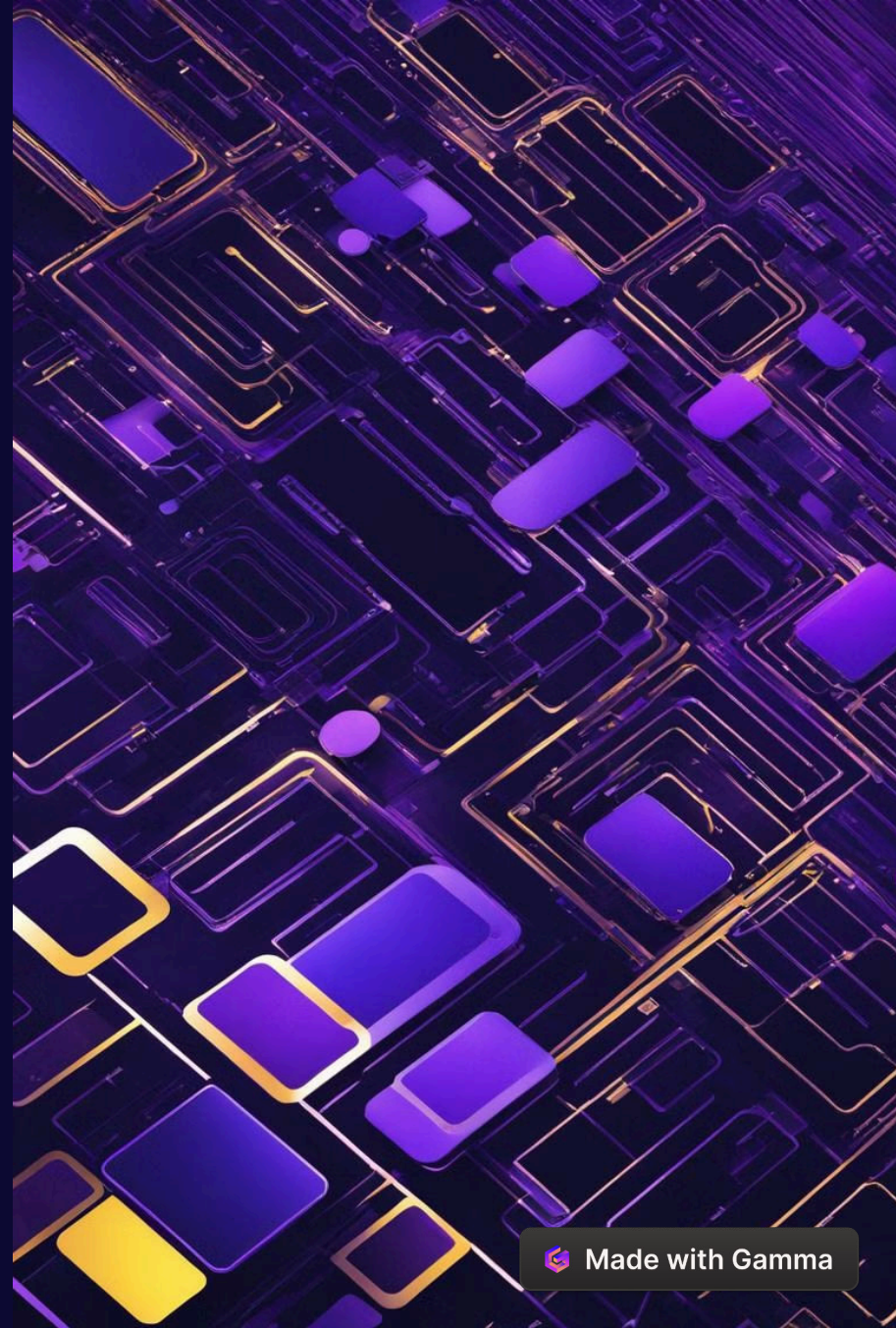
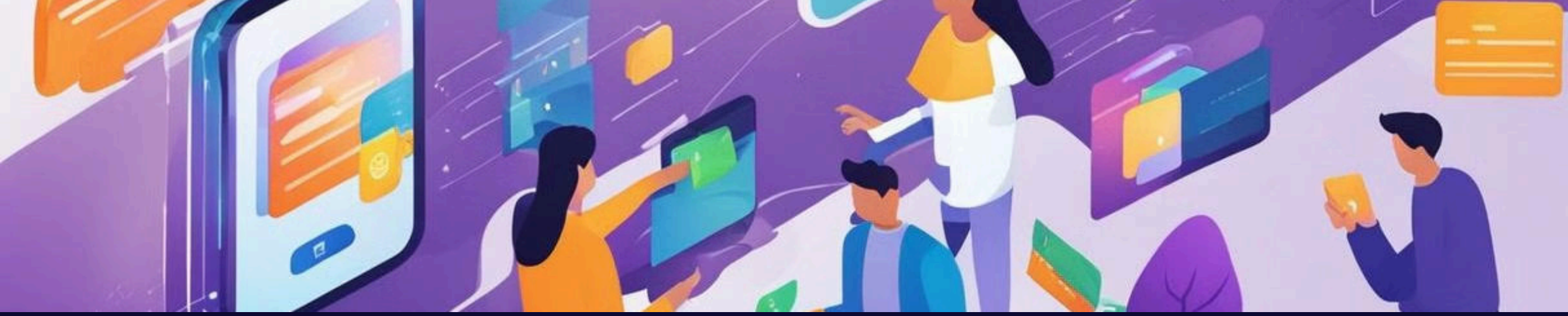
CREDIT CARD FRAUD DETECTION





CREDIT CARD FRAUD DETECTION

Automated Fraud Detection

Our project proposes an automated credit card fraud detection system using machine learning.

Transaction Analysis

We'll analyze transaction features like amount, location, and habits to distinguish fraud.

Adaptive Machine Learning

Machine learning adapts to new tactics, offering superior accuracy and automation.

Presented by:

Ayush Pathak

21052747

Divyesh Kulshreshtha

21052757

Ankita Kumari

21053235

Jhanvi Jain

21053233

Somya Sinha

21053365

CONTENT

INTRODUCTION

Our project proposes an automated credit card fraud detection system using machine learning.

We'll analyze transaction features like amount, location, and habits to distinguish fraud.

Machine learning adapts to new tactics, offering superior accuracy and automation.

Through feature engineering, model selection, and real-time integration, we aim to build a robust shield against evolving fraud.

SCOPE

Credit card fraud detection is crucial in finance due to increasing digital transactions.

Machine learning analyzes transaction data to proactively identify and prevent fraudulent activities.

By leveraging historical data, ML models distinguish between legitimate and fraudulent transactions, enhancing security and minimizing financial losses.

ML-based solutions track patterns to prevent abnormal transactions, addressing common fraud techniques like counterfeit cards and CNP fraud.

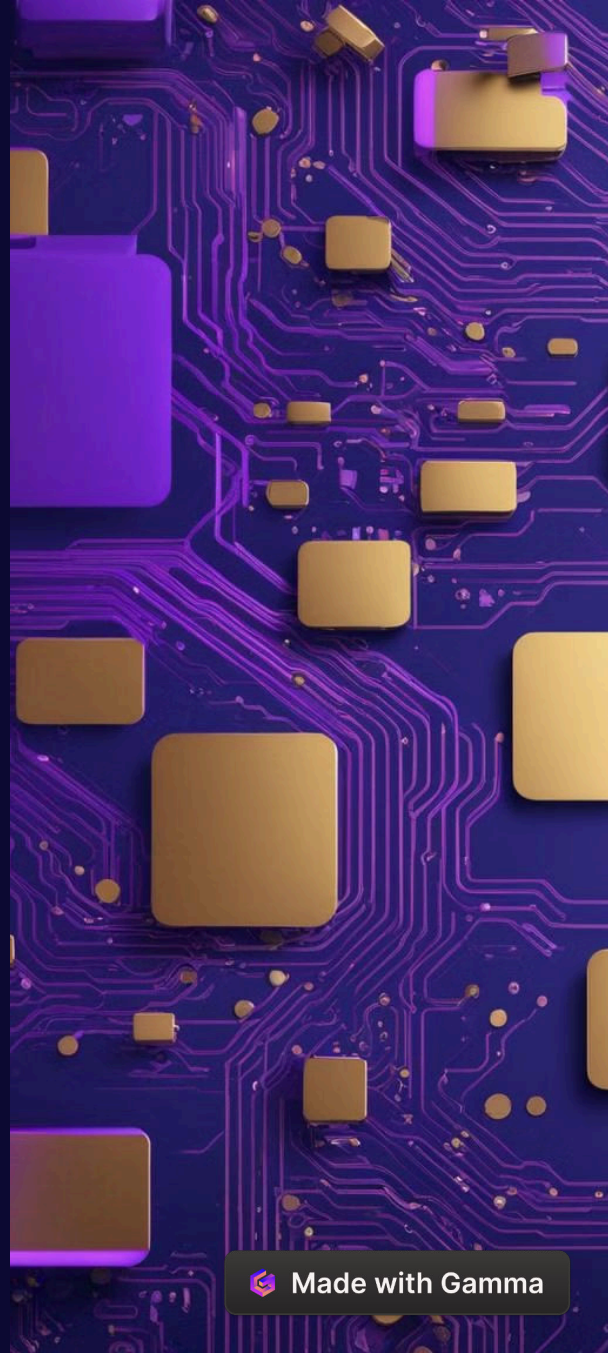
PROBLEM STATEMENT

Developing an advanced machine learning model

Developing an advanced machine learning model for credit card fraud detection presents a critical challenge in financial security.

Rising threats of fraudulent activities

Addressing the rising threats of fraudulent activities, this project aims to implement innovative algorithms to accurately identify and prevent unauthorized transactions, safeguarding both consumers and financial institutions.



REQUIREMENTS AND SPECIFICATION

PROJECT PLANNING

Define the goals and data requirements for building a credit card fraud detection model using machine learning.

PROJECT ANALYSIS

Evaluate the strengths, weaknesses, opportunities, and threats associated with the current approach to credit card fraud detection.

SYSTEM DESIGN

Plan the architecture and components of the system for implementing the machine learning model for fraud detection.

PROJECT PLANNING

Goal : Build a credit card fraud detection model using machine learning.
Data : "creditcard.csv" containing transaction information.
Clean and pre-process data.
Split features (X) and target (y). Train Logistic Regression model.
Evaluate model performance (accuracy, report, confusion matrix). Visualize results.
Outcomes: Trained model, performance metrics, visualizations.
Contingency : Address data quality, explore alternative models if needed.
Libraries: pandas, scikit learn, matplotlib

PROJECT ANALYSIS

- **Strengths:** Clear goals, transparent code, baseline performance, modular design.
- **Weaknesses:** Limited data handling, simple algorithm, basic features, narrow evaluation, deployment not considered.
- **Opportunities:** Enhance data quality, explore better algorithms, advanced feature engineering, comprehensive evaluation metrics, deployment possibilities.
- **Threats:** Data availability, achieving high accuracy, computational demands, privacy concerns.
- **Recommendations:** Focus on data quality, explore alternatives, leverage feature engineering, use diverse evaluation metrics, consider deployment, ensure data privacy.

SYSTEM DESIGN

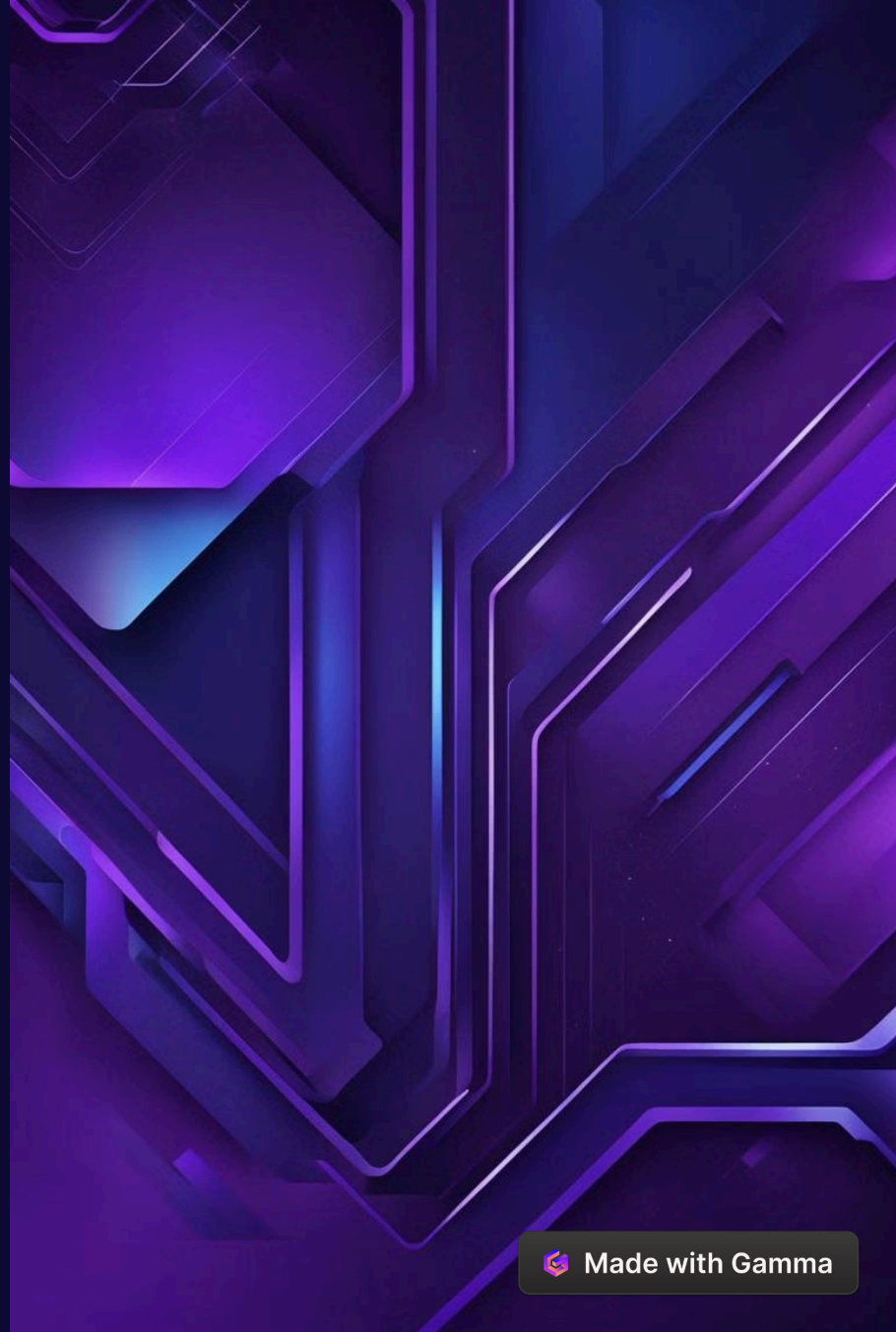
Strengths: Clear goals, transparent code, baseline performance, modular design.

Weaknesses: Limited data handling, simple algorithm, basic features, narrow evaluation, deployment not considered.

Opportunities: Enhance data quality, explore better algorithms, advanced feature engineering, comprehensive evaluation metrics, deployment possibilities.

Threats: Data availability, achieving high accuracy, computational demands, privacy concerns.

Recommendations: Focus on data quality, explore alternatives, leverage feature engineering, use diverse evaluation metrics, consider deployment, ensure data privacy.



IMPLEMENTATION

DATA ACQUISITION/ COLLECTION

Data acquisition refers to the process of gathering the raw data that will be used to train and evaluate machine learning models. This data can come from various sources and may require different techniques for collection and preparation.

The project exclusively deals with numerical data.

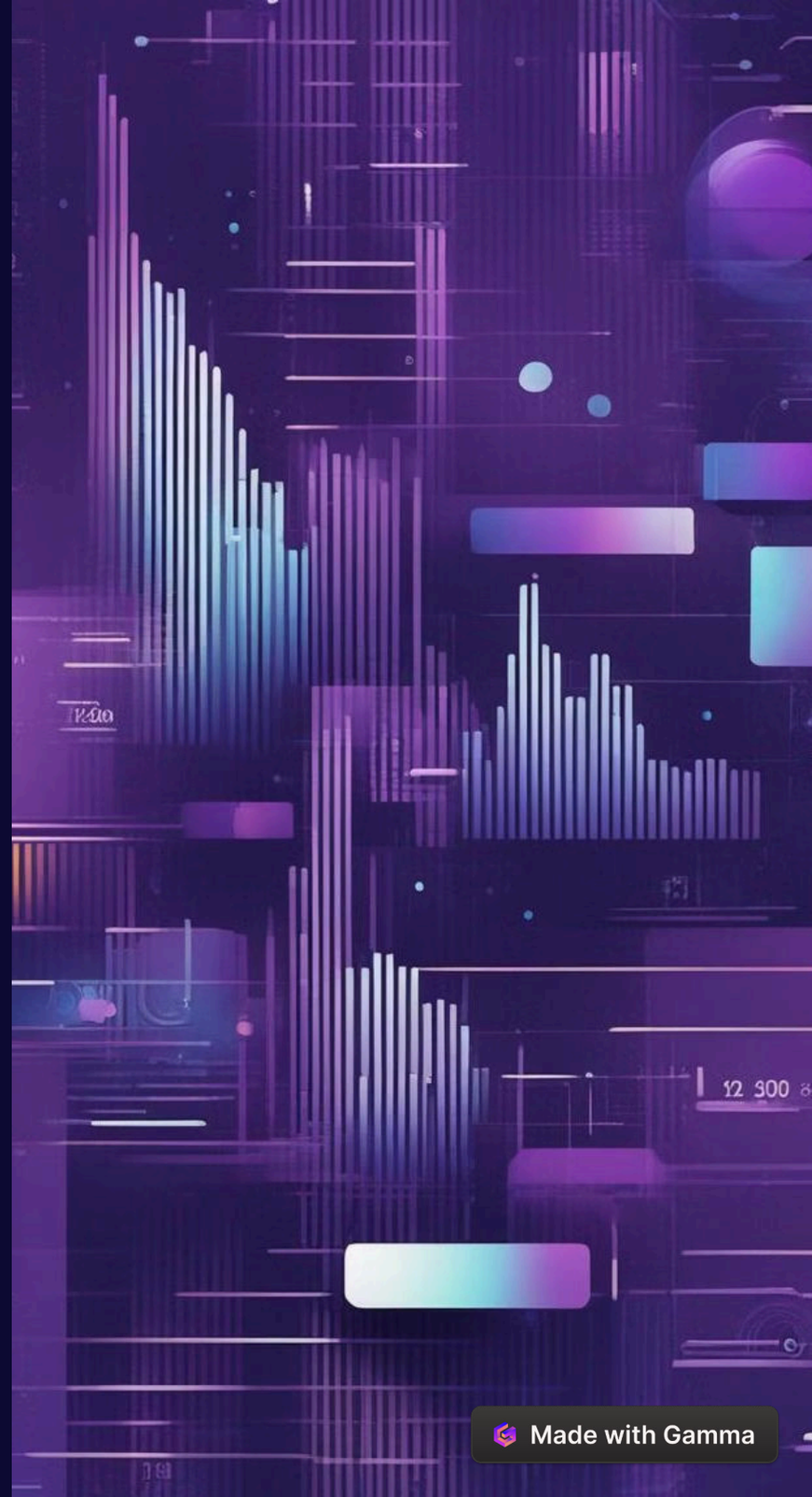
Principal Component Analysis (PCA) is utilized to transform the features.

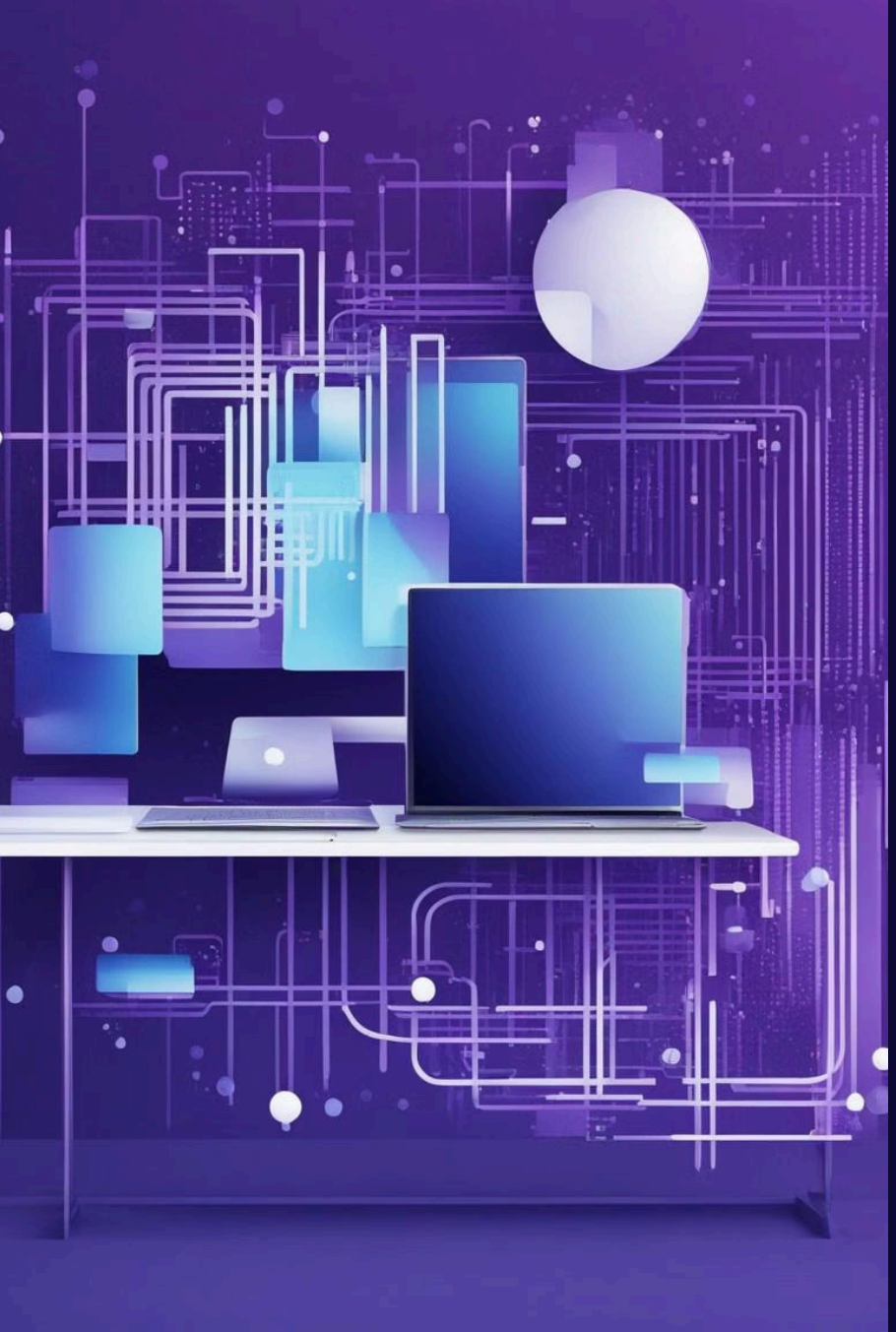
Features V1 - V28 are representations of the principal components derived from PCA.

Additionally, there are two raw features: Time (measured in elapsed seconds) and Amount (transaction value).

The response class categorizes transactions as either fraudulent (1) or non-fraudulent (0).

Due to confidentiality reasons, only limited information about the data is accessible.





DATA PRE-PROCESSING

- Load data & remove duplicates
- Separate features (X) and target variable (y)
- Split data into training and testing sets (80%/ 20%)
- No scaling/ normalization needed (PCA transformed data)
- Train a Logistic Regression model
- Evaluate model performance (accuracy, classification report)
- Visualize predicted frauds vs actual frauds
- Create and visualize confusion matrix

MODELING TRAINING

Separates features (X) from target variable (y)

Splits data into training (X_train, y_train) and testing sets (X_test, y_test)

Creates a Logistic Regression model (linear classifier)

Trains the model using training data (X_train, y_train)

This training process learns the relationship between features and fraudulent transactions

No separate scaling/normalization needed (data assumed preprocessed)

MODEL EVALUATION

Model evaluation in machine learning (ML) is the crucial process of assessing a model's performance. It majorly involves:

Metrics: You use quantitative measures like accuracy, precision, recall, or F1-score to assess the model's performance. The choice of metrics depends on your specific problem.

Testing data: A separate set of data (not used for training) is used for evaluation. This ensures the model isn't simply memorizing the training data.

CONCLUSION

- This project proposes a machine learning system (Python) to fight credit card fraud.
- It will analyze past transactions to identify patterns of fraud and flag suspicious activity in real-time.
- This will help financial institutions catch fraudsters and protect their customers.
- Even with a large amount of data and the challenge of imbalanced data (mostly legitimate transactions), the system aims for high accuracy with minimal false positives.
- This will create a more secure financial environment for everyone

THANK YOU