Your Group

PREDICTION OF CUTOMER CHURN: SYRIATEL TELECOMMUNICATIONS

Presented by: Joyce Chepngeno

Table Of Contents

Overview

SyriaTel, a telecommunications company, faces a significant challenge in reducing customer churn, which can negatively impact their revenue and profitability. Customer churn refers to the phenomenon where customers discontinue their services with a company, often switching to competitors or simply discontinuing the service altogether. some of the factors contributing significantly to customer churn are Poor service experience and customer service, making it easy for customers to switch providers and poor customer experiences, such as multiple contacts for issue resolution. These factors highlight the importance of addressing service quality and customer satisfaction to reduce churn rate.

Business and data understanding

Predicting customer churn is essential for SyriaTel to retain its customer base and minimize revenue loss. By identifying customers at risk of churning, SyriaTel can proactively engage with them through incentives, personalized offers, or enhanced customer service. The high cost of acquiring new customers compared to retaining existing ones motivates companies like SyriaTel to focus on churn reduction strategies. Leveraging predictive modelling allows for the detection of behavioural patterns indicating a high likelihood of churn. This proactive approach aims to enhance customer retention rates and overall business performance

The dataset consist of 3333 rows and 21 columns. The column include a mix of data types that is object, integer, float and boolean. There are no missing values and duplicates. The columns for the dataset are as follows:

“state', 'account length', 'area code', 'phone number', 'international plan', 'voice mail plan', 'number voicemail messages', 'total day minutes', 'total day calls', 'total day charge', 'total eve minutes', 'total eve calls', 'total eve charge','total night minutes', 'total night calls', 'total night charge', 'total international minutes', 'total intl calls', 'total international charge','customer service calls', 'churn'

Objectives

- Develop a robust predictive model using machine learning algorithms to accurately forecast customer churn in the telecommunications company.

- Identify predictive patterns and to evaluate the model's performance using appropriate metrics and validate its effectiveness.

- To determine the Percentage rate of customer churn per area code in the company

- To determine the factors that mostly contribute to customer churn in the company

Modelling

1. Logistic regerssion: Baseline model
2. Decision trees model
3. K- Nearest Neighbor(KNN)
4. Random Forest model
5. Gradient Forest model

Evaluation: performance of each model

	Model	Train Accuracy	Train Precision	Train ROC AUC	Test Accuracy	Test Precision	Test ROC AUC
0	Logistic Regression	86.2	55.7	58.2	84.4	42.9	54.6
1	Decision Trees	95.4	94.8	85.6	93.5	89.0	81.7
2	KNN	100.0	100.0	100.0	88.0	76.6	63.6
3	Random Forest	95.4	94.2	85.8	93.5	88.2	82.0
4	Gradient Boosting	97.2	99.0	90.7	94.8	93.6	84.8

Based on the provided metrics in the table above, the Gradient Boosting model consistently outperforms the other models on both the training and test sets, achieving the highest accuracy, precision, and ROC AUC score. Therefore, the Gradient Boosting model is deemed the best among the five models considered for this classification task. The advantages of Gradient model are high accuracy, robustness to overfitting, handling Non-linear relationships and can handle various types of data numerical and categorical features. In conclusion, the gradient boosting model, after thorough evaluation and validation, emerged as the most effective predictive model for identifying customer churn in the SyriaTel telecommunications company dataset. Its high accuracy and robustness make it a valuable tool for informing strategic decisions and implementing retention strategies to reduce customer churn and improve business outcomes.

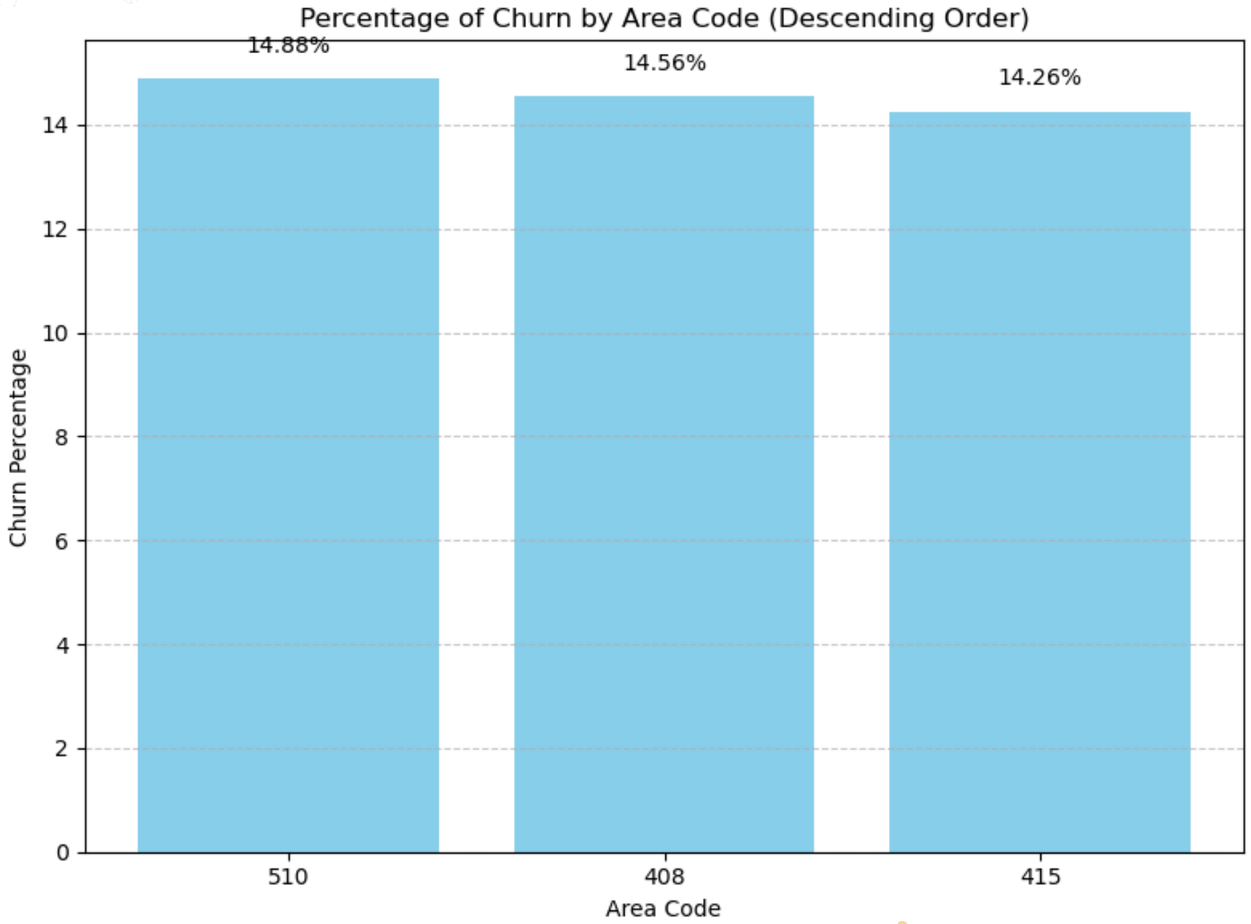
Evaluation: performance of each model

	Model	Train Accuracy	Train Precision	Train ROC AUC	Test Accuracy	Test Precision	Test ROC AUC
0	Logistic Regression	86.2	55.7	58.2	84.4	42.9	54.6
1	Decision Trees	95.4	94.8	85.6	93.5	89.0	81.7
2	KNN	100.0	100.0	100.0	88.0	76.6	63.6
3	Random Forest	95.4	94.2	85.8	93.5	88.2	82.0
4	Gradient Boosting	97.2	99.0	90.7	94.8	93.6	84.8

Based on the provided metrics in the table above, the Gradient Boosting model consistently outperforms the other models on both the training and test sets, achieving the highest accuracy, precision, and ROC AUC score. Therefore, the Gradient Boosting model is deemed the best among the five models considered for this classification task. The advantages of Gradient model are high accuracy, robustness to overfitting, handling Non-linear relationships and can handle various types of data numerical and categorical features. In conclusion, the gradient boosting model, after thorough evaluation and validation, emerged as the most effective predictive model for identifying customer churn in the SyriaTel telecommunications company dataset. Its high accuracy and robustness make it a valuable tool for informing strategic decisions and implementing retention strategies to reduce customer churn and improve business outcomes.

Percentange rate of customer churn per area state

The percentange of customer churn by area code is approximately between 14.26% - 14.88%. The leading area code in terms of customer churn is 510(14,88%), followed by 408(14.56%) and finally 415(14.26%).



Conclusion

Based on the provided metrics in the table above, the Gradient Boosting model consistently outperforms the other models on both the training and test sets, achieving the highest accuracy, precision, and ROC AUC score. Therefore, the Gradient Boosting model is deemed the best among the five models considered for this classification task.

The advantages of Gradient model are high accuracy, robustness to overfitting, handling Non-linear relationships and can handle various types of data numerical and categorical features.

In conclusion, the gradient boosting model, after thorough evaluation and validation, emerged as the most effective predictive model for identifying customer churn in the SyriaTel telecommunications company dataset. Its high accuracy and robustness make it a valuable tool for informing strategic decisions and implementing retention strategies to reduce customer churn and improve business outcomes.

The 510 area code had the highest percentage rate of 14.88%, the 408 area code had a rate of 14.56%, and the 415 area code had a rate of 14.26%.

Feature importance analysis provided insights into the factors contributing to customer churn, allowing for targeted interventions

The most influential factors to customer churn include the total minutes spent on day calls, the frequency of customer service calls, the charges associated with day calls, the presence of international plans, and the charges for international calls.

Recommendations

- Utilize Gradient Boosting Model: Given its superior performance in accurately predicting customer churn, the company should prioritize the implementation and deployment of the Gradient Boosting model for ongoing monitoring and prediction of churn.

- Strategic Decision Making: The insights gained from the feature importance analysis highlight specific areas that the company can focus on to mitigate customer churn. For instance, efforts can be directed towards improving customer service quality to reduce the frequency of customer service calls, optimizing international plan offerings, and managing charges associated with day calls.

- Proactive Retention Strategies: Leveraging the predictive power of the Gradient Boosting model, the company can proactively identify customers at risk of churn and implement targeted retention strategies. This may include personalized offers, loyalty programs, or proactive outreach to address customer concerns and enhance satisfaction.

- Continuous Monitoring and Optimization: Customer preferences and behaviors may evolve over time, necessitating continuous monitoring and optimization of the predictive model. Regular updates and refinements based on new data and changing business dynamics will ensure the model remains effective and relevant in predicting churn.

Next steps

- Ensure that the model is integrated into existing model systems for ongoing monitoring
- Develop strategies and initiatives aimed at addressing main factors contributing to customer churn
- Design and implement targeted retention strategies such personalized offers, loyalty programs and proactive outreach to retain these customers and enhance satisfaction
- Gather feedback from customers and stakeholders to assess the impact of interventions and identify areas for further improvement
- Ensure the relevant teams are trained in using and interpreting the gradient Boosting model and its insight. Foster a culture of data-driven decision-making and empower teams to leverage the model for proactive churn management

THANK YOU