

# Data Science PROJECT SUMMARY

Client: No-Churn Telecom | Category: Telecom – Churn Rate ML

-by,

Shubham Makreja

## Business Case:

No-Churn Telecom is an establish Telecom operation in Europe with more than a decade in Business. Due to new players in the Market, telecom industry has become very competitive and retaining customers becoming challenge.

In spite of No-Churn initiative for reducing tariffs and promoting more offers, the churn rate (percentage of customers migrating to competitors) is well above 10%.

No-Churn wants to explore possibility of Machine Learning to help with following use cases to retain competitive edge in the industry.

## PROJECT GOAL

Help No-Churn with their use cases with ML

1. Understanding the variables that influencing the customers to migrate.

2. Creating Churn risk scores that can be indicative to drive

3. Exporting the trained model with prediction capability for

CHURN-FLAG Highlights the flag (with input variables

documents) that can be integration with internal application

help to identify possible CHURN-FLAG YES customers and provide

more attention in customer touch point areas, including customer

care support, request fulfilment, auto categorizing tickets as high

priority for quick resolutions any questions they may have etc.,

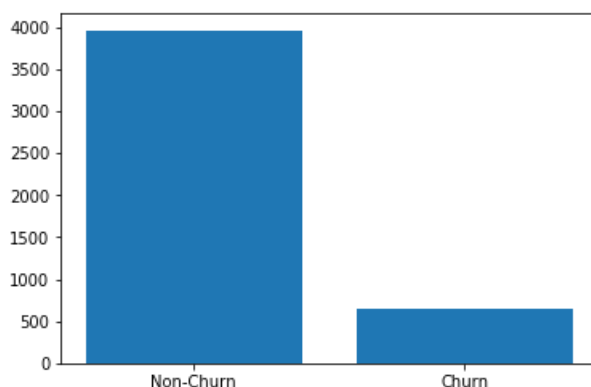
## INTRODUCTION

The objective of this project is to identify factors involved in The customer migration to other Telecom companies in Europe, from almost a decade old Company named NO-CHURN , and make a relevant model for predicting whether a customer will migrate.

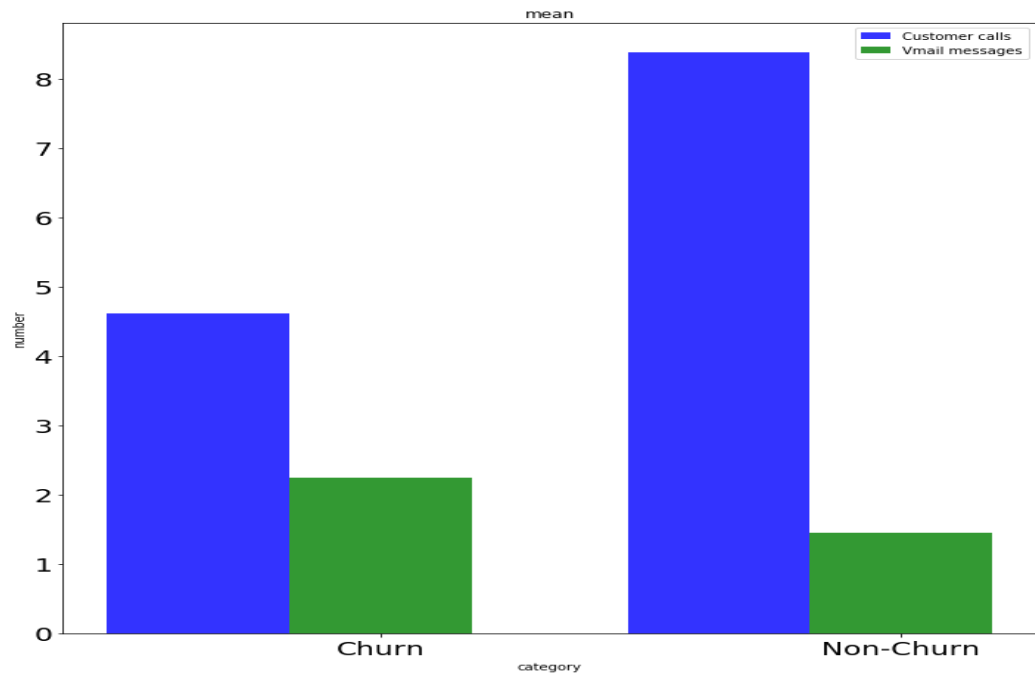
In the project , various Exploratory and visualization analysis techniques were used to understand the reason behind the churning of the customers from the Telecom Industry.

## Summary

- The project starts with importing the given dataset.
- Label encoding is done to convert certain necessary alphabetic values to numeric ones like in 'Churn ' , 'yes' is converted to 1 and 'no' to 0.
- Then dataset is divided into 2 parts based on 'yes' or 'no' Values of churn.
- 

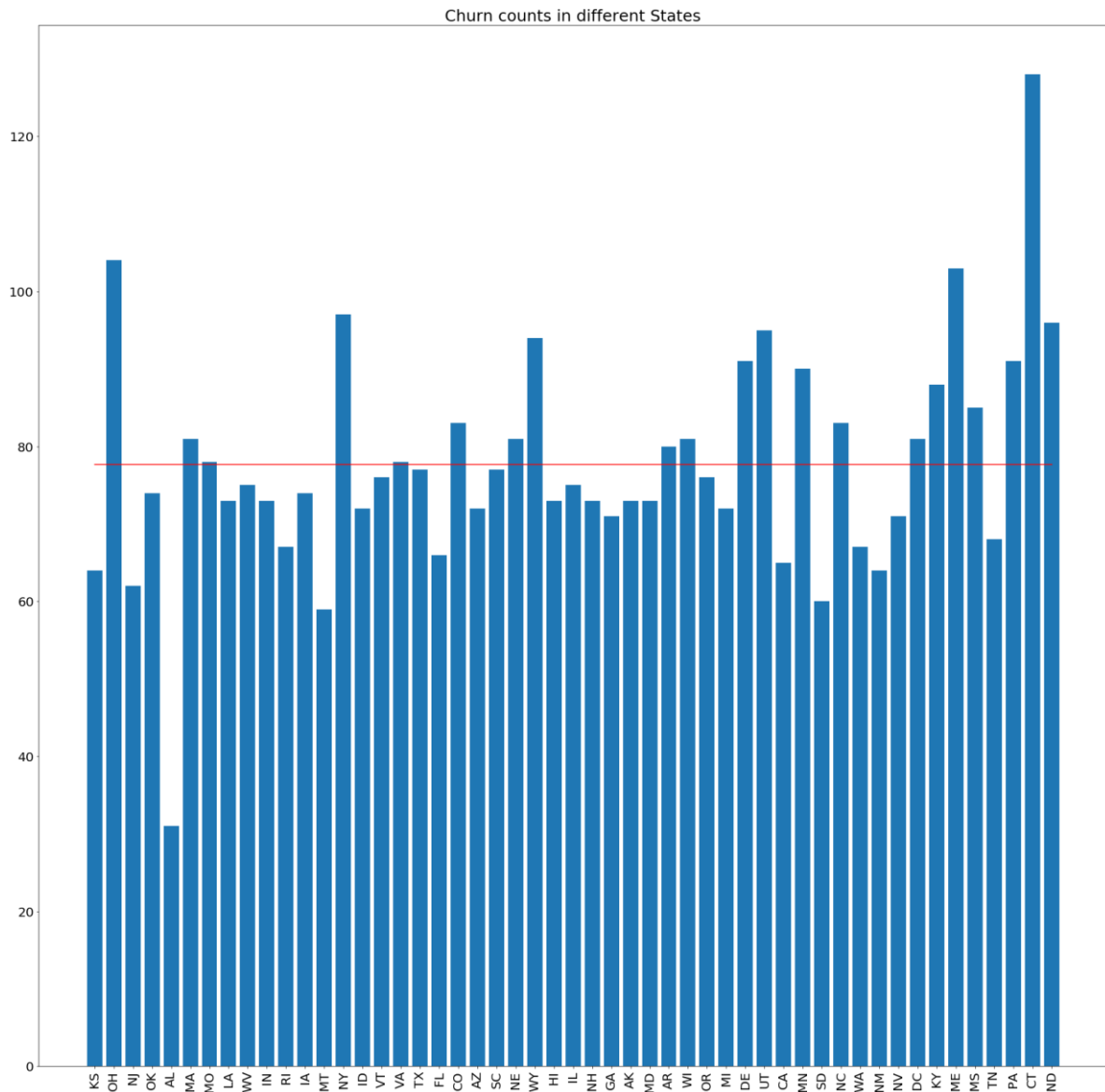


- The graph shows number of churns are far less than non-churns.



The above image shows the customer calls and vmail messages of churns and non churns.

- Here, the number of Customer calls of the churns are almost half the size of that of non-Churns . Whereas the vmail messages of churns are higher than that of non-churns. Hence, it can be said that customers leave the company because they don't get proper response from the Customer Care Executives. Also , that they call Customer Support as they face some issues in sending vmail messages or the calls . thus, these services must be reviewed. The customers could churn Due to poor quality of calls/vmail messages , or must have been overcharged.
  - However , it must not be forgotten that some customers could also do it without any valid or logical reason, based on their



The above image shows the number of churns in different states. The red line indicates the average of total number of churns in all the states.

➤ The derived level chart shows nuber of churns in the different states are :

-AL  
-CO  
-CT  
-ID  
-KS

-MA

-MI

-MN

-NH

-NJ

-NY

-OH

-OR

-RI

-TX

-UT

-VA

-VT

-WI

-WV

-WY

However, Three countries have hyeghest number of churns i.e.

\*WV : 128

\*AL: 108 and

\*VA : 103

### ML Model Summary

The Machine learning model was created with “Churn “ as target variable, and the rest of the dataset was used for training the model.

Different Classifiers were tried for accomplishing the task of churn prediction.They include :

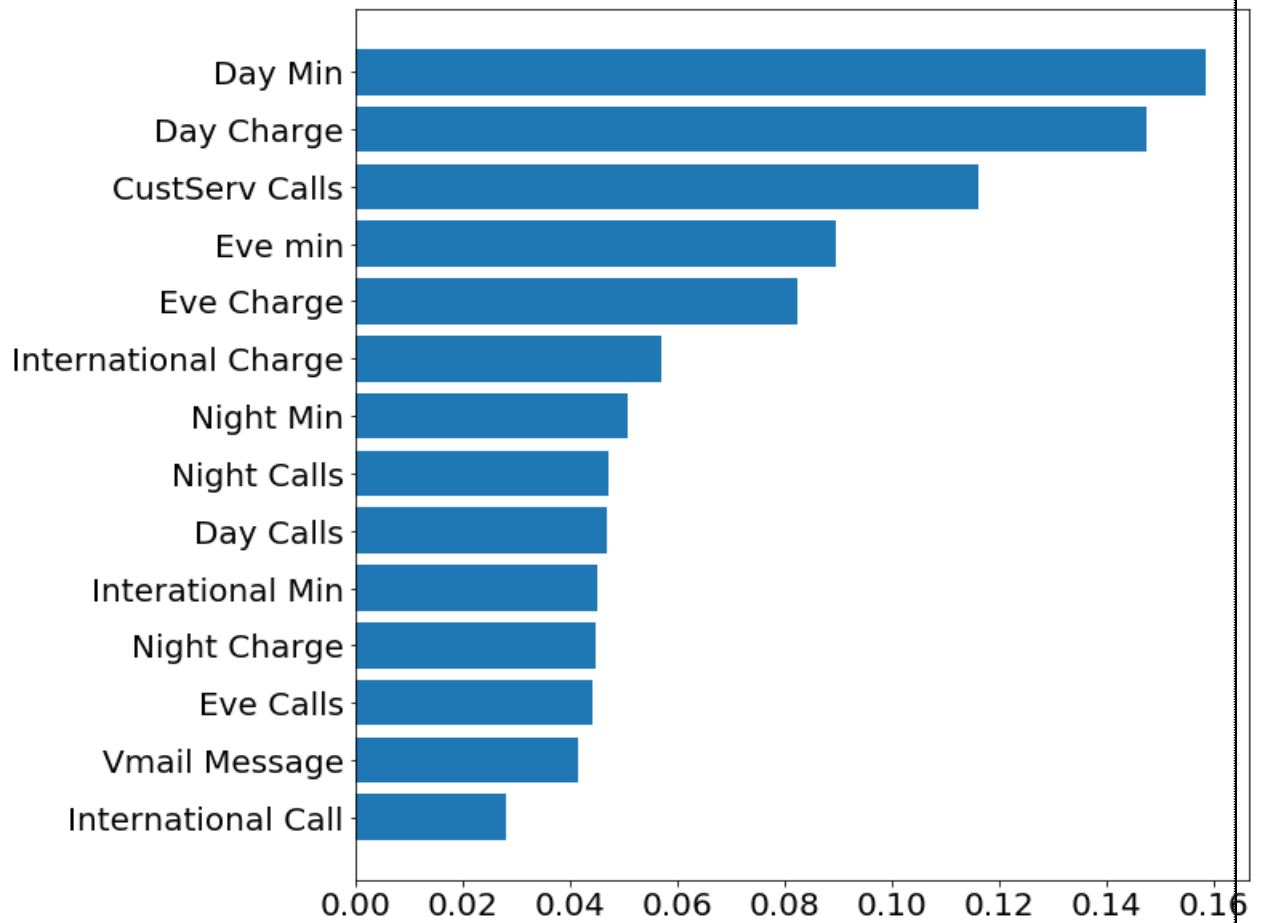
- Support Vector Machine
- Decision Tree Classifier
- Logistic Regression
- Random Forest .

The accuracy scores for all the models were calculated , and the model with the best score was selected , which turned out to be Random Forest Classifier ,so , the rest of the processing was continued with this training model.

For FINE TUNING the model, ,the number of parameters and the values for hyper parameters were selected using “GridSearchCV”

However, “RandomizedSearchCV” was not used to avoid any complications.

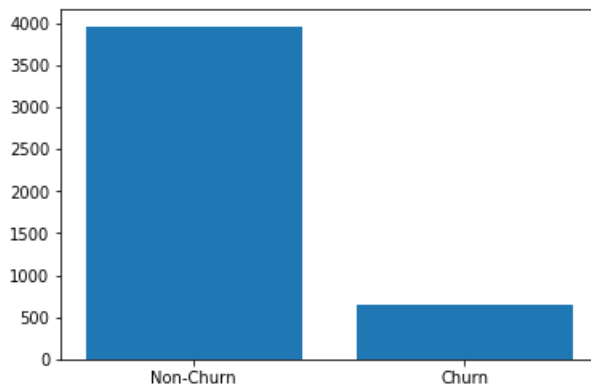
The “feature\_importances\_ “ was used to get the priority of the features based on which , the model predicts the Churn outcome.this sequence of the features is as follows:



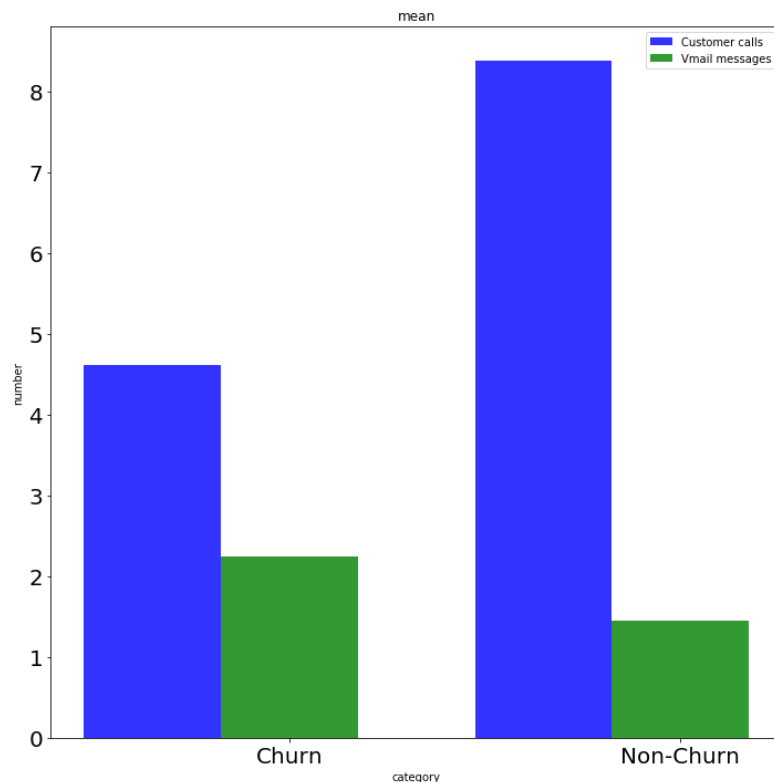


## Conclusion

The project explores various features responsible for the Churns of the customers from the telecomm industry. The key explorations of this project include:



- The number of churns are considerably less than the number of non-churns, which is definitely good for the telecom company, but however, the churns were due to some major problems that, if left unattended, might create a reasonable rise in the rate of the Churns.
- Most of the stats were almost similar for the Churns and the non-Churns, however there was a relatively higher difference found in the mean of the Vmail messages as well as Customer Service Calls, which is depicted below:



➤ F

➤ From all the states , 21 states were found to have the Churn rate above the average , which is 77. But , out of those , three states had highest Churn rate . these include:

D

State	Churn
WV	128
AL	104
VA	103

Therefore , these are the states that company should focus more on improving the services that were mentioned above i.e. Vmail messages, and customer calls.

For the future predictions of the customer churns , a model was built. After testing the different types of models, which included:

- Support Vector Machine

-Decision Tree Classifier

-Logistic Regression

-Random Forest Classifier.

The one with the best accuracy score was selected , i.e “ Random Forest Classifier”.

- The model was fine tuned using “GridSearchCV”. -
- The optimized model was exported to a file named “shubhs\_md1” for deployment and further predictions.
- The priorities based on which the model predicts the churn are as follows:

