

SYRIATEL TELECOMMUNICATIONS COMPANY CHURN PROJECT

BY KELSEY MAINA



PROJECT OUTLINE

- OVERVIEW
- BUSINESS AND DATA UNDERSTANDING
- MODELLING
- EVALUATION
- RECOMMENDATIONS
- NEXT STEPS



OVERVIEW

- This project aims to identify patterns and factors that will enable SyriaTel telecommunications company to anticipate customer behavior and take proactive steps to reduce churn rates. High churn rates impact revenue and customer satisfaction. A way to reduce churn is by enhancing customer service by, for example; improving service quality and personalizing services through data analytics.

BUSINESS UNDERSTANDING

BUSINESS PROBLEM;

SyriaTel is looking to obtain predictive machine learning models that can predict whether customers are likely to churn or not by leveraging data from the SyriaTel dataset that contains data on customers' usage patterns and services information.

OBJECTIVE 1

To build machine learning models that will predict how likely a customer will churn by analyzing customer features.

OBJECTIVE 2

To identify the specific features that influence the rate of customer churn.

OBJECTIVE 3

To determine the most accurate model in predicting the classification of churn/non-churn customers.

DATA UNDERSTANDING

- This project analysis uses previous customer data from SyriaTel telecommunications company.
- The dataset contains 3333 rows and 21 columns which include: demographic data , customer usage patterns and service information.
- The dataset is complete as it has no null values or duplicates.

MODELLING

MODEL 1;

LOGISTIC REGRESSION
MODEL

- Cross validation – 0.783
- Accuracy – 0.80
- Precision – 0.40
- Recall – 0.70
- F1 score – 0.51

Score:

0.7961019490254873

Cross validation:

[0.74015748 0.79002625 0.78843627 0.77135348 0.78449409]

Class report:

	precision	recall	f1-score	support
False	0.94	0.81	0.87	566
True	0.40	0.70	0.51	101
accuracy			0.80	667
macro avg	0.67	0.76	0.69	667
weighted avg	0.86	0.80	0.82	667

ROC_AUC:

0.75784557254312

MODEL 2; DECISION TREE CLASSIFIER

- Cross validation – 0.848
- Accuracy – 0.88
- Precision – 0.56
- Recall – 0.83
- F1 score – 0.67

Score:

0.8755622188905547

Cross validation:

[0.85433071 0.82939633 0.85151117 0.83442838 0.84362681 0.86070959]

Class report:

	precision	recall	f1-score	support
False	0.97	0.88	0.92	666
True	0.56	0.83	0.67	101
accuracy			0.88	667
macro avg	0.76	0.86	0.80	667
weighted avg	0.91	0.88	0.88	667

ROC_AUC:

0.8575376972326207

MODEL 3; RANDOM FOREST

- Cross validation – 0.88
- Accuracy – 0.90
- Precision – 0.70
- Recall – 0.57
- F1 score – 0.63

Score:

0.8980509745127436

Cross validation:

[0.87270341 0.89370079 0.89224704 0.87516426 0.89224704 0.90670171]

Class report:

	precision	recall	f1-score	support
False	0.93	0.96	0.94	566
True	0.70	0.57	0.63	101
accuracy			0.90	667
macro avg	0.81	0.77	0.79	667
weighted avg	0.89	0.90	0.89	667

ROC_AUC:

0.7650439072175769

MODEL 4; KNN

- Cross validation – 0.90
- Accuracy – 0.84
- Precision – 0.48
- Recall – 0.68
- F1 score – 0.56

Score:

0.8380809595202399

Cross validation:

[0.90419948 0.89107612 0.90932983 0.8935611 0.89093298 0.91852825]

Class report:

	precision	recall	f1-score	support
False	0.94	0.87	0.90	566
True	0.48	0.68	0.56	101
accuracy			0.84	667
macro avg	0.71	0.77	0.73	667
weighted avg	0.87	0.84	0.85	667

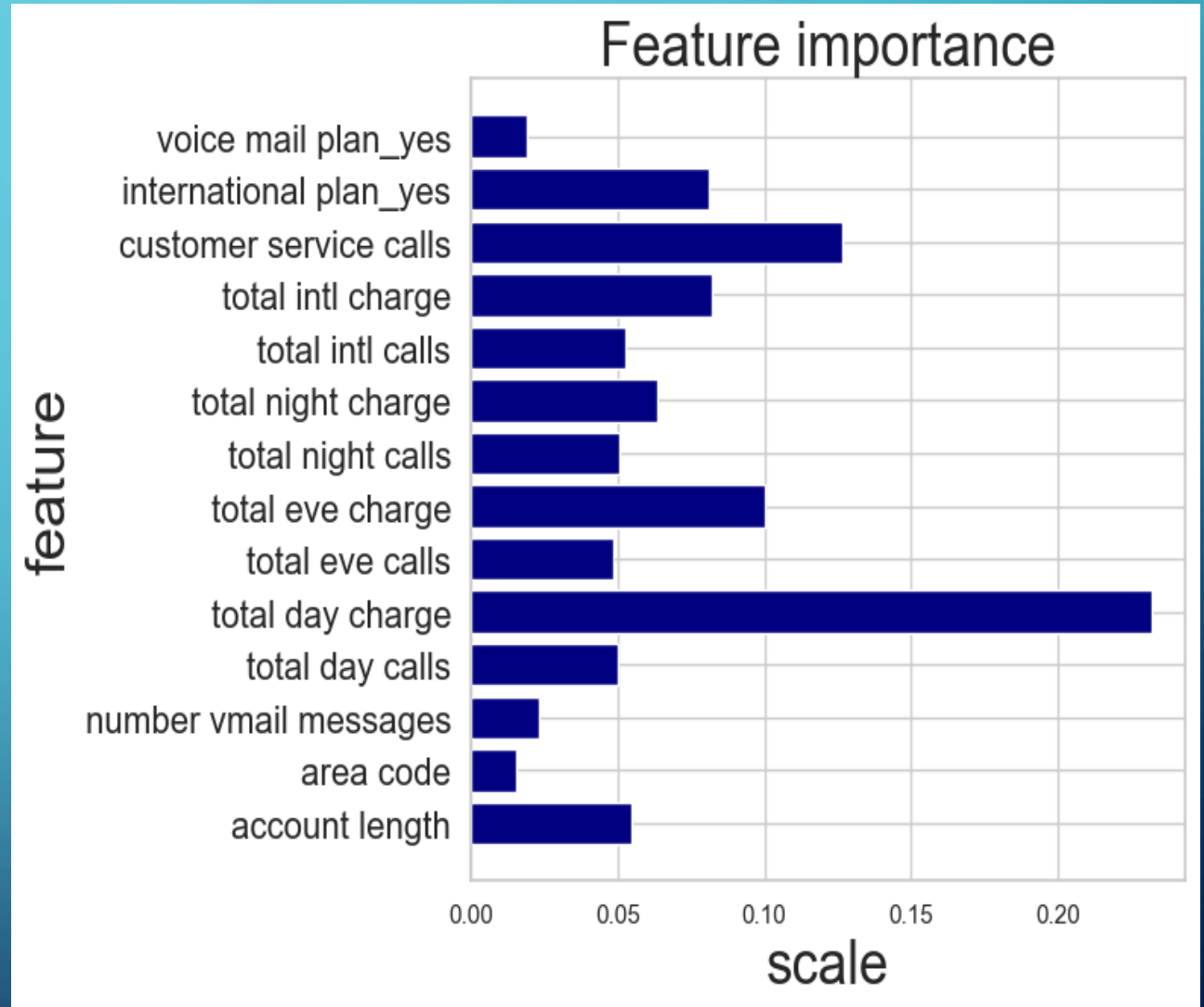
ROC_AUC:

0.7744463492285625

FINDING THE MOST IMPORTANT FEATURES;

The top features that influence the rate of customer churn (in order) are:

- total day charge
- customer service calls
- total eve charge



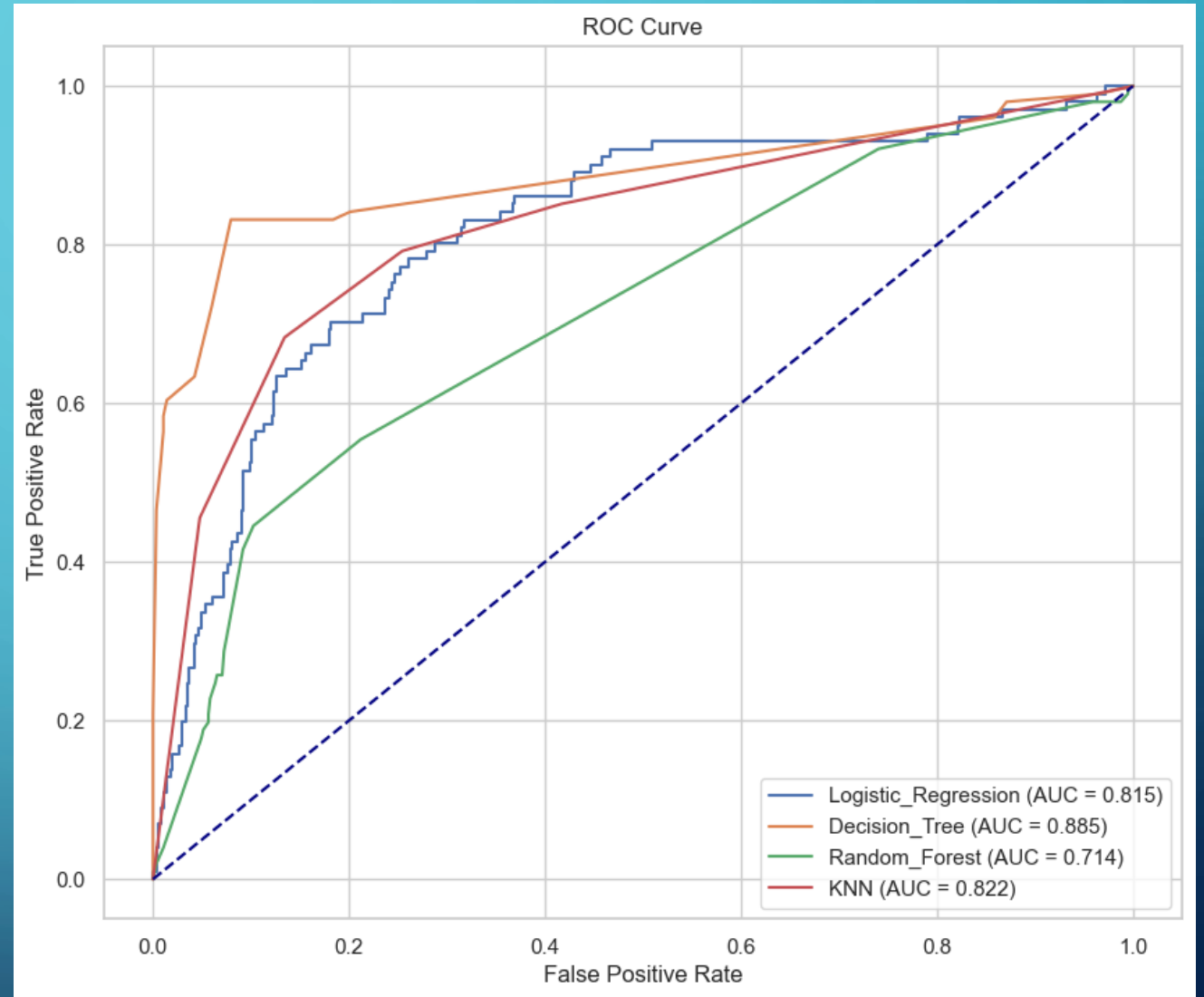
EVALUATION

- The **best** model is the **Decision tree** with a recall score of 0.832. For the Decision Tree, this means that the model is able to correctly identify around 83.2% of the churned customers in the test set.
- In the context of predicting customer churn for a telecommunications company, it is generally more important to focus on recall scores rather than accuracy.

MODEL	RECALL	ACCURACY
LOGISTIC REGRESSION	0.703	0.796
DECISION TREE	0.832	0.876
RANDOM FOREST	0.248	0.832
KNN	0.683	0.838

ROC CURVE ANALYSIS

The best model with the highest AUC is the **Decision Tree**, as it is the curve at the top left corner of the plot. This curve represents the model with the highest True Positive Rate (TPR) and the lowest False Positive Rate (FPR), which corresponds to the highest AUC (Area Under the Curve) of 0.885



RECOMMENDATIONS

1. Reduce the total day charge on calls.
2. Improve the service quality on customer service calls.
3. Reduce the total eve charge on calls.
4. Focus on offering International plans to customers.

NEXT STEPS

1. Develop incentive programs for day calls reduction.
2. Enhance customer service experience.
3. Explore strategies for eve charge reduction.
4. Expand and introduce appealing international plans.

THANK YOU!

CONTACT: [WWW.LINKEDIN.COM/IN/KELSEY-MAINA-7249222BB](https://www.linkedin.com/in/kelsey-maina-7249222bb)

