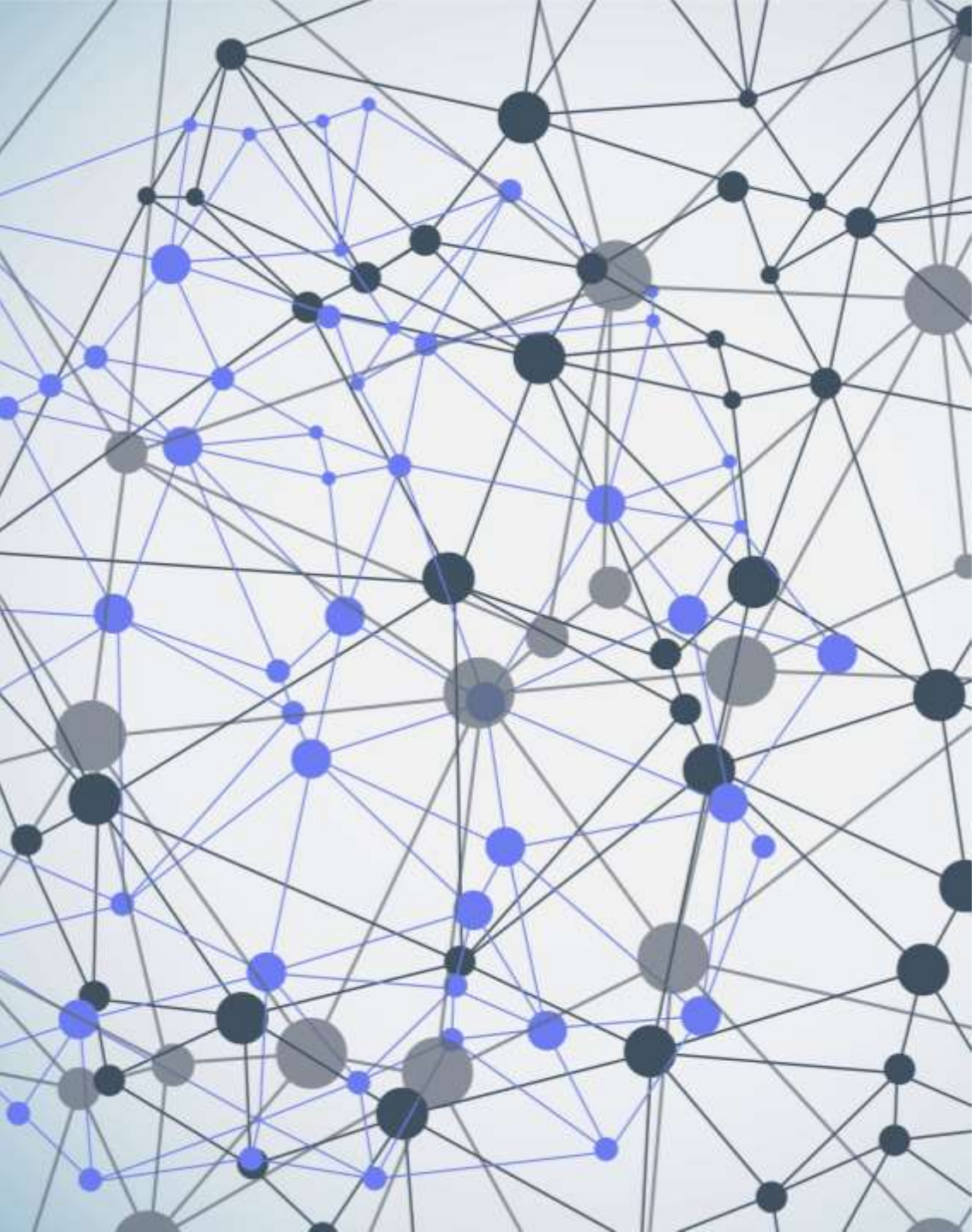
Teesside University

DISSERTATION VIVA PRESENTATION

PREDICTIVE MODELLING FOR CREDIT CARD FRUAD DETECTION USING MACHINE LEARNING

UMAR SHEHZAD B1514261

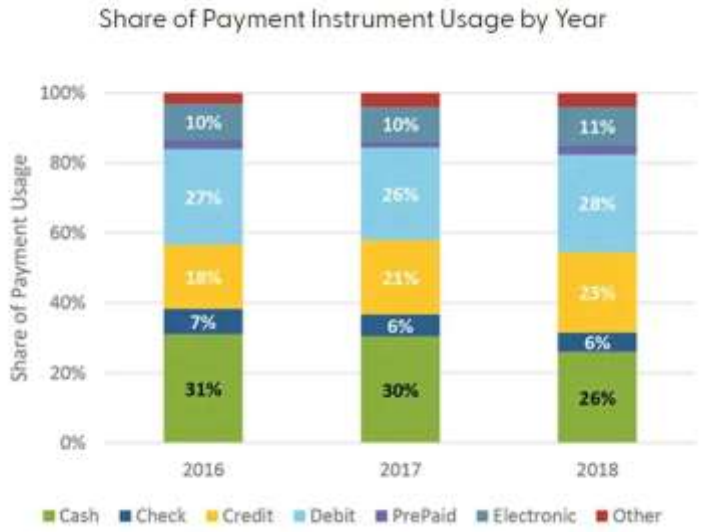


AGENDA

- Introduction
- Research Objectives
- Method
- Objective Solution
- Key Findings
- Key Achievements
- Conclusion
- Q&A Session

INTRODUCTION

- Credit card usage has surged globally, making it the second preferred payment method. However, this rapid expansion has led to a significant challenge in credit card fraud, resulting in significant financial losses and consumer trust loss. This highlights the need for robust systems.
- Financial fraud has significantly increased due to the rise of computer technology and the digital era. A recent study found that 65% of US adults have experienced credit card fraud, with online shopping potentially increasing the risk of fraud.
- This study proposes an innovative approach to credit card fraud detection using machine learning techniques. It aims to develop a predictive model that accurately distinguishes fraudulent transactions from legitimate ones, contributing to the advancement of secure financial transactions and a safer financial landscape.



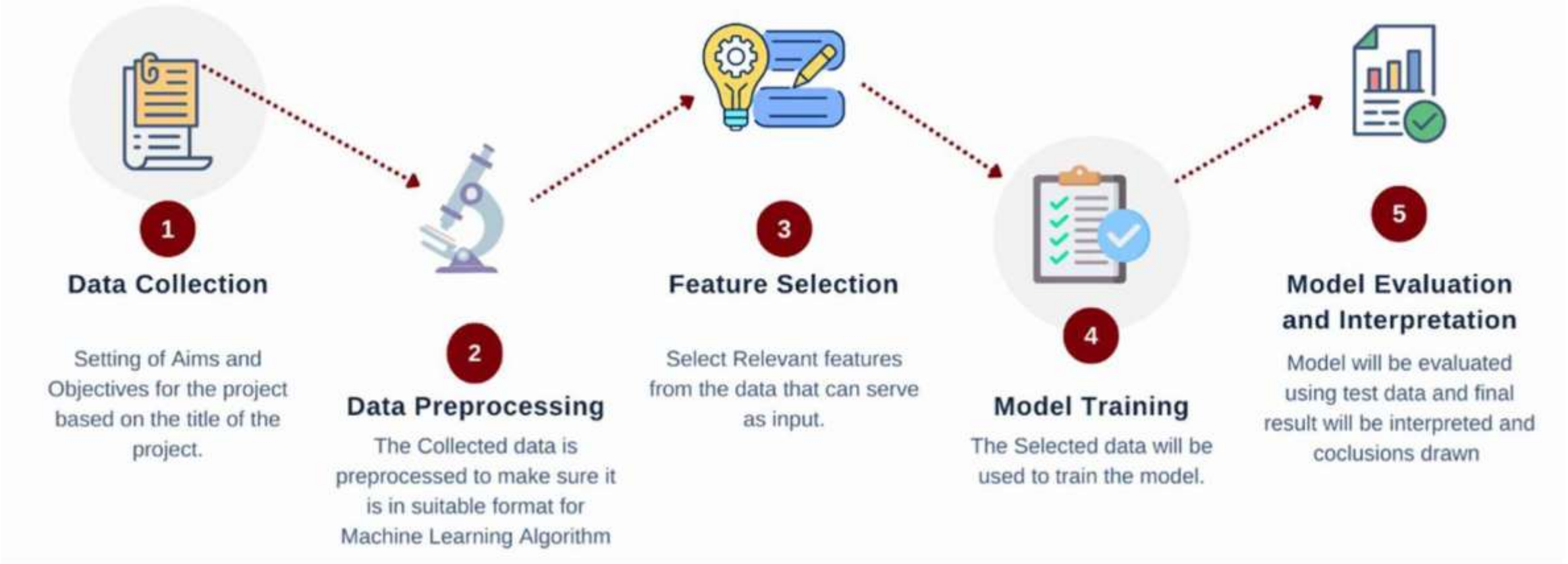
RESEARCH OBJECTIVES

To explore the effective utilization of data science and machine learning techniques for the detection of credit card fraud. The specific focus will be on:

Identifying the most suitable machine learning algorithms and approaches	Enhancing the overall detection accuracy of fraudulent activities in credit card transactions.

This investigation aims to:- Assess the current landscape of data science and machine learning applications in credit card fraud detection.- Identify challenges and limitations in existing methods.- Provide recommendations and insights for improving the efficacy of fraud detection systems in the financial domain.

METHOD



1

Data Collection

Setting of Aims and Objectives for the project based on the title of the project.

2

Data Preprocessing

The Collected data is preprocessed to make sure it is in suitable format for Machine Learning Algorithm

3

Feature Selection

Select Relevant features from the data that can serve as input.

4

Model Training

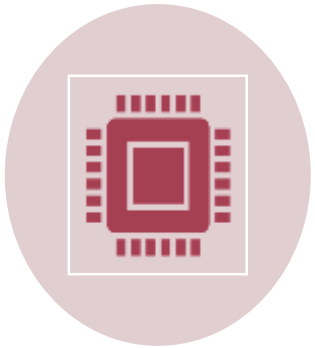
The Selected data will be used to train the model.

5

Model Evaluation and Interpretation

Model will be evaluated using test data and final result will be interpreted and coclusions drawn

OBJECTIVES SOLUTION



The study explores the use of data science and machine learning in credit card fraud detection, highlighting how these technologies are transforming fraud prevention strategies. Mastercard uses machine learning to predict transaction risks, improving detection and reducing false declines. Challenges include imbalanced data, confidentiality laws, incomplete information, and overlapping data. The findings suggest continuous exploration of feature engineering techniques, hyperparameter tuning, ensemble approaches and real-world validation.

The research used a systematic approach to identify suitable machine learning algorithms for classification problems. After meticulous testing, the Random Forest algorithm emerged as the most effective solution, achieving a balanced accuracy of 96%. This strategic identification is crucial for developing an advanced credit card fraud detection system.

The dataset was transformed to improve accuracy in detecting fraudulent activities in credit card transactions. This involved aligning input features with suitable data types and tuning the model for optimal results. This synergistic approach ensures a robust credit card fraud detection system.

A robust solution was developed using data science and machine learning methodologies. The preprocessing phase involved removing duplicates, cleaning data, and transforming diverse data types. Outliers were addressed to ensure data integrity. The pivotal step was converting data into a suitable input for machine learning algorithms.

KEY FINDINGS

The analysis reveals peak fraud occurrences during late-night transactions, lower incidence during early morning hours, monthly variations, and a weekly distribution of moderate activity on Mondays and higher rates on Tuesdays.

Geospatial visualization of fraudulent transactions across geographical locations reveals hotspots, highlighting areas with high activity. This helps identify areas requiring targeted attention for fraud prevention, enhancing proactive measures and resource allocation.

The study analyzes transaction distribution across states, identifying top cities and states with the highest fraudulent activity. Visual aids help identify regions with higher or lower fraudulent occurrences, aiding in strategic decision-making and developing geographically tailored fraud prevention strategies.

The study analysed transaction characteristics and identified top merchants with the highest number of fraudulent transactions. 'grocery_pos' and'shipping_net' were identified as the most susceptible categories. This information helps develop targeted fraud detection measures and focuses resources on these merchants. The analysis also examined the correlation between transaction amounts and fraudulent activities across various job roles, identifying outliers and comparing median amounts, thereby aiding in fraud prevention strategies.

The Random Forest model outperformed four other machine learning models for credit card fraud detection, with a 96% accuracy rate. The model's high precision and recall, along with its robust F1-score and ROC AUC, suggest its potential for diverse application contexts.

KEY ACHIEVEMENTS

The study developed a sophisticated credit card fraud detection system using advanced algorithms, including the Random Forest model, which achieved a balanced accuracy of 96%.

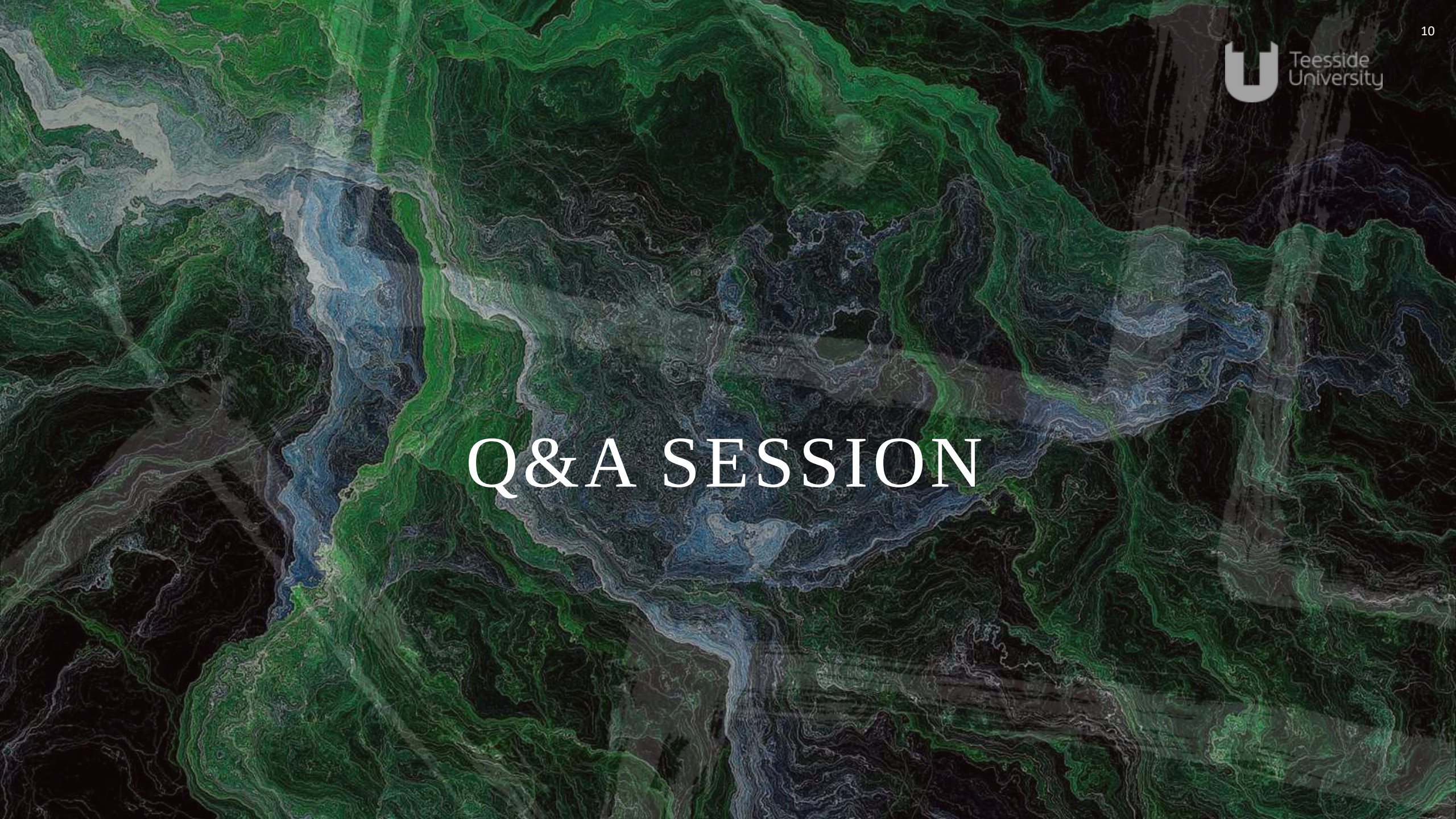
The system also underwent meticulous data preprocessing, addressing imbalanced datasets and removing outliers. The Random Forest model demonstrated superior precision and recall, reducing false positives and false negatives.

The study also provided strategic recommendations for system enhancement, including continuous feature engineering, hyperparameter tuning, and ensemble approaches.

The model's real-world validation was proposed to ensure its applicability and generalization in the ever-evolving financial landscape.

CONCLUSION

- The study explores the field of credit card fraud detection, focusing on data science and machine learning applications. It reveals a significant achievement in the development of a system using advanced algorithms, with a balanced accuracy of 96%. The system's robustness is attributed to meticulous data preprocessing strategies. The study also offers actionable recommendations for enhancing fraud detection systems, including feature engineering, hyperparameter tuning, and ensemble approaches. The findings provide valuable guidance for practitioners seeking reliable approaches to combat credit card fraud in a complex digital landscape.



Q&A SESSION