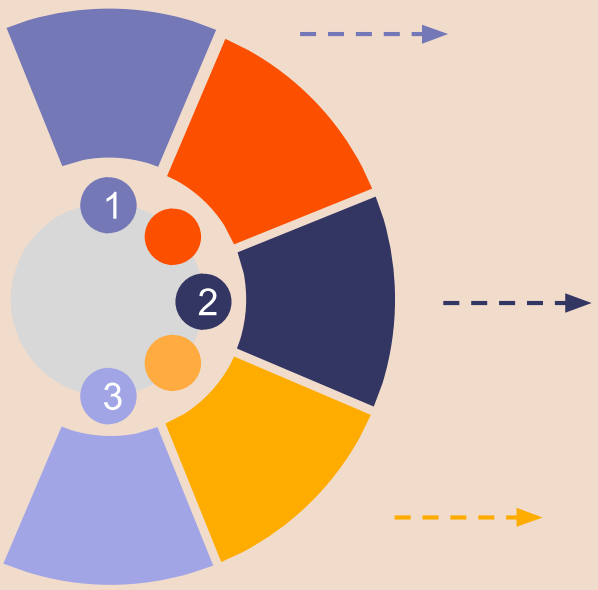
Classification Models to Predict Customer Churn for SyriaTel, a Telecommunications Company

Presentation By Denis Omondi Ochieng

Table of contents



01

Beginning

About the Project

01 Overview

- The project, focusses on developing classification models to predict customer churn for SyriaTel, a telecommunications company.
- Churn prediction is crucial for businesses as it helps them identify customers at risk of leaving and take proactive measures to retain them.
- Our goal is to explore whether there are predictable patterns in the data that can help SyriaTel reduce customer churn and the associated financial losses.

02 Business and Data Understanding

- Our analysis started with a deep dive into the data, examining various features such as account length, international plan, voice mail plan, and more. We looked for patterns and relationships that could help us predict churn.
- The data is clean and does not contain null values hence fit for training our models.

Background

SyriaTel is facing a challenge in retaining its customers, leading to financial losses and a potential threat to its market position. The business needs to develop a predictive model to identify customers who are at a high risk of churning, enabling proactive actions to retain these customers and reduce overall churn rates.

Once potential churners are identified, SyriaTel can implement tailored customer retention strategies. These strategies could include targeted promotions, enhanced customer support, or personalized offers to retain high-risk customers and improve overall customer satisfaction.

Customer churn, or the loss of customers, is a significant concern for SyriaTel. Churn impacts the company's revenue, profitability, and market share. To address this issue, the company aims to develop a solution that can predict which customers are likely to churn in the near future.

02

Middle

Key Arguments Related to the Choice of Models

Brief

- We built three different classification models: Logistic Regression, Random Forest, and a Tuned Random Forest.
- These models were chosen because they are well-suited for binary classification tasks.
- While the technical details of model implementation aren't necessary for a non-technical audience, what's important is that we used these models to predict whether a customer is likely to churn or stay with SyriaTel.

2.0 Random Forest (More-Complex Model):

Complex Patterns: Random Forest is a more complex model that can capture complex patterns and interactions among features. In cases where customer churn is influenced by intricate relationships, Random Forest can outperform simpler models.

Handling Non-linearity: Random Forest can handle non-linear relationships effectively. Customer behavior leading to churn often involves non-linear patterns, making this model a valuable choice.

Imbalanced Data Handling: Random Forest has built-in mechanisms for handling imbalanced datasets. It can assign different weights to classes, which is crucial for predicting churn, a typically imbalanced problem.

1.0 Baseline Model (Logistic Regression):

Simplicity and Interpretability: Logistic Regression was chosen as the baseline model due to its simplicity and high interpretability. It provides a clear understanding of how each feature impacts the prediction of churn, which is valuable for business stakeholders.

Initial Performance Benchmark: It serves as an initial benchmark to assess how well a straightforward model performs in predicting customer churn. This aids in setting a baseline level of model performance.

Continuation

3.0 Tuned Random Forest (Hyperparameter-Tuned Version of Random Forest):

Hyperparameter Tuning: Hyperparameter tuning was applied to the Random Forest model to optimize its performance. By systematically searching for the best combination of hyperparameters, we aimed to improve its predictive power.

Balanced Precision and Recall: The tuned Random Forest model achieves a balance between precision and recall. It captures a significant portion of actual churn cases while maintaining high precision, making it a valuable choice for the business.

Best Overall Performance: The hyperparameter-tuned Random Forest model achieved the highest accuracy, precision, and F1-score among the models. Its performance, when fine-tuned, outperformed the other models in our specific problem context.

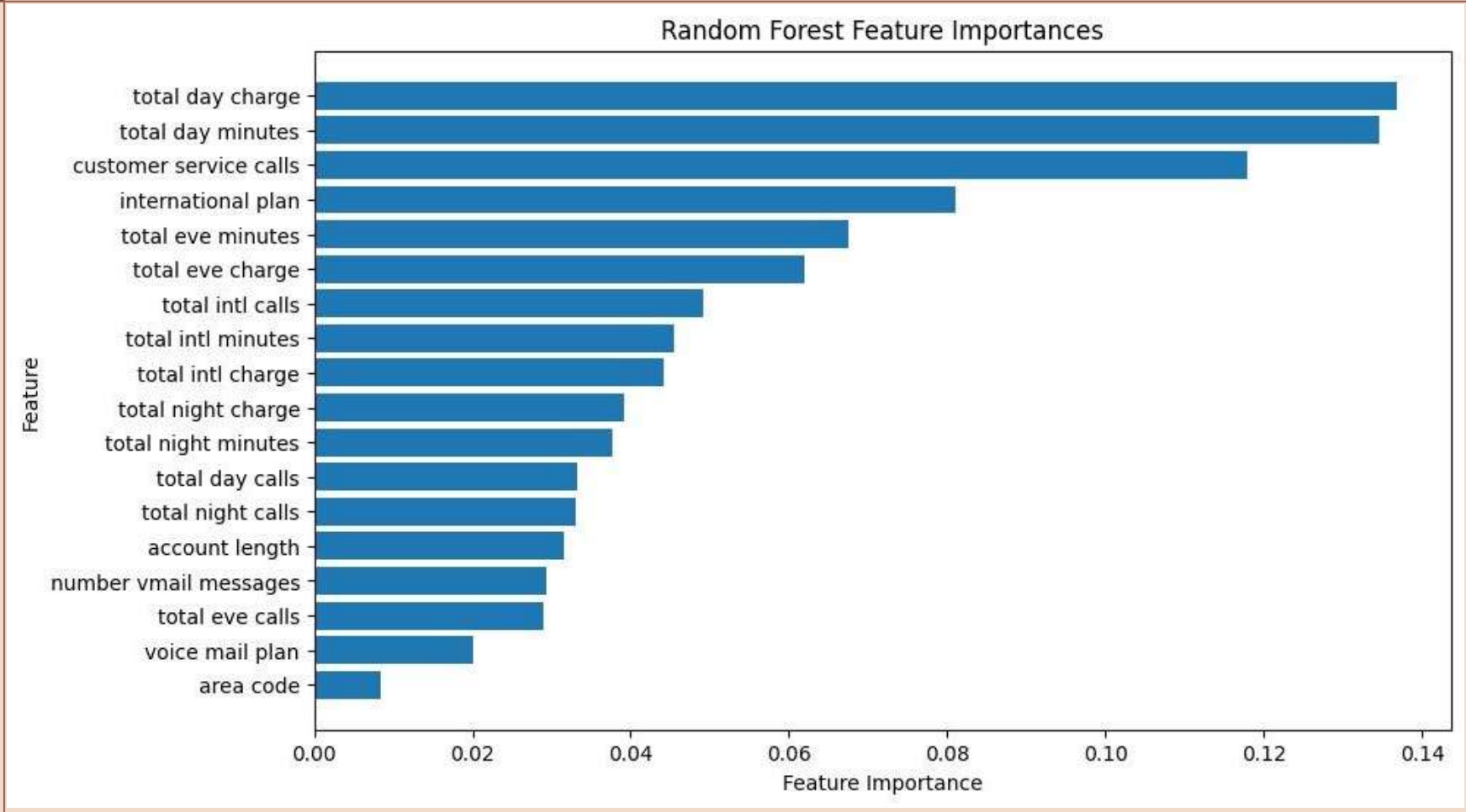
Model Evaluation

Model evaluation ensures we choose the best model, optimize its performance, and ensure it aligns with business objectives. It plays a fundamental role in the development and deployment of machine learning models in real-world applications.

We evaluated the models using various metrics, such as accuracy, precision, recall, and F1-score.

These metrics help us understand how well the models are performing. For example, accuracy tells us the percentage of correct predictions, while precision tells us how many of the positive predictions are accurate.

Recall informs us about the percentage of actual positive cases the model correctly predicted. The F1-score is a balance between precision and recall. For specific numbers, you can refer to the respective sections for each model.


Random Forest Feature Importances
total day charge
total day minutes
customer service calls
international plan
total eve minutes
total eve charge
total intl calls
total intl minutes
total intl charge
total night charge
total night minutes
total day calls
total night calls
account length
number vmail messages
total eve calls
voice mail plan
area code
Feature
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Feature Importance

The Best Model Among The Three

Summary

The best-performing model among the three is Model 3: Tuned Random Forest, that showed an accuracy of 0.95 and the highest precision for predicting churn.

Model 3 also maintained a good balance between precision and recall.

The tuned hyperparameters have improved the model's ability to correctly identify churn cases.

Gaps Identified

Model 1: Logistic Regression – This model exhibits a significant gap in recall for predicting churn (1), which is only 0.04. It captures a very small proportion of actual churn cases. This gap indicates that Model 1 is not effective at identifying customers who are likely to churn.

Model 2: Random Forest – Model 2, the Random Forest model, shows improved performance compared to Model 1, with higher accuracy and a better balance between precision and recall. However, there is still a gap in recall (0.69) that indicates the model does not capture all churn cases. This gap could be reduced to achieve better predictive results.

03

End

Recommendations

Contexts Where the Model May Be Useful

1. Customer Retention Strategies: Predictions made by the model are valuable for customer retention strategies. In cases where the model identifies customers with a high likelihood of churning, SyriaTel can implement targeted retention efforts, such as offering discounts, personalized services, or loyalty programs to retain these customers.

2. Resource Allocation: The model can assist in optimizing resource allocation. SyriaTel can direct its customer service and marketing resources more efficiently, focusing on customers at higher risk of churning.

3. Pricing Strategies: Predictions inform pricing strategies. SyriaTel can modify pricing plans to align with the preferences of different customer segments, ensuring that pricing is competitive and attractive to retain existing customers and attract new ones.

Contexts Where the Model May Not be Useful

1. Short-Term Focus: If SyriaTel primarily focuses on short-term revenue and is not concerned with long-term customer relationships, the model may not be as beneficial. The model's recommendations are geared toward fostering customer loyalty over time.

2. Neglect of Customer Service: If the business does not prioritize customer service quality, the impact of the model's recommendations might be limited. Neglecting the number of customer service calls could hinder the effectiveness of retention strategies.

3. Unrealistic Expectations: If SyriaTel expects immediate results from the model without understanding that customer loyalty strategies may take time to yield significant outcomes, the model's utility could be underestimated.

Next Steps

Enhance Customer Service Quality

Given the significance of **customer service calls** as a predictor of churn, SyriaTel should focus on enhancing customer service quality. Promptly resolving customer issues and offering effective support can reduce the need for excessive service calls.

Offer Personalized Incentives

To improve results in reducing churn, SyriaTel can provide personalized incentives based on the model's predictions. Offering targeted discounts or bonuses to high-risk customers can enhance their loyalty and retention.

Feature Importance

From the feature importance analysis, the organization will need to play close look at this variables in order of importance they play in determining the churning rate.

In summary we have inferred several features related to call usage, such as total day charge and total day minutes, as highly influential in predicting churn. Additionally, the number of customer service calls appears to be a critical factor contributing to customer churn.

These insight suggests the importance of focusing on these aspects to reduce churn rates.

Thanks!

Do you have any questions?

ochieng40denis@gmail.com
+254 720 664 599

Denis Omondi