

TelCo Customer Churn Prediction

Devanshu Khurma (260894480)

Jiajun Huang (260629217)

Charlie Cai (...)

Evelyn Sun (260915480)



Problem Definition & Core Tasks



1

Problem Definition

- Predicting (i.e., classifying) whether a current customer will churn from TelCo or not
- 7043-observation (customer) with 19 predictors dataset

2

Benefits

- Marketing and Customer Retention team - Model Prediction
- Higher-Level Managers - Interpretability Report

3

Core Tasks

- Trained multiple Python-based Machine Learning models
- Programmed a H2O.ai-based Auto Machine Learning model
- Tested causal inference on binary predictors such as: gender, SeniorCitizen, Partner, Dependents, PhoneService, and PaperlessBilling
- Assembled a model interpretability and explainability analysis using SHAP package
- Composed a Machine Learning Bias Report based on the SHAR analysis

4

Other Considerations

- Run offline periodically (at most once a day)
- Batch processing system using Apache Spark



Python Based Model

Model	Accuracy Score	F1 Score
KNN	0.772903	0.564356
Decision Tree	0.792688	0.541825
Random Forest	0.792688	0.560219
Tuned Random Forest	0.849032	0.560218
Logistic Regression	0.80086	0.59135
XG Boost	0.807742	0.588018





H2O AtuoML

H2O's AutoML can be used for automating the machine learning workflow, which includes automatic training and tuning of many models within a user-specified time-limit. It has made it easy for non-experts to experiment with machine learning to set a benchmark.

H2O AutoML Model Results

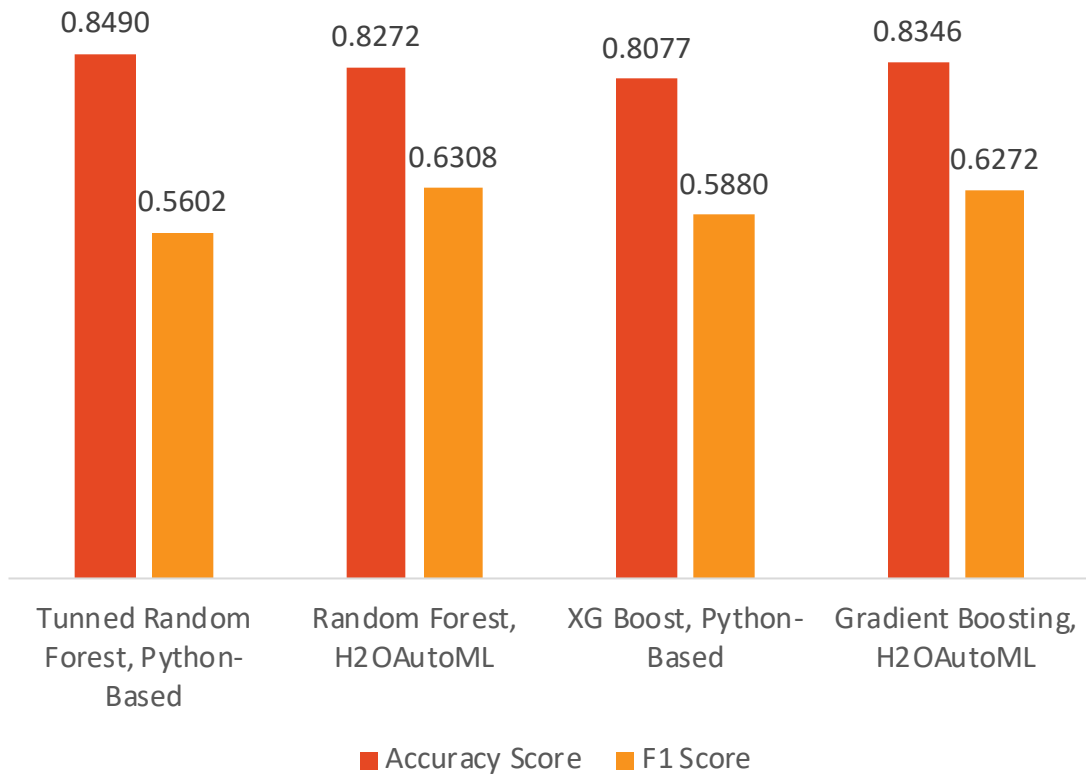
Model	Accuracy Score	F1 Score
Random Forest	0.8272	0.6308
Gradient Boosting	0.8346	0.6272
Neutral Network	0.8257	0.6193

10 Best Performing Models Selected by H2O AutoML

model_id	auc	mse
StackedEnsemble_BestOfFamily_AutoML_20200216_115946	0.850544	0.133636
StackedEnsemble_AllModels_AutoML_20200216_115946	0.850509	0.133678
GBM_5_AutoML_20200216_115946	0.848467	0.13321
GLM_1_AutoML_20200216_115946	0.847801	0.133984
GBM_grid__1_AutoML_20200216_115946_model_1	0.842607	0.148619
GBM_1_AutoML_20200216_115946	0.841779	0.136873
GBM_2_AutoML_20200216_115946	0.841048	0.136614
GBM_3_AutoML_20200216_115946	0.838457	0.138417
DeepLearning_1_AutoML_20200216_115946	0.83469	0.13864
GBM_4_AutoML_20200216_115946	0.832629	0.140648
XRT_1_AutoML_20200216_115946	0.832395	0.140071
DRF_1_AutoML_20200216_115946	0.828687	0.141467



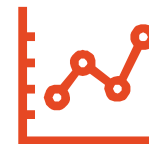
Model Results Comparison



H2O AutoML generates better accuracy score



H2O AutoML generates better F1 score



In general, Random Forest and Gradient Boosting from H2O AutoML deliver better performance

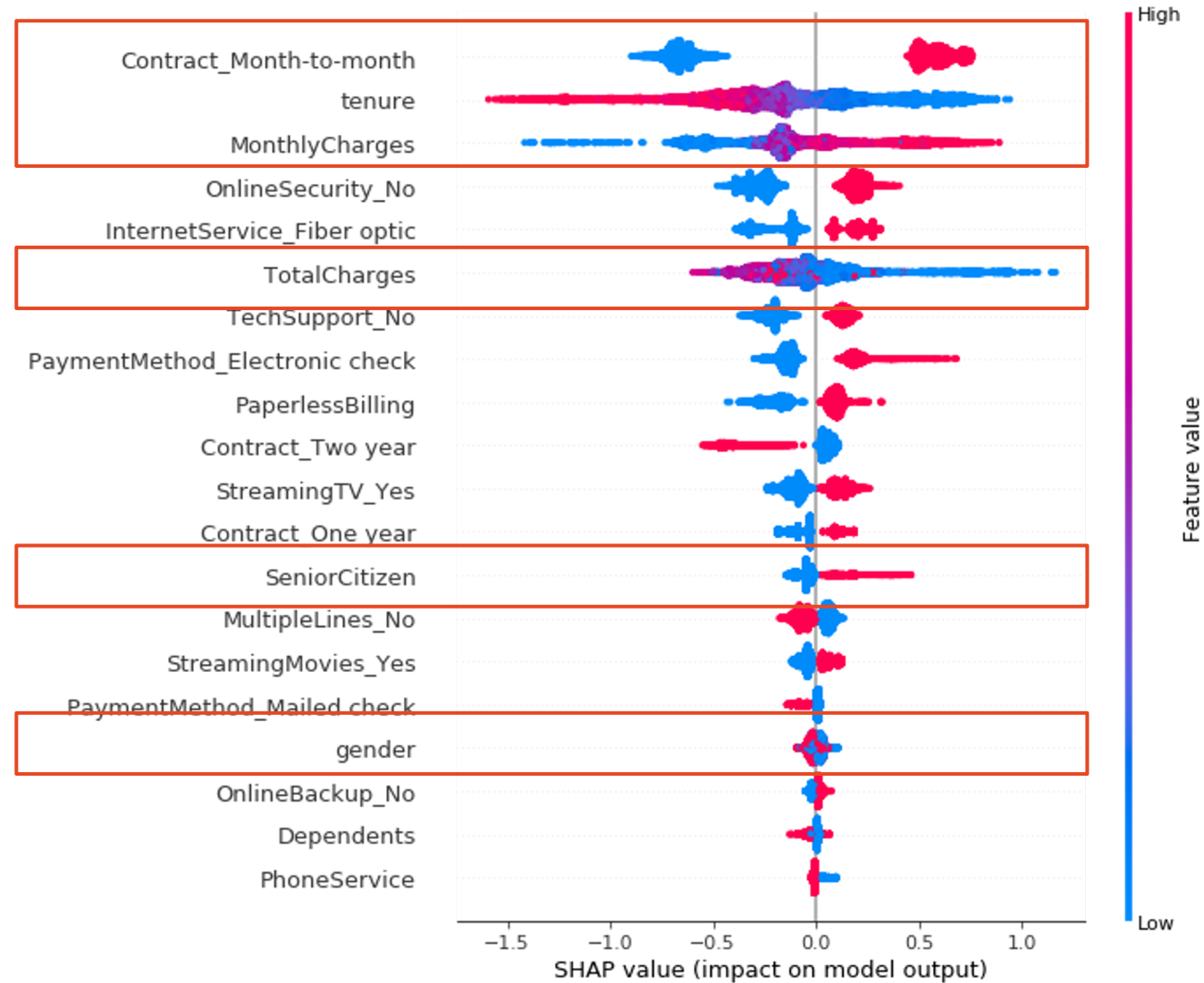


Causal Inference

Variables	P-Value	Causal Estimate
Contract Month-to-month	<0.001	0.05997
Paperless Billing	<0.001	0.04491
Gender	0.358	-0.00329
Senior Citizen	0.044	0.04420
Partner	0.464	-0.00122
Dependents	0.065	-0.02092
Phone Service	0.424	0.00988



Model Interpretability and Explainability





Conclusion

1

Model Bias Report

- Gender,
- Senior citizen (i.E., Whether the customer is a senior citizen or not),
- Partner (i.E., Whether the customer has a partner or not)
- Dependents

2

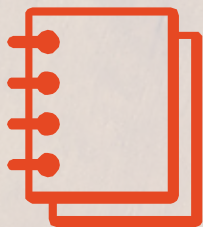
Action Recommended

- **Drop:** Gender
- **Keep:** partner, dependents, senior citizen

3

Next Step

- Marketing campaign and offers based on model interpretability report and causal inference
- Find the unobserved variables



THANK YOU

