Introduction

Title: Syriatel Telecom Phase 3 Project
Subtitle: Predicting Customer Churn
Presenter: Phoebe Wawire
Class: DSF-PT7

Problem Statement

- **Objective:**
  Predict whether a customer of SyriaTel will stop doing business with them (churn) based on various factors.

- **Audience:**
  Telecom businesses interested in understanding potential revenue loss due to customer churn.

- **Target Variable:**

- Categorical with two classes:

  - 1 = Churn

  - 0 = No Churn

Data Collection

- **Libraries Used:**
  - ✓ pandas, numpy, seaborn
  - ✓ sklearn: LogisticRegression, confusion_matrix, etc.
- imbalanced-learn: SMOTE, NearMiss
- **Dataset Details:**
- Dimensions: 3333 instances, 21 variables
- Types of Variables:
  - Boolean: 1
  - Floating Point: 8
  - Integer: 8
  - Object (String): 4
- **Target Variable: Churn**

Data Preparation

Data Types & Categories:

▶ Categorical: state, phone number, international plan, voice mail plan

▶ Numerical: Floating point, Integer

▶ Boolean: 1 column

Steps Taken:

▶ Converted categorical variables to numeric

▶ Checked and imputed missing values and outliers

▶ Feature engineering and scaling applied

Data Modelling using Logistic Regression

Metrics:

- True Negatives (TN): 566
- False Positives (FP): 0
- False Negatives (FN): 91
- True Positives (TP): 10

Performance:

- Precision (Churn): 0.77
- Recall (Churn): 0.10
- F1-Score (Churn): 0.18

Data Modelling using Logistic Regression continuation

Overall Accuracy: 86%

Insights:

▶ High accuracy for non-churn (class 0)

▶ Low recall for churn (class 1)

Recommendation:

▶ Collect more data for churn class

Data Modelling using Decision Trees

Metrics:

- True Negatives (TN): 566
- False Positives (FP): 51
- False Negatives (FN): 42
- True Positives (TP): 59

Performance:

- Precision (Churn): 0.55
- Recall (Churn): 0.59
- F1-Score (Churn): 0.57

Data Modelling using Decision Trees continuation

Overall Accuracy: Similar to Logistic Regression

▶ **Insights:**

▶ Better performance on churn class compared to Logistic Regression

▶ High precision and recall for non-churn (class 0)

Recommendation:

▶ Collect more data and explore additional modeling techniques

Conclusion

Comparison:

- **Decision Tree:** Outperforms Logistic Regression in predicting churn with better recall and F1-Score.

- **Logistic Regression:** Performs well on non-churn but struggles with churn.

Key Findings:

- Decision Tree's flexibility helps in better classifying churn.

- Logistic Regression showed a performance drop for churn class.

Next Steps:

- Validate Decision Tree with more data or cross-validation.

- Improve data collection for churn cases.

Recommendations

For Model Improvement:

▶ **Data Collection:** Increase data for churn cases to balance the dataset.

▶ **Model Enhancement:** Explore advanced techniques and ensure robust validation.

For Business:

▶ Focus on understanding churn customers using all features.

▶ Develop strategies to reduce churn clients from moving to competition by solving the raised concerned within set timelines, have a 24 hour contact center for quick customer issues sorting e.t.c.

Questions & Discussion

- **Thank You!**
  Questions and feedback are welcome