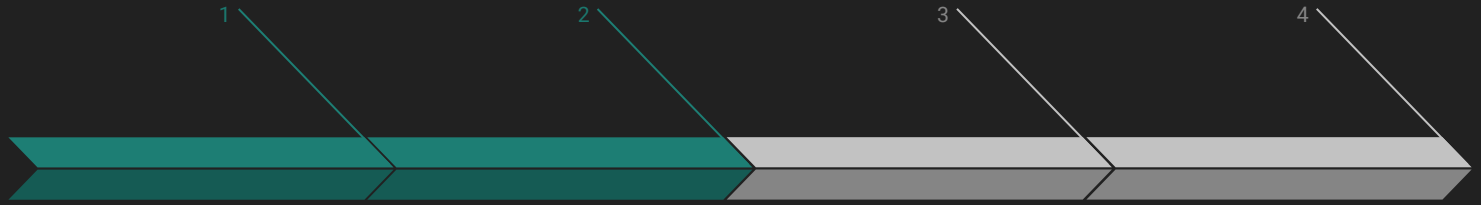
SyriaTel

Customer Retention

T. Hale Robert



Analysis Intent

Identify customers within SyriaTel's existing customer base that are likely to churn

Methods

Explored churn rate as a function of available data features, selected classification models

Recommendations

Suggested approaches to continue this work into the future

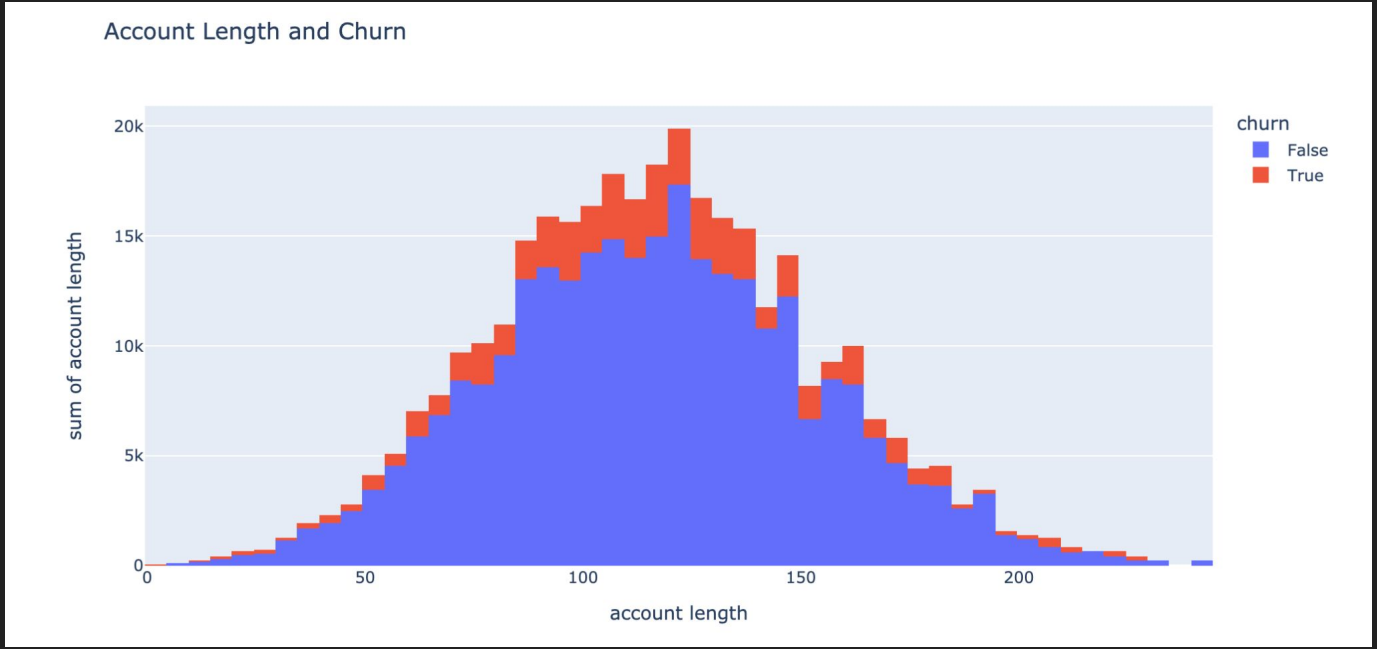
Conclusion

Growth areas and limitations of the existing project

About the data:

- 3300 customer records
- Mix of categorical & continuous data types
- US based, 51 states represented

Exploratory Data Analysis

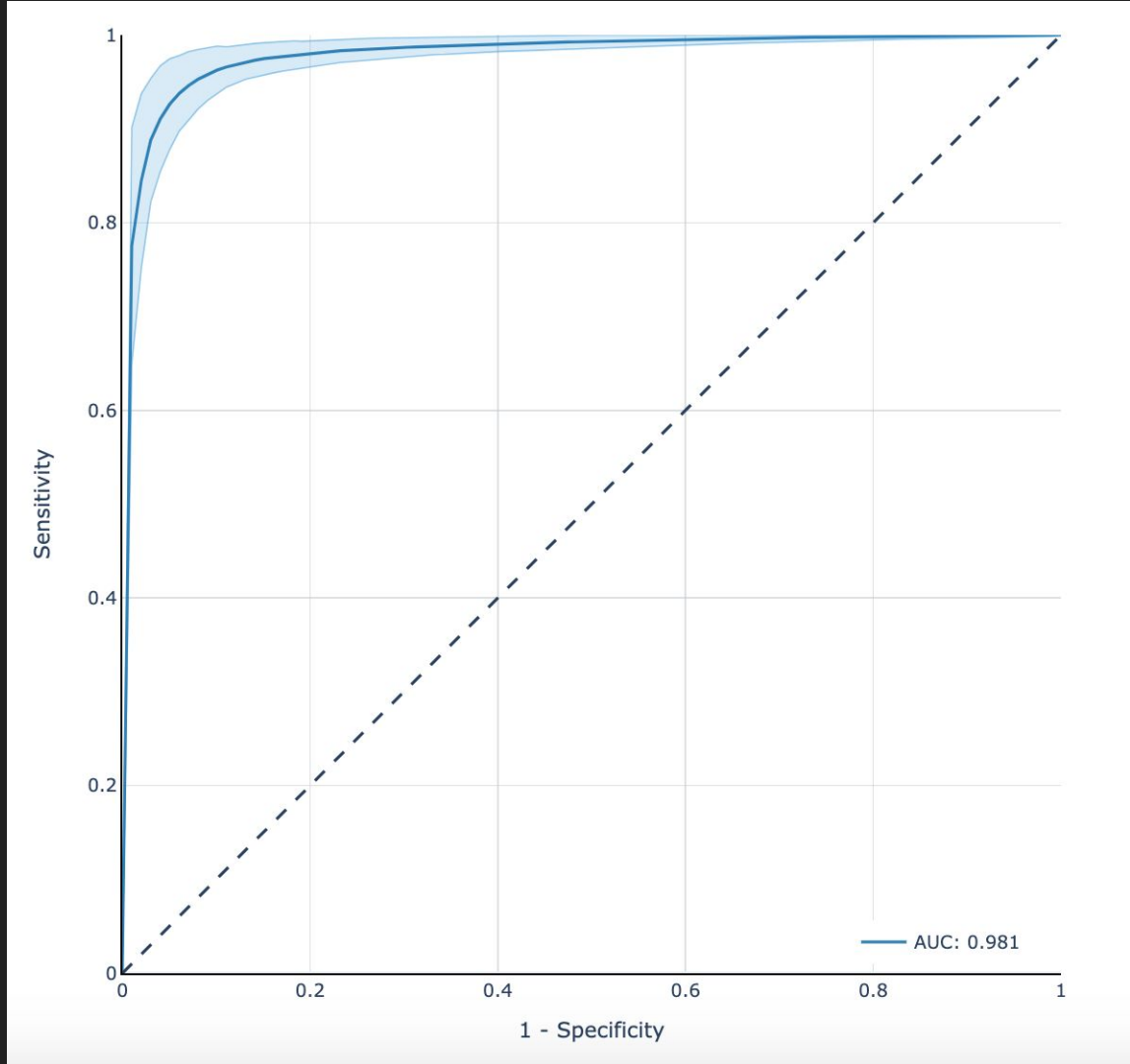


Initial Modeling

Choosing the direction forward based on initial performance

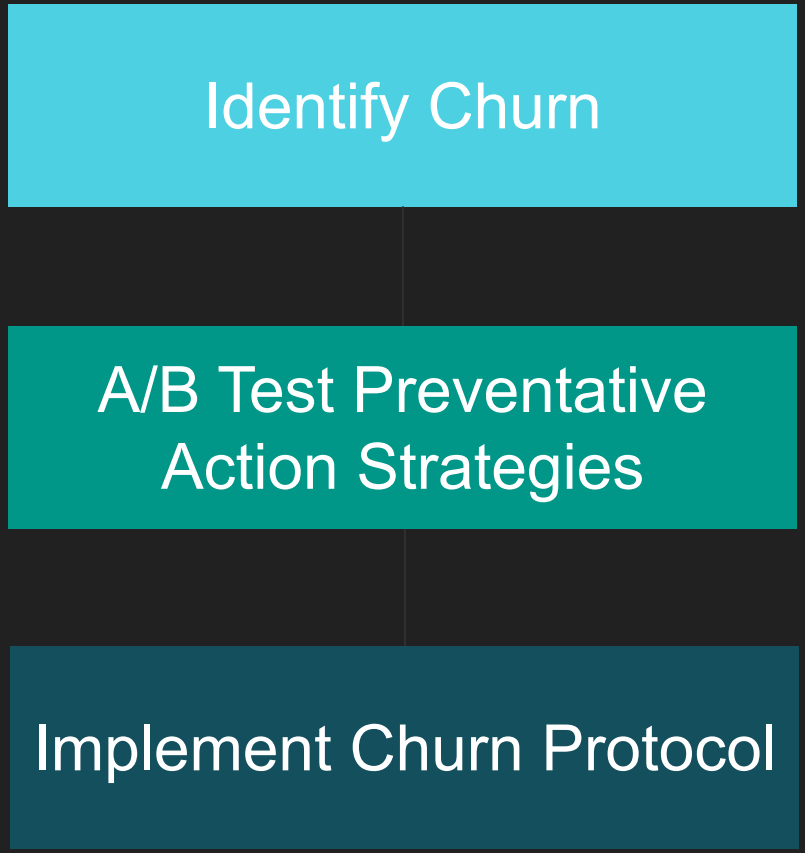
	classifier	train accuracy	train precision	train recall	train f1 score	test accuracy	test precision	test recall	test f1 score	test time
0	Logistic Regression	0.854	0.807	0.854	0.803	0.859	0.821	0.859	0.808	0.09
1	Nearest Neighbors	0.904	0.904	0.904	0.887	0.855	0.810	0.855	0.812	0.30
2	Naive Bayes	0.607	0.817	0.607	0.665	0.570	0.791	0.570	0.634	0.03
3	Linear SVM	0.855	0.876	0.855	0.789	0.857	0.734	0.857	0.791	35.48
4	RBF SVM	1.000	1.000	1.000	1.000	0.857	0.734	0.857	0.791	2.01
5	Decision Tree	0.952	0.951	0.952	0.949	0.936	0.933	0.936	0.932	0.03
6	Random Forest	0.855	0.876	0.855	0.789	0.857	0.734	0.857	0.791	0.05
7	Gradient Boost	0.972	0.972	0.972	0.971	0.949	0.949	0.949	0.945	0.82
8	AdaBoost	0.895	0.885	0.895	0.886	0.885	0.871	0.885	0.871	0.31
9	XGBoost	1.000	1.000	1.000	1.000	0.956	0.955	0.956	0.954	0.65

Cross Validation + ROC/AUC



Conclusion

Running this model continuously could identify customers likely to churn.



THANK YOU

For your time and attention.

Project Repository: GitHub repo link

Authors presenting today can be contacted using the following information:

Taylor Hale Robert *taylorhalerobert@gmail.com*