Customer Churn Prediction

Presented By: Madhumita Chaudhary

AGENDA OF THE PRESENTATION

- Business Problem Understanding
- Deploying Machine Learning Models
- Recommendations on Marketing Campaigns and Strategies

Business Problem Understanding

This project belongs to Madhumita Chaudhary

Our task/objective as an analytics firm -

- To be able to predict customer churn
- Highlight variables or features contributing to the customer attrition
- Deploying machine learning algorithms to build a robust predictive model while identifying suitable model for the business
- Building a recommendation & marketing plan based on the data analysis to ensure lower churn or customer attrition

Overview of the Methodology

Exploratory Data Analysis (EDA)

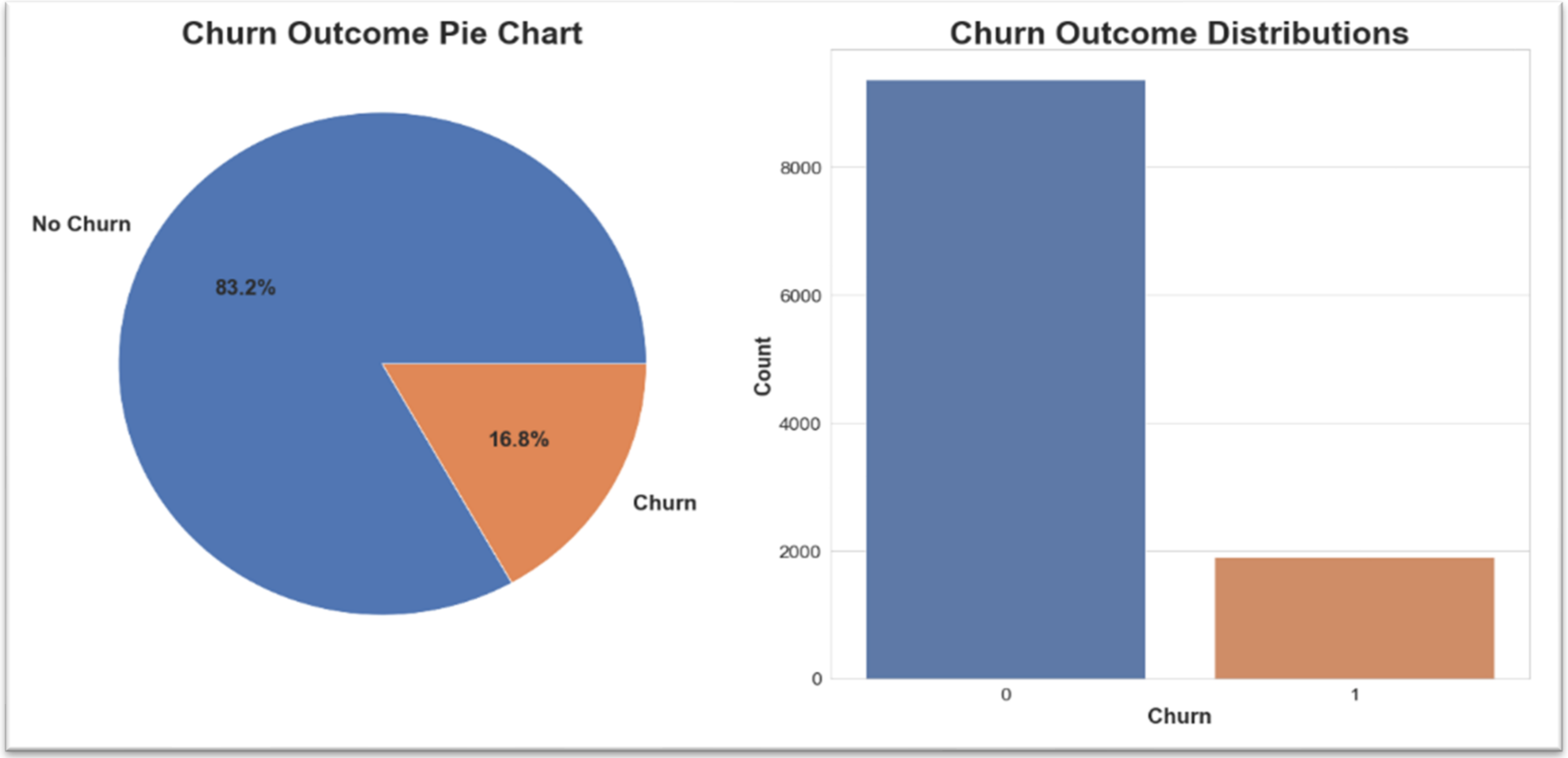
Building Classification Models

Formulating Strategies & Recommendations

Data Overview & Insights

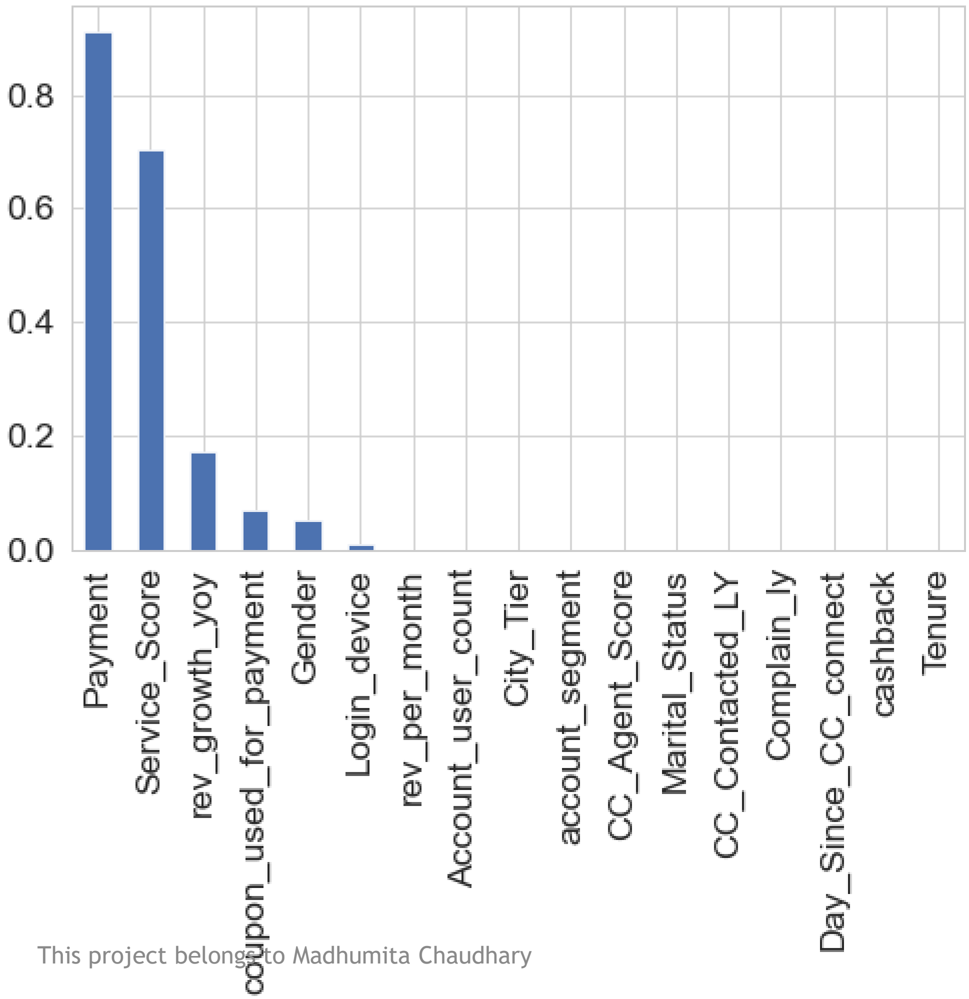
This project belongs to Madhumita Chaudhary

Target Customer Churn



Feature Correlation To Churn

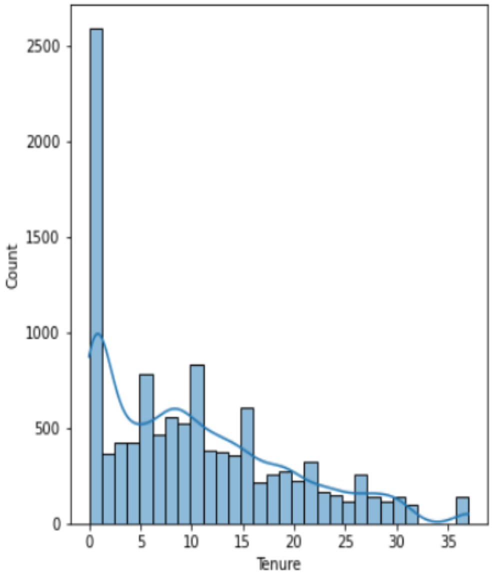
Least to most correlated

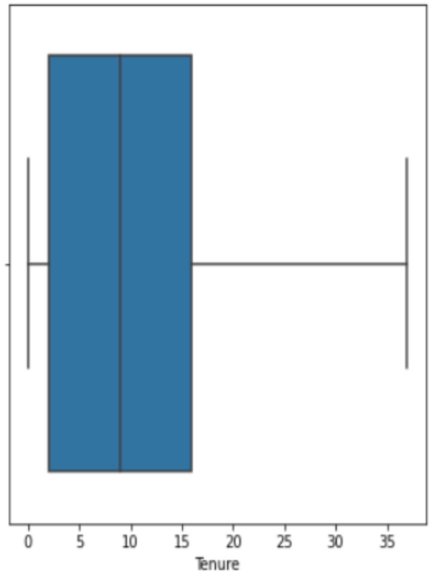


Most to least correlated

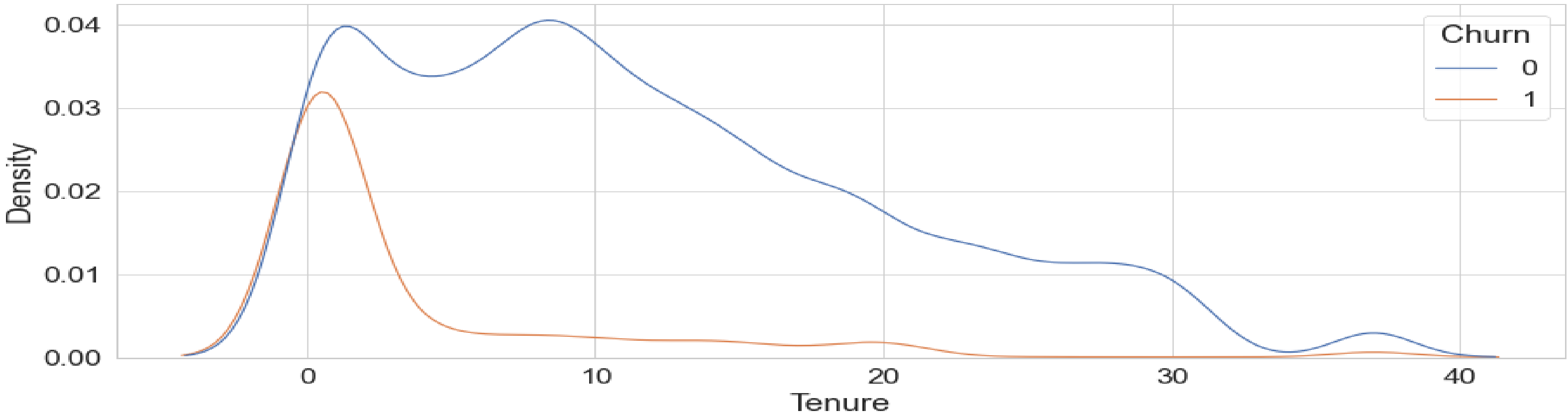
	Features	Score
0	Tenure	6789.36293
15	cashback	2182.30929
14	Day_Since_CC_connect	494.57281
11	Complain_ly	343.97039
2	CC_Contacted_LY	178.61295
9	Marital_Status	60.99436
8	CC_Agent_Score	51.86449
7	account_segment	26.96012
1	City_Tier	24.99334
6	Account_user_count	18.86355
10	rev_per_month	14.70173
16	Login_device	5.02118
13	coupon_used_for_payment	3.27528
4	Gender	1.89676
12	rev_growth_yoy	0.39053
5	Service_Score	0.06349
3	Payment	0.00188

Tenure Insights



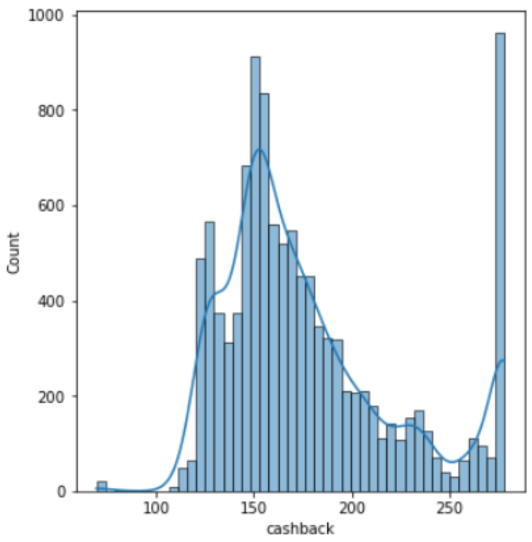


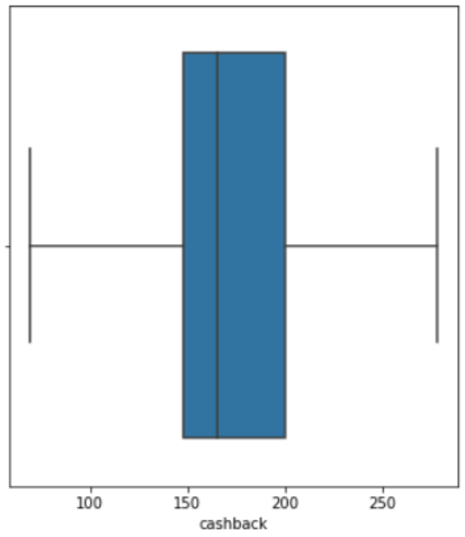
- The tenure for most customers has been less than one month
- On an average an account tenure with DTHExpress has been between 5 to 10 months



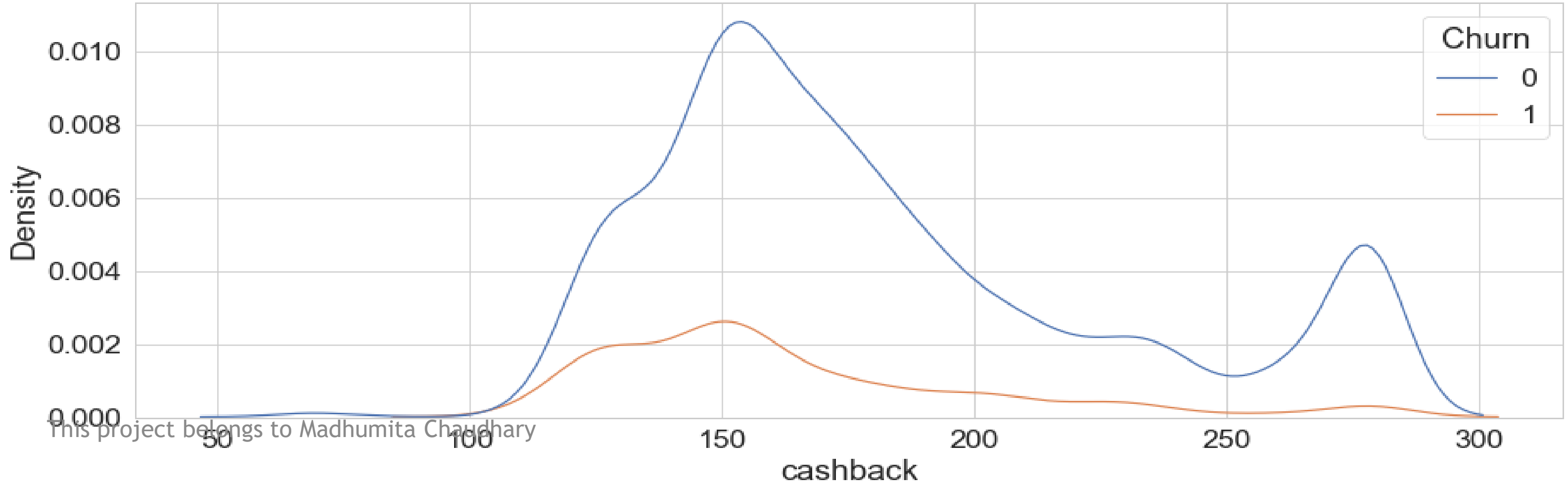
This project belongs to Madhumita Chaudhary

Cashback Insights

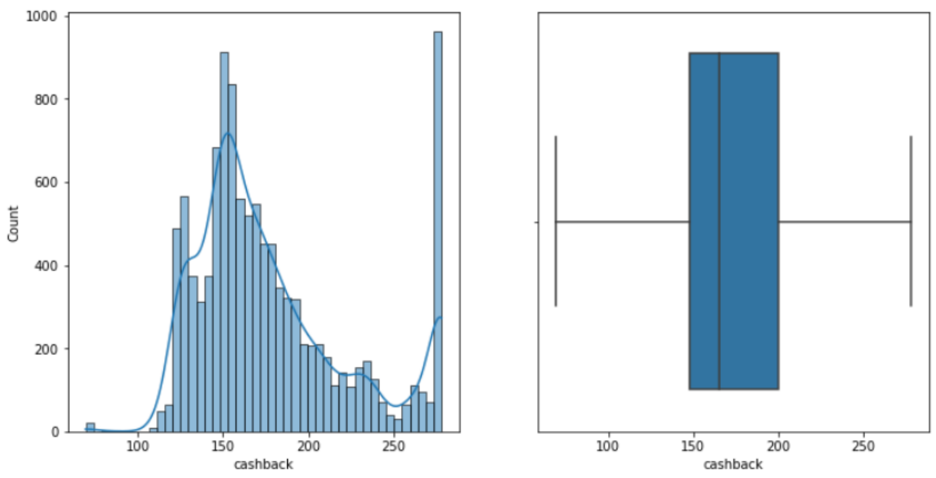




- On an average a customer has generated a cashback between 150-200 during last year.
- Customer cashback high indicates less churn

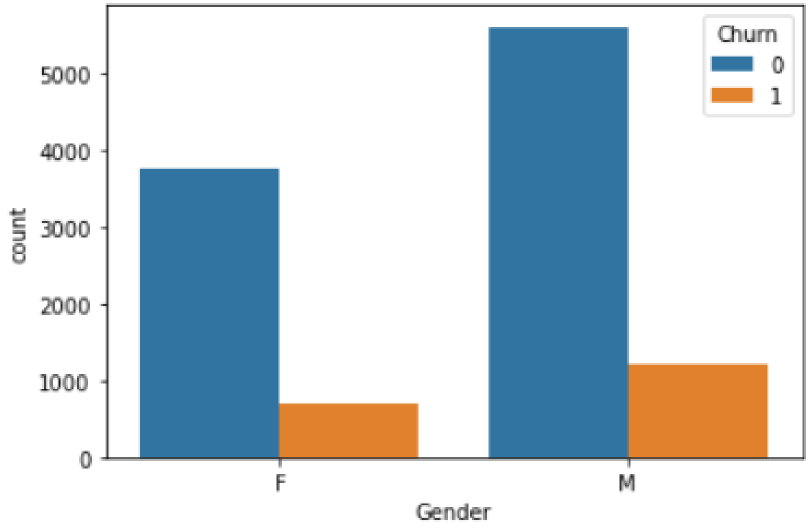


Cashback Insights

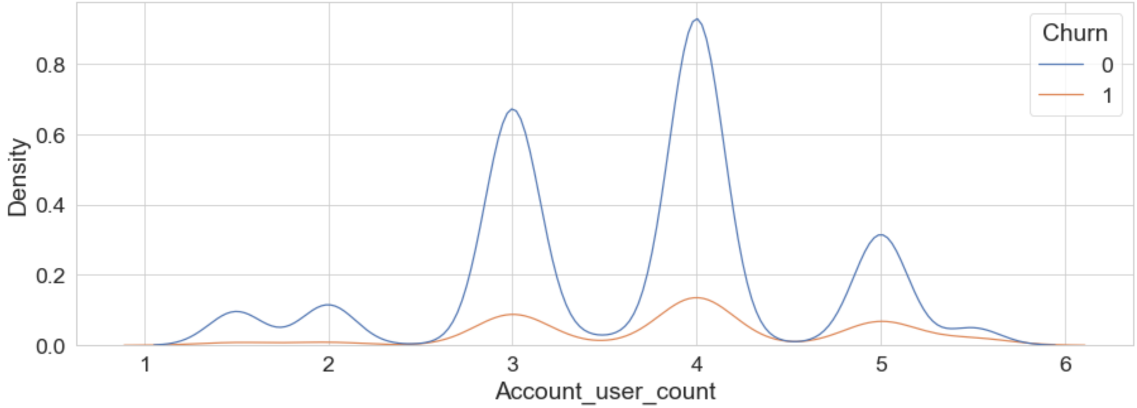


- On an average a customer has generated a cashback between 150-200 during last year.
- This could be a good indicator, suggesting that the customer is active

Gender Insights



- Number of male customers is considerably high; however male customers are also highest churners



Quick Dataset Overview

- We have the data for 11,260 customers for ExpressDTH
- The dataset gives us details on the several attributes related to a customer including their –
  - Gender
  - City_Tier,
  - preferred mode of payment
  - service satisfaction rating
  - preferred device of login
  - marital status, **among others.**

Data Modelling

This project belongs to Madhumita Chaudhary

Models Considered

- This is a classification problem, i.e., the predictions are made for two distinct classes (Churn: 1 or Loss: No Churn)
- For classification, the following models are suitable *(not an extensive list)*:
  - Logistic Regression
  - Linear Discriminant Analysis (LDA)
  - Random Forest
  - Artificial Neural Network (ANN)
  - SVM

Modeling Approach Used

- Of all models, the Support Vector Machine and KNN model was observed to be performing consistently on the training and test sets.
- To enhance the classification, hyperparameter optimization/tuning and SMOTE was performed on the Support Vector Machine model and KNN Model
- It was observed that the performance of the model improved after SMOTE. The scores of accuracy, precision, and recall witnessed an increase

Which Performance Metric Is Suitable?

- In churn analytics, both false negatives and false positives are not desirable. However, the cost associated with false negatives is higher, i.e., if a model predicts customer will not churn (0)
- Is Classification Accuracy a viable evaluation metrics to compare baseline our data set? The answer is No

Confusion Matrix		Actual	
		Will Churn (1)	Not Churn (0)
Predicted	Will Churn (1)	True Negative	False Positive
	Not Churn (0)	False Negative	True Positive

Key Metrics Considered *(In the Order of Preference)*

Preference	Performance Metric	Formula
1	Recall	True Positive / (True Positive + False Negative)
2	Precision	True Positive / (True Positive + False Positive)
3	F1 Score	2 * [(Precision * Recall) / (Precision + Recall)]

<table>
<tr><th colspan="2" rowspan="2">Confusion Matrix</th><th colspan="2">Actual</th></tr>
<tr><th>Will Churn (1)</th><th>Not Churn (0)</th></tr>
<tr><td rowspan="2">Predicted</td><td>Will Churn (1)</td><td>True Negative</td><td>False Positive</td></tr>
<tr><td>Not Churn (0)</td><td>False Negative</td><td>True Positive</td></tr>
</table>

Model Selected for Implementation (1/3)

We selected the SVM Model and KNN model for implementation in the real world

Train Data

KNN

	precision	recall	f1-score	support
0	0.89	0.97	0.93	6575
1	0.75	0.39	0.51	1307
accuracy			0.88	7882
macro avg	0.82	0.68	0.72	7882
weighted avg	0.87	0.88	0.86	7882

Support Vector Machine

	precision	recall	f1-score	support
0	0.93	0.99	0.96	2789
1	0.91	0.65	0.76	589
accuracy			0.93	3378
macro avg	0.92	0.82	0.86	3378
weighted avg	0.93	0.93	0.92	3378

Test Data

KNN

	precision	recall	f1-score	support
0	0.89	0.98	0.93	2789
1	0.78	0.41	0.54	589
accuracy			0.88	3378
macro avg	0.83	0.69	0.73	3378
weighted avg	0.87	0.88	0.86	3378

Support Vector Machine

	precision	recall	f1-score	support
0	0.94	0.99	0.97	6575
1	0.96	0.69	0.80	1307
accuracy			0.94	7882
macro avg	0.95	0.84	0.88	7882
weighted avg	0.94	0.94	0.94	7882

Model Selected for Implementation (2/3)

Test Data - Before SMOTE

KNN

	precision	recall	f1-score	support
0	0.89	0.97	0.93	6575
1	0.75	0.39	0.51	1307
accuracy			0.88	7882
macro avg	0.82	0.68	0.72	7882
weighted avg	0.87	0.88	0.86	7882

Support Vector Machine

	precision	recall	f1-score	support
0	0.93	0.99	0.96	2789
1	0.91	0.65	0.76	589
accuracy			0.93	3378
macro avg	0.92	0.82	0.86	3378
weighted avg	0.93	0.93	0.92	3378

Test Data - After SMOTE

KNN

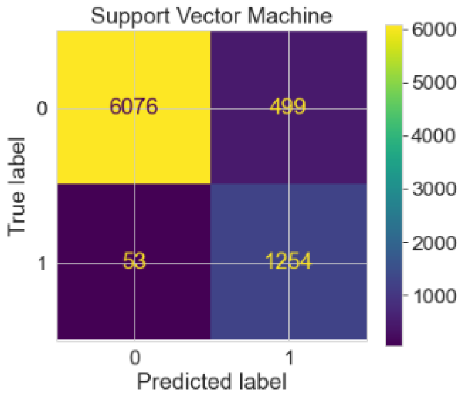
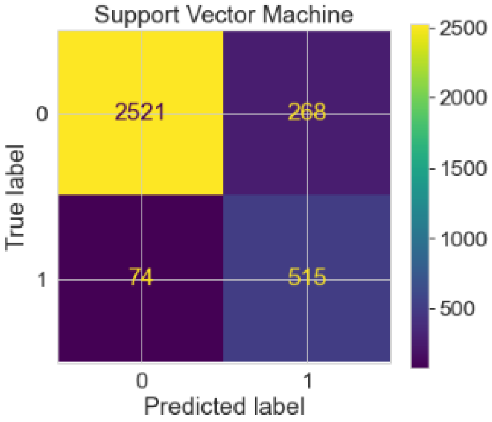
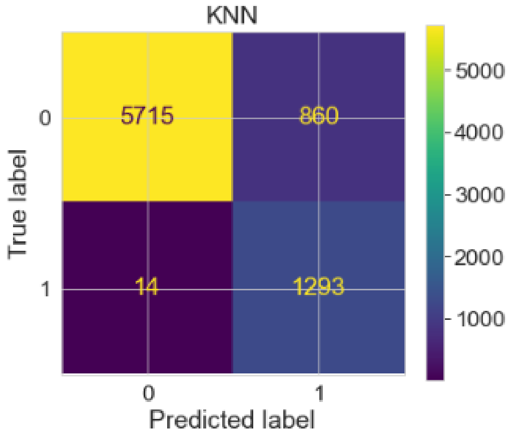
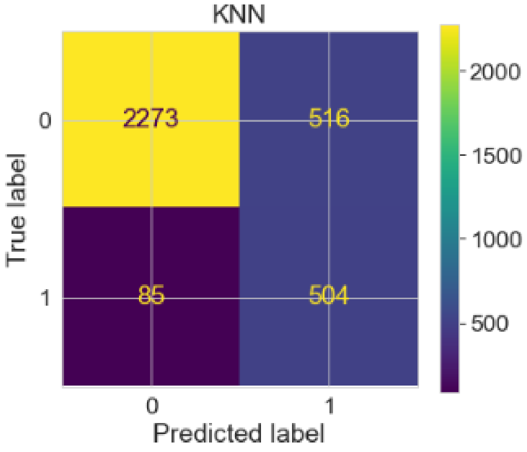
	precision	recall	f1-score	support
0	0.94	0.74	0.83	2789
1	0.39	0.78	0.52	589
accuracy			0.75	3378
macro avg	0.66	0.76	0.67	3378
weighted avg	0.84	0.75	0.77	3378

Support Vector Machine

	precision	recall	f1-score	support
0	0.97	0.90	0.94	2789
1	0.66	0.87	0.75	589
accuracy			0.90	3378
macro avg	0.81	0.89	0.84	3378
weighted avg	0.92	0.90	0.90	3378

Recall has improved drastically post SMOTE

Model Selected for Implementation (3/4)

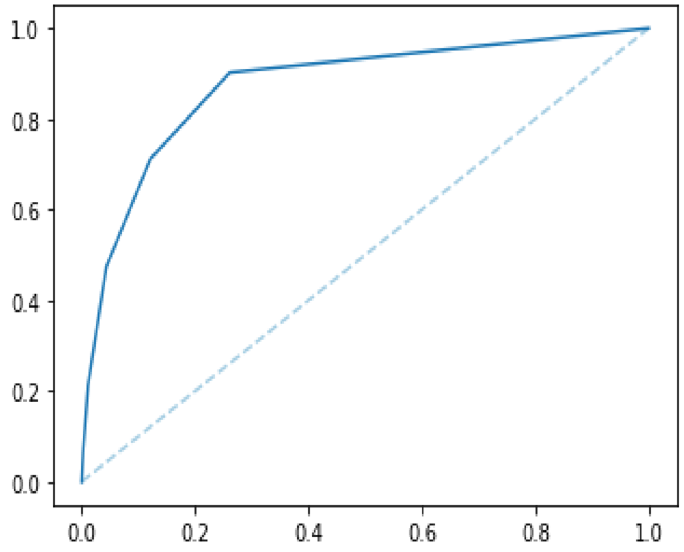
<table>
<tr><th>Model</th><th>Training Set</th><th>Test Set</th></tr>
<tr><td>SVM</td><td>
</td><td>
</td></tr>
<tr><td>KNN</td><td>
</td><td>
</td></tr>
</table>

Model Selected for Implementation (4/4)

SVM

AUC: 0.874

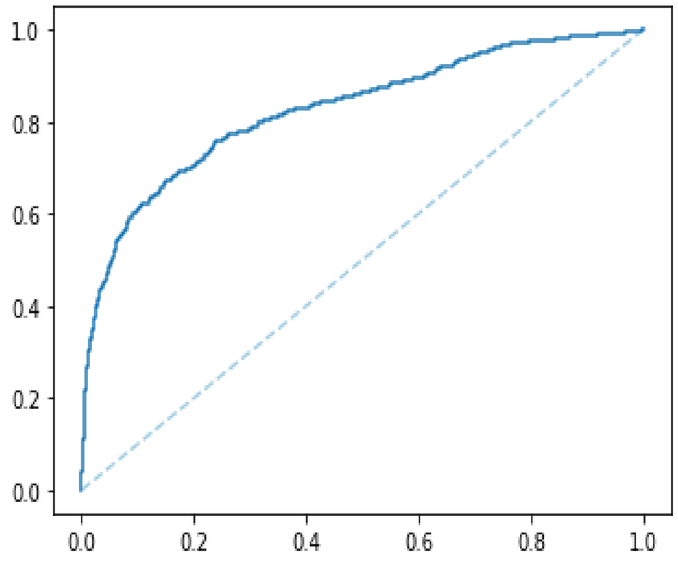
[<matplotlib.lines.Line2D at 0x1567b593ee0>]



KNN

AUC: 0.826

[<matplotlib.lines.Line2D at 0x15646019730>]



Insights, Recommendations and Predicting Future Results

This project belongs to Madhumita Chaudhary

INSIGHTS OBTAINED FROM EDA

- The churn rate for customer with higher tenure is less, whereas customers with less than 10month tenure with the company as account holder have a higher churn rate.
- The churn rate is highest for single customers with lower tenure. We need to check if there are certain concerns of issues resulting in higher churn for singles
- The percent of churners with agent service score above 3 also show a propensity to churn. We need to understand the underlying factors resulting in churn despite high agent service ratings

Recommendations

Recommendations

- Customer complaints should be taken up seriously since the number of calls received has been on a higher side. This could also help the company improve its satisfaction scores in terms of service and account service
- Offering reasonably priced family plans with add-on benefits to each user tagged to a primary account will help create revenue streams
- Focus on customers with long term tenure – offering them discounts, premium services at a discounted price. Along with DTH service offer other service plans
- Customer retention is the most significant revenue source so the focus should be on retaining existing customers
- Pay extra attention to complains and improve on key concern areas based on customer feedback

THANK YOU!

ANY QUESTIONS?