

Customer Churn Analysis and Prediction

Using Machine Learning to Enhance
Customer Retention



Overview

SyriaTel, a telecom company, is losing customers who switch to other providers, causing revenue loss and increased acquisition costs. To tackle this, we're building a predictive model to spot potential churners early. Understanding why customers leave helps SyriaTel keep them satisfied and loyal. This means less revenue loss and happier customers overall.



Business Understanding

Problem Statement:

SyriaTel is dealing with a big problem: customers leaving, which means they're losing money and spending more to get new customers. This hurts their business and makes it tough to keep customers happy. When lots of people leave, it could mean they're not happy with SyriaTel's services, or they found better deals somewhere else. So, SyriaTel needs to figure out why customers are leaving and find ways to keep them around for the long haul.



Data Understanding

I chose the SyriaTel dataset because it gives us a lot of info about customers in the telecom industry. It tells us things like who the customers are, how they use the services, and details about their accounts. This info helps us understand why customers might leave. For telecom companies like SyriaTel, keeping customers happy is super important because it costs less than finding new ones. So, by looking at this dataset, we can figure out what makes customers leave and come up with ways to keep them happy and stick around.



Objectives

To Gather Data: We will collect and clean relevant customer data to make sure it's accurate and ready for analysis.

To Engineer Features: We'll look at different aspects of customer behavior, like how they use the service and interact with customer service, to create features that help us understand why customers might leave.

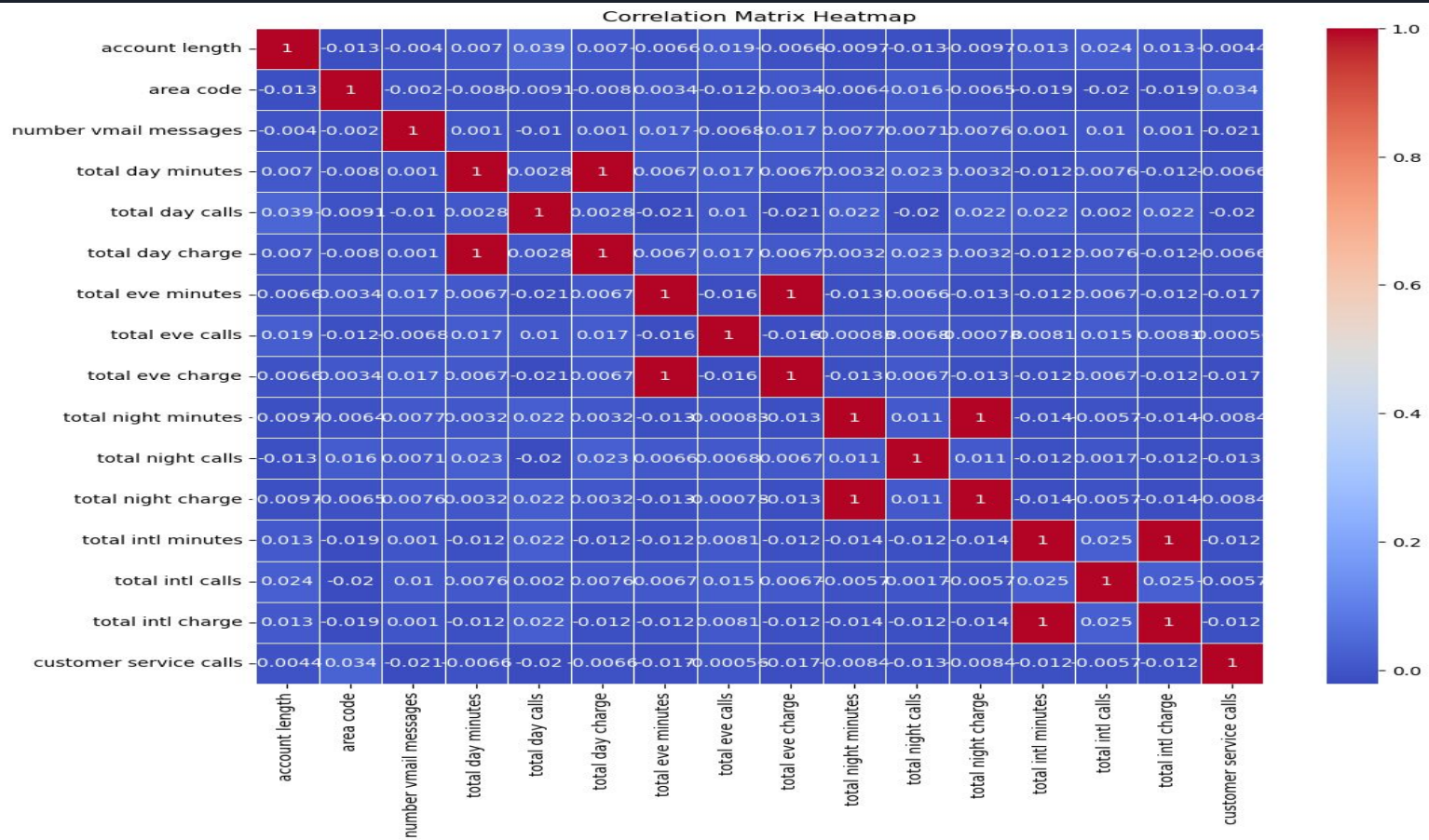
To Build Models: We'll use machine learning to build models that predict if a customer is likely to leave. We'll try out different algorithms to find the best one.

To Evaluate Models: We'll use metrics like accuracy, precision, recall, and F1 score to see how well our models are doing. Our goal is to have models that are at least 80% accurate.

To Implement the Solution: Once we have the best model, we'll integrate it into SyriaTel's systems. This will help them identify customers who might leave so they can take action to keep them.

To Monitor and Improve: We'll keep an eye on how well the model is doing over time. If we see it's not performing as well as before, we'll update it to make sure it stays accurate.

Analysis of Data





Analysis of data

High Positive Correlations: Minutes and Charges: When customers use more minutes during the day, evening, or night, their charges increase accordingly. This means that the more they talk, the more they pay for those calls.

Call Counts: The number of calls made by customers also shows a strong relationship with the minutes they spend talking and the charges they incur.

Moderate Positive Correlations: Features like total evening minutes and total evening calls have a moderate positive relationship. This means that as customers spend more time talking in the evening, the number of calls they make tends to increase, but not necessarily in a perfectly proportional way.

Low or No Correlations: Area code and number of voicemail messages don't seem to have much impact on other features. Similarly, customer service calls don't show strong relationships with other metrics, suggesting they might not affect usage patterns as much.



Modeling and Evaluation

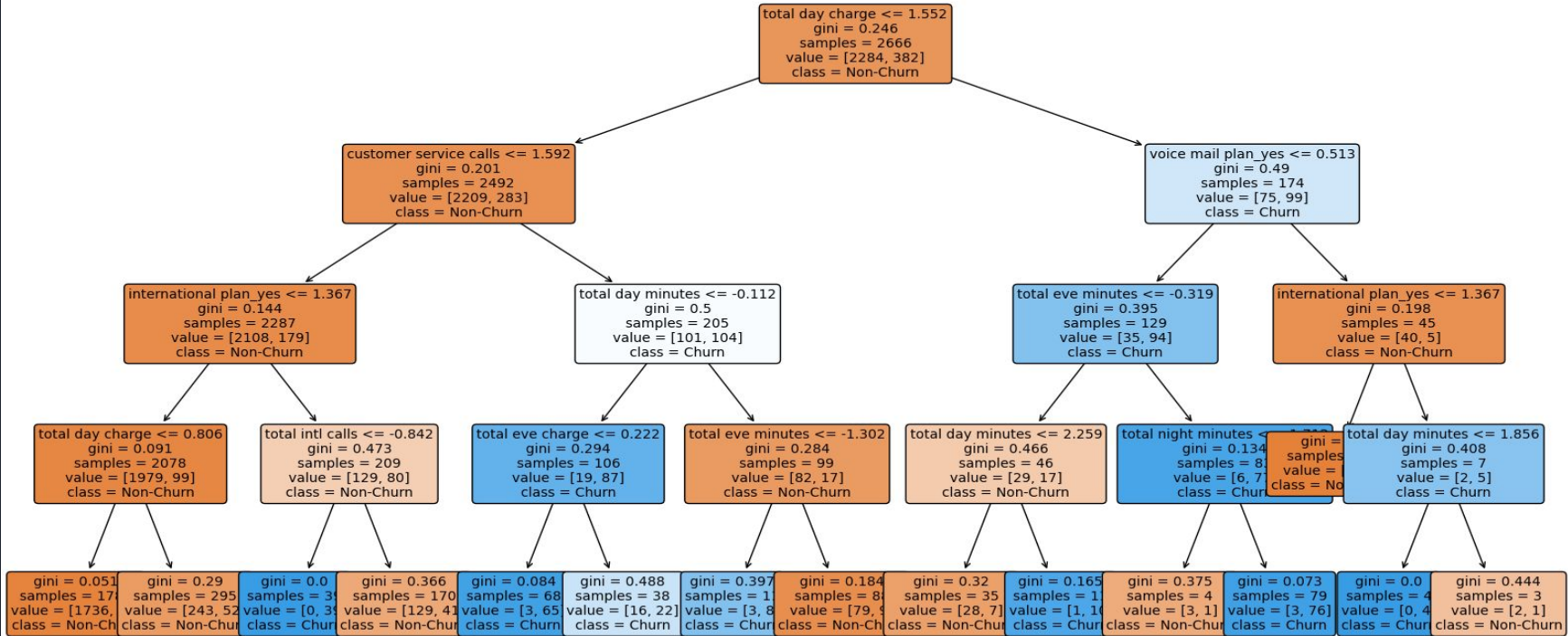
Logistic Regression insights:

Interpretation: Our model does really well on training data, correctly identifying customers who are likely to leave. But on new data, it struggles to spot these customers, which means it might miss some people who are unhappy and thinking of leaving.

Recommendations:

- **Focus on Keeping Customers:** Pay more attention to customers our model says might leave. Offer them special deals or better support to convince them to stay.
- **Improve the Model:** Work on our model to make it better at spotting customers who might leave. This might mean trying different ways of training it or adding more information.
- **Listen to Customers:** Talk to customers who our model thinks might leave to understand why they're not happy. Use this feedback to make our services better and keep more customers.

Decision tree model





Decision tree model

Key Changes in the Decision Tree Model:

- **Balanced Class Handling:** Adjusted to better handle imbalanced classes.
- **Improved Parameter Tuning:** Fine-tuned settings for better pattern capture.
- **Enhanced Validation:** Tested thoroughly for reliable performance on new data.

Analysis of the Model:

- **High Accuracy:** Identifies churners and non-churners with 95% accuracy.
- **Balanced Performance:** Works well for both types of customers.
- **Strong Prediction Power:** Effectively differentiates likely churners.
- **Improved Handling of Imbalance:** Adjustments enhance accuracy.

Recommendations for Stakeholders:

- **Prioritize Keeping Customers:** Focus on at-risk customers with special offers.
- **Act Early:** Reach out proactively to prevent churn.
- **Improve Service Quality:** Address issues to enhance satisfaction.
- **Keep Improving:** Monitor and update the model regularly for accuracy.



Random forest model

Interpretation: The Random Forest model shows strong performance with accuracy of 93% (0.928) in accurately classifying customers as churners or non-churners.

Recommendations:

1. **Focus on Retention Strategies:** Implement targeted strategies to retain customers predicted to churn, such as personalized incentives or proactive outreach.
2. **Continuous Model Evaluation:** Regularly monitor the model's performance and update it as needed to ensure accurate predictions.
3. **Feature Importance Analysis:** Identify key features driving churn predictions to better understand customer behavior and preferences.
4. **Enhance Customer Experience:** Invest in initiatives to improve overall customer satisfaction and reduce churn.
5. **Iterative Model Refinement:** Continuously refine the model based on feedback and evolving business needs to improve predictive performance over time.



Random forest identified as the best model

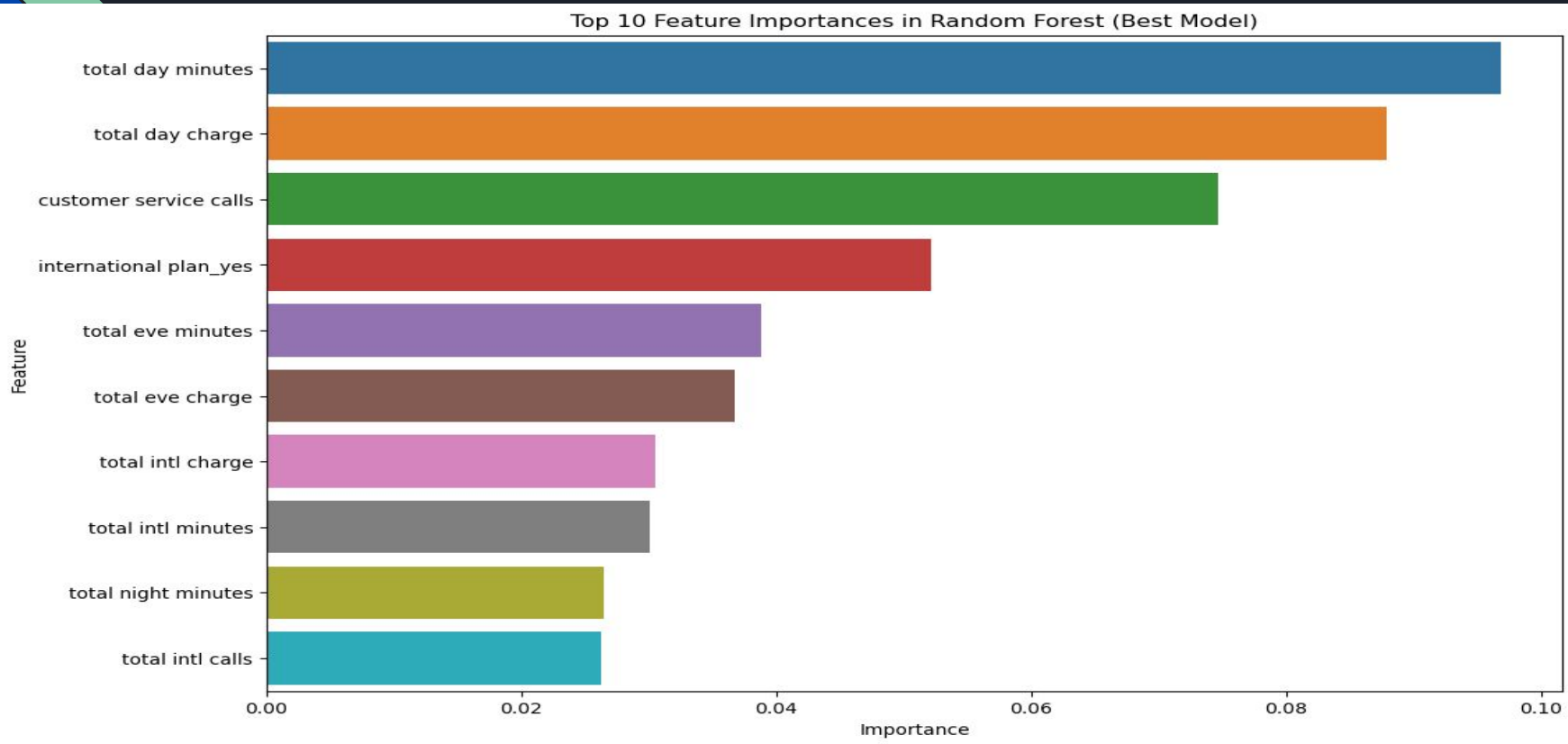
Among the models evaluated (Logistic Regression, Decision Tree, and Random Forest), the Random Forest model emerges as the top performer for predicting customer churn. Here's why:

Best Performance: The Random Forest model consistently performed better than the others in accuracy, precision, recall, and ROC-AUC score, achieving the highest score of 0.928. This indicates its strong ability to understand which customers are likely to leave and which are likely to stay.

Generalization: Despite its complexity, the Random Forest model works well on new data, showing that it can be trusted to make accurate predictions even in situations it hasn't seen before.

Insights into Customer Behavior: By identifying which factors most strongly influence whether a customer will leave, the Random Forest model gives valuable insights into what makes customers unhappy. This can help the company make changes to keep more customers happy and lo

Important Features





Explaining the feature importance graph

Feature Importance: The importance of a feature is determined by how much it contributes to the model's decision-making process. Features with higher importance scores are considered more influential in predicting customer churn.

Top Features: The top 10 features with the highest importance scores are displayed on the y-axis of the graph. These features are ranked from most to least important based on their importance scores.

Importance Scores: The length of each bar (on the x-axis) represents the importance score of the corresponding feature. Longer bars indicate higher importance, meaning that those features have a greater impact on the model's predictions.



Recommendations to the stakeholders

1. **Customer Segmentation:** Utilize the churn prediction model to segment the customer base into different risk categories based on their likelihood of churning. High-risk customers can be identified through factors such as frequent customer service calls, low usage patterns, or recent downgrades in service plans.
- 2.
3. **Real-Time Predictions:** Implement the churn prediction model in SyriaTel's operational environment to generate real-time predictions for individual customers. This enables immediate identification of customers at high risk of churning, allowing SyriaTel to trigger timely retention actions such as offering discounts or personalized messages.
- 4.
5. **Integration with Marketing Campaigns:** Incorporate churn predictions into marketing campaigns to personalize communications and offers for at-risk customers. Tailored messages and incentives, based on individual churn probabilities, can significantly increase the effectiveness of retention efforts.
- 6.
7. **Feedback Loop:** Establish a feedback loop to monitor the effectiveness of the churn prediction model and retention strategies. By analyzing outcomes such as customer retention rates and satisfaction scores, SyriaTel can refine predictive models and intervention strategies to improve overall retention rates.



End

Thank you!