Welcome to the Presentation!

Objective:

- Addressing Revenue Loss from Customer Churn
- Welcome esteemed stakeholders to today's presentation.
- We are here to discuss a crucial initiative undertaken by SyriaTel.
- The primary goal: Addressing revenue loss attributed to customer churn.

About SyriaTel

- SyriaTel is a leading telecommunications company in the industry.
- With a strong presence, our aim is to provide top-tier services.
- Today, we delve into a critical analysis to curb revenue loss.

Business Understanding

Objective:

- Build a classifier to predict customer churn for SyriaTel, a telecommunications company.

Problem Type:

Binary Classification

- Predict whether a customer is likely to discontinue their services soon.

•**Audience**:

SyriaTel, the telecommunications company, aiming to minimize revenue loss due to customer churn.

•**Business Goal**:

Reduce financial losses attributed to short-lived customer relationships by identifying predictable patterns of churn.

Data Pre-processing

Dataset: Customer information from SyriaTel's telecommunications services.

Data Cleaning:

- Checked for missing values to ensure data integrity.

Feature Engineering:

- Calculated total charges for day, evening, night, and international usage.

Categorical Variables:

- Encoded categorical variables ('state', 'international plan', 'voice mail plan') using one-hot encoding.

Irrelevant Features:

- Removed 'phone number' column.

Logistic Regression Model

Model Selection:

- Chose Logistic Regression as the baseline model.

Rationale:

- Logistic Regression is selected for its simplicity and interpretability.

Performance Metrics: The Logistic Regression model yielded the following performance results:

- **Accuracy**: 86.06%
- **Precision**:
  - For non-churn cases: 86%
  - For churn cases: 83%
- **Recall**:
  - For non-churn cases: 100%
  - For churn cases: 10%
- **F1-Score**:
  - For non-churn cases: 92%
  - For churn cases: 18%

Interpretability: One of the notable strengths of Logistic Regression is its ability to provide clear insights into the importance of features, aiding in the understanding of the underlying factors influencing the predictions.

Random Forest Model

- **Model Selection**:
  - Chose Random Forest for its ability to capture complex, non-linear relationships and interactions.

Performance Metrics: The Random Forest model demonstrated exceptional performance:

- **Accuracy**: 93.25%
- **Precision**:
  - For non-churn cases: 93%
  - For churn cases: 97%
- **Recall**:
  - For non-churn cases: 100%
  - For churn cases: 57%
- **F1-Score**:
  - For non-churn cases: 96%
  - For churn cases: 72%

Interpretation: Random Forest's strength lies in its capacity to discern intricate patterns and interactions in the data. It provides robust insights into the factors influencing customer churn.

Rationale for Using Machine Learning

Why Machine Learning?

- **Complex Relationships**:
  - Traditional statistical methods may struggle with capturing intricate, non-linear patterns within the data.
- **Telecom Industry Complexity**:
  - Numerous interrelated factors contribute to customer churn, often exhibiting non-linear dependencies.

Advantages of Machine Learning:

- **Random Forest**:
  - Demonstrates exceptional ability to discern complex patterns and interactions.
- **Enhanced Predictive Power**:
  - Ability to capture subtle nuances in customer behaviour leads to more accurate churn predictions.
- **Complementary Models**:
  - Models like K-Nearest Neighbours (KNN) and Naive Bayes provide additional perspectives for comprehensive insights.

Results and Classification Metrics

- **Key Metrics Considered**:
  - **Precision**:
    - Proportion of correctly predicted churn cases out of all positive predictions.
  - **Recall**:
    - Ability of the model to identify all actual churn cases, minimizing potential misses.
  - **F1-Score**:
    - Harmonic mean of precision and recall, providing a balanced assessment of performance.
  - **Accuracy**:
    - Overall correct predictions out of the total.

- **Implications for SyriaTel**:
  - **Precision**:
    - Identifying high-risk customers, crucial for targeted retention strategies.
  - **Recall**:
    - Ensuring no potential churner goes unnoticed, minimizing revenue loss.
  - **F1-Score**:
    - Balancing between identifying potential churners and minimizing false alarms.

Limitations

- **Subset Accuracy**:
  - Certain customer segments may exhibit less accurate predictions due to unique behaviours not well-represented in the training data.

- **Evolution of Customer Behaviour**:
  - The model's performance may need periodic retraining or fine-tuning as customer behaviour evolves over time.

Recommendations and Next Steps:

Continuous Monitoring:

- Implement a robust monitoring system to track the model's performance in real-time.

Periodic Retraining:

- Regularly update the model with new data to adapt to changing customer patterns.

Feedback Loop with Experts:

- Establish a feedback loop with domain experts to review model predictions and refine retention strategies.

Explore Advanced Techniques:

- Investigate advanced modelling techniques to further improve churn prediction accuracy.