

LCR2022 / Phase_3_project

Q

<> CodeIssuesPull requestsActionsProjectsWikiSecurityInsightsSettings

👁

🔗

★

★ 0 stars🔗 2 forks👁 1 watching🔗 1 Branch📁 0 Tags📈 Activity

🌐 Public repository

🔗 main1 Branch📁 0 Tags🔗📁


Q Go to filetGo to file+Add file<> Code...

🌐 LCR20221 minute ago

📁 content	Adding data	14 hours ago
📄 Phase 3 Project Final Notebook....	uploading final changes to documents	1 minute ago
📄 README.md	Update README.md	14 hours ago
📄 SyriaTel_app.py	phase 3 project upload	15 hours ago
📄 Telecom_Churn_Presentation.pdf	uploading final changes to documents	1 minute ago
📄 Telecom_Churn_Presentation.p...	uploading final changes to documents	1 minute ago
📄 dsc_phase_3_project.ipynb	uploading final changes to documents	1 minute ago
📄 image.jpg	phase 3 project upload	15 hours ago
📄 model.pkl	phase 3 project upload	15 hours ago
📄 repository.pdf	uploading final changes to documents	1 minute ago
📄 syriaTel_customer.csv	phase 3 project upload	15 hours ago
📄 syriatel.png	phase 3 project upload	15 hours ago

📖 README

✎☰



Understanding Customer Churn, A Case of SyriaTel

Communications

Description

In the dynamic telecommunications industry, customer retention is paramount to sustained success. SyriaTel acknowledges the growing challenge of customer churn and is committed to leveraging data-driven insights to mitigate it. This README outlines our approach, objectives, and anticipated outcomes in addressing this critical issue.

Problem Statement

SyriaTel is grappling with escalating customer churn rates, which not only impact revenue but also inflate acquisition costs. Identifying churn drivers and at-risk customers is crucial for implementing targeted retention strategies.

Objectives

- 1. Customer Segmentation:** Utilize clustering techniques to segment SyriaTel's customer base, facilitating targeted interventions.
- 2. Impact Assessment:** Evaluate the influence of various service features on customer retention through predictive modeling techniques.
- 3. Model Validation:** Validate the performance of predictive models using appropriate evaluation metrics and refine them for improved accuracy and generalizability.
- 4. Retention Strategy Evaluation:** Assess the long-term effectiveness of implemented retention strategies by monitoring and analyzing churn rates over time.
- 5. Strategy Formulation:** Formulate and recommend bespoke retention strategies based on analytical findings, aiming to mitigate churn and foster business growth.

Data Usage Strategy:

SyriaTel, a prominent telecommunications provider, is confronted with escalating customer churn rates, impacting both revenue and market share. To address this challenge, an in-depth analysis was conducted, leveraging a combination of customer usage data, service plans, and interaction history. This data served as the foundation for exploring patterns and predictors of churn through statistical analysis and machine learning models.

Procedure Overview:

- 1. Data Collection:** Comprehensive datasets encompassing customer usage patterns, demographics, and service interactions were gathered.
- 2. Exploratory Data Analysis (EDA):** Initial exploration of the data revealed trends, correlations, and potential churn indicators.
- 3. Clustering Analysis:** Utilizing clustering techniques, the customer base was segmented into distinct groups based on various factors such as usage patterns and demographics.
- 4. Predictive Modeling:** Advanced predictive models were developed to assess the impact of different service features on customer retention and to identify customers at risk of churning.
- 5. Model Validation and Refinement:** The predictive models underwent rigorous validation using appropriate evaluation metrics. Refinements were made to enhance accuracy and generalizability.
- 6. Evaluation of Retention Strategies:** Long-term effectiveness of implemented retention strategies was evaluated through continuous monitoring and analysis of churn rates over time.

7. **Recommendations:** Based on the analytical findings, bespoke retention strategies were formulated and recommended to increase customer satisfaction, foster loyalty, and ultimately reduce churn rates.

Methodology.

Churn is our independent variable.

Target variable insights : These are the main factors to be investigated in the study:

Numeric Features:

- Subscription Period
- Number of voice Mail messages
- Total day Minutes
- Total day Calls
- Total evening minutes
- Total evening calls
- Total Night minutes
- Total Night calls
- Total International calls
- Total International minutes
- Customer Service calls

Categorical features:

- International Plan
- Voice Mail Plan

Procedures used to analyse the dataset

1. Outliers Detection:

- Outliers are data points that significantly deviate from the rest of the dataset.
- Detecting outliers helps in identifying unusual or anomalous behavior within the dataset, which could potentially skew statistical analysis or machine learning models.
- For SyriaTel, outlier detection could reveal extreme values in customer behavior metrics (e.g., call duration, data usage) or financial metrics (e.g., monthly charges), providing insights into irregular customer activities or billing discrepancies.

2. Multicollinearity Analysis:

- Multicollinearity refers to the high correlation between predictor variables in a regression analysis.
- Identifying multicollinearity helps in understanding how independent variables relate to each other and how they collectively influence the dependent variable.
- In the context of SyriaTel, multicollinearity analysis could reveal correlations between customer attributes (e.g., age, income) or service features (e.g., international plan, voice mail plan), helping to refine predictive models and understand the drivers of customer churn more accurately.

3. Univariate Analysis:

- Univariate analysis involves examining the distribution and characteristics of a single variable.
- It helps in understanding the central tendency, variability, and shape of the data distribution.

- In the SyriaTel dataset, univariate analysis could be used to explore the distribution of key variables such as customer tenure, monthly charges, customer service calls, and churn status.
- This analysis provides insights into the distributional properties of individual variables, enabling a deeper understanding of customer behavior and churn patterns.

Modelling outcomes and interpretation

1. Baseline Model: Logistic Regression:

The logistic regression model yielded the following performance metrics on the test set:\

Accuracy: 85.3%

Precision: 55.2%

Recall: 15.8%

F1 Score: 24.6%

- The Logistic Regression model demonstrates moderate accuracy but relatively low precision, recall, and F1-score.
- Addressing the imbalance between precision and recall is crucial to enhance the model's predictive power and reliability.
- SyriaTel can use the model's predictions as a starting point for targeted retention efforts, focusing resources on customers identified as high-risk by the model.

2. Decision Tree Classifier:

The Decision Tree classifier significantly improved the model's performance on the test set compared to the initial logistic regression model, with the following metrics:

Accuracy: 94.3%

Precision: 77.4%

Recall: 88.1%

F1 Score: 82.4%

- The model effectively identifies churn cases while maintaining a relatively low rate of false positives, indicating its reliability in predicting customer churn.
- SyriaTel can use the Decision Tree Classifier's predictions to guide targeted retention efforts, focusing resources on customers identified as high-risk by the model.
- The high recall rate suggests that the model can capture a large proportion of churn cases, enabling SyriaTel to proactively intervene and retain valuable customers.
- Leveraging the insights provided by the Decision Tree Classifier, SyriaTel can implement personalized retention strategies tailored to the specific characteristics and behaviors of at-risk customers.

3. Random Forest Classification Model:

The Random Forest classifier significantly improved the model's performance on the test set compared to the logistic regression and Decision Tree models, with the following metrics:

Accuracy: 98.1%

Precision: 100.0%

Recall: 87.1%

F1 Score: 93.1%

- The high accuracy, precision, recall, and F1 score of the Random Forest model indicate its effectiveness in predicting customer churn for SyriaTel.
- By accurately predicting churn, SyriaTel can optimize resource allocation for retention efforts. They can focus their resources on retaining high-risk customers identified by the model, maximizing the impact of their retention initiatives.
- Leveraging the insights provided by the Random Forest model, SyriaTel can implement targeted and personalized retention strategies to reduce churn and enhance customer loyalty. This can lead to improved customer satisfaction, increased customer lifetime value, and sustained business growth for SyriaTel.

4. Gradient Boosting Classifier Model

The Gradient Boosting Classifier yielded the following results:

Testing Accuracy: 98.1%

Precision: 100.0%

Recall: 87.1%

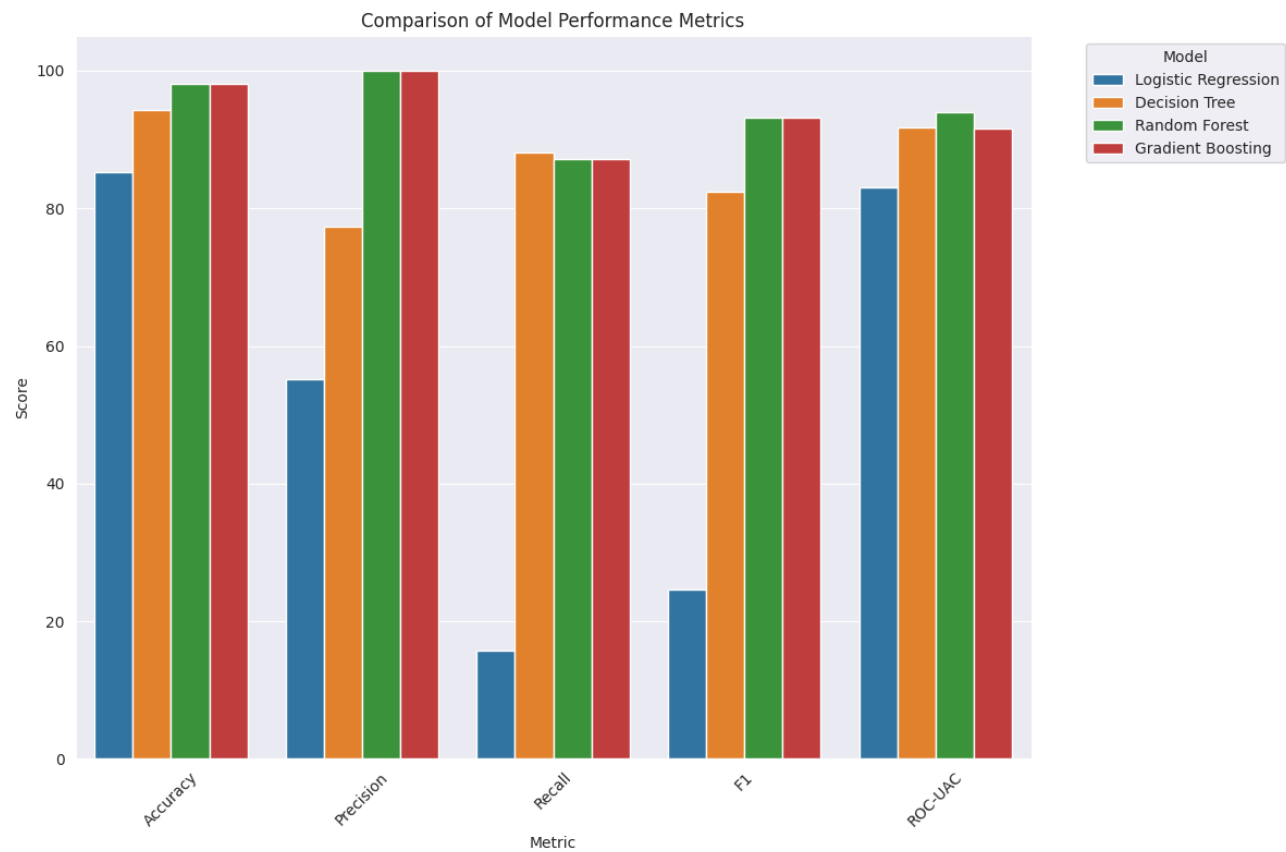
F1-Score: 93.1%

ROC-AUC Score: 91.6%

- SyriaTel can use this model to identify customers at risk of churning with high confidence, facilitating proactive retention strategies.
- They can focus their resources on retaining high-risk customers identified by the model, maximizing the impact of their retention initiatives.
- Leveraging the insights provided by the Gradient Boosting Classifier, SyriaTel can implement targeted and personalized retention strategies to reduce churn and enhance customer loyalty. This can lead to improved customer satisfaction, increased customer lifetime value, and sustained business growth for SyriaTel.

Comparative analysis of the models used

Comparison of Results Compiling the results into a coherent format for comparison by assuming the variables `accuracy_log_reg`, `roc_auc_log_reg`, `accuracy_rf`, `roc_auc_rf`, `accuracy_gb`, and `roc_auc_gb` hold the results from the above evaluations. These are the findings:



Model Accuracy Precision Recall F1 ROC-AUC

0 Logistic Regression 85.3 55.2 15.8 24.6 83.1

1 Decision Tree 94.3 77.4 88.1 82.4 91.8

2 Random Forest 98.1 100.0 87.1 93.1 93.9

3 Gradient Boosting 98.1 100.0 87.1 93.1 91.6

Model Selection:

-Random Forest and Gradient Boosting models demonstrate superior performance across multiple metrics compared to Logistic Regression and Decision Tree.

- SyriaTel may consider deploying Random Forest or Gradient Boosting models for churn prediction, as they offer higher accuracy, precision, recall, F1 score, and ROC-AUC score.

Resource Allocation:

- Leveraging the insights provided by Random Forest or Gradient Boosting models, SyriaTel can efficiently allocate resources for targeted retention efforts, focusing on customers identified as high-risk by these models.

Improved Customer Retention:

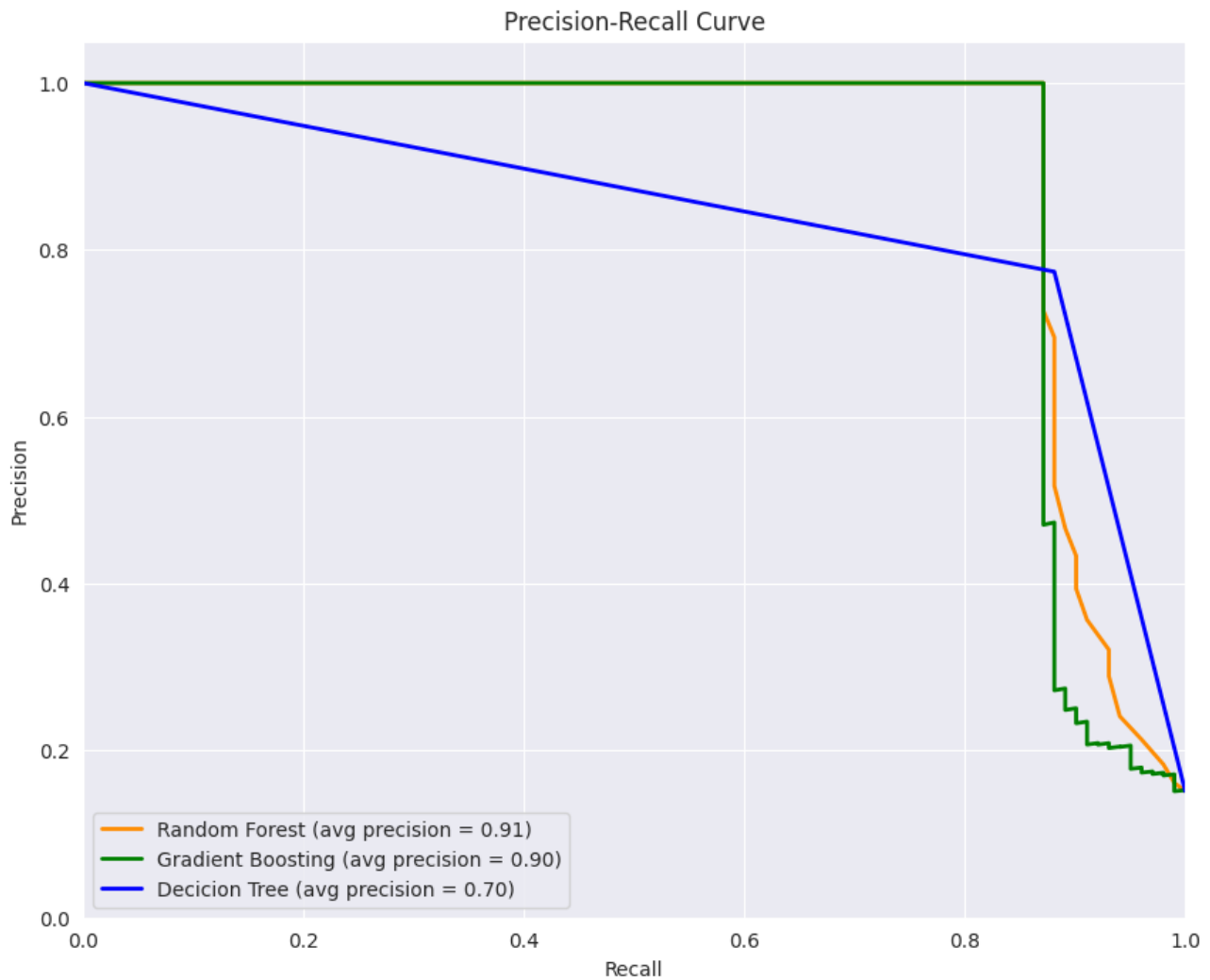
- By deploying advanced predictive models such as Random Forest or Gradient Boosting, SyriaTel can implement personalized retention strategies tailored to the specific characteristics and behaviors of at-risk customers. This can lead to improved customer satisfaction, increased customer lifetime value, and sustained business growth for SyriaTel.

ROC Curve Comparison



- **Model Selection:** Understanding the performance of different classification models, such as Random Forest, Gradient Boosting, and Decision Tree, helps SyriaTel in selecting the most suitable model for predicting customer churn or other relevant business outcomes.
- **Predictive Accuracy:** The high AUC values indicate that these models have strong predictive accuracy in identifying customers who are likely to churn. This means that SyriaTel can rely on these models to effectively target and intervene with at-risk customers to mitigate churn.
- **Strategic Decision-making:** Armed with reliable predictive models, SyriaTel can make more informed strategic decisions regarding customer retention efforts. They can allocate resources more efficiently by focusing on customers identified as high-risk by these models, thus maximizing the impact of their retention strategies.
- **Competitive Advantage:** Utilizing advanced predictive analytics tools like Random Forest and Gradient Boosting can provide SyriaTel with a competitive edge in the telecommunications market. By proactively addressing customer churn, they can enhance customer satisfaction, loyalty, and ultimately, their market position.
- **Continuous Improvement:** Regular monitoring and evaluation of model performance allow SyriaTel to continuously refine and improve their predictive models. This iterative process ensures that their models remain accurate and relevant over time, adapting to changing customer behavior and market dynamics.

Precision - Recal Curve



The Precision-Recall curves for the models offer insights into the trade-off between precision (the ability of the classifier not to label a negative sample as positive) and recall (the ability of the classifier to find all the positive samples).

- **Model Selection:**

Random Forest and Gradient Boosting models demonstrate higher average precision scores compared to Decision Tree, indicating their superior performance in correctly identifying churn cases. SyriaTel may prioritize the deployment of Random Forest or Gradient Boosting models for churn prediction, as they offer higher precision and more reliable predictions.

- **Resource Allocation:**

Leveraging the insights provided by Random Forest or Gradient Boosting models, SyriaTel can efficiently allocate resources for targeted retention efforts, focusing on customers identified as high-risk by these models. These models offer higher precision, ensuring that retention efforts are directed towards customers who are most likely to churn.

- **Continuous Improvement:**

SyriaTel can continuously monitor and evaluate model performance, refining predictive models to enhance precision and improve the effectiveness of churn prediction over time. By incorporating feedback and iteratively improving predictive models, SyriaTel can stay ahead in managing customer churn and fostering long-term customer relationships.

Conclusion

Based on the analysis conducted on SyriaTel's customer churn data, several conclusions and recommendations can be drawn to address the challenges faced by the telecommunications company:

1. Churn Analysis:

The analysis revealed that the churn rate for SyriaTel is approximately 14.5%, indicating a significant portion of customers leaving the service. Factors such as having an international plan and the number of customer service calls show a moderate positive correlation with churn, suggesting that these factors may influence customer retention.

2. Customer Behavior Insights:

Customers with an international plan are more likely to churn compared to those without one, indicating a need to investigate the reasons behind this trend. The number of customer service calls shows a weak positive correlation with churn, suggesting that higher engagement with customer service may not necessarily lead to improved retention.

3. Geographical Patterns:

The distribution of churned and non-churned customers across different states reveals potential geographical variations in churn rates, which could be further explored to understand regional customer behavior.

6. Service Usage Analysis:

Analysis of service usage patterns, such as total day minutes and total evening minutes, did not show significant correlations with churn. However, further exploration may be warranted to understand the impact of these factors on customer retention.

Recommendation

1. Targeted Retention Strategies:

Implement targeted retention strategies specifically tailored to address the needs of customers with international plans. This could involve personalized offers or incentives aimed at improving satisfaction and loyalty among this segment.

2. Enhanced Customer Support:

Focus on improving the quality of customer service to address issues and concerns proactively, potentially reducing the need for multiple customer service calls and improving overall customer satisfaction.

3. Geographically Targeted Campaigns:

Develop geographically targeted marketing campaigns and promotions to address regional variations in churn rates. Understanding the unique needs and preferences of customers in different states can help in tailoring retention strategies effectively.

4. Customer Feedback Mechanisms:

Establish robust mechanisms for collecting and analyzing customer feedback to identify pain points and areas for improvement in service delivery. Actively soliciting and acting upon customer feedback can help in enhancing customer experience and reducing churn.

5. Continuous Monitoring and Analysis:

Implement a system for continuous monitoring of churn indicators and conducting regular analyses to identify emerging trends and patterns. This proactive approach will enable SyriaTel to adapt and refine its retention strategies in real-time.

6. Data-driven Decision Making:

Emphasize the importance of data-driven decision-making throughout the organization, leveraging insights from churn analysis and customer behavior to inform strategic decisions related to product offerings, pricing, and customer service.

7. Collaboration Across Departments:

Foster collaboration between marketing, sales, customer service, and data analytics teams to ensure alignment in retention efforts and a holistic approach to addressing churn.

By implementing these recommendations, SyriaTel can improve its customer retention efforts, reduce churn rates, and enhance overall customer satisfaction and loyalty in the highly competitive telecommunications market. It's essential for SyriaTel to continuously adapt and refine its strategies based on ongoing data analysis and customer feedback to stay competitive and sustain long-term growth.

Recommendation For further Analysis

1. Customer Segmentation

High-Value vs. Low-Value Customers: By analyzing usage patterns, plan subscriptions, and service calls, it's possible to identify high-value customers who contribute significantly to revenue. Strategies can be tailored to ensure their retention.

Risk Profiles: Segmenting customers based on their churn risk can help in prioritizing retention efforts. Customers identified with higher risk can be targeted with personalized retention campaigns.

2. Usage Patterns

Peak vs. Off-Peak Usage: Understanding when customers use services the most could guide promotional offers or adjustments in service plans to enhance satisfaction.

Service Utilization: Analysis of how different services (voice, international, voicemail) influence churn can help tailor service offerings to match customer preferences and needs.

3. Customer Feedback and Satisfaction

Service Calls as Feedback: High numbers of customer service calls might indicate areas where service improvements are needed. Detailed analysis of the reasons behind these calls can uncover specific pain points.

Surveys and Feedback Loops: Incorporating customer feedback directly through surveys or feedback mechanisms can provide qualitative insights that complement the quantitative findings from the dataset.

4. Competitive Analysis

Market Competition: States with higher churn rates might be experiencing more aggressive competition. A deeper market analysis could reveal competitive pressures influencing churn.



Releases

No releases published
[Create a new release](#)


Packages

No packages published
[Publish your first package](#)

Languages

Suggested workflows


Based on your tech stack



Pylint

Lint a Python application with pylint.


Configure



Python Package using Anaconda

Create and test a Python package on multiple Python versions using Anaconda for package management.

Configure



Python application

Create and test a Python application.

Configure

[More workflows](#)

Dismiss suggestions