

OVERVIEW

This project has been undertaken to determine how SyriaTel Company can predict if a customer will soon end their subscription to the telephone service.





BUSINESS UNDERSTANDING

• The stakeholder will be SyriaTel as they assess whether they can determine that a customer will opt-out of their service soon and the predictors that they will opt out. Then I will construct a predictive model that will give recommendations as to how they can reduce their costs incurred by customers who do not stay very long. This could be from reducing capital expenditure or revenue expenditure on certain services they offer.

KEY QUESTIONS

- 1. What are the features or determinants of customer churn?
- 2. What is the customer churn rate and how can it be improved upon based on the data?
- 3. What is the core services they
 offer and can costs of other services
 be reduced in order to maximize the
 core service?



MODELLING

METHODS USED

- 1. Simple Logistic Regression
- 2. SMOTE (Oversampling)
- 3. RandomUndersampling
- 4. Logistic Regression with Cross validation
- 5. Recursive Feature Elimination
- 6. SMOTE + Cross Validation

CHART TITLE

EVALUATION

Summary Table:

Model	Precision	Recall	F1-score	ROC-AUC	Notes
Baseline Logistic Regression	0.54	0.25	0.34	0.61	Biased to majority class
SMOTE	0.36	0.78	0.49	0.76	Better recall, lower precision
Undersampling	0.38	0.78	0.50	0.77	Slight ROC-AUC improvement
Cross Validation	0.66	0.17	0.27	0.58	Higher precision, lower recall
SMOTE + CV	0.38	0.78	0.52	0.78	Best F1-score
RFECV	0.36	0.03	0.06	0.51	Feature selection

GENERAL RECOMMENDATION

• No single model is perfect: There is always a trade-off between catching more churners (recall) and not bothering loyal customers (precision).

- Best model: The best model for the business use case would be SMOTE
 + CV as it is a better overall model in terms of F1 score
- **Next steps:** Consider trying more advanced models (e.g., Random Forest, XGBoost) and further tuning.

