SyriaTel: Classifying Customers Based on Churn

May 2023

Business Goal Overview

To leverage machine learning processes in building a model that helps SyriaTel increase brand loyalty and identify consumer behaviors signaling a potential churn
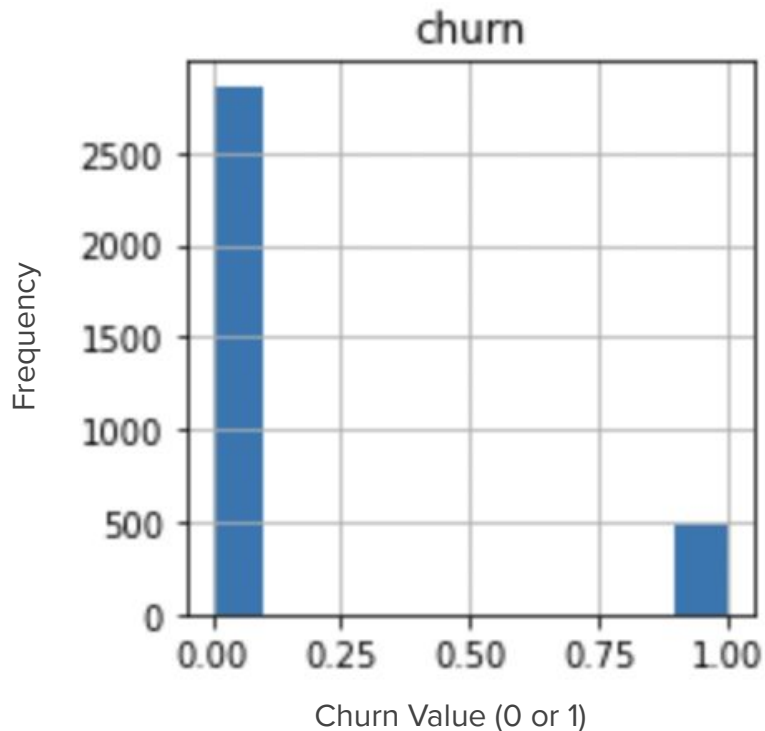
The Data: Raw Dataset Overview

	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 calls	total eve charge	total night minutes	total night calls	total night charge	total intl minutes	total intl calls	total intl charge
0	KS	128	415	382-4657	no	yes	25	265.1	110	45.07	...	99	16.78	244.7	91	11.01	10.0	3	2.70
1	OH	107	415	371-7191	no	yes	26	161.6	123	27.47	...	103	16.62	254.4	103	11.45	13.7	3	3.70
2	NJ	137	415	358-1921	no	no	0	243.4	114	41.38	...	110	10.30	162.6	104	7.32	12.2	5	3.29
3	OH	84	408	375-9999	yes	no	0	299.4	71	50.90	...	88	5.26	196.9	89	8.86	6.6	7	1.78
4	OK	75	415	330-6626	yes	no	0	166.7	113	28.34	...	122	12.61	186.9	121	8.41	10.1	3	2.73

5 rows × 21 columns

- The raw dataset contains 21 columns and 3,333 rows
  - The target variable that we're classifying customers based on is "churn"
  - The remaining 20 columns describe customer behavior (i.e. "total day calls", "state", etc.)

The Data: Target Variable "Churn"

- The bar plot shows the distribution of "churn" data

- 0 = no churn; 1 = churn

- The current churn rate is around 14.49%

- We further account for this imbalance in our analysis

The Data: Further Processing

- Since some columns come in different formats, we standardized the data before analyzing

- We also created a heatmap to remove one of any pairs of columns that are highly correlated

- For our analysis, we use 16 of the original 21 columns and the following dataset is used for analysis:

	state	account length	area code	international plan	voice mail plan	number vmail messages	total day calls	total day charge	total eve calls	total eve charge	total night calls	total night charge	total intl calls	total intl charge	customer service calls	churn
0	16	128	415	0	1	25	110	45.07	99	16.78	91	11.01	3	2.70	1	0
1	35	107	415	0	1	26	123	27.47	103	16.62	103	11.45	3	3.70	1	0
2	31	137	415	0	0	0	114	41.38	110	10.30	104	7.32	5	3.29	0	0
3	35	84	408	1	0	0	71	50.90	88	5.26	89	8.86	7	1.78	2	0
4	36	75	415	1	0	0	113	28.34	122	12.61	121	8.41	3	2.73	3	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3328	3	192	415	0	1	36	77	26.55	126	18.32	83	12.56	6	2.67	2	0
3329	49	68	415	0	0	0	57	39.29	55	13.04	123	8.61	4	2.59	3	0
3330	39	28	510	0	0	0	109	30.74	58	24.55	91	8.64	6	3.81	2	0
3331	6	184	510	1	0	0	105	36.35	84	13.57	137	6.26	10	1.35	2	0
3332	42	74	415	0	1	25	113	39.85	82	22.60	77	10.86	4	3.70	0	0

3333 rows × 16 columns

Results

Final Model: Decision Tree Classifier

Throughout our analysis, we tested three algorithms. Our final Decision Tree model shows that the most relevant columns are:
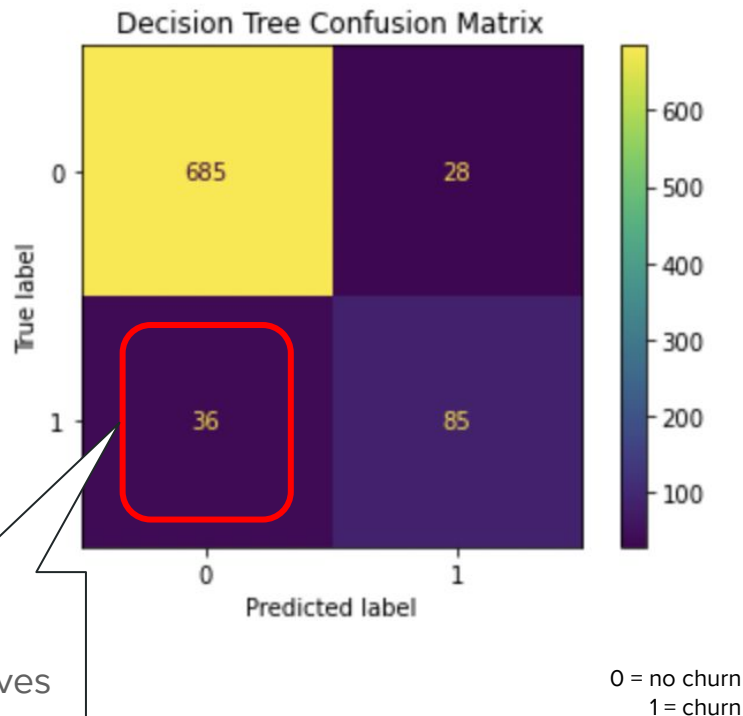
- ❖ Total day charge
- ❖ Total eve charge
- ❖ Customer service calls

Results: Decision Tree Model

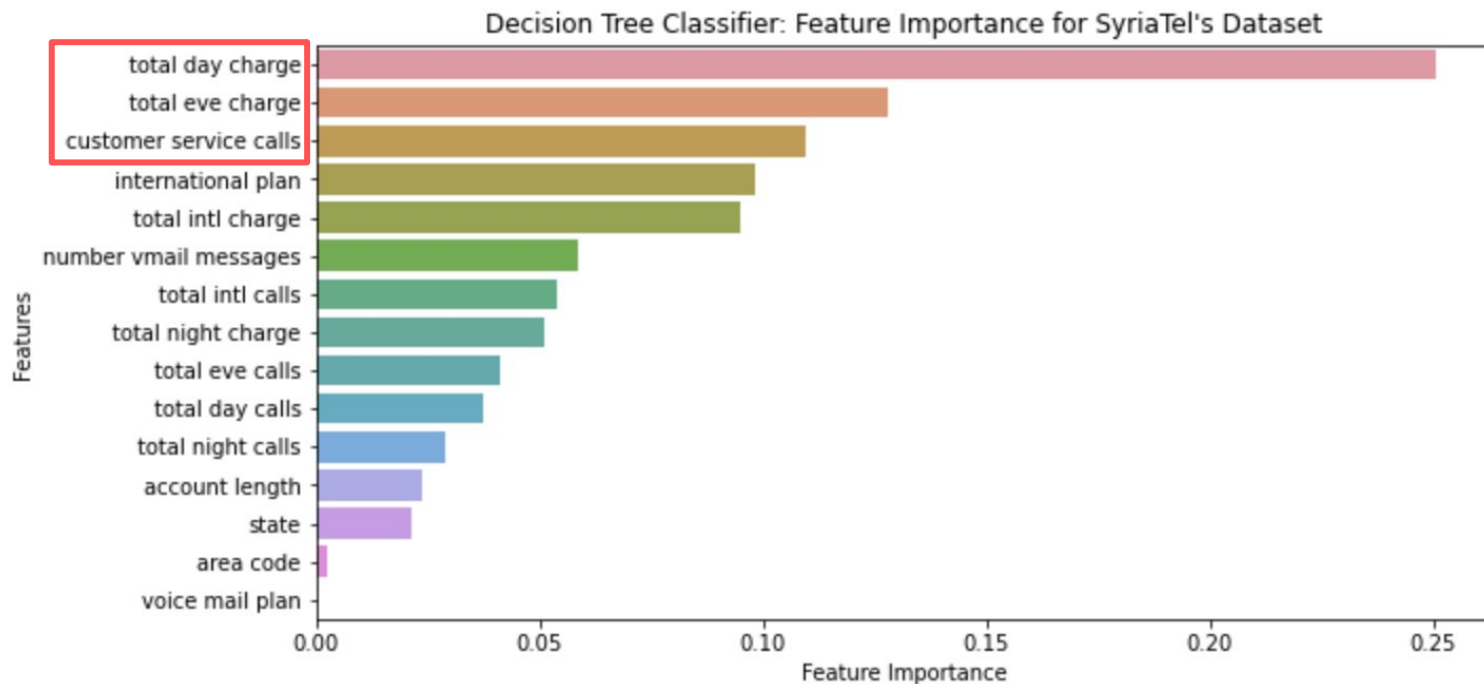We evaluated our model's performance against four metrics:

- ❖ Accuracy: 0.92
- ❖ Precision: 0.75
- ❖ Recall: 0.70
- ❖ F1 (combination of recall & precision): 0.73

The model is tuned to maximize recall, because doing so reduces false negatives. In other words, we'd rather overestimate how many clients churn than how many don't.

Decision Tree Confusion Matrix

We have 36 false negatives

0 = no churn
1 = churn

Results: Feature Importances

Decision Tree Classifier: Feature Importance for SyriaTel's Dataset

High day and evening charges along with increased customer service calls are indicators of customers who are likely to churn

Results: Feature Importances

	Total day charge	Total eve charge	Customer service calls
25th percentile	24.43	14.16	1.00
50th percentile	30.50	17.12	1.00
75th percentile	36.79	20.00	2.00

SyriaTel can use the percentiles of the top three features as a benchmark for which customers to target when working to increase brand loyalty and retention

Takeaway and Recommendation

SyriaTel can start by applying these strategies towards customers near the 75th percentile:

- Added value discounts
- Rewards programs
- Omnichannel support

SyriaTel must also ensure price transparency so customers understand the value of services being paid for

Email: kezia.setokusumo@stern.nyu.edu
Github: github.com/keziasetokusumo/p3_project
LinkedIn: www.linkedin.com/in/keziasetokusumo/