

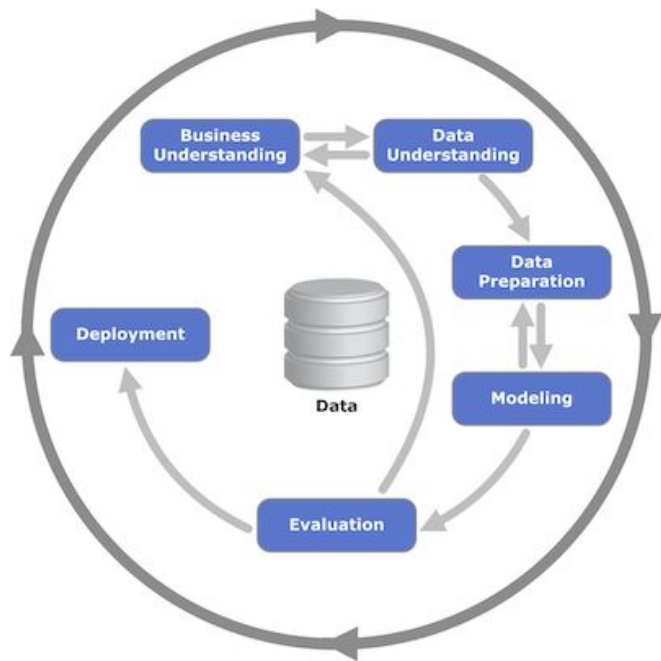
Lukas Sykora

Get loans with Home Credit

Kaggle competition

(<https://www.kaggle.com/c/home-credit-default-risk>)

Workflow



Business Understanding

Target

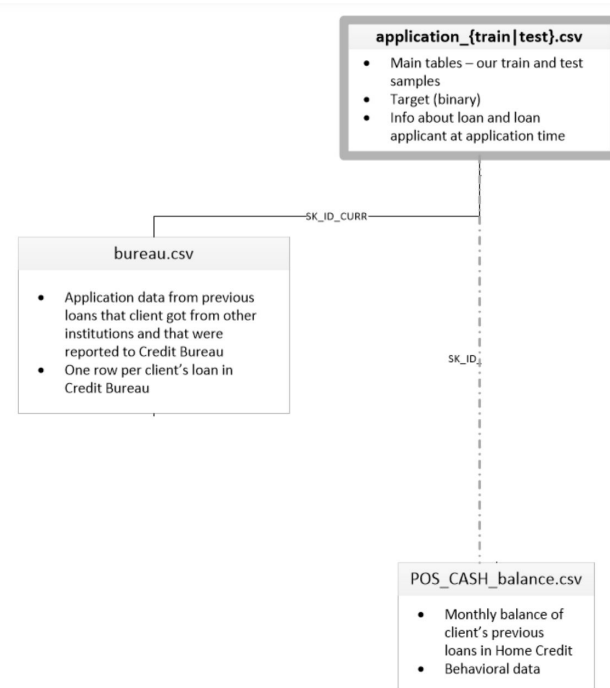
Ensure that clients capable of repayment are not rejected and that loans are given.

Data Preparation

Null Values

Null values replaced with average value.

Data Understanding



Modeling

Gradient Boosting

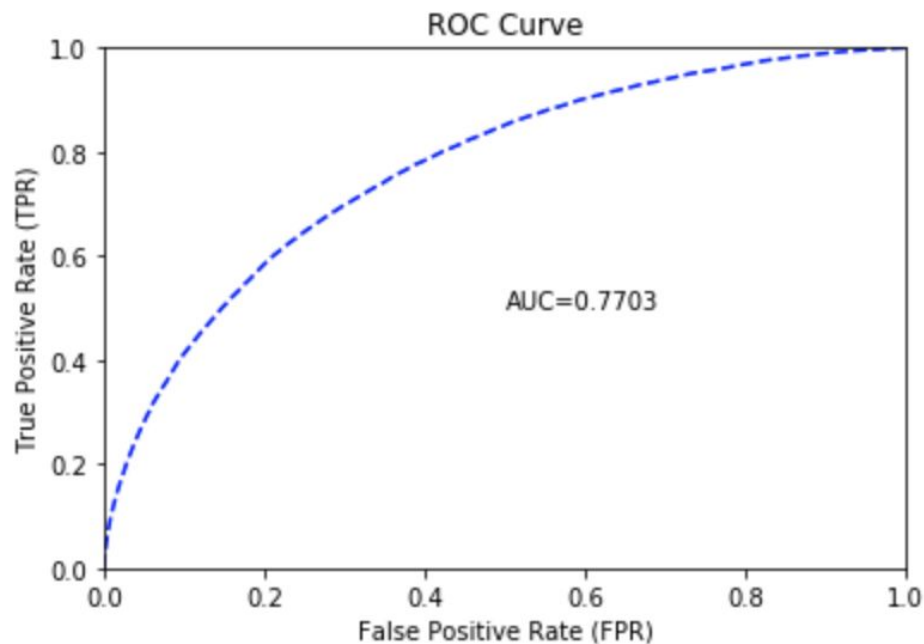
Gradient boosting is a machine learning technique, which produces a prediction model in the form of an ensemble of weak prediction models, typically decision trees. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Auto ML

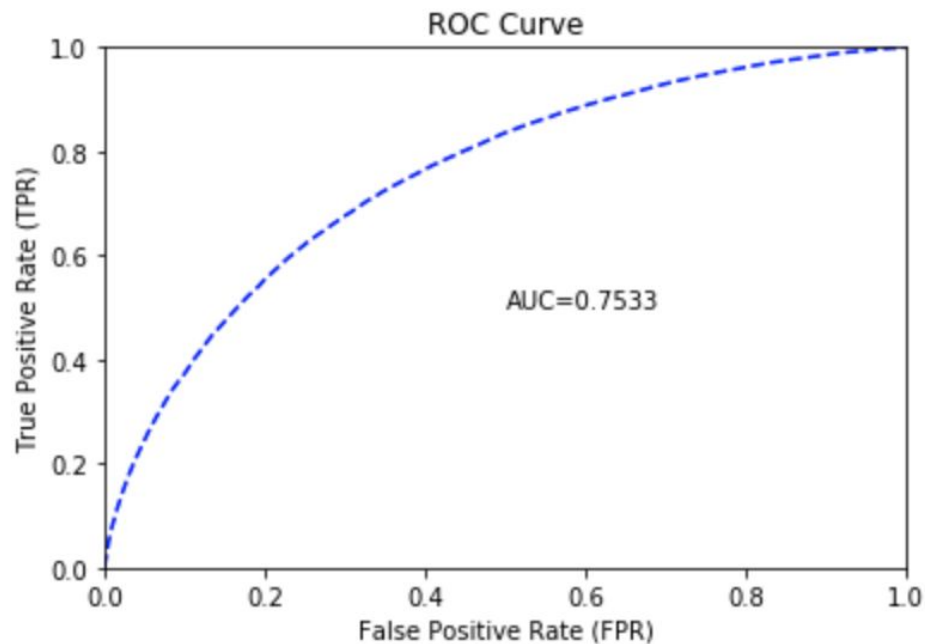
H2O's AutoML can be used for automating the machine learning workflow, which includes automatic training and tuning of many models.

Evaluation - Gradient Boosting - AUC

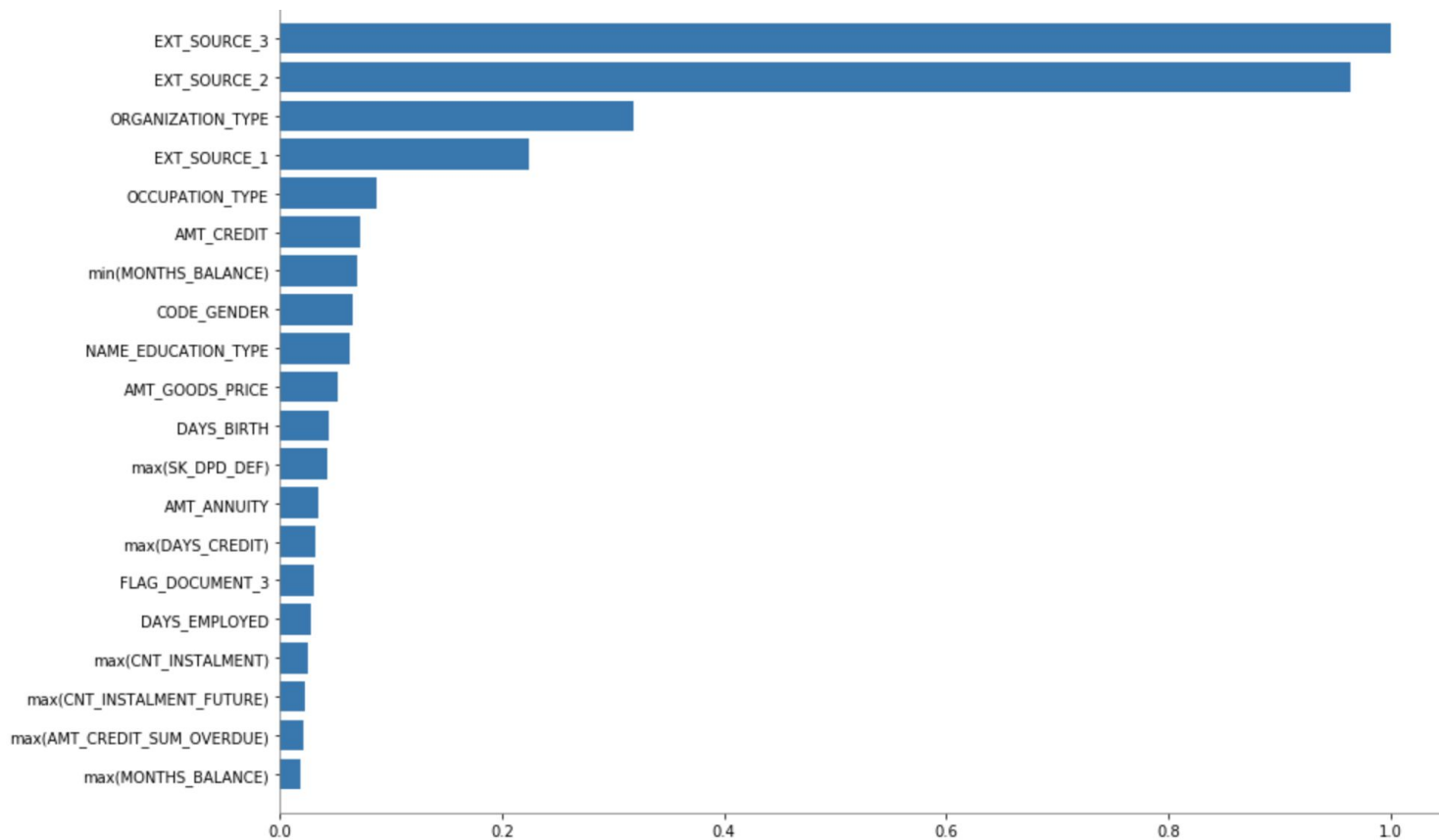
Training Data



X-Val



Evaluation - Gradient Boosting - Variable Import.



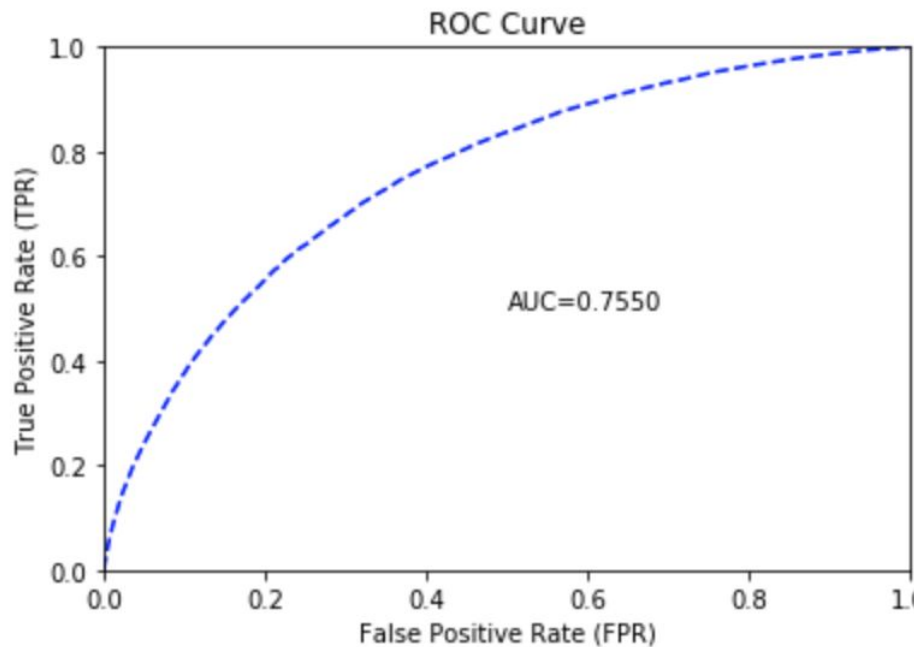
Evaluation - Auto ML - Leaderboard

	model_id	auc	logloss
GLM_grid_0_AutoML_20200410_085008_model_0		0.751253	0.248761
StackedEnsemble_AllModels_0_AutoML_20200410_085008		0.750832	0.25169
StackedEnsemble_BestOfFamily_0_AutoML_20200410_085008		0.750832	0.25169
DRF_0_AutoML_20200410_085008		0.719212	0.256384
XRT_0_AutoML_20200410_085008		0.7163	0.257949

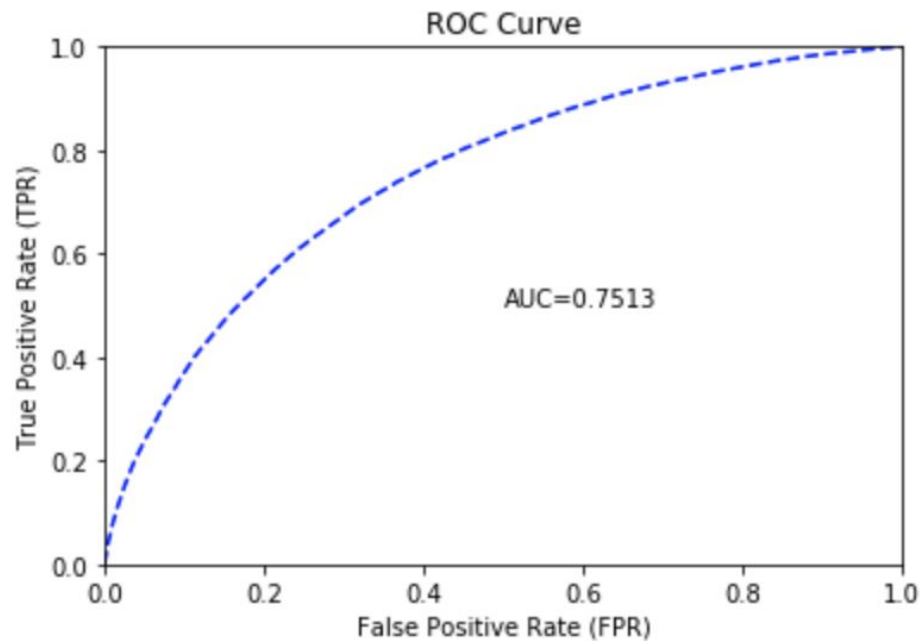
Winner: Generalized Linear Models (GLM), a statistical technique for linear modeling.

Evaluation - Auto ML (GLM) - AUC

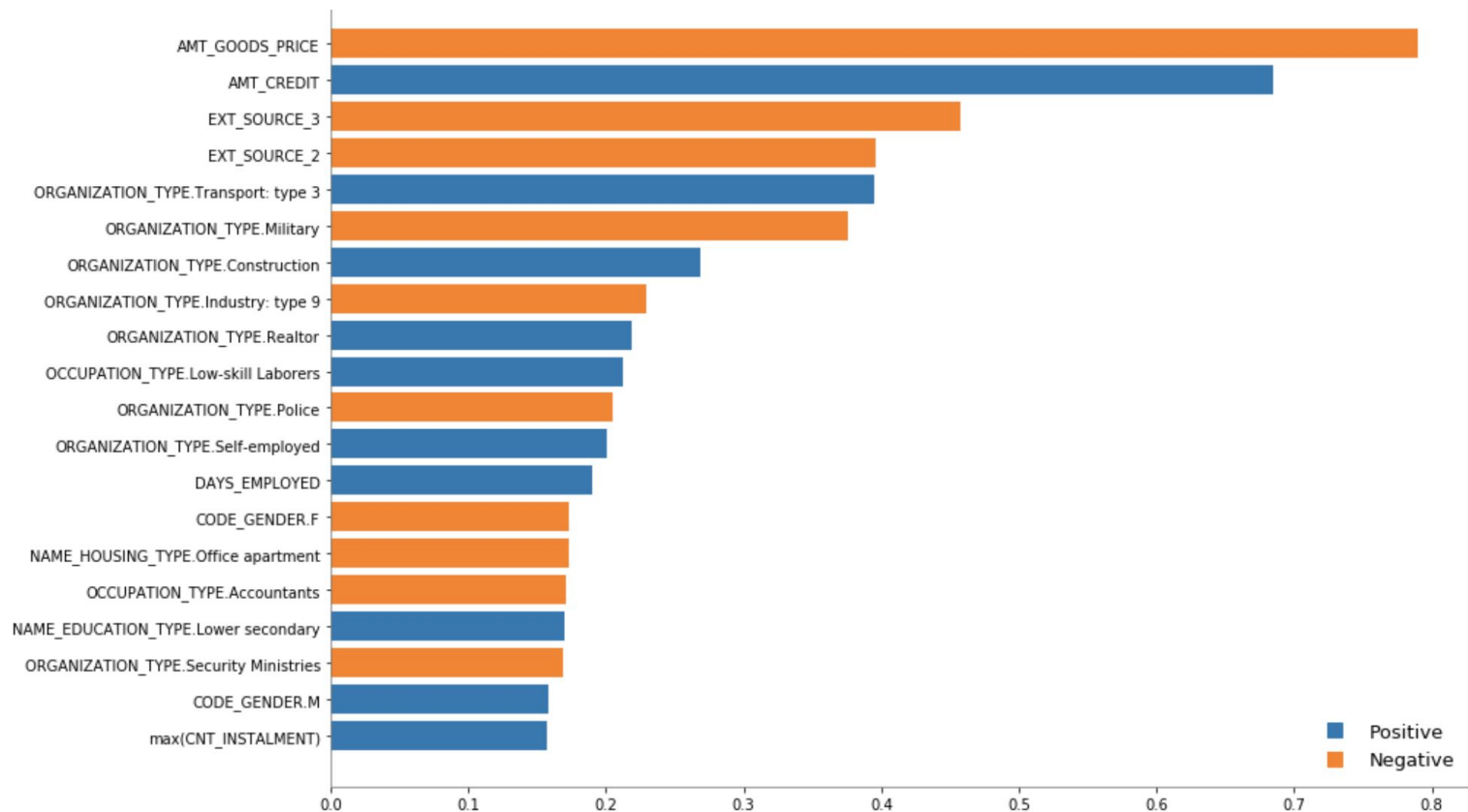
Training Data



X-Val



Evaluation - Auto ML (GLM) - Variable Import.



Deployment

Kaggle Result

Submission and Description	Private Score	Public Score	Use for Final Score
automlpred.csv a few seconds ago by Syky Auto ML prediction	0.62776	0.62544	<input type="checkbox"/>
gradientpred.csv a minute ago by Syky Gradient Boosting	0.62318	0.60816	<input type="checkbox"/>