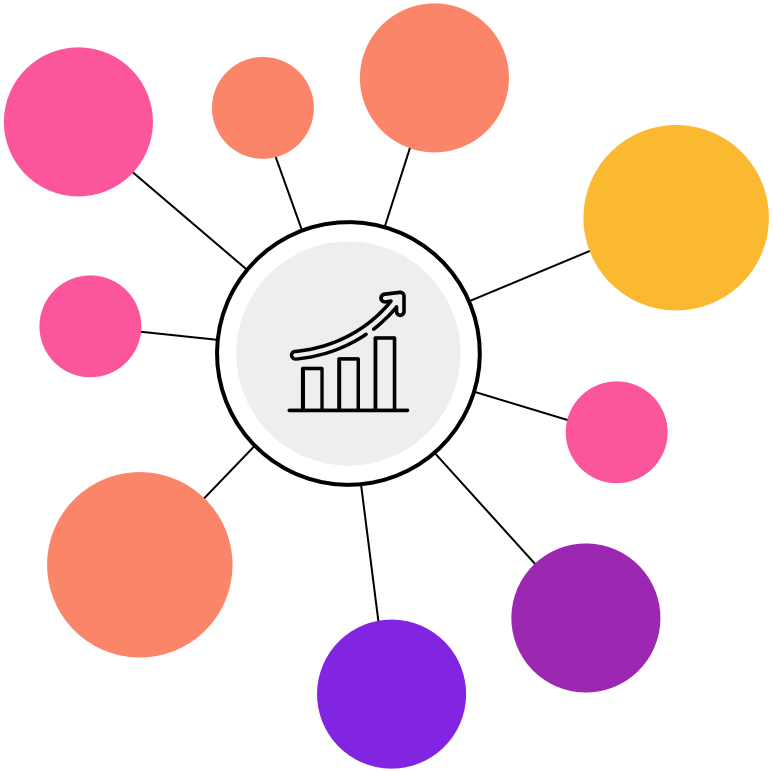
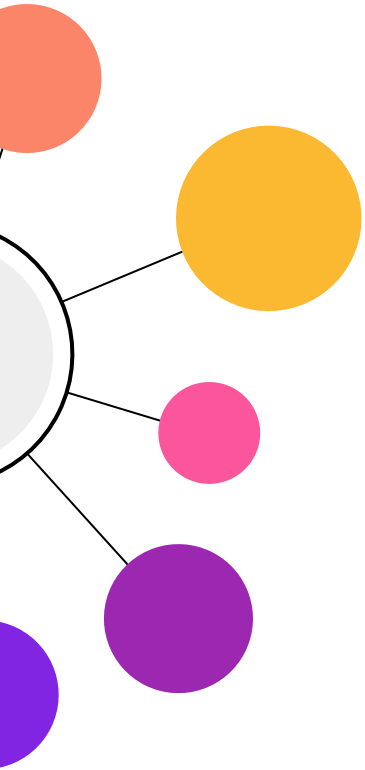


SyriaTel Telecom Data Analysis

By Daniel Oselu

Overview



Building a classifier to predict whether a customer will stop doing business with SyriaTel.

Business Problem

Assisting the Customer Acquisition and Management team, and the Product Development team with insight into customer behaviour.

Particularly, project we will be checking for churn, that is, detecting if a client might terminate subscription, we'll also check the dataset to suggest some new products for the Product development team.

Data Understanding

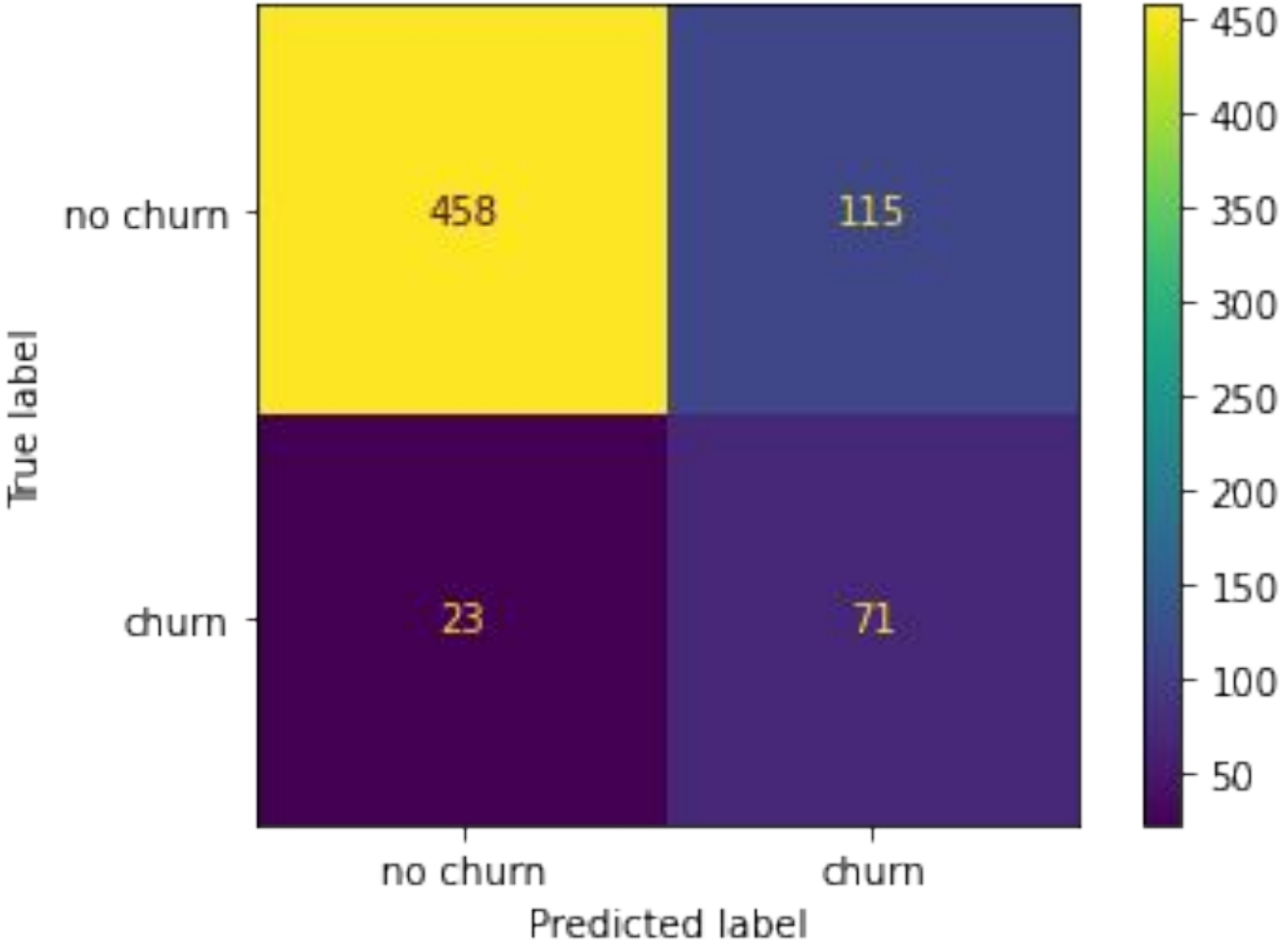
- The dataset is contained in a CSV(comma separated Values) file.
- The target used is the churn column, that indicates to us if the client churned or not.
- Our dataset contains 3333 datapoints/rows and 21 columns giving details about each datapoint.
- The dataset can be downloaded from Kaggle at SyriaTel Dataset

Modelling

We used different algorithms to create models that predicted the classification problem being dealt with;

- Logistic Regression
- K-Nearest Neighbors

First a vanilla base model was created, then **model iteration was done**, with tuning being done at the different stages to come up with a final model.


no churn
churn
True label
458
115
23
71
no churn
churn
Predicted label
450
400
350
300
250
200
150
100
50

Evaluation

PRECISION	RECALL	ACCURACY	F1
			
39%	77%	80%	52%
How precise our predictor was	Of classes we are interested in were captured	Total number of predictions our model got right	Balance between our precision and recall

Next Steps

- What makes a state perform better than its counterparts?
- What are some of the indicators of high expenditure by an account?

Thank You

Daniel Oselu
doselu3@gmail.com
0712345678