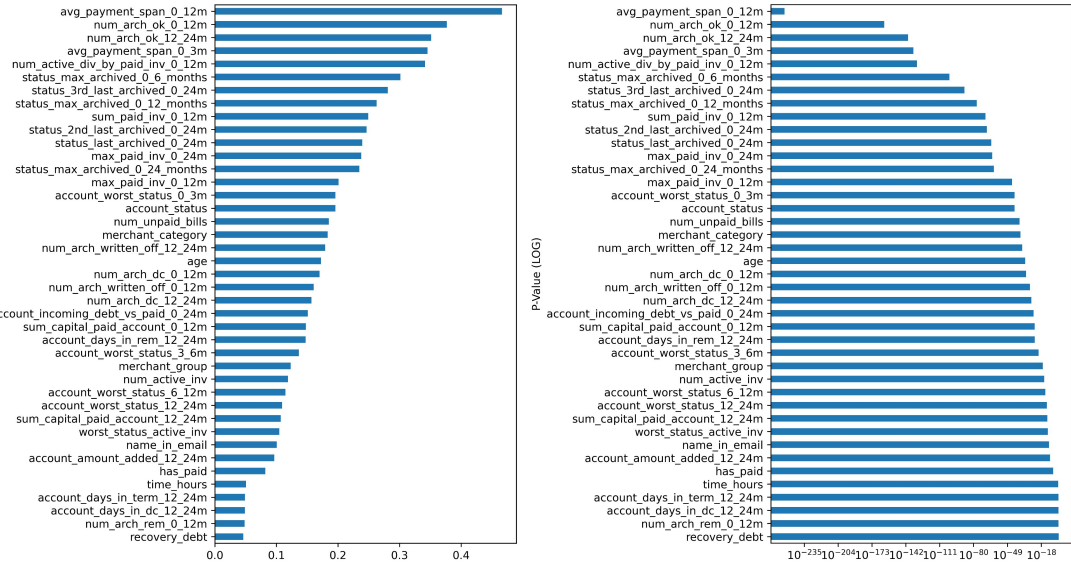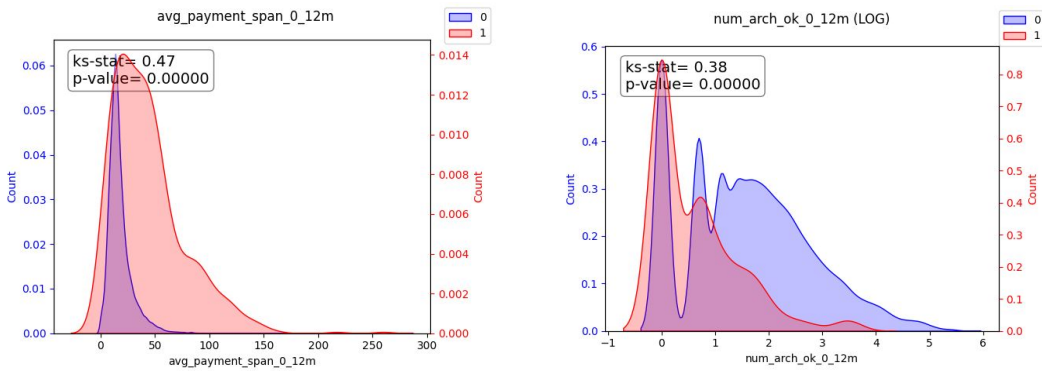# Klarna.

Default Forecasting Credit Model

Data

- Sample Customers Data
- ~100k Datapoints
  - 90k training
  - 10k validation
- Unbalanced dataset
  - 1.4% Default = 1
- 41 Features
  - 25 Numerical
  - 16 Categorical
- Exploratory Data Analysis to explore the dynamics of the data
  - KS-Test statistics
  - Feature X Default Distributions plotting

Model

- Data pre-processing: RawData -> Features
- Automated Feature Selection (Lasso Regression)
  - Drop 5 least correlated features
- Train/Test Split: 70/30
- Multiple models tested:
  - Logistic Regression
  - Random Forest Classification
  - Catboost Classification
  - XGBoost Classification (selected model)
- Hyperparameter tuning performed on all challenger models
- Model predictions calibrated using Platt calibration
- All steps integrated using scikit-learn pipelines

MLOps

Fully scalable platform built on top of AWS for automatic model (re)training and online model serving.
Features:

- CI/CD pipelines: AWS Codebuild
- Platform and Model code versioning: Github
- Resources Storage: AWS S3
- DataLake: AWS Athena
- Orchestration: Kubernetes (on AWS EC2)
- Workflow Management: Airflow
- Serverless Model Training: AWS Sagemaker
- Online Model Serving: KServe
- LoadBalancing: AWS ELB
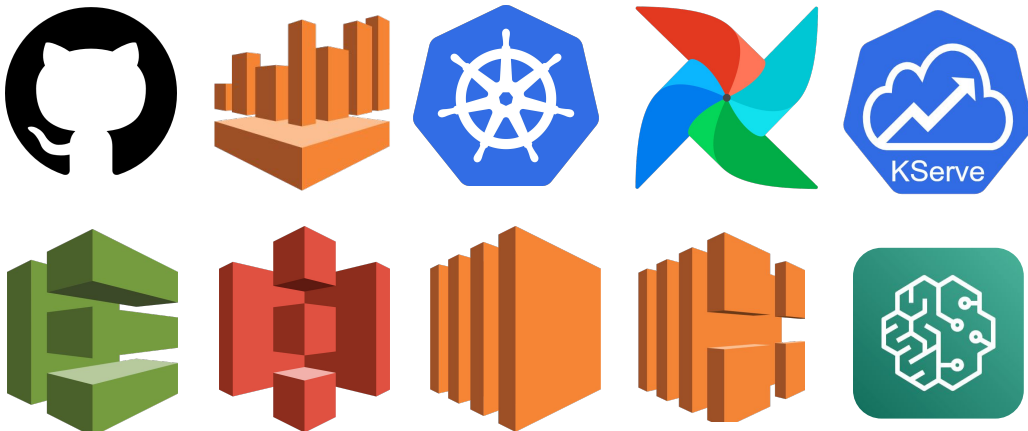
Results

- Selected Model: XGBoost Classifier
- Model Tuning target: Highest F1
- Model Selection target: ROC AUC
- Model Calibration: Platt Method
- Model Endpoint:
  - http://kubernetes-loadbalancer-1324688419.us-east-1.elb.amazonaws.com/v2/models/credit-model/infer

Model	ROC AUC	F1	Recall	Precision
XGB	0.70	0.28	0.42	0.21
CBC	0.68	0.28	0.37	0.22
RFC	0.65	0.31	0.31	0.30
LR	0.65	0.24	0.33	0.20
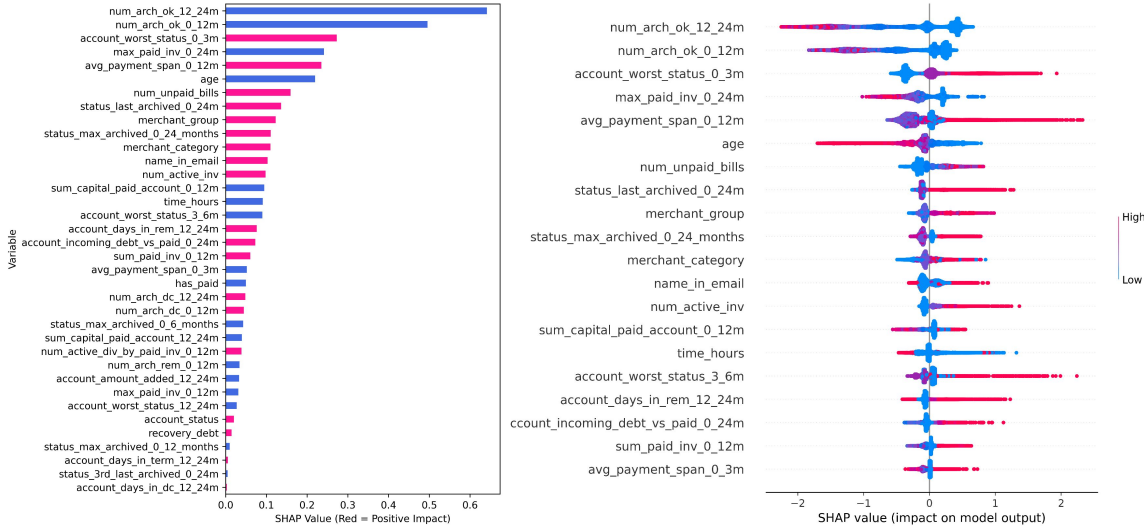
Contact

✉ otaviolcals@gmail.com
in linkedin.com/in/otaviocals