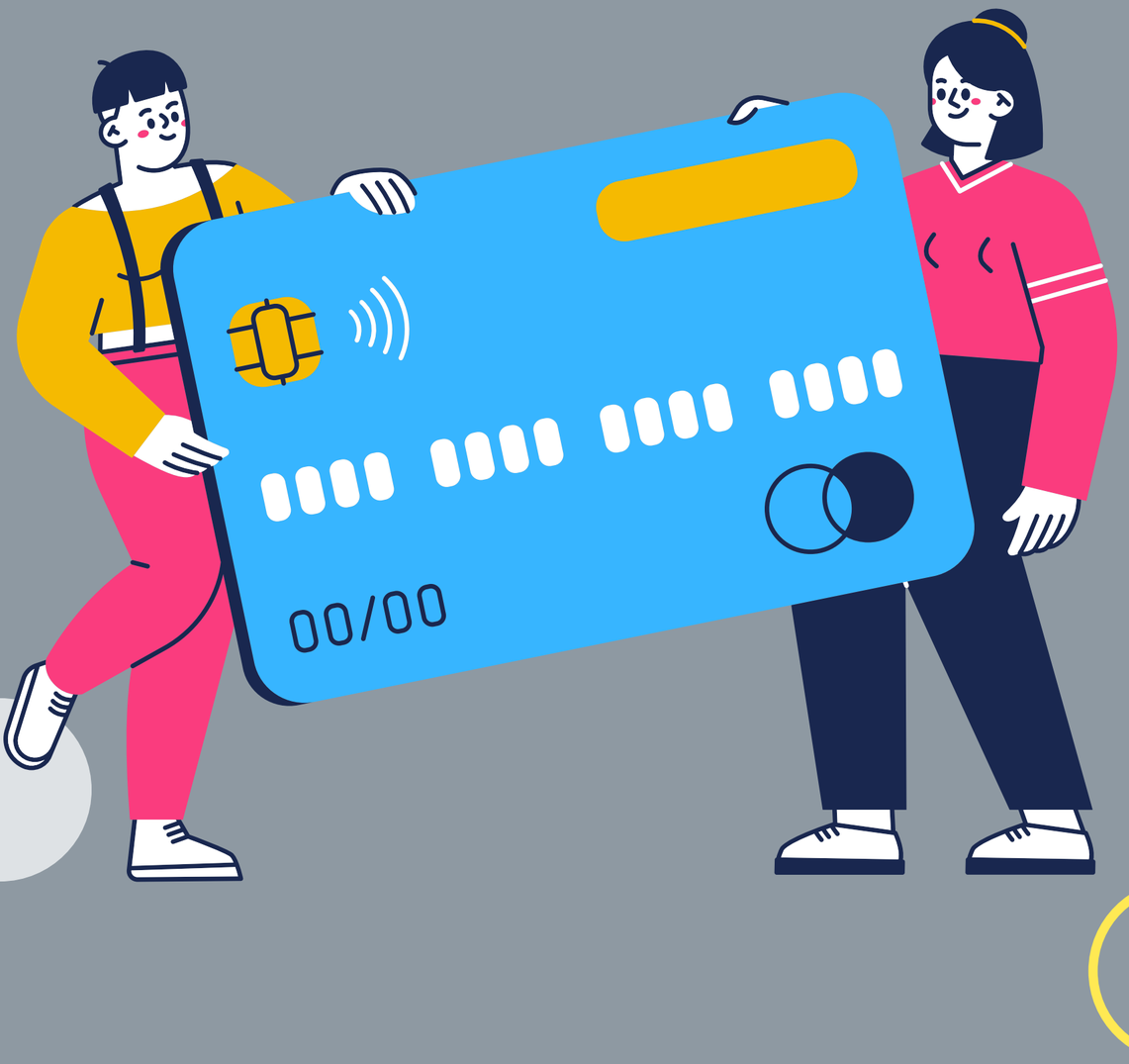
EXPLORING THE WORLD OF CREDIT CARD FRAUD

CREDIT CARD FRAUD DETECTION

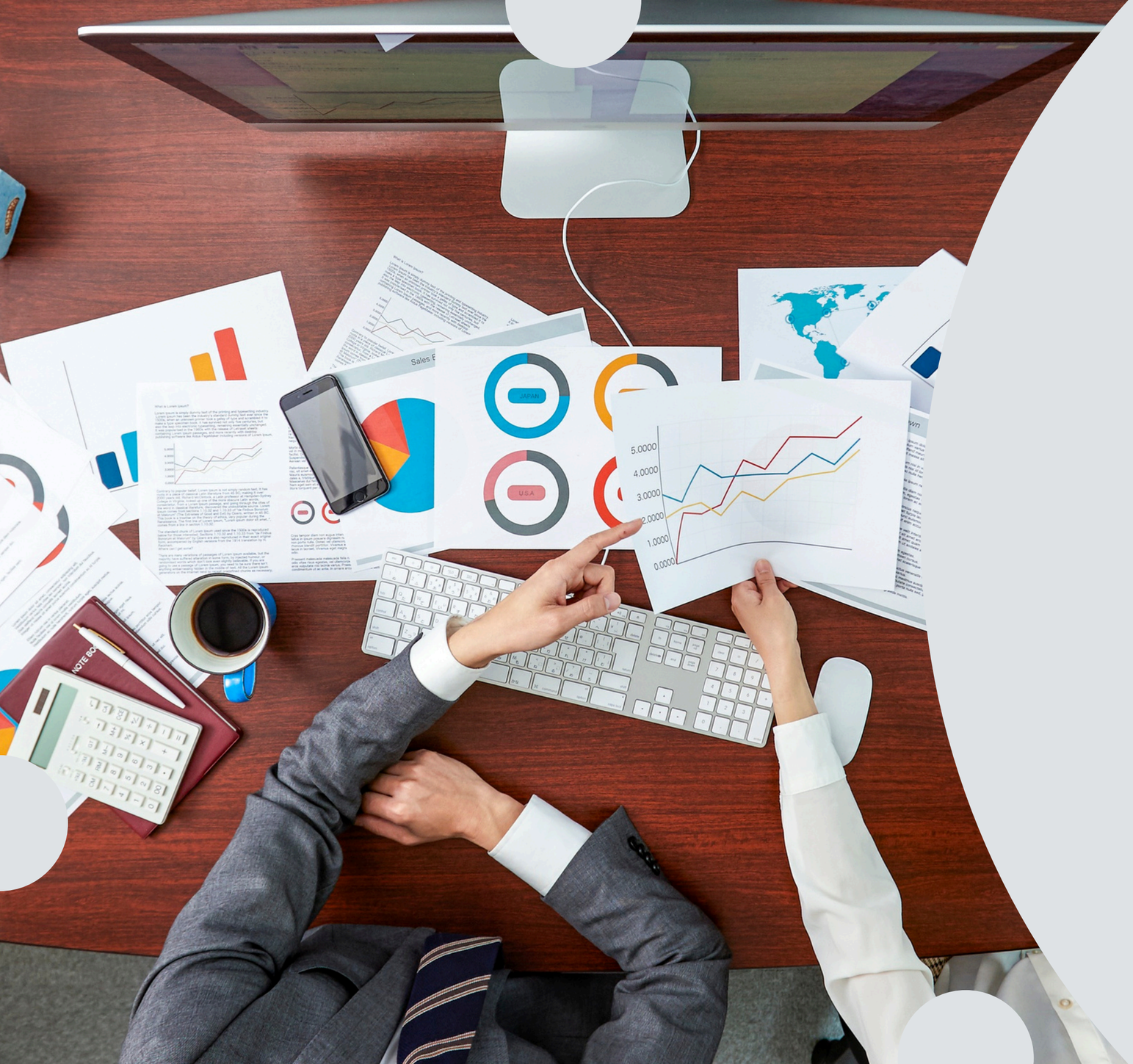
This presentation covers the use of machine learning to predict fraudulent transactions, leveraging a dataset of over 284,000 transactions to ensure robust fraud detection in financial technology.

PART OF CODSOFT DATA SCIENCE INTERNSHIP
BY NISHANT SHARMA



OBJECTIVE OF THE PROJECT

The goal of this project is to predict whether a credit card transaction is fraudulent or legitimate using machine learning techniques. By analyzing various features of the transaction data, we aim to build a model that accurately identifies fraudulent activities, thereby assisting financial institutions in minimizing losses due to fraud. The importance of timely detection cannot be overstated, as it protects consumers and maintains trust.



DATASET OVERVIEW AND IMBALANCE

The dataset comprises 284,807 transactions with features V1 to V28, along with Time and Amount. Each entry is classified as either legitimate (0) or fraudulent (1), highlighting a significant imbalance between classes. With fraud cases being rare, effective detection requires sophisticated techniques to address this disparity and ensure accurate predictions for legitimate transactions.

TOOLS AND LIBRARIES USED

DATA MANIPULATION

Pandas is essential for data manipulation and analysis, allowing for efficient handling of large datasets during the preprocessing stages of the project.

BOOSTING ALGORITHMS

XGBoost is a high-performance library designed for boosting algorithms, particularly effective for imbalanced datasets like credit card transactions.

MACHINE LEARNING

Scikit-learn provides powerful tools for machine learning algorithms, enabling model training and evaluation with various metrics for assessing performance.

PROJECT WORKFLOW OVERVIEW

1. Data loading is the first step where we import the dataset into our environment, ensuring we have access to all the necessary features for analysis.

2. Exploratory Data Analysis (EDA) follows, which allows us to visualize the data, understand relationships, and identify patterns or anomalies that may exist within the transactions.

3. After EDA, we apply SMOTE to balance the dataset, enhancing the model's ability to accurately classify transactions as legitimate or fraudulent through effective training.

MODEL COMPARISON: PERFORMANCE METRICS EXPLAINED

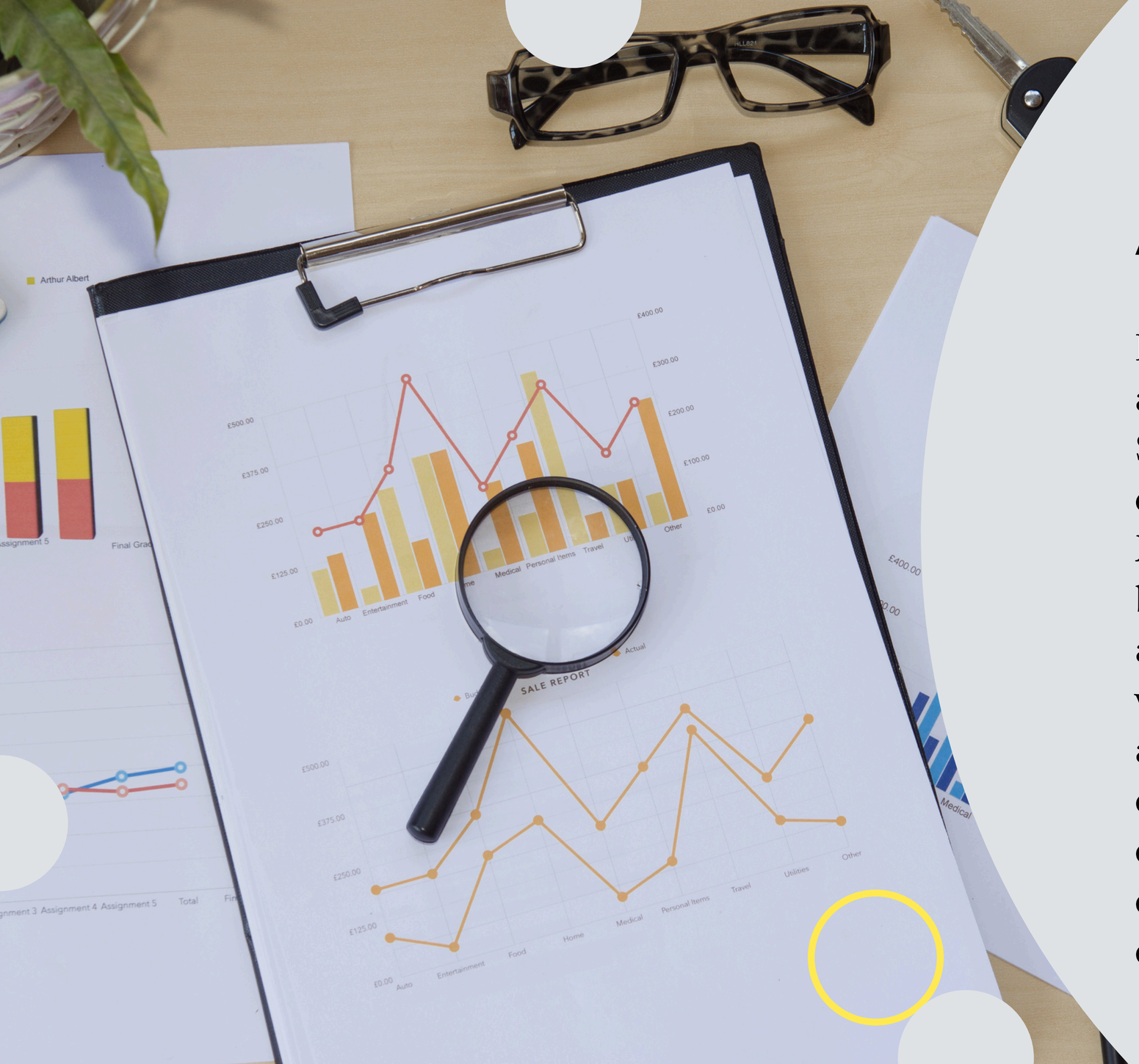
This page showcases the effectiveness of different algorithms.

PERFORMANCE METRICS FOR EACH ALGORITHM

- Logistic Regression: Precision of 0.85, Recall of 0.72, F1-score of 0.78.

INSIGHTS GAINED FROM METRICS

- Logistic Regression performs adequately, but lacks in recall.



KEY FINDINGS AND INSIGHTS

In this project, we effectively addressed the class imbalance using SMOTE, which significantly improved our model's performance. The XGBoost algorithm emerged as the best performer, showcasing superior accuracy and reliability. Additionally, we utilized SHAP for interpretability, allowing us to understand feature contributions to predictions, which enhances trust in the model's decisions and aids in further development for fraud detection.

THANK YOU FOR YOUR ATTENTION!

Dataset Used: kaggle
LinkedIn: Nishant Sharma
Git Hub: Nishantksh