DSC - Phase 3 Project

Pamela Godia, Moringa School

10.05.2025

1. Business understanding

Project Goal

- To predict whether there is a pattern of customers who will ("soon“) stop (“churn”) doing business with SyriaTel, a telecommunications company.

Objectives

- To determine if there is a predictive pattern of customers who will ("soon") stop doing bussiness with SyriaTel.

Target Audience: Telecom business staff

2. Data understanding

Churn in Telecom's dataset

- 3333 entries
- 20 columns

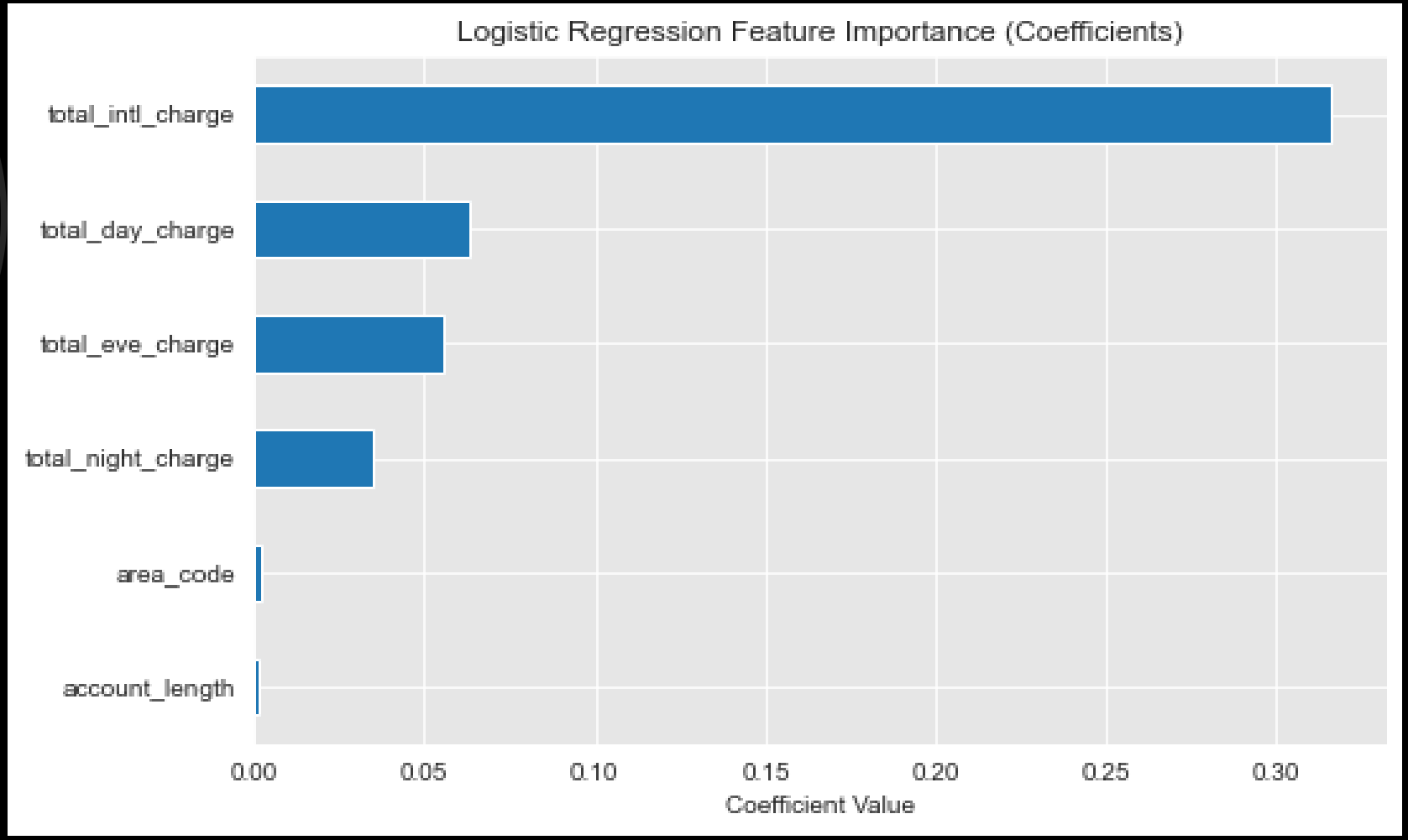
Target (y)

- Churn [Yes = 1], [No = 0]

Predictors

- ['account_length',
- 'area_code'
- 'total_day_charge'
- 'total_eve_charge'
- 'total_night_charge'
- 'total_intl_charge']

Logistic regression features (coefficients)



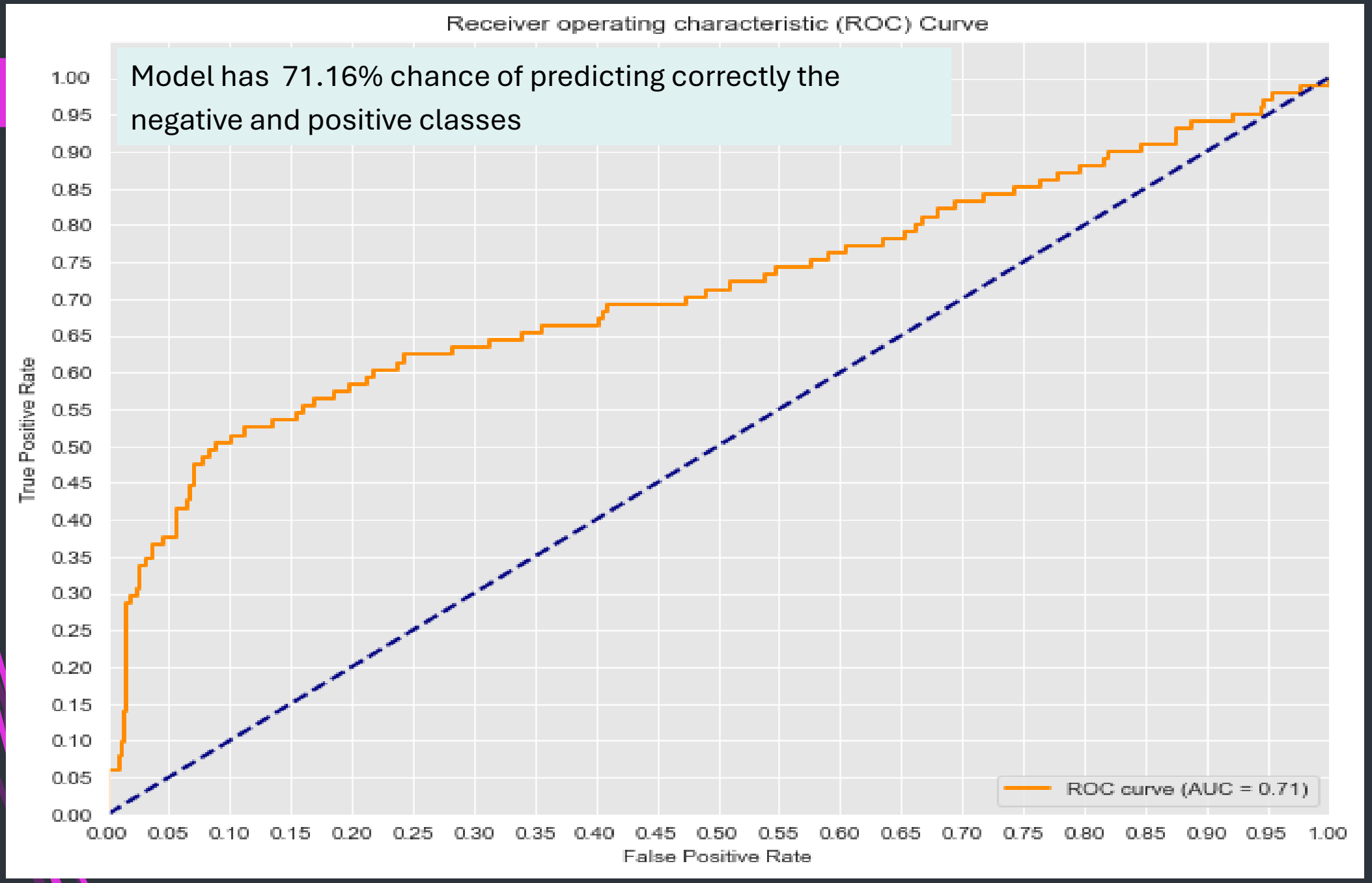
- ✓ The most important feature is the total_intl_charge
- ✓ A unit increase in international charge increases the chance of churn by 31%

3. Modelling using scikit-learn

Churn	Precision	Recall	F1 score	Support
0	0.86	1.00	0.92	2284
1	1.00	0.01	0.03	382
Accuracy			0.86	2666

- **The f1 score for class 0 is 0.92.**
- **The f1 score for class 1 is very low 0.02.**

Problem - data imbalances: 2284 instances for class 0: compared to 382 instances for class 1

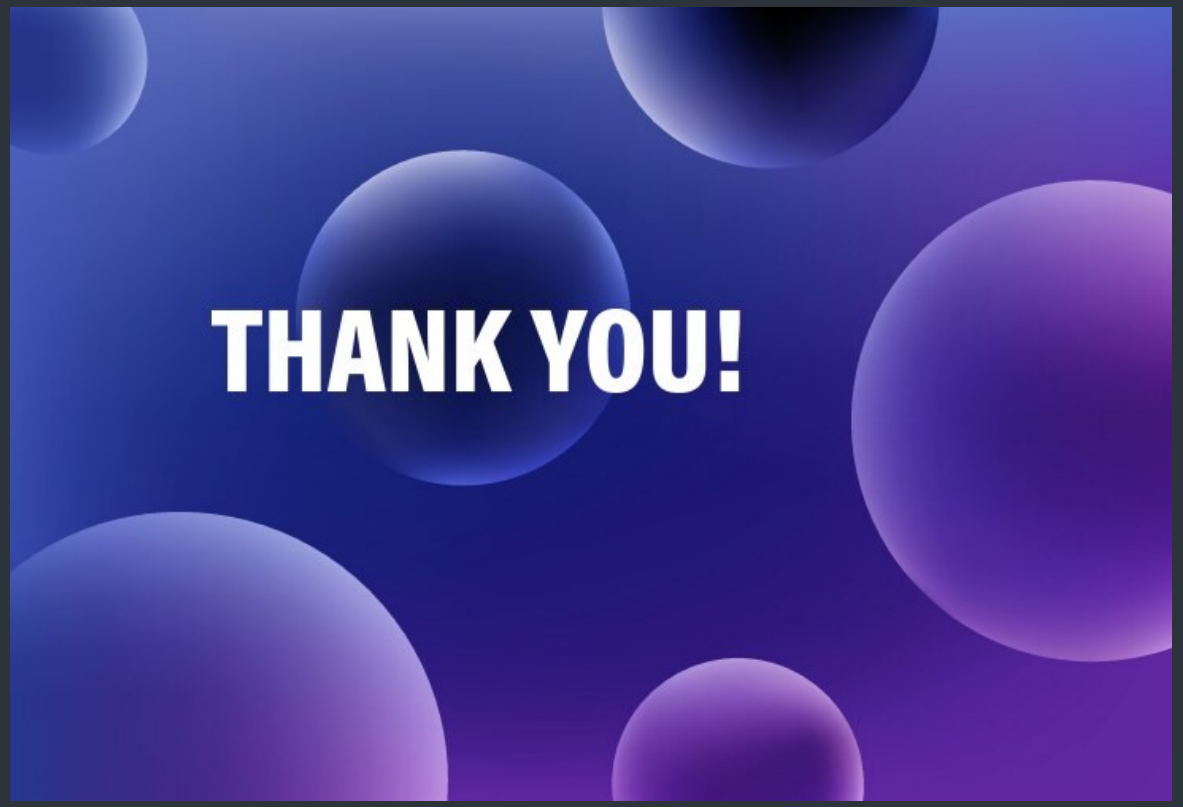

Receiver operating characteristic (ROC) Curve
Model has 71.16% chance of predicting correctly the negative and positive classes
True Positive Rate
False Positive Rate
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
ROC curve (AUC = 0.71)

Recommendations

- International Calls seems to be the main reason why customers may want to churn SyriaTel company
- It may be good to explore other variables such compare the international tariffs by other companies in the same sector
- One can also explore the local costs for the competitors in the market

Next steps

- ✓ SMOTE (Synthetic Minority Over-sampling Technique)
- ✓ Regularization - Apply L1 (Lasso) or L2 (Ridge)


THANK YOU!