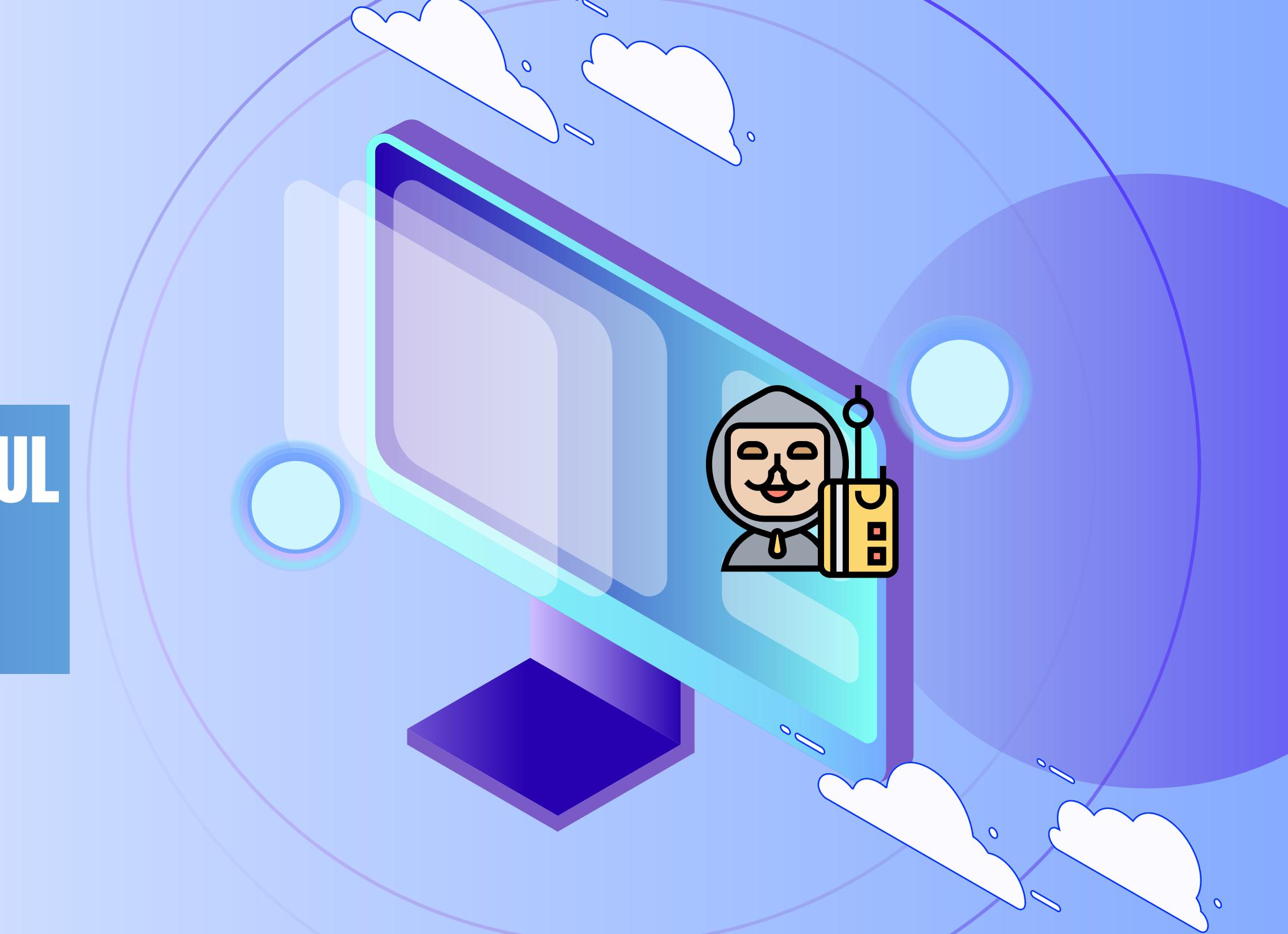


FRAUD TRANSACTIONS DETECTION USING AI

DR. MOHAMMAD SHAHIR ABDUL
MAJED SHIKH

ARTIFICAL INTELLIGENCE &
MACHINE LEARNING

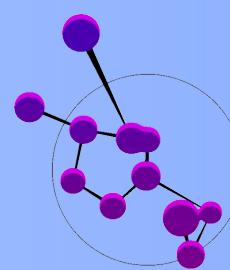
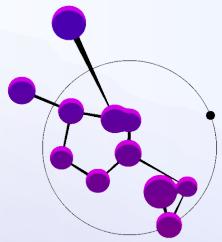
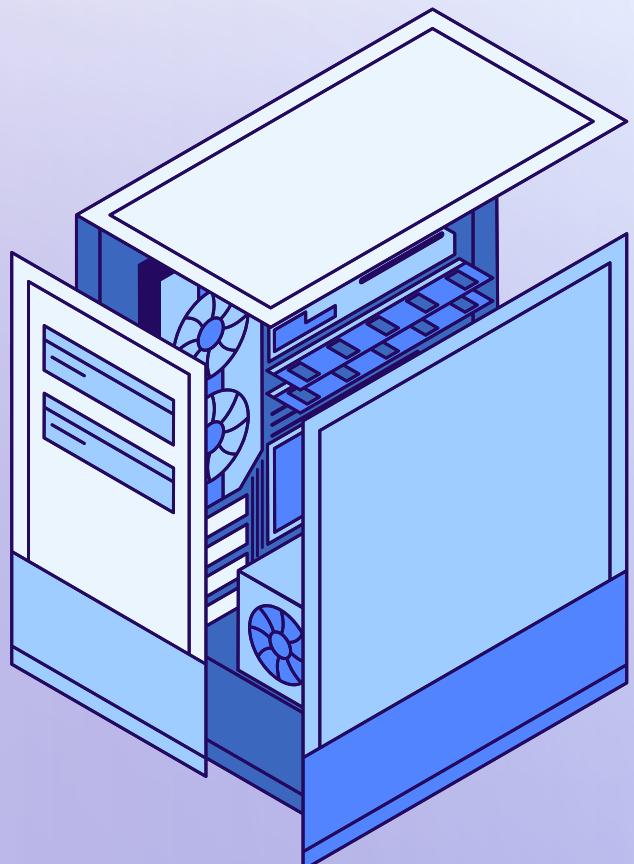
FARZAM NAZARI 400367200



CONTENTS

- Introduction
- Problem Statement
- AI/ML Solution Design and Concept
- KNN Algorithm Overview
- Application of KNN in Fraud Detection
- Case Study
- Project Timeline and Governance
- Conclusion
- Q&A

4
5
6
10
8
9
10
11
12



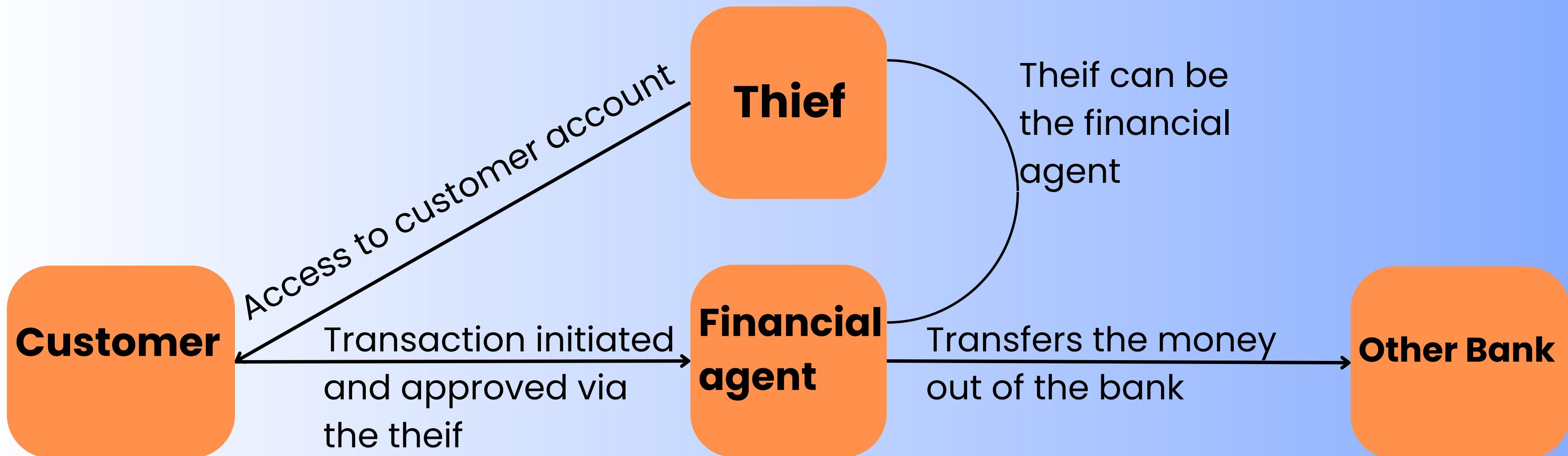
INTRODUCTION

- Fraud detection is crucial in financial services.
- Rising credit card fraud impacts institutions and consumers.
- In the US, consumers reported losses totaling over \$10 billion to fraud in 2023.
- In the UK Over £1.2 billion stolen through unauthorized activities in 2022 (UK Finance Report).
- Germany: Credit card fraud cases and cybercrime are increasing, with a notable rise in fraud attempts on financial institutions in recent years (BusinessWire).



PROBLEM STATEMENT

- Challenges: Financial losses, reputational damage, legal implications.
- Impact: Significant financial and operational costs.
- Current Solutions: Existing methods often have limitations in accuracy and speed.



AI/ML SOLUTION DESIGN AND CONCEPT



Focus Areas:

- Abnormal Activities: Detect unusual patterns and behaviors in transactions.
- Abnormal Spending: Identify transactions that deviate from typical spending patterns.
- Geolocation of Transactions: Monitor the locations where transactions occur to identify anomalies.

Supervised Learning: Uses labeled data to predict outcomes.

Unsupervised Learning: Uses unlabeled data to find patterns.



CHOOSING THE BEST MODELS FOR THE SOLUTION

K-Nearest Neighbors (KNN) is a simple, effective algorithm for classification.

How it Works:

- Instances represented as points in a Euclidean space.
- The nearest neighbor is defined in terms of Euclidean distance.
- An unknown sample is assigned to the most common class among its k nearest neighbors.

Advantages:

- Simplicity
- Effective in pattern recognition
- Adaptable to various datasets

CHOOSING THE BEST MODELS FOR THE SOLUTION

Neural Networks are computational models inspired by the human brain, used for both classification and regression tasks.

How it Works:

- Consists of layers of interconnected nodes (neurons).
- Neurons process input data and pass it through activation functions.
- The network learns by adjusting weights based on error through backpropagation.

Advantages:

- High accuracy
- Capable of learning complex patterns
- Scalable to large datasets

CHOOSING THE BEST MODELS FOR THE SOLUTION

K-Means Clustering is a popular unsupervised learning algorithm for clustering data.

How it Works:

- Partitions data into k clusters.
- Assigns each data point to the nearest cluster center (centroid).
- Iteratively updates centroids until convergence.

Advantages:

- Simple to implement
- Efficient for large datasets
- Works well with distinct clusters

CHOOSING THE BEST MODELS FOR THE SOLUTION

Principal Component Analysis (PCA) is a dimensionality reduction technique.

How it Works:

- Transforms data into a set of orthogonal (uncorrelated) components.
- Components are ordered by the amount of variance they explain.
- Reduces the number of dimensions while preserving most of the data's information.

Advantages:

- Reduces computational cost
- Helps visualize high-dimensional data
- Removes noise and redundancy



APPLICATION OF KNN IN FRAUD DETECTION

Process:

1-Data Collection

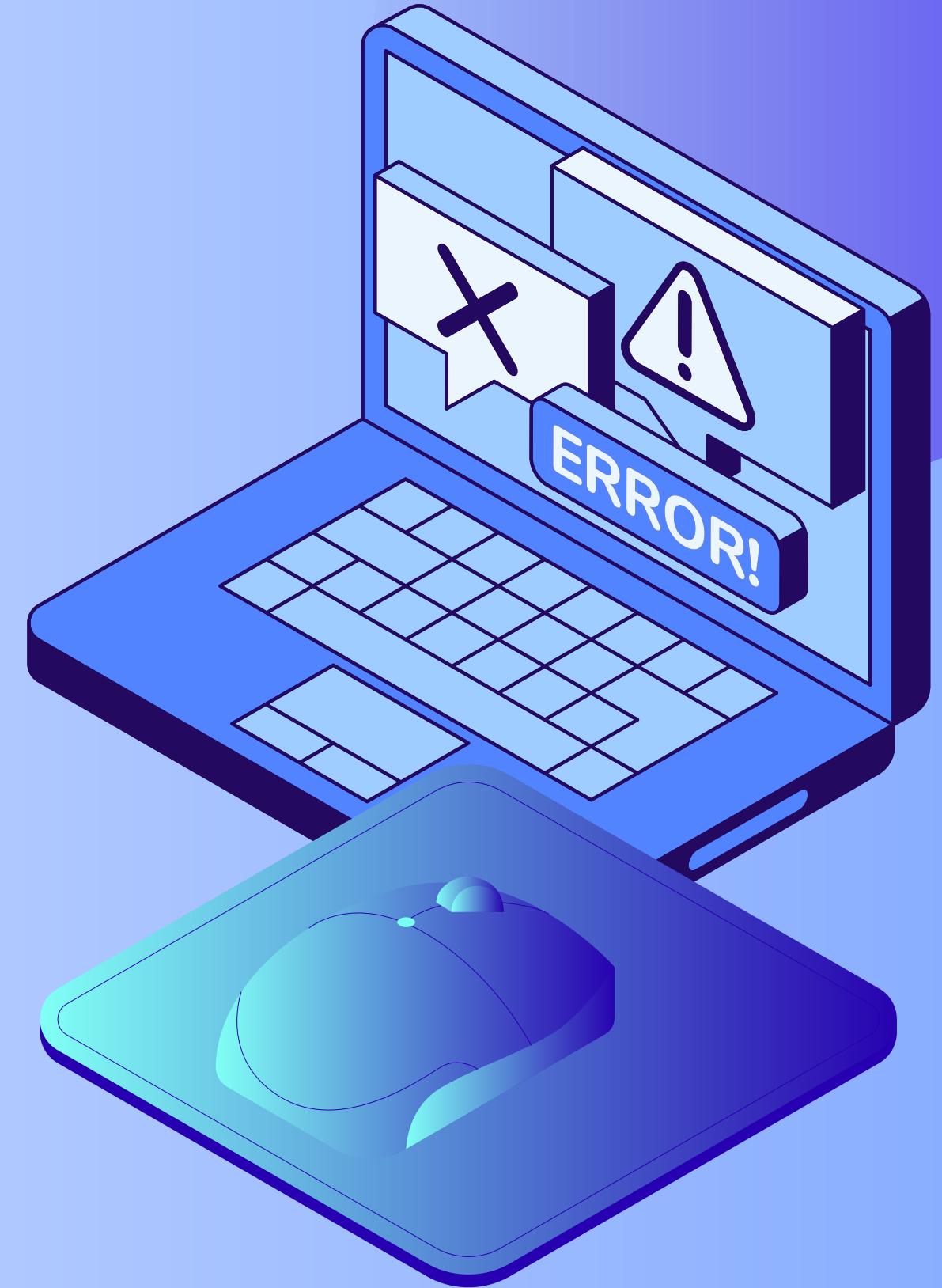
- Sources: Transaction data, card details, user data
- Methods: Database extraction, APIs, web scraping

2-Data Preprocessing

- Data Cleaning
- Data Normalization

3-Data Analysis and Visualization

- Descriptive Statistics
- Visualization: Histograms, scatter plots, box plots



4-Feature Selection

- Correlation Analysis
- Feature Importance

5-Splitting the Dataset

- Training Set
- Testing Set

6-Applying Algorithm

- Choosing k
- Distance Metrics
- Model Training
- Hyperparameter Tuning

7-Decision Making

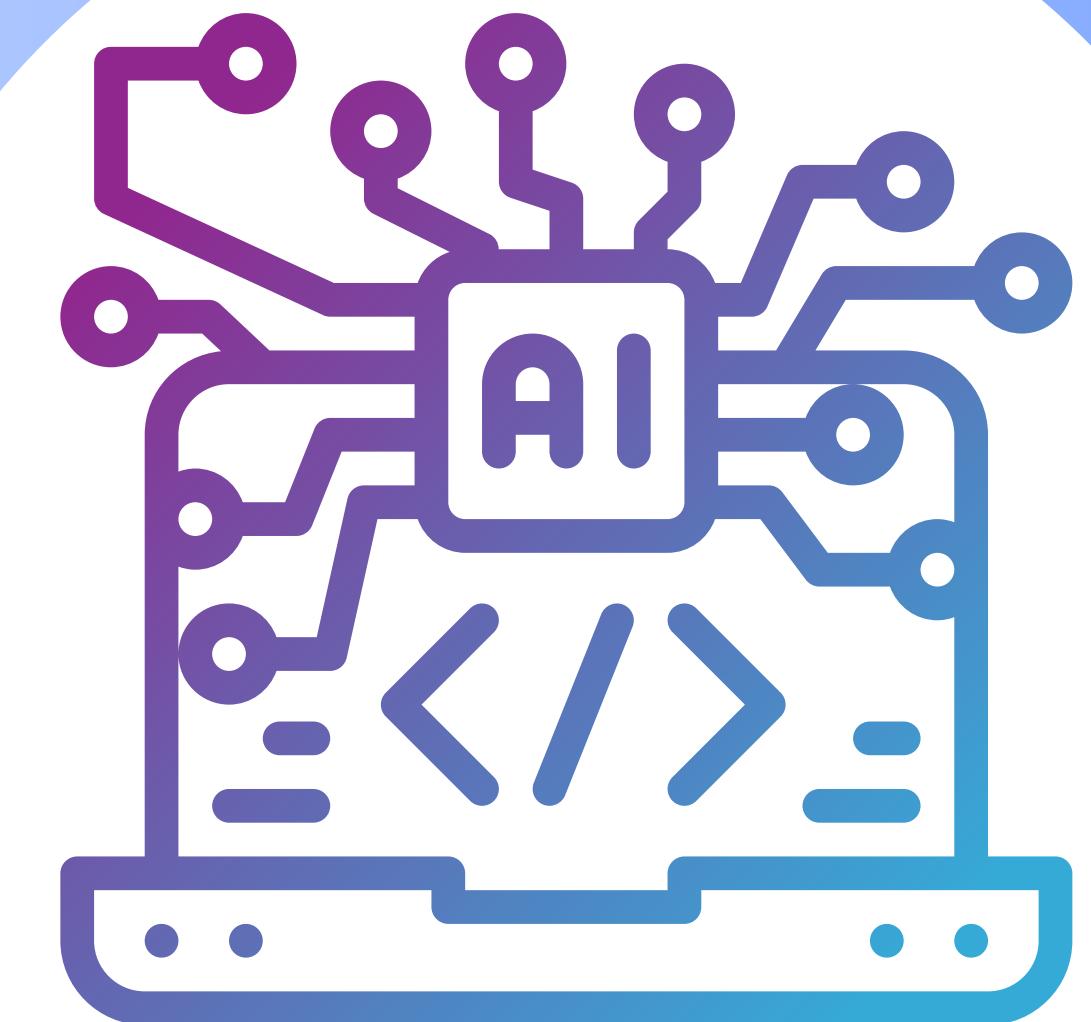
- Predicting Fraudulent Transactions
- Threshold Setting

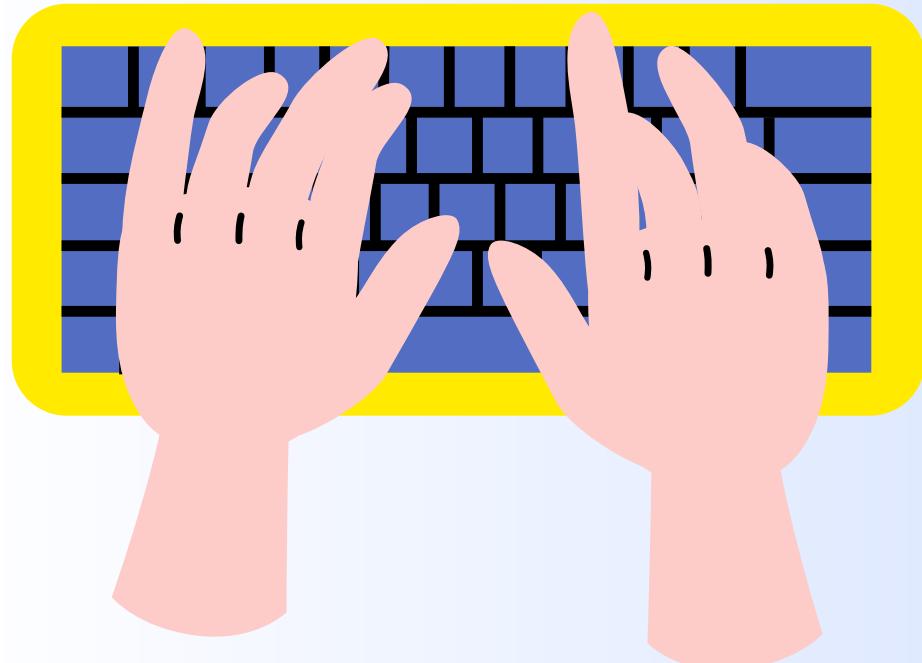


8-Evaluation

- Confusion Matrix
- Accuracy, Precision, Recall, F1 Score
- ROC Curve and AUC

9-Comparison: Supervised (KNN, Neural Networks) vs. Unsupervised (K-Means, PCA) techniques.





CASE STUDY

Github Repository :
<https://github.com/Farzam1372/Fraud-Transactions-Detection>

CONCLUSION

Compared to K-Means, both KNN and Neural Network have higher F1-scores, meaning they offer a better balance of precision and recall. This suggests KNN and Neural Network are likely to perform better in tasks where both correct identification (precision) and capturing all relevant instances (recall) are important.

K-Means, while having the highest precision, significantly underperforms in recall, leading to its lowest F1-score. This makes it a less desirable choice unless the task specifically prioritizes precision over recall.

	Model	Precision	Recall	F1-Score
0	KNN	0.946668	0.944162	0.944223
1	Neural Network	0.935350	0.934010	0.934078
2	K-Means	0.997284	0.473879	0.641325

REFERENCES

- Sahithi, G.L.; Roshmi, V.; Sameera, Y.V.; Pradeepini, G. Credit Card Fraud Detection using Ensemble Methods in Machine Learning. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 28–30 April 2022; pp. 1237–1241.
- Federal Trade Commission. CSN-Data-Book-2022. no. February 2023. Available online: https://www.ftc.gov/system/files/ftc_gov/pdf/CSN-Data-Book-2022.pdf
- UK Finance. Annual Report and Financial Statements 2022. Available online: <https://www.ukfinance.org.uk/annual-reports>
- Gupta, P.; Varshney, A.; Khan, M.R.; Ahmed, R.; Shuaib, M.; Alam, S. Unbalanced Credit Card Fraud Detection Data: A Machine Learning-Oriented Comparative Study of Balancing Techniques. Procedia Comput. Sci. 2023, 218, 2575–2584.
- Mondal, I.A.; Haque, M.E.; Hassan, A.-M.; Shatabda, S. Handling imbalanced data for credit card fraud detection. In Proceedings of the 2021 24th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 18–20 December 2021
- Khalid, A.R.; Owoh, N.; Uthmani, O.; Ashawa, M.; Osamor, J.; Adejoh, J. Enhancing Credit Card Fraud Detection: An Ensemble Machine Learning Approach. Big Data Cogn. Comput. 2024, 8, 6
- Shah, R. Introduction to k-Nearest Neighbors (kNN) Algorithm. Available online: <https://ai.plainenglish.io/introduction-to-k-nearest-neighbors-knn-algorithm-e8617a448fa8>
- Jadhav, S.D.; Channe, H.P. Comparative study of K-NN, naive Bayes and decision tree classification techniques. Int. J. Sci. Res. 2016, 5, 1842–1845
- Randhawa, K.; Loo, C.K.; Seera, M.; Lim, C.P.; Nandi, A.K. Credit card fraud detection using AdaBoost and majority voting. IEEE Access 2018, 6, 14277–14284
- Yee, O.S.; Sagadevan, S.; Malim, N.H.A.H. Credit card fraud detection using machine learning as data mining technique. J. Telecommun. Electron. Comput. Eng. 2018, 10, 23–27



The background features a large, semi-transparent white circle centered on the left side. Overlaid on it are several concentric circles in shades of purple, blue, and light blue, creating a layered effect. A second, smaller semi-transparent white circle is positioned on the right side, also with concentric circles overlaid.

THANK YOU!