CERTIFICATE

This is to certify that the Real-Time Research Project Report entitled

'Detection of Account Takeover Using Machine Learning Models' being submitted by [Student Names and Roll Numbers]

in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in

Computer Science and Engineering to the Jawaharlal Nehru Technological University, Hyderabad,

is a record of bonafide work carried out under my guidance and supervision during the Academic Year 2023 - 24.

The results embodied in this thesis have not been submitted to any other University or Institute for the

award of any other degree or diploma.

ABSTRACT

This project focuses on detecting account takeovers using machine learning techniques.
The dataset consists of login records, IP addresses, device types, and authentication results. Various
machine learning models were trained and evaluated, including Logistic Regression, Decision Tree,
Random Forest, AdaBoost, and XGBoost. Feature engineering was performed by encoding categorical
variables, converting IP addresses into numerical values, and extracting login time information. The models
were assessed using accuracy, AUC-ROC, and precision-recall metrics. The results highlight the effectiveness
of ensemble models like XGBoost in detecting suspicious login activities.

TABLE OF CONTENTS

1. INTRODUCTION

Account takeovers pose a significant cybersecurity threat, leading to unauthorized access and fraudulent transactions. Detecting such activities requires robust machine learning models trained on

real-time login data.

2. LITERATURE SURVEY

Existing research highlights various approaches to fraud detection, including statistical methods and deep learning techniques. Traditional rule-based systems often fail to detect new attack

patterns, making machine learning models a preferable alternative.

3. ANALYSIS AND DESIGN

Dataset Description:

- Login Timestamp: Time of login
- User ID: Unique identifier for each user
- IP Address: Source of login attempt
- Device Type: Desktop/Mobile
- Login Successful: Boolean flag indicating success or failure
- Is Attack IP: Label indicating if an IP is associated with attacks
- Is Account Takeover: Target variable

Data Preprocessing:

- Handling missing values
- Encoding categorical variables
- Feature engineering (e.g., extracting login hour)

4. IMPLEMENTATION

Machine Learning Models Used:

- Logistic Regression
- Decision Tree
- Random Forest
- AdaBoost
- XGBoost

Model Training and Evaluation:

- Train-test split (80%-20%)
- Feature scaling and encoding
- Pipeline creation using scikit-learn
- Hyperparameter tuning

5. TESTING AND RESULTS

Model Performance Metrics:

Model	Accuracy	AUC Score
Logistic Regression	99.99%	0.867
Decision Tree	99.99%	0.625
Random Forest	99.99%	0.857
AdaBoost	99.99%	0.978
XGBoost	99.99%	0.985

6. CONCLUSION

This study demonstrated the effectiveness of machine learning models in detecting fraudulent
account takeovers. While traditional models like Logistic Regression provided decent results, ensemble models,
especially XGBoost, significantly outperformed them. Future work may involve integrating deep learning techniques
and real-time detection systems for improved security.

7. REFERENCES

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

2. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.

3. Scikit-learn Documentation. https://scikit-learn.org/stable/