

SyriaTel Customer Churn Analysis

By Yesim Cebeci

What is Customer Churn?

Customer churn is the percentage of customers who stopped purchasing your business's products or services during a certain period of time





BUSINESS
UNDERSTANDING

DATA
UNDERSTANDING

METHODS

01

02

03

TABLE OF CONTENTS

04

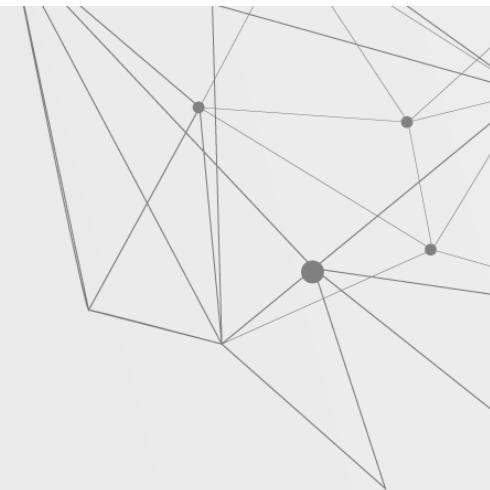
MODEL

05

CONCLUSION &
RECOMMENDATIONS

06

FUTURE WORK



01

BUSINESS UNDERSTANDING





Syriatel is telecommunication company that has concerns regarding churn rate.

Using customer account data:

- We will analyze what features from the data are most important in predicting customer churn and
 - Whether or not they will leave the company
-



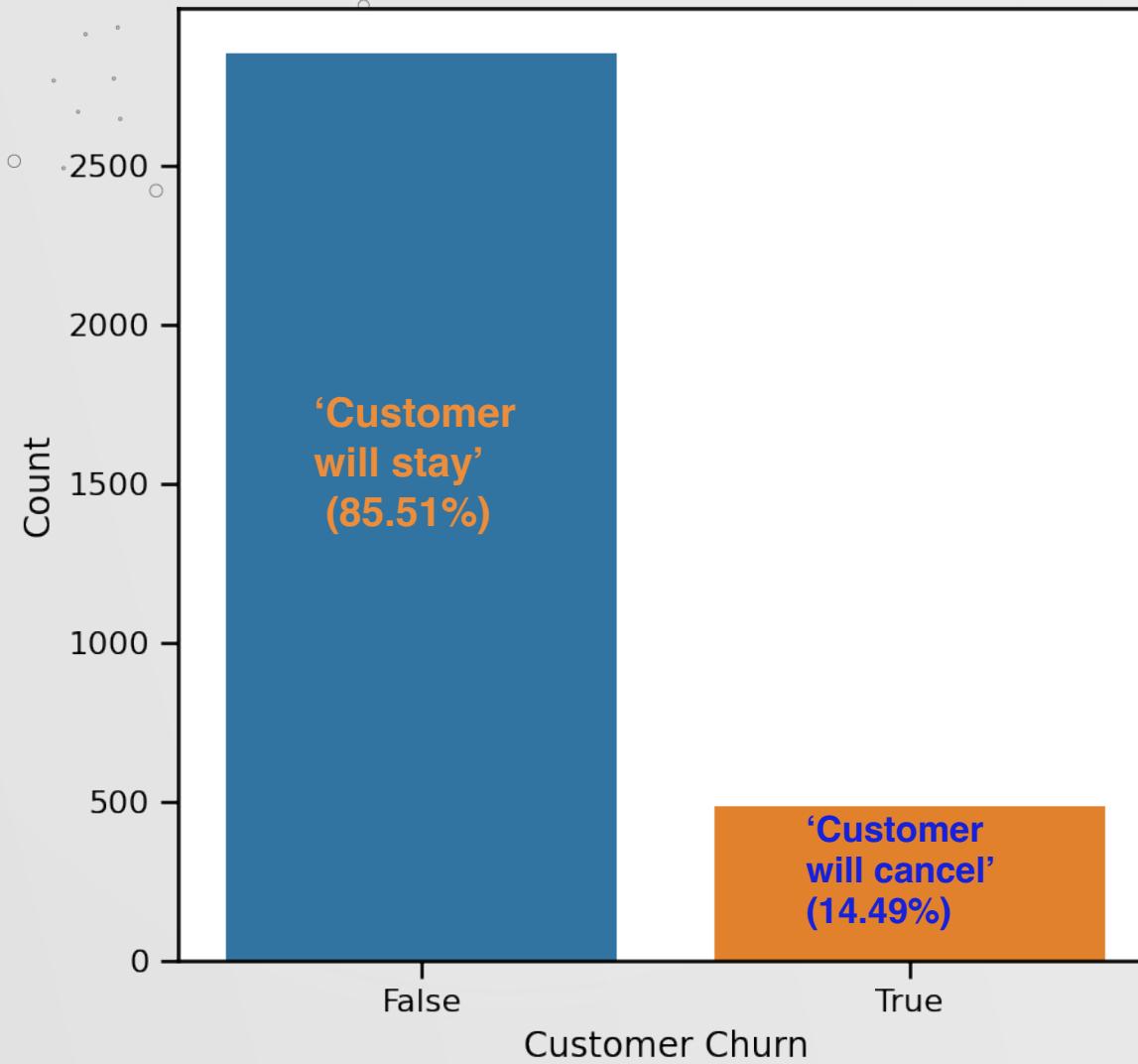
A large, abstract network graph is positioned on the left side of the slide. It consists of numerous small, light-gray dots connected by thin gray lines, forming a complex web of triangles and polygons. Some dots are larger and darker, indicating they are central hubs in the network.

02 DATA UNDERSTANDING



- The dataset contains 3333 SyriaTel customers account information.
- We will be focus on following features while modeling:
 - Account length
 - Day / Evening / Night / International minutes
 - Day / Evening / Night / International calls
 - Day / Evening / Night / International charges
 - Customer Service Calls
 - Number of Voice Mail Messages
 - Voice Mail Plan
 - International Plan
 - States

Distribution of Customer Churning



The background of the slide features a complex, abstract network graph composed of numerous small, dark gray dots connected by thin, light gray lines. This graph is organized into several distinct clusters of nodes, with some nodes having many connections and others having fewer. The overall effect is one of a dense, interconnected system.

03

METHODS



Using data to create classification models
to predict customer churn



Analyzing model according to recall metric



Analyzing best performing model for top feature
importance



Conclusion and Business Recommendation

Metric to Use : RECALL

- The Recall is calculated as the ratio between the number of **Positive** samples correctly classified as **Positive** to the total number of **Positive** samples.

$$Recall = \frac{True_{positive}}{True_{positive} + False_{negative}}$$



Metric to Use : RECALL

- Recall would be a better metric for this dataset because:

- Misidentifying someone as cancelling and hitting them with a strategy to keep them engaged would be more beneficial than
- Missing someone who cancelled and not hitting them with a strategy to keep them subscribed to the service.



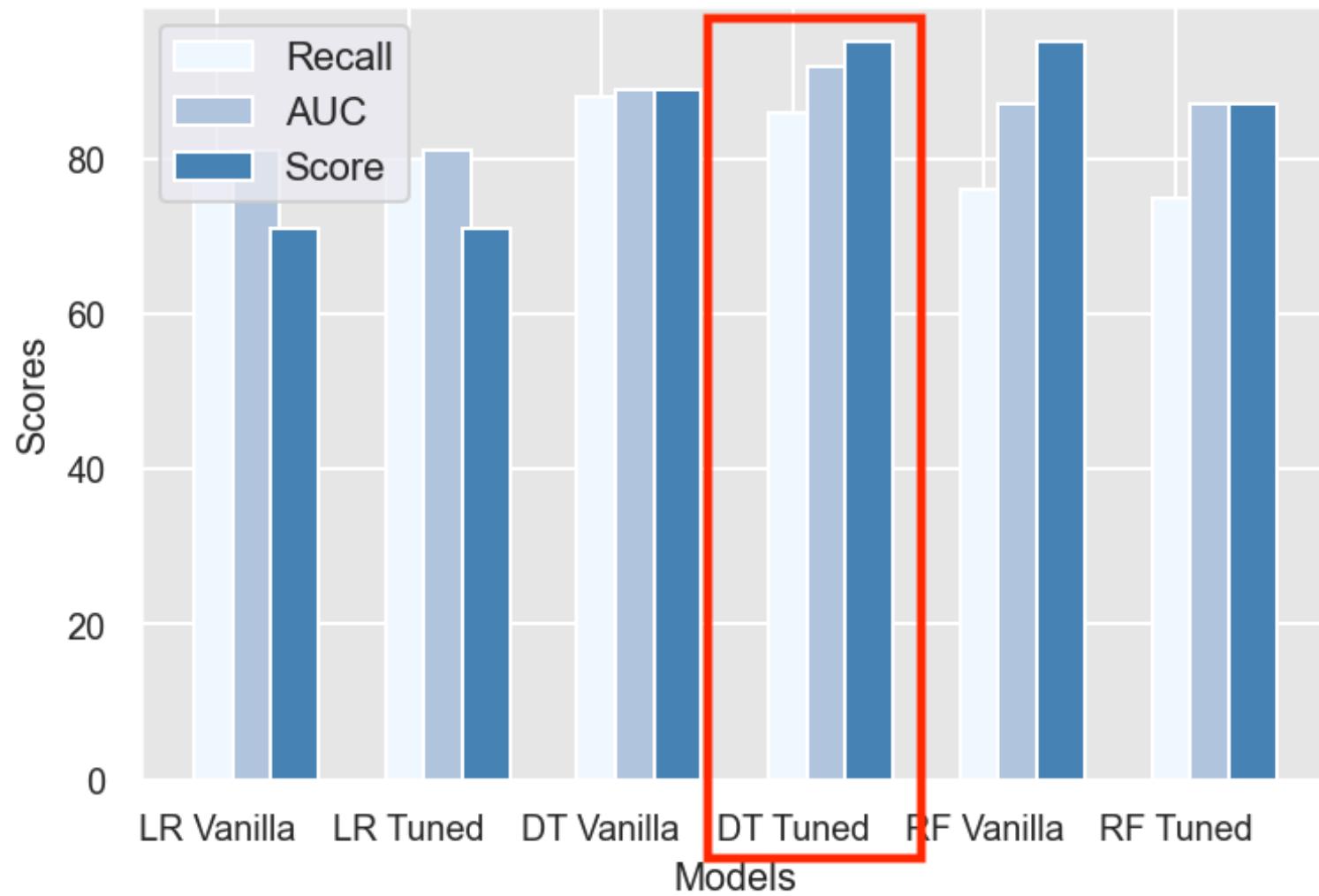
04

BEST PERFORMING MODEL

(DECISION TREE MODEL(TUNED PARAMETERS))



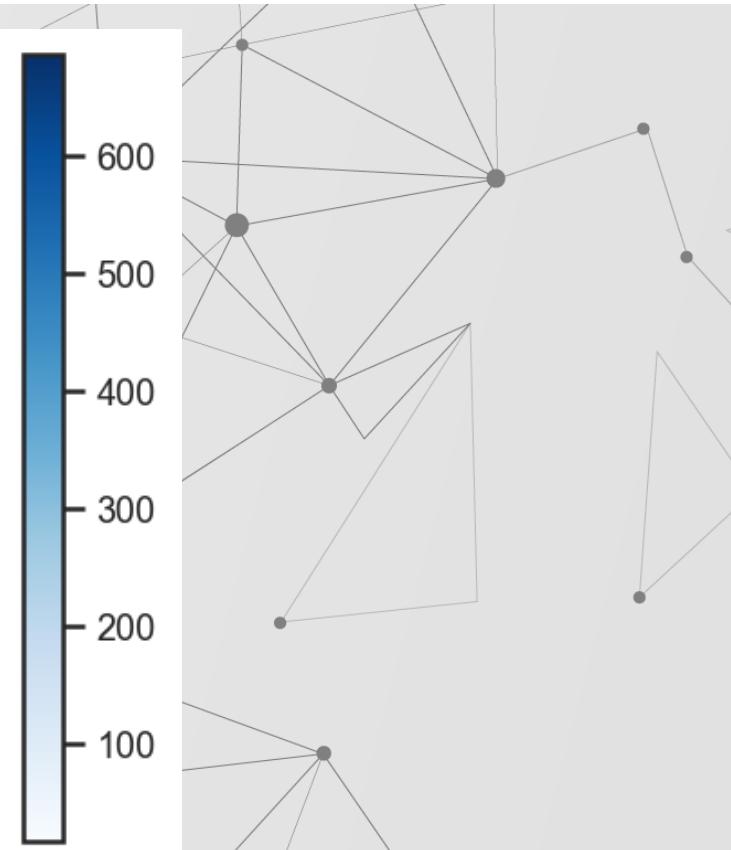
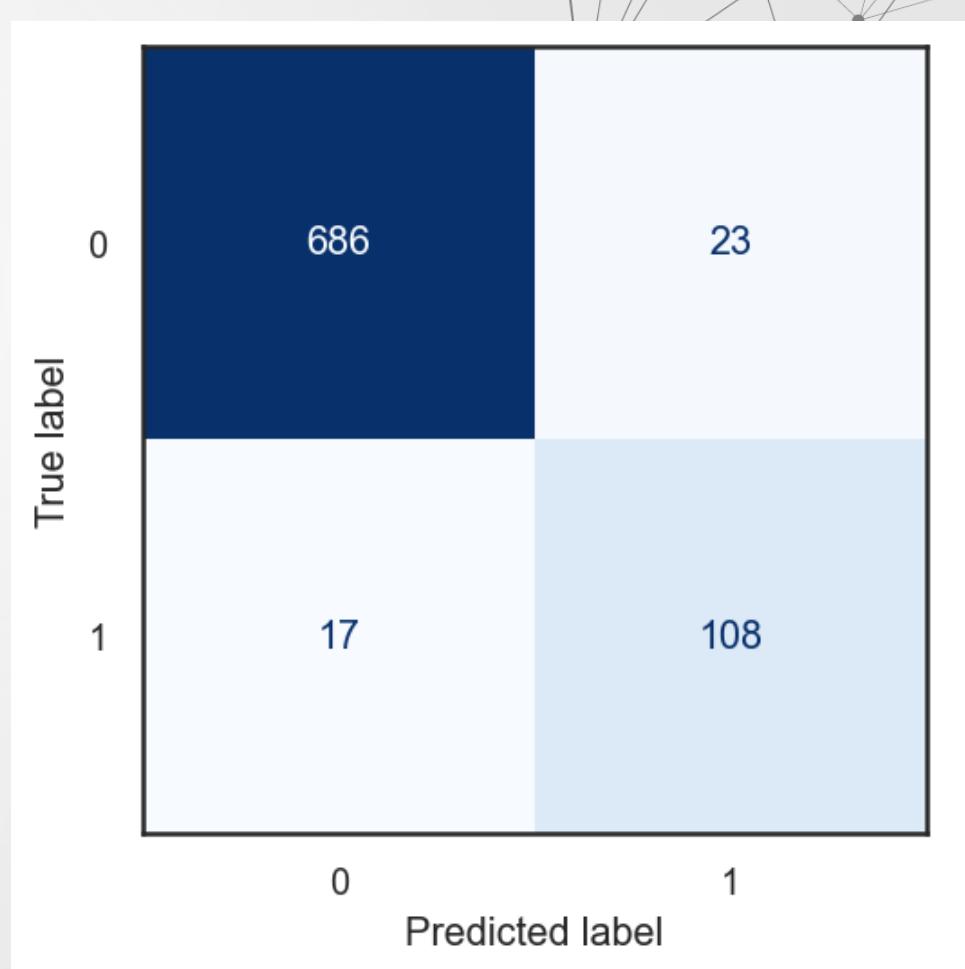
Comparing Scores



Confusion Matrix Analysis

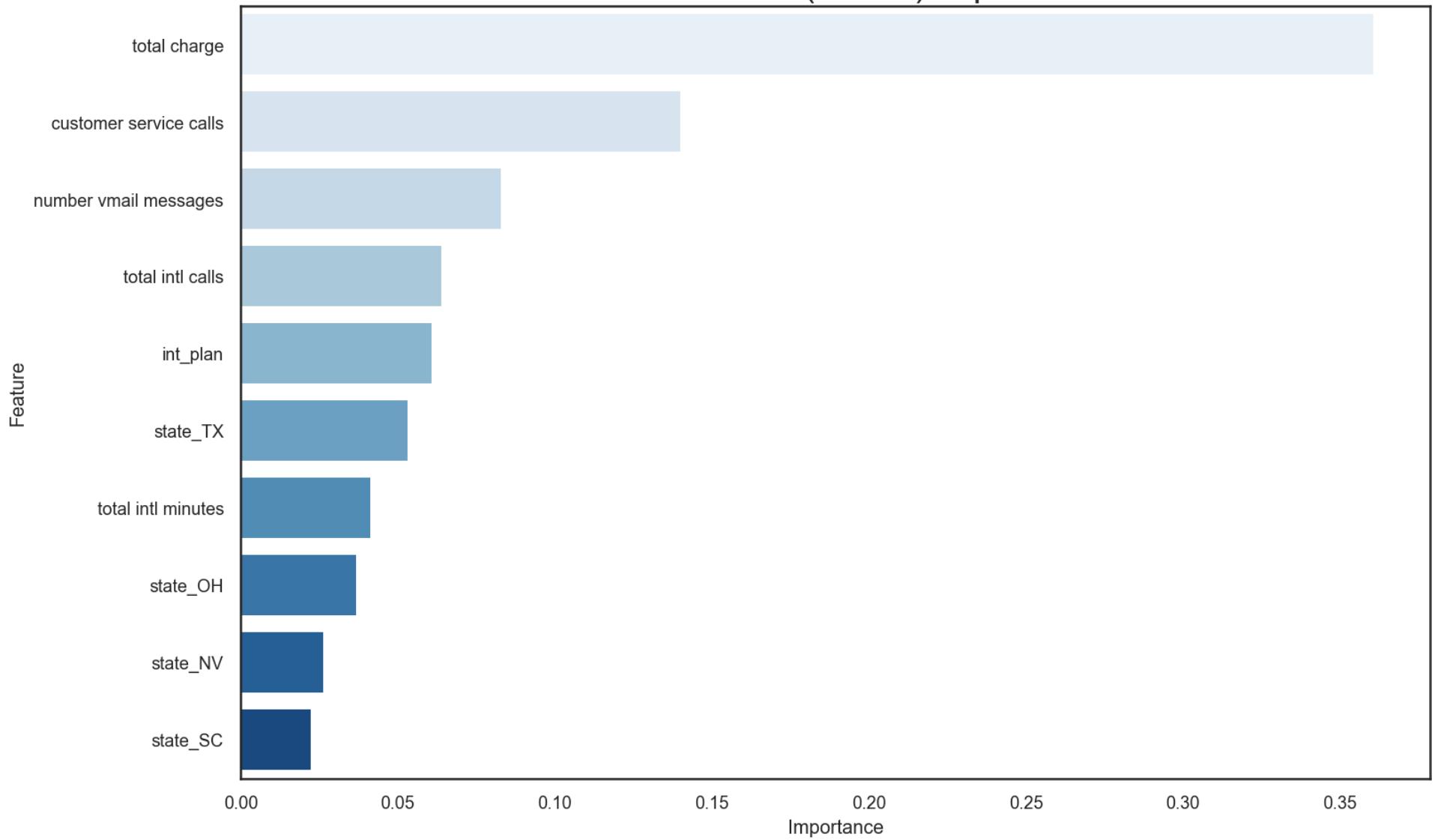
- We want to **reduce** False Negatives (Type II Error) means our model is predicting that the customer isn't going to cancel but they really are.

		Predicted Class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)



- **Decision Tree (Tuned) Model Confusion Matrix result shows that we have lowest False Negatives**

Decision Tree Model(Tuned) Top 10 Features



05

CONCLUSION & RECOMMENDATIONS





Work on total charge and making promotions on this feature especially those who call customer service.

Improving customer service. The more customer service calls, the more people leave the plan.

Focus on international plan and prevent international customer From leaving.

06

FUTURE WORK





Get more competitors' datasets to see if the reason of higher churn because of the competition at these states TX , OH , NV , SC.

Test on more models such as Naive Bayes , K-Nearast Neigbor and Gradient Boosting

Get more data on what specific reason customers called customer service for

THANKS



ysmcebeci@gmail.com

+1 929 342 89 42

www.linkedin.com/in/yesim-cebeci



<https://github.com/yesimcebeci>

