Data Report

SyriaTel Telecom. Customer Churn

Business Understanding.

Syriatel has been leading the Syrian mobile telecommunication market since 2000. The company has successfully established its reputation by focusing on customer satisfaction and social responsibility.Understanding customer behavior is top of mind for every business today. Gaining insights into why and how customers buy can help grow revenue. But losing customers (also called customer churn) is always a risk, and insights into why customers leave can be just as important for maintaining revenues and profits.

Research Question.

The goal of this project is to understand customer behavior by building models to predict churn.

Objectives.

Main Objective.

To build a model that predicts whether a customer will churn or not..

Specific Objectives.

- To determine the relationships between customer churn and various predictor variables.
- To determine the top features affecting customer churn.
- Advice the client on best strategy

Data Understanding.

Data Source.

The dataset was sourced from Kaggle

https://www.kaggle.com/becksddf/churn-in-telecoms-dataset

Data Description.

The dataset has 3333 rows and 21 columns. The columns are as follows:

- state
- account length
- area code
- phone number
- international plan
- voice mail plan
- number vmail messages
- total day minutes
- total day calls
- total day charge

- total eve minutes
- total eve calls
- total eve charge
- total night minutes
- total night calls
- total night charge
- total intl minutes
- total intl calls
- total intl charge
- customer service calls
- churn

Data Preparation.

Loading the data.

The data set was then loaded into the jupyter notebook where it was worked. The data frame was then displayed to show the data it contains and how the variables relate to each other.

Tidying the dataset.

The data was checked for null values and duplicates whereby there was no missing data.
I proceeded to drop unnecessary columns i.e state, account length, area code, phone number.

Encoding Categorical Columns.

The voice mail plan column and International plan column have two unique values which is yes or no.

Import OrdinalEncoder and encode the two columns to numerical columns.

Handling imbalance

There was an imbalance in the target variable, I resampled the data using SMOTE then proceeded to split the features into train and test set.

Modeling.

I checked for the best performing model n terms of accuracy score, precision score and f1 score. The best performing model was Random Forests.

Accuracy and Precision score of different models
Random Forests has the highest precision and Accuracy score

	Accuracy score			Precision score	
	Models	ACC		Models	PRE
0	LR	0.757895	0	LR	0.751351
1	SVC	0.914912	1	SVC	0.927239
2	KNN	0.914035	2	KNN	0.866774
3	DT	0.908772	3	DT	0.900533
4	RF	0.955263	4	RF	0.966667
5	GBC	0.922807	5	GBC	0.955166

Implementing the Random Forests Model.

The hyperparameters were then tuned to improve model accuracy.

Using random search, I carried out several iterations to get the best conditions for random forest and run them.

The best parameters were:{'n_estimators': 400, 'max_features': 'auto', 'max_depth': None}

The classification report was as follows:

```
=== Confusion Matrix ===
[[570  15]
 [ 38 517]]


=== Classification Report ===
              precision    recall  f1-score   support

           0       0.94      0.97      0.96       585
           1       0.97      0.93      0.95       555

    accuracy                           0.95      1140
   macro avg       0.95      0.95      0.95      1140
weighted avg       0.95      0.95      0.95      1140



=== All AUC Scores ===
[0.97674361 0.98383503 0.99509388 0.9952293  0.99713143 0.99377039
 0.99508156 0.99356725 0.99127116 0.99112342]


=== Mean AUC Score ===
Mean AUC Score - Random Forest:  0.9912847029855341
```

Conclusion

- The top 10 important features are voice mail plan,total night charge, total night minutes, total international calls,number of voice mail messages,total intl minutes, total intl charge,total day calls, total night calls and total eve calls
- When customer service calls are more than 2 then the customer tends to leave.
- Customers with international plan churn at a higher rate than the customers without international plan churn.

Recommendation

- SyriaTel should improve the customer service.
- Syriatel should reduce charges for international calls in order to retain customers with international plan.