SYRIATEL CUSTOMER CHURN PREDICTION

PHASE 3 PROJECT

Presentation by Christine Njeri Muhoro

OVERVIEW

- SyriaTel is a telecommunications company experiencing customer churn, where customers stop doing business with the company.
- Customer churn leads to significant revenue loss and reduced long-term profitability.
- Therefore, early identification of customer churns allows intervention with prevention of churn strategies.
- This project aims to build a model that predicts churn and minimizes chances of missing a churn.

OBJECTIVES

1. Develop a predictive model that classifies customers as likely to churn or not, allowing early churn prevention.
2. Identify factors(features) that highly influence a customer to churn.
3. Minimize chance of missing a customer who will churn i.e False Negatives. Getting a Recall Score of 81%.
4. Create a highly accurate model. Getting an Accuracy Score of 94%.

DATA UNDERSTANDING

The dataset used contained the churn target and the following customer details and factors affecting churn that were used in the modeling process:

- State where customer lives.
- account length - number of days customer has used the telecommunication company.
- international plan - Has customer subscribed to international calling plan?
- voice mail plan - Does customer have an active voicemail service?
- Number of voicemail messages customer currently has.
- Daytime Usage, Evening Usage, Night Usage and International Usage each having total number of minutes and total number of calls.
- Number of times customer contacted customer service.

MODELING AND EVALUATION

OBJECTIVE 1 :BUILDING THE CLASSIFIER

Logistic Regression Model

- Trained a baseline model for logistic regression. This model had no overfitting and underfitting. It also had a 76% recall and a 76% accuracy.
- Tuned the baseline logistic regression model to control and balance regularization. This model had a 77% recall and a 77% accuracy.

Decision Tree Classifier

- Trained a baseline decision tree classifier. This model had a 77% recall and a 91% accuracy.
- Tuned the baseline decision tree classifier with the best hyperparameters from a grid search cross-validation. This model had a 81% recall and a 94% accuracy.

From this model building and evaluation, the best model according to both recall and accuracy was the tuned decision tree classifier.

MODELING AND EVALUATION

OBJECTIVE 2 : IDENTIFY FEATURES THAT HIGHLY INFLUENCE CHURN

- The best model so far being the tuned decision tree classifier, its highest feature importance values were displayed.
- These highest values indicate strongest influence features(factors) on churn.
- The features influencing churn greatly are *total anytime minutes*, *customer service calls* and *international plan yes*.

MODELING AND EVALUATION

OBJECTIVE 3 : BEST RECALL SCORE I.E MINIMIZE FALSE NEGATIVES(MISSING A CHURNING CUSTOMER)

- Evaluated the 4 models created using accuracy, recall, precision scores and f1-score and created the table below with these results.

	Model	Accuracy	Recall	Precision	F1 Score
0	Baseline Logistic Regression	0.765	0.762238	0.351613	0.481236
1	Tuned Logistic Regression	0.768	0.769231	0.355987	0.486726
2	Baseline Decision Tree	0.914	0.769231	0.674847	0.718954
3	Tuned Decision Tree	0.946	0.811189	0.811189	0.811189

- From the table, the recall score increased as model building continued and the model with less false negative classification was the tuned decision tree with recall score of 81%.

MODELING AND EVALUATION

OBJECTIVE 4 : BEST ACCURACY SCORE I.E HIGHLY ACCURATE MODEL

	Model	Accuracy	Recall	Precision	F1 Score
0	Baseline Logistic Regression	0.765	0.762238	0.351613	0.481236
1	Tuned Logistic Regression	0.768	0.769231	0.355987	0.486726
2	Baseline Decision Tree	0.914	0.769231	0.674847	0.718954
3	Tuned Decision Tree	0.946	0.811189	0.811189	0.811189

- From the same table, the accuracy score also increased as model building continued and the highly accurate model was the tuned decision tree with an accuracy of 94%.

RECOMMENDATIONS

1. A classifier that predicts churn well was created and can hence be used for early identification of churn cases. I recommend deploying this model in real-time systems so that churns are flagged immediately.
2. Factors(features) that highly influence churn are such as customer service calls, total anytime minutes and international plan. I recommend focusing churn prevention strategies on these high impact factors.
3. Since high recall reduces false negatives i.e chance of missing a churn, I recommend prioritizing recall as a key performance metric when tuning future models.

CONCLUSION

- This project successfully addressed the business problem of customer churn at SyriaTel by developing a strong and highly accurate predictive model.
- Through model building and evaluation, the study identified key churn drivers and achieved strong performance particularly in recall and accuracy.
- The insights provided provide SyriaTel with the ability of early identification of customer churns and to implement churn prevention strategies.
- This project therefore helps reduce revenue loss and support the business's growth in the telecommunications industry.