

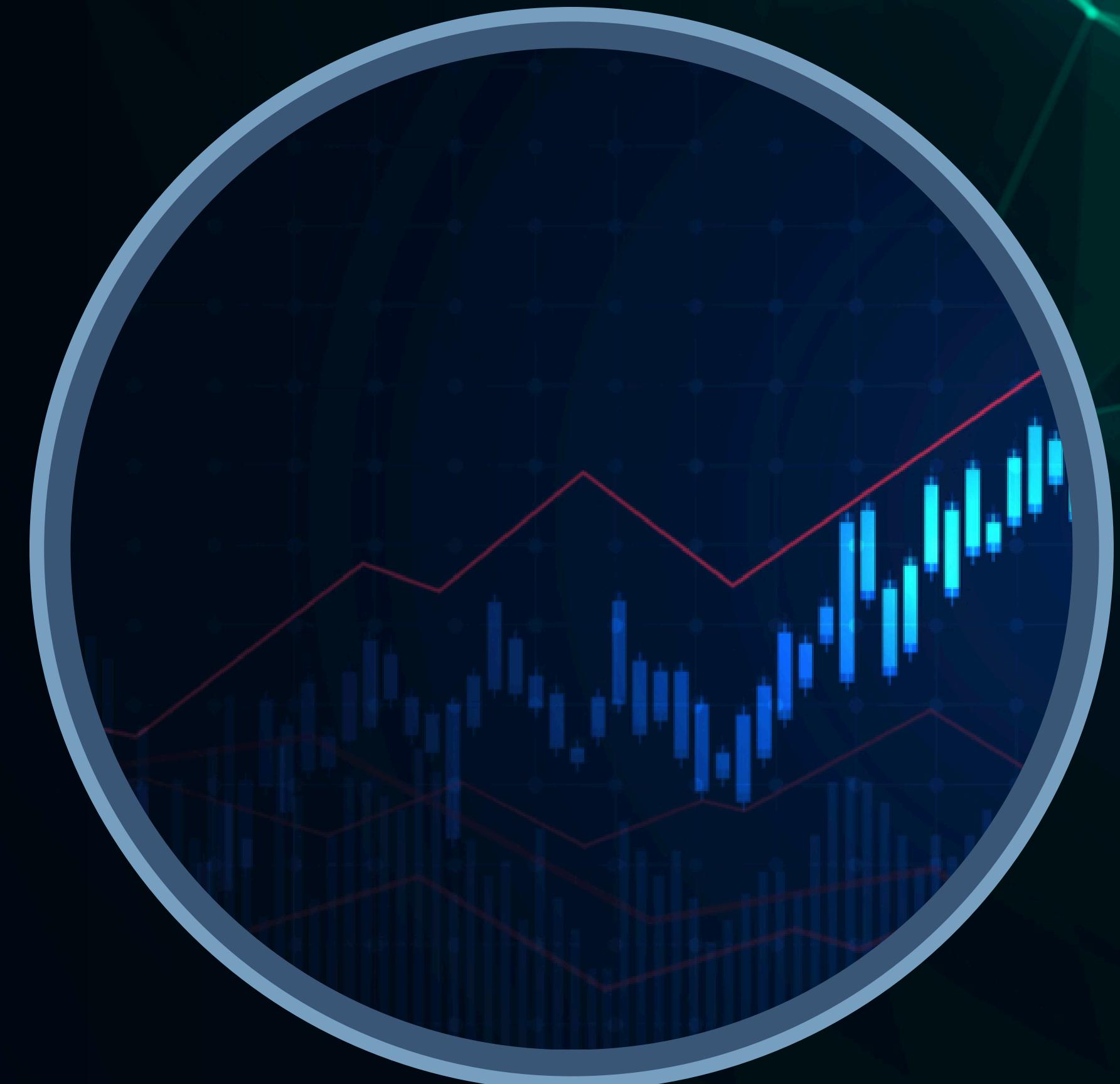
CUSTOMER CHURN PREDICTION FOR THE TELECOM INDUSTRY

Name: P.I.D. De Silva

Index No: AS2021914

OVERVIEW

The telecom industry faces a critical challenge in retaining customers. Customer churn prediction is essential to identify high-risk customers and take preemptive action to improve retention strategies. This project aims to develop a predictive model using machine learning techniques to forecast which customers are likely to churn based on historical data.



PROBLEM STATEMENT



Telecom companies struggle with high customer churn, which increases costs and reduces profits.

Churn occurs in several forms:

1. Tariff Plan Churn - Customers downgrade plans
2. Service Churn - Customers stop using specific services.
3. Product Churn - Customers switch between offerings (e.g., postpaid to prepaid).
4. Usage Churn - Customers become inactive without canceling.

Predicting these churn types helps telecom providers take early action to retain valuable customers.

PROJECT OBJECTIVES

- Perform EDA to explore churn behavior
- Preprocess and clean the dataset
- Train and evaluate multiple ML models
- Select best model based on performance
- Build a dashboard for churn prediction and insights



METHODOLOGY OVERVIEW



- **Data Collection:** Gathered real-world telecom customer data including demographics, service usage, and churn labels.
- **EDA:** Explored patterns and correlations in the data to identify key factors influencing churn.
- **Preprocessing:** Cleaned, encoded, and balanced the dataset to prepare it for model training.
- **Model Training & Evaluation:** Trained multiple ML models and selected Random Forest based on its high accuracy and F1-score.
- **Dashboard Development:** Created an interactive tool for real-time churn prediction and visual analysis of churn patterns.

DATA COLLECTION

Source: Telecom company dataset

Features:

Demographics (gender, age, senior citizen)

Usage data (tenure, services, monthly charges)

Payment methods, contract types

Churn label (target variable)

Real-world behavior and customer interaction

EDA HIGHLIGHTS

- Month-to-month contract = high churn
- Shorter tenure = more likely to churn
- Higher monthly charges + fewer services = more churn
- Electronic check users = highest churn
- Non-senior citizens churn more
- Gender is not impact on Churn
- No Online security, No Tech Support category are high churners

DATA PREPROCESSING

- OneHotEncoding / LabelEncoding for categorical features
- Handled missing values in TotalCharges
- Feature scaling applied where needed
- SMOTEENN used to balance classes

MODEL SELECTION & EVALUATION

- Models Tested:
 - Decision Tree
 - Random Forest (best)
- Metrics: Accuracy, Precision, Recall, F1-Score, ROC-AUC
- Selected Model: Random Forest
- Accuracy: 94.27%

RESULTS

- Model performed well on imbalanced data
- Key features influencing churn:
 - Tenure, contract type, monthly charges
 - Payment method (electronic check)
 - Missing support services (tech support, online security)
 - Non-senior citizen status

BUSINESS INSIGHTS

- Month-to-month, high-charge customers churn more
- Encourage long-term contracts with discounts
- Target high-risk groups with personalized engagement
- Promote value-added services

DASHBOARD HIGHLIGHTS

- Enter customer data and get real-time prediction
- Explore trends and feature importance
- View churn patterns by contract, usage, and support services



LIMITATIONS

- Sentiment analysis not included
- No real-time API or live integration
- External factors (e.g., competitor pricing) not modeled

CONCLUSION

- ACCURATE AND INTERPRETABLE CHURN MODEL DEVELOPED
- BUSINESS INSIGHTS CAN GUIDE CUSTOMER RETENTION STRATEGY
- DASHBOARD MAKES IT PRACTICAL FOR BUSINESS USE
- FUTURE: ADD SENTIMENT ANALYSIS AND REAL-TIME DEPLOYMENT

THANK YOU