CUSTOMER CHURN MODEL

This is a predictive LR model built using python 3 in google colab.

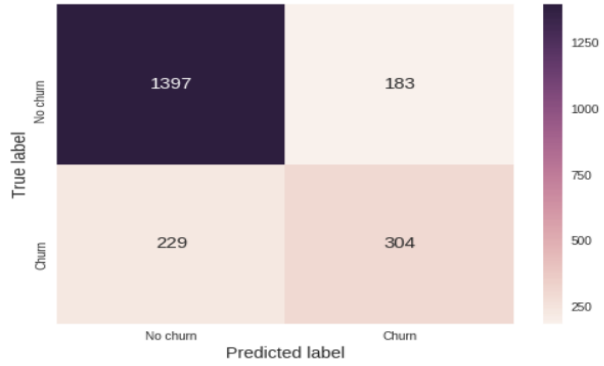
DATA INSIGHTS

1. The dataset has about 7043 rows. Each row gives details for all the individual customers like tenure, service type, monthly charges, total charges, etc.
2. The target for prediction is the churn column. This column indicates whether the customer cancelled their service.
3. From the dataset, it is determined that 'TotalCharges' column has 11 NaN/blank entries.
4. These blank 'TotalCharges' column values correspond to 0 months tenure. So, these values are adjusted to 0.
5. CustomerID column is dropped from the dataset as it is not used in the predictive model.
6. Correlations between churn and other customer features – applying dataframe.corr() and heatmap plotting, we find the following results.
   a. 'TotalCharges' is strongly correlated with 'MonthlyCharges' and 'Tenure' (with correlation value of 0.65 and 0.83). To avoid instability of coefficients, 'TotalCharges' variable is dropped from the dataset.
7. The dataset has categorical values, which are converted into numeric values, so the LR algorithm can process the data. For instance, gender['male','female'] entries are converted to [0,1], partner['yes','no'] entries are converted to [1,0] and so on.

CONFUSION MATRIX

```
confusion_matrix_df = pd.DataFrame(confusion_matrix, ('No churn', 'Churn'), ('No churn', 'Churn'))
heatmap = sns.heatmap(confusion_matrix_df, annot=True, annot_kws={"size": 15}, fmt="d")
plt.ylabel('True label', fontsize = 14)
plt.xlabel('Predicted label', fontsize = 14)
```

```
Text(0.5, 28.5, 'Predicted label')
```



ACCURACY

- The model gives the accuracy of 81%.
- But the precision and recall values are relatively low for the predictions in the positive class i.e., churn class. This implies that the dataset may be imbalanced.