

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 business with SyriaTel.

Target Audience: Telecom business staff

2. Data understanding

Churn in Telecom's dataset

- ▶ 3333 entries
- ▶ 20 columns

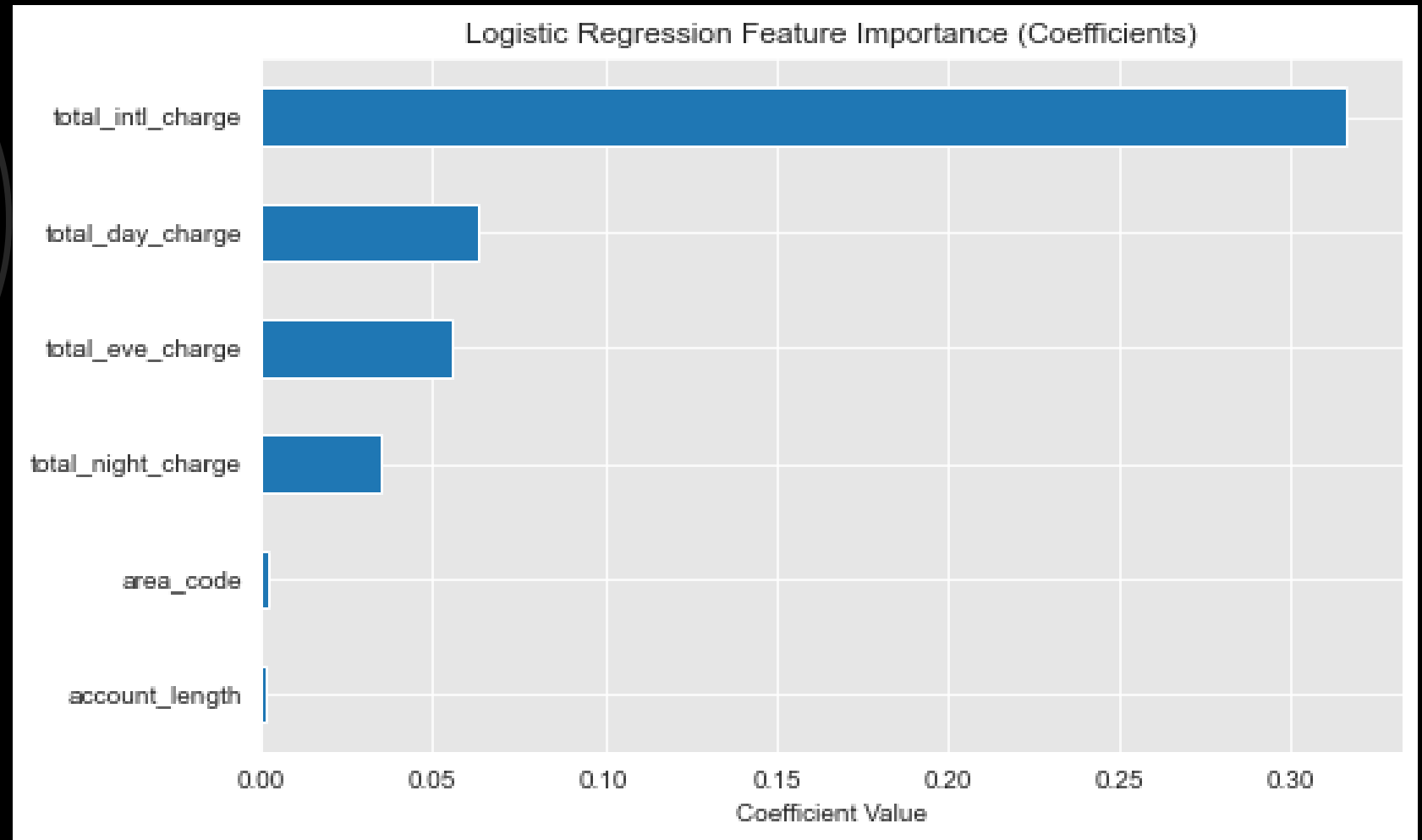
Target (y)

- ▶ 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%

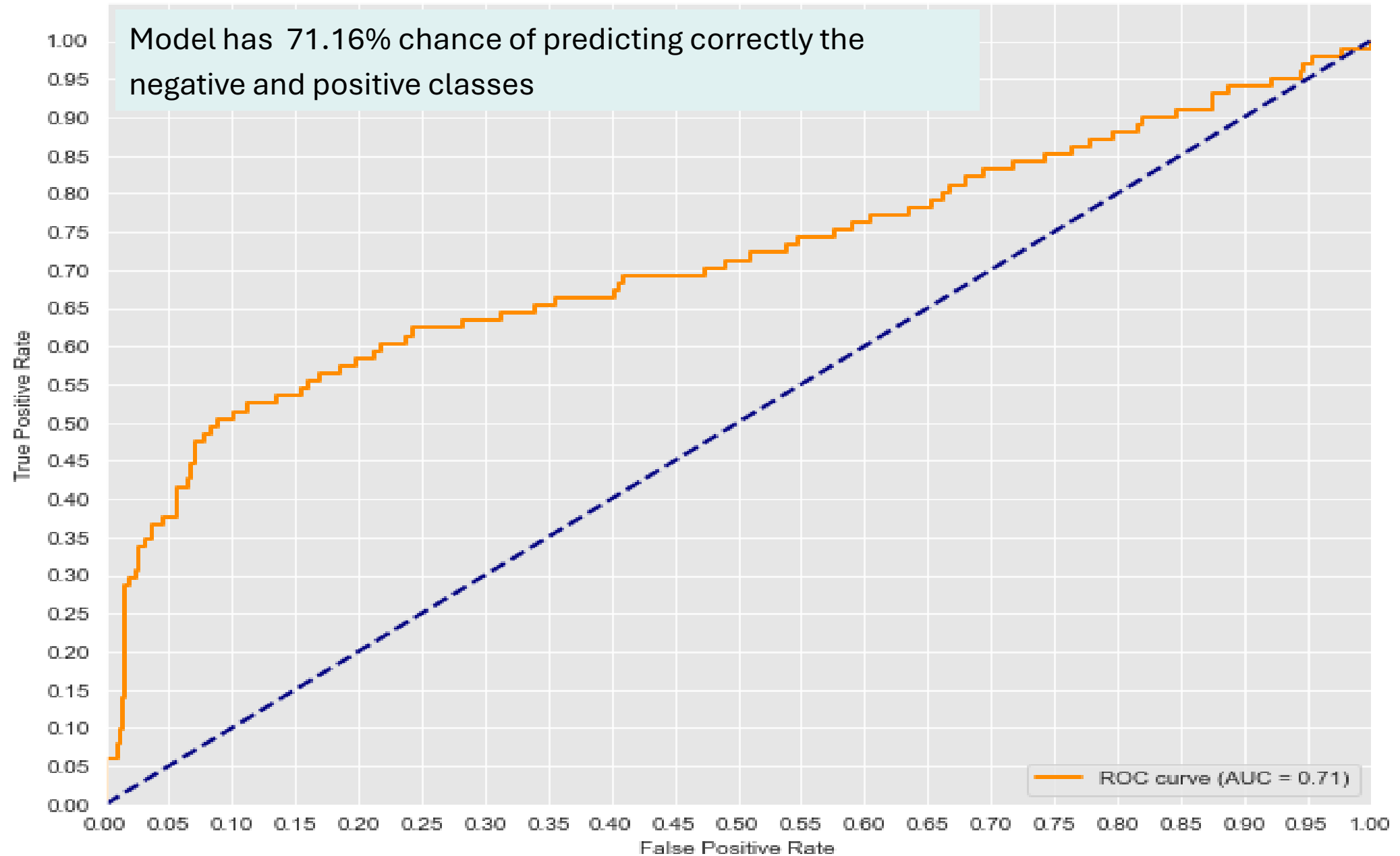
3. Modelling using scikit-learn

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 Improvement

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



THANK YOU!