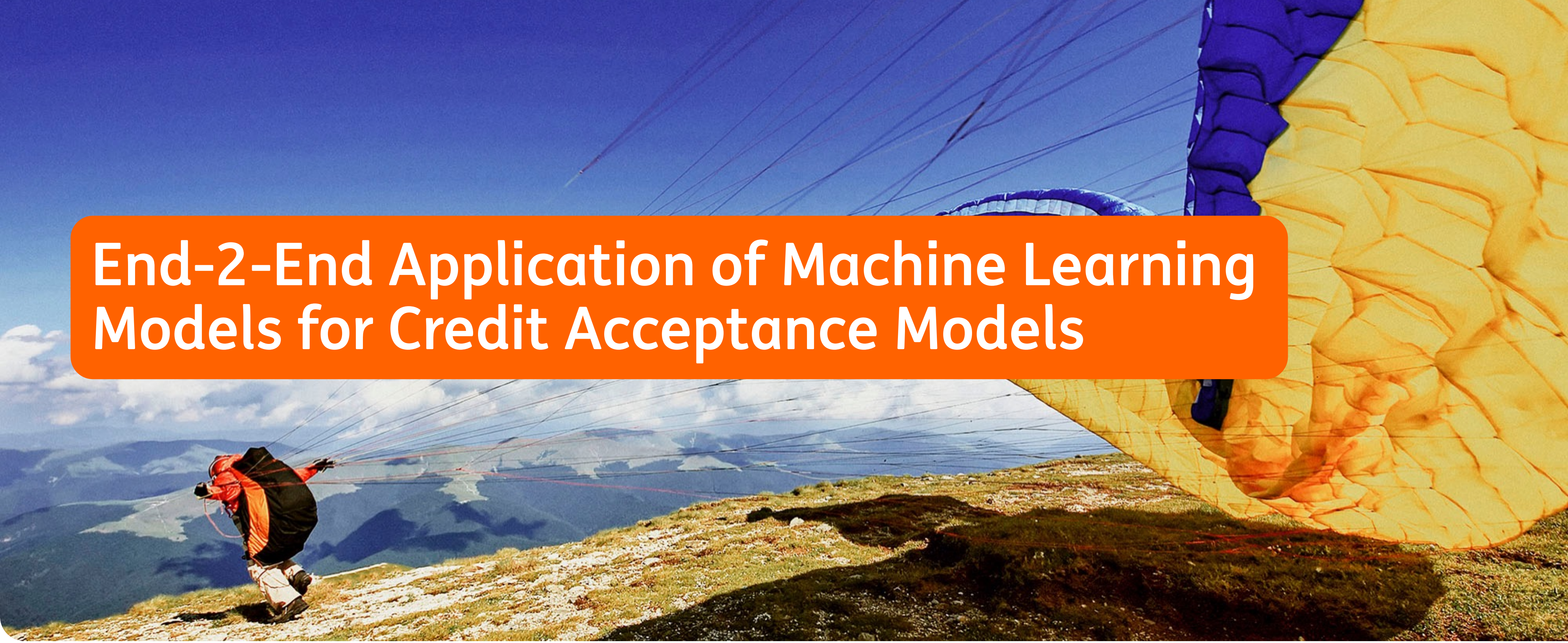


End-2-End Application of Machine Learning Models for Credit Acceptance Models

Artur Usov
TopQuants
November 1, 2023

do your thing

Introduction

Artur Usov

- **Principal Data Scientist** with 11 years of analytical experience, current focus on instant lending.
- MSc in Economics & MSc in Statistics

ING

INGA
STRENGTH IN NUMB3RS

RBA
Retail Banking Analytics

Retail Banking Analytics Tribe

- Focusing on analytics products in **lending, pricing, collection and personalization**

Building analytical capabilities on top of transactional data is crucial for the realization of ING's instant lending ambitions

Risk Assessment

Current Lending

Process
- Credit Risk assessment with traditional scorecard based on application data

Challenges
- Outdated financial data provided by client might not reflect current financial situation
- Poor performance & high rejection rate N2B clients

TX data & ML models

Instant Lending

Process
- Credit Risk assessment with traditional scorecard and transactional data of client

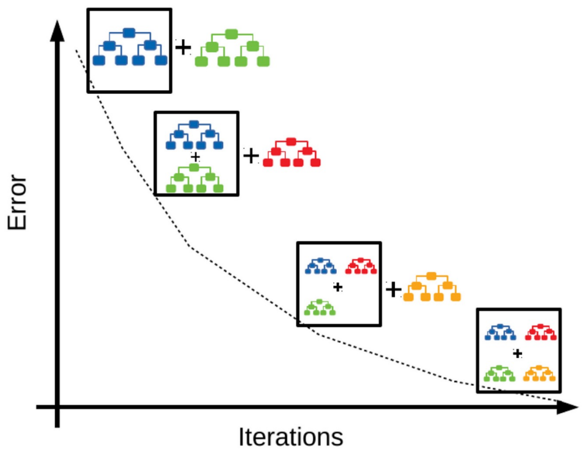
Opportunities
- Most up to date financial situation of client
- Improved model performance in estimating risk
- Pre-score existing clients or service instantly
- Risk estimation of N2B clients with PSD2 data

- Transactional data is typically of high quality and granularity
- Transactional data contains both, non-linear effects & interaction terms, which are easier captured by ML models as opposed to classical statistical models

! <u>**Right solutions for given problem**</u> instead of doing for the hype

We use ML Models because they provide interactions and non-linear effects out of the box

- Most commonly used algorithms:
  Gradient Boosting Tree ensembles:
  - XGBoost
  - Lightgbm



XGBoost

LightGBM

- Binning of the risk drivers is performed by the tree algorithm

- At the same time, every tree encodes interactions between features

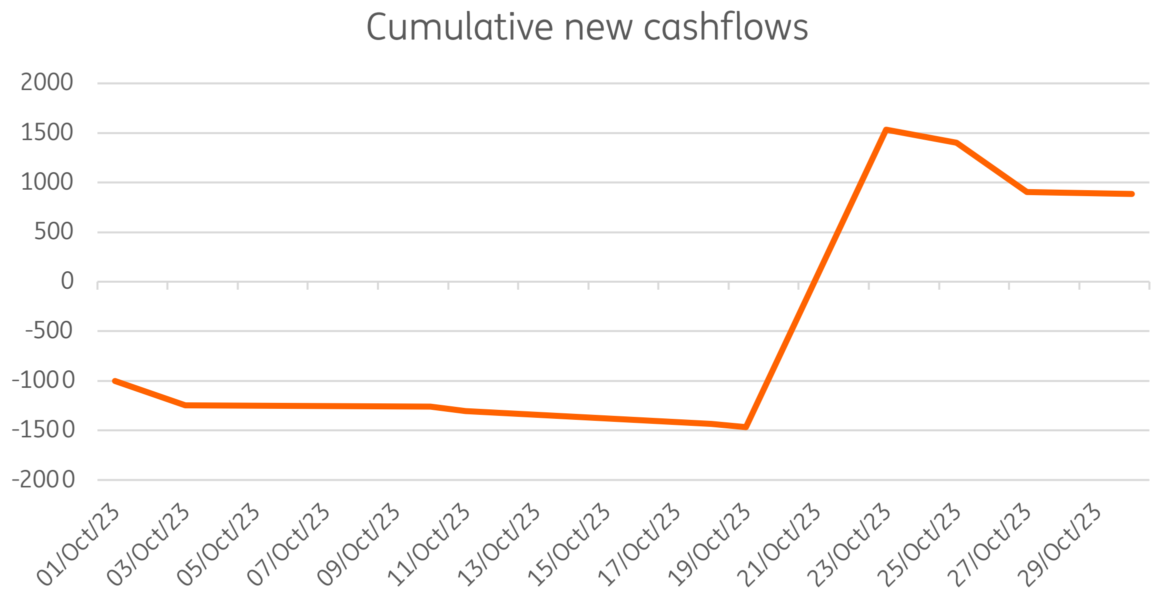
- Multiple weak learners working together to generate a strong learner: every subsequent tree is using residuals from previous tree as modelling target

- Non parametric models

Transactional data is both simple and complex

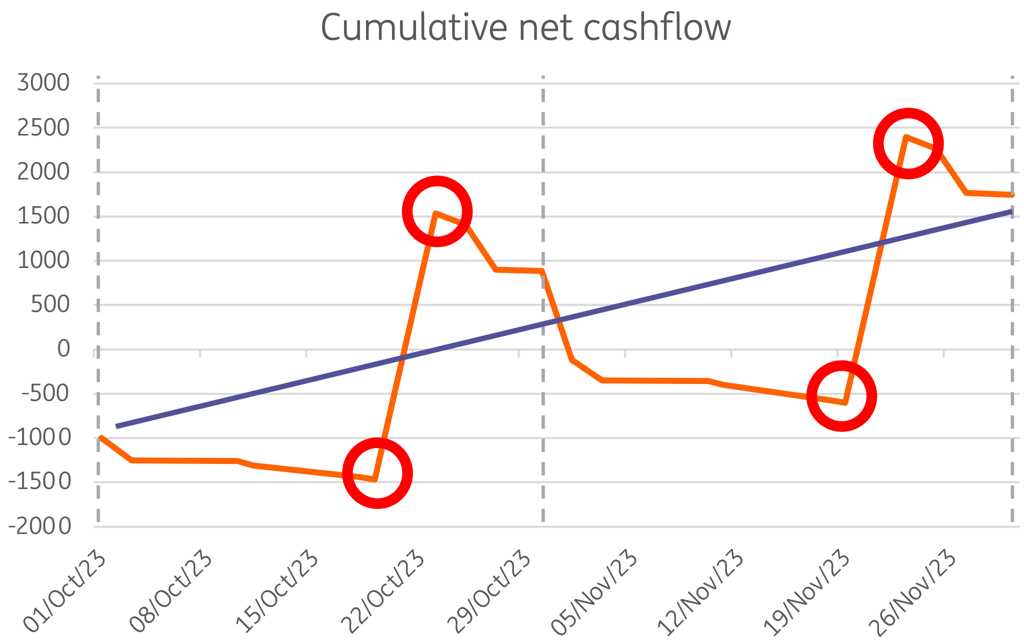
Hypothetical Example

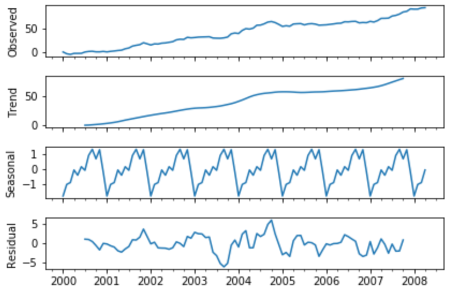
Transaction Date	Amount	Remaining Balance
01/Oct/23	-1000	10000
03/Oct/23	-250	9750
10/Oct/23	-12	9738
11/Oct/23	-45	9693
18/Oct/23	-130	9563
19/Oct/23	-30	9533
23/Oct/23	3000	12533
25/Oct/23	-130	12403
27/Oct/23	-500	11903
30/Oct/23	-20	11883



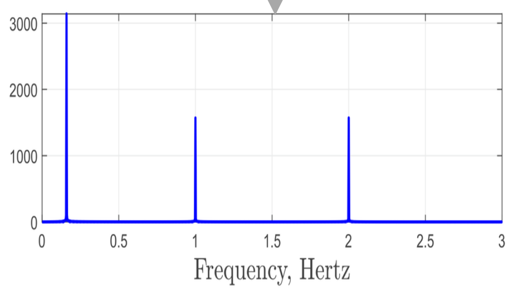
Notes:
- Above data is hypotehtical
- All data usage in modelling phase should always be within the legal framework and approvals

With some creativity, one can extract a lot of relevant signals





Timeseries Decomposition



Signal Processing

Risk Driver Design:

- Risk drivers are computed 1/2/3/6/12m prior to application date
- Simple summary statistics of the amounts (net, credits, debits, balances)
- Ratios: Debits/Credit, debits in first week vs last week, etc.
- Intervals: days between maximum debit and credit, how long to you remain with negative balance, how fast do you come back to negative balance
- Time series decomposition: Trend & Seasonality
- Signal Processing: Fourier and Wavelet transform
- Etc……

Considerations:

- Computationally extensive procedures are not always feasible to use due to size of transactional data
- Non-stationarity and multiple currencies might can an issue

Final Pool of Potential Risk Drivers:

- Typically 3000+ potential risk drivers for modelling
- Sky (and cloud memory) is the limit

We start with a large pool of potential risk drivers, but need to reduce to a stable and reasonable size

High Missingness Features

Correlated Features

Low Variance Features

Population Stability Index (PSI)

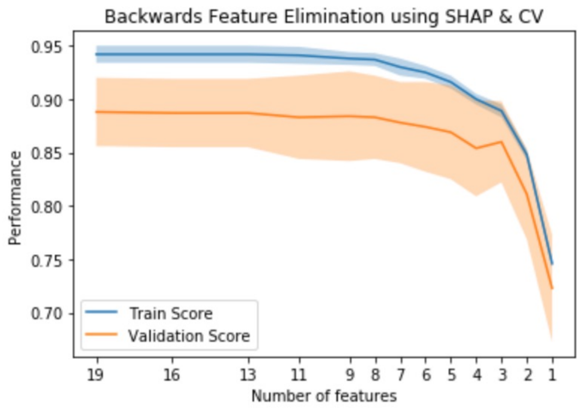
Low Model Performance Contribution

- Removing features with high level of missing values
- Removing highly correlated features
- Removing features with low variance
- Removing unstable features based on PSI
- Fit a simple model and remove features with low/none contribution to the model performance

→ Unstable Features (Resemblance Model)

- Splitting the data into in-time and out-of-time sets
- Fit a model to predict to which set observation belongs (resemblance model)
- Assess AUC of res model and most predictive features which predicts to which sample observation belongs
- Remove top-N most predictive features for the res model but least predictive with original target (minimize information loss, maximize shift reduction)
- Repeat until min AUC on res model is reached (0.70 usually)

→ Recursive Backward Feature Selection

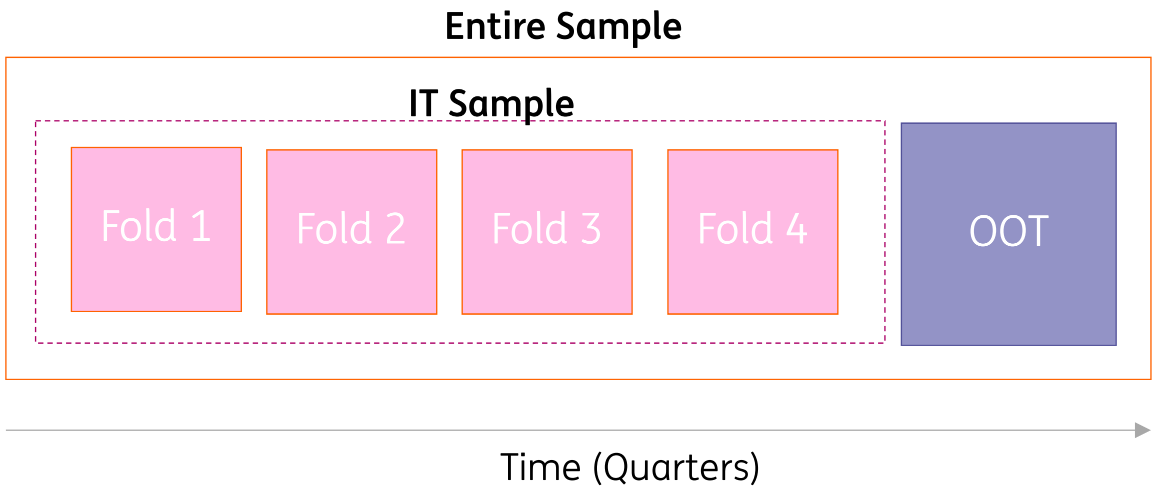


- Removing N features at a time based on feature importance (SHAP) until no further or small improvement in AUC are observed
- Select top N features when the increase in AUC stabilizes (elbow method)

Model Stability and tuning is of most importance....

Temporal Cross Validation:

- Samples to assess the model:
  - **Out-of-Time**: most recent data, used for final model assessment
  - **In-Time:** used for model training and tuning
  - **Out-of-Sample:** used for model evaluation
- The IT sample is split into K time-dependent folds, the model is trained on K-1 folds and evaluated on the hold out fold. Process repeated K time and model performances is reported across all K steps.



Hyper paramter tuning:

- ML models has a vast variety of hyperparameters, checking all of combinations is computationally heavy
- Random grid search: could result in local minima, but not global
- Bayesian approach (Optuna):
  - Tree-structured Parzen Estimator for hyper parameter tuning
  - Start with a random sample of parameters from a given grid search
  - Continue in direction which minimizes the loss
  - Stop when a minimum delta loss is achieved
  - Drawback: one parameter at a time

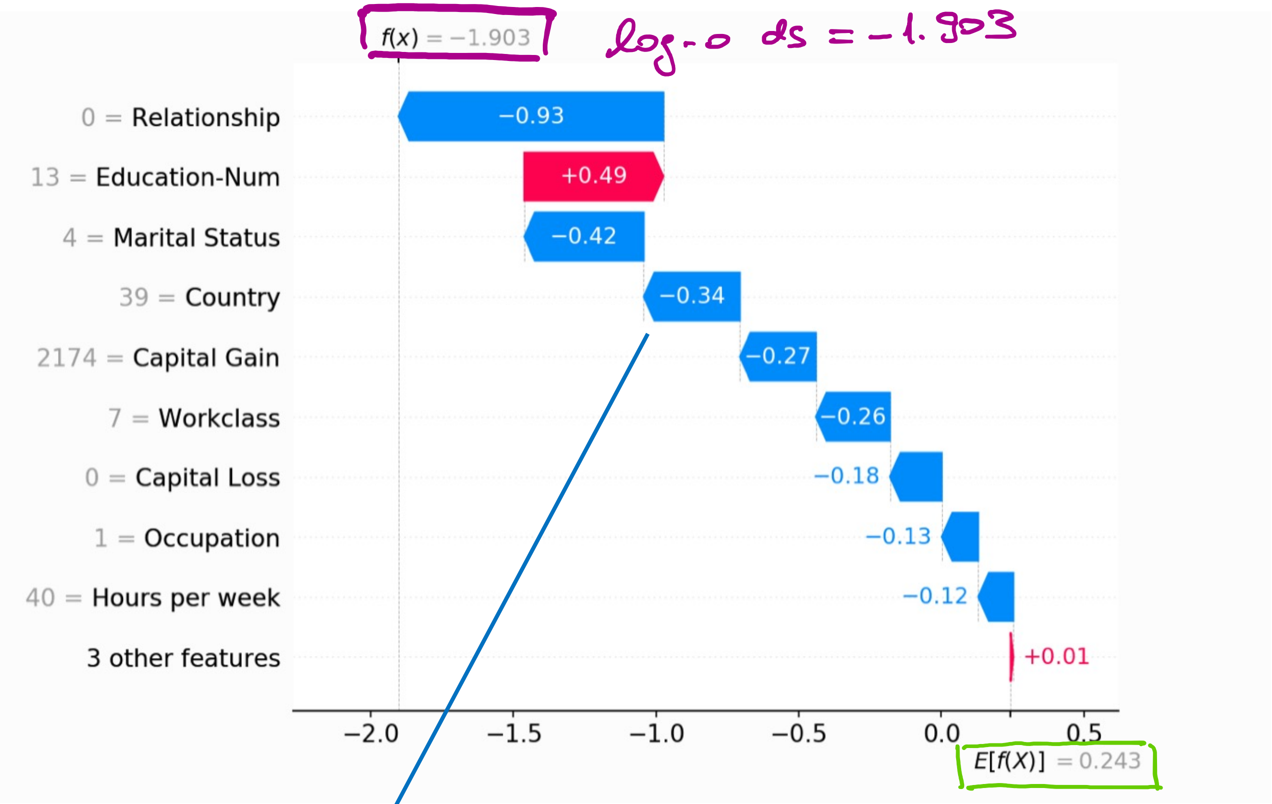
We need to be able to explain our models (hypothetical example)

SHAP values (SHapley Additive exPlanations) is a method based on cooperative game theory and used to increase transparency and interpretability of machine learning models.

Individual shap values represent the marginal contribution of a feature in terms of log-odds.

The contribution is always expressed relative to the average odds of the sample

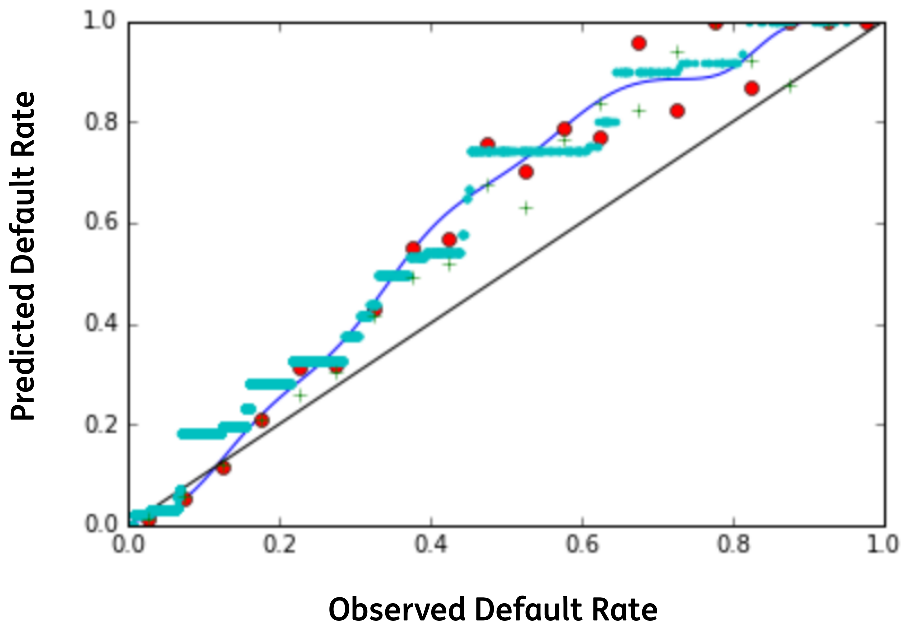
prediction for the instance log-odds = -1.903



average log-odds of the sample 0.243

Shap value:
Country contributes -0.34
to the difference in log odds

Model Calibration is needed if the model is used for decision making



Notes:

- Model probability needs to be calibrated if it is used for decision making
- Calibrated model has a mean PD = ODR, overall and per PD buckets (diagonal in the figure)
- Calibration options: Isotonic Regression or Plat Scaling (LR)

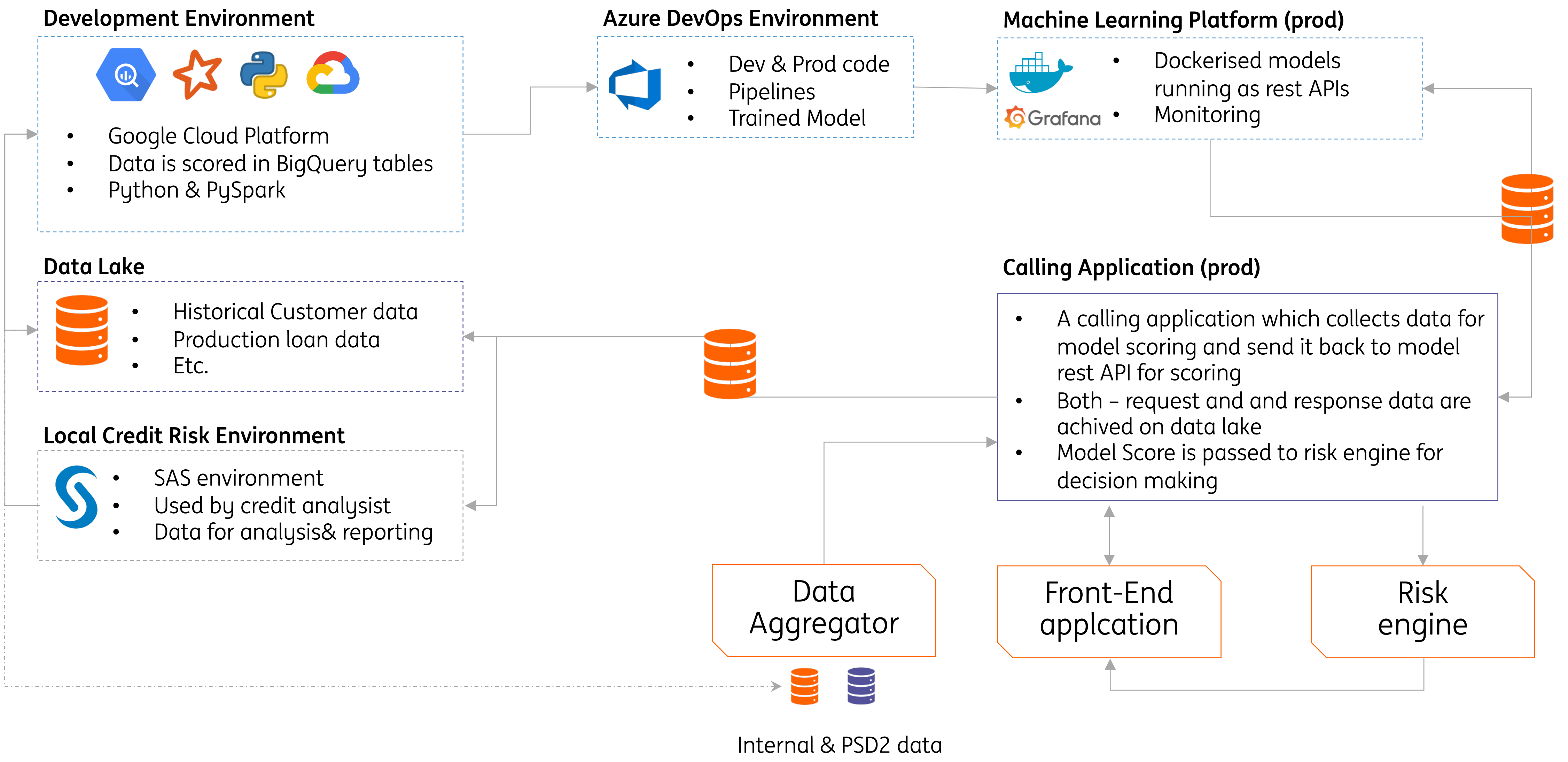
Isotonic Regression:

- Monotonically increasing step function
- Nonparametric method
- Works poorly with low number of defaults, interpolates constant PD values for buckets where no defaults are observed

Platt Scaling:

- Fitting a Sigmoid function between ODR and PD values
- Able to interpolate missing buckets well

Model Deployment & Assessment (Monitoring). It takes time to build the IT capabilities and resources to utilize ML models.



High Level Recap: Model Development cycle

Raw Data Acquisition	Risk Driver Design	Risk Driver Selection	Model Development	Model Deployment
Transaction data	Summary Statistics	High Missingness Features	Hyper Parameter Tuning	Deployment
Balance data	Seasonal Signals	Correlated Features	Model Calibration	Monitoring
Customer Info	Trends & Ratios	Low Variance Features	Uncertainty Estimation	
	Temporal Signals	Unstable Features (PSI)	Model Explainability	
	Cashflow Concentration	Unstable Features (Resemblance Model)		
	Etc.	Recursive Backward Feature Selection		

Thank you for your attention!

Questions?

do your thing