

1 minute ago  16

[View code](#)

As a data scientist, I have been hired to develop a robust model that can effectively predict which customers are likely to discontinue their services with SyriaTel.

## Data Understanding

SyriaTel has provided me with that includes information about their customers. The dataset contains information of 33,333 of SyriaTel customers and the features are as follows:

- `state` - Client's residence.
- `account length` - How long they have had the subscription.
- `area code` - Client's area code.
- `phone number` - Client's phone number.
- `international plan` - Is the client subscribed to the international plan?(yes/no).
- `voice mail plan` - Is the client subscribed to the voice mail plan?(yes/no).
- `number vmail messages` - The number of the voicemail messages.
- `total day minutes, calls, charge` - the client's daily minutes, calls, and charges.
- `total eve minutes, calls, charge` - the client's evening minutes, calls, and charges.
- `total night minutes, calls, charge` -the client's night minutes, calls, and charges.
- `total intl minutes, calls, charge` - the client's tital international minutes, calls, and charges.
- `customer service calls` - how many times the customer service line was called.
- `churn` - The response variable we will be targeting.

## Modelling

---

Various models are created using logistic regression, decision tree and random forest algorithms. Hyperparameter tuning is applied to ldecision tree and random forest algorithms. A pipeline is used to prevent data leakage. Data is scaled in the pipeline. The image below shows a summary of the models and their performance.

	accuracy	f1	precision	recall	name
21	0.973621	0.944115	0.984952	0.912000	Scaled Random Forest
22	0.972422	0.941359	0.984290	0.908000	Tuned Random Forest(scaled)
19	0.970381	0.970368	0.971224	0.970381	Tuned Bagged Tree
17	0.964029	0.929422	0.929422	0.929422	Scaled Bagged Trees
18	0.954866	0.954840	0.955914	0.954866	Resampled Bagged Trees
14	0.935120	0.935120	0.935123	0.935120	Resampled Decision Tree
15	0.933709	0.933709	0.933709	0.933709	Tuned Decision Tree
13	0.926859	0.870204	0.840911	0.910855	Scaled Decision Tree
2	0.816643	0.816550	0.817290	0.816643	Resampled Logistic Regression
3	0.815233	0.815229	0.815255	0.815233	Tuned Logistic Regression
1	0.780576	0.675225	0.657201	0.749038	Scaled Logistic Regression
8	0.770983	0.516604	0.518743	0.516056	Original Naive Bayes
5	0.738609	0.607514	0.599801	0.655165	Scaled KNN
6	0.684767	0.682218	0.690892	0.684767	Resampled KNN
7	0.671368	0.664698	0.686183	0.671368	Tuned KNN
11	0.650212	0.648533	0.653137	0.650212	Tuned Naive Bayes
10	0.650212	0.648533	0.653137	0.650212	Resampled Naive Bayes
9	0.579137	0.494655	0.538527	0.574550	Scaled Naive Bayes
20	0.394484	0.382328	0.550271	0.581264	Original Random Forest
16	0.149880	0.130344	0.074940	0.500000	Original Bagged Trees
12	0.149880	0.130344	0.074940	0.500000	Original Decision Tree
4	0.149880	0.130344	0.074940	0.500000	Original KNN
0	0.149880	0.130344	0.074940	0.500000	Original Logistic Regression

## Evaluation

### 1. Accuracy:

Accuracy represents the proportion of correctly classified instances out of the total number of instances. Higher accuracy values indicate better performance.

In this case, the scaled random forest, the tuned random forest and the tuned bagged tree achieve the highest accuracy scores of above 0.97, indicating that they have the highest overall classification accuracy among the models.

	accuracy	f1	precision	recall	name
21	0.973621	0.944115	0.984952	0.912000	Scaled Random Forest
22	0.972422	0.941359	0.984290	0.908000	Tuned Random Forest(scaled)
19	0.970381	0.970368	0.971224	0.970381	Tuned Bagged Tree
17	0.964029	0.929422	0.929422	0.929422	Scaled Bagged Trees

## 2. F1-Score:

The F1-score is the harmonic mean of precision and recall. It provides a balanced measure between precision (ability to correctly identify positive instances) and recall (ability to correctly identify all positive instances). Similar to accuracy, higher F1-scores indicate better performance.

In this case, the tuned bagged tree, the resampled bagged tree and the scaled random forest achieve the highest f1 scores of above 0.94, indicating that they have the highest overall classification f1 among the models.

	accuracy	f1	precision	recall	name
19	0.970381	0.970368	0.971224	0.970381	Tuned Bagged Tree
18	0.954866	0.954840	0.955914	0.954866	Resampled Bagged Trees
21	0.973621	0.944115	0.984952	0.912000	Scaled Random Forest
22	0.972422	0.941359	0.984290	0.908000	Tuned Random Forest(scaled)

## 3. Precision:

Precision represents the proportion of true positive predictions out of all positive predictions. It measures the model's ability to avoid false positives. Higher precision values indicate fewer false positives.

In this case, the scaled random forest, the tuned random forest and tuned bagged tree have the highest precision scores, above 0.97.

	accuracy	f1	precision	recall	name
21	0.973621	0.944115	0.984952	0.912000	Scaled Random Forest
22	0.972422	0.941359	0.984290	0.908000	Tuned Random Forest(scaled)
19	0.970381	0.970368	0.971224	0.970381	Tuned Bagged Tree
18	0.954866	0.954840	0.955914	0.954866	Resampled Bagged Trees

## 4. Recall:

Recall (also known as sensitivity or true positive rate) represents the proportion of true positive predictions out of all actual positive instances. It measures the model's ability to identify positive instances correctly. Higher recall values indicate fewer false negatives.

Tuned bagged tree, resampled bagged tree and resampled decision tree achieve the highest recall scores, above 0.93.

	accuracy	f1	precision	recall	name
19	0.970381	0.970368	0.971224	0.970381	Tuned Bagged Tree
18	0.954866	0.954840	0.955914	0.954866	Resampled Bagged Trees
14	0.935120	0.935120	0.935123	0.935120	Resampled Decision Tree
15	0.933700	0.933700	0.933700	0.933700	Tuned Decision Tree

## Recommendations

1. The company should closely analyze the performance of states with low performance and determine if the issue lies in network coverage. If network coverage is found to be inadequate, we will develop a strategy to improve coverage by deploying additional boosters in those states.
2. SyriaTel's focus should be on investigating whether a more attractive international calling plan can encourage customers to consider its international plan while they are traveling.
3. SyriaTel should evaluate it's customer service and what it might be able to do better to assist the customer.
  - Additional training for customer service staff might be needed, but it also could be related to other factors.
4. SyriaTel should explore the possibility of engaging with different vendors or establishing temporary partnerships to provide incentives and promotions aimed at increasing customer satisfaction and reducing churn among dissatisfied customers.

## Next Step

### 1. Network Coverage Improvement:

- \* Conduct a detailed analysis of poorly performing states to identify specific areas with network coverage issues.
- \* Evaluate the feasibility and cost-effectiveness of deploying additional boosters or infrastructure to enhance network coverage in those areas.
- \* Collaborate with the network operations team to plan and implement the expansion of network coverage.
- \* Monitor the performance and impact of the network coverage improvements.

### 2. International Calling Plan Enhancement:

- \* Conduct market research and competitor analysis to understand customers' preferences and demands regarding international calling plans.
- \* Design and develop a more attractive international calling plan, considering factors such as competitive pricing, flexible options, and value-added features.
- \* Communicate the new plan effectively to customers through targeted marketing

campaigns and personalized offers.

- \* Monitor customer response and adoption rates of the enhanced international calling plan.

### 3. Vendor Partnership and Customer Incentives:

- \* Research potential vendors and partnership opportunities that can provide additional value or incentives for dissatisfied customers.
- \* Establish collaboration or partnership agreements with selected vendors to offer exclusive promotions, discounts, or rewards to targeted customer segments.
- \* Develop customer retention programs that leverage these partnerships and incentives to improve customer satisfaction and loyalty.
- \* Monitor customer feedback, satisfaction levels, and churn rates to evaluate the effectiveness of the vendor partnerships and customer incentives.

### 4. Implementation and Evaluation:

- \* Implement the proposed initiatives, including network coverage improvements, enhanced international calling plans, and vendor partnerships.
- \* Monitor customer churn rates, customer satisfaction levels, and key performance indicators (KPIs) to assess the impact of the implemented strategies.
- \* Calculate the cost associated with customer retention efforts and evaluate the return on investment (ROI).
- \* Continuously analyze and optimize the implemented strategies based on the obtained results and customer feedback.



## Releases

No releases published

[Create a new release](#)

## Packages

No packages published


[Publish your first package](#)

## Languages

● Jupyter Notebook 100.0%

### Suggested Workflows


Based on your tech stack



Actions Importer

Automatically convert CI/CD files to YAML for GitHub Actions.


Set up



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