i-POP : Interpretable - Probability of Payment model

Sunkara Yadavendra Babu Ph.D.*
R. Shyaam Prasadh Ph.D.*

Credit risk analytics / GDIA / Ethics in AI & Robust AI (Explainable AI)

Abstract

- The Probability of Payment (POP) model is a risk scoring model that calculates a POP estimate for each Ford credit contract at origination. POP values are used to predict baseline expected defaults and expected credit loss (ECL)
- Recently, Credit analytics team (GDIA) used interpretable XGBoost to predict the raw POP scores. It imposes both monotonicity and interaction constraints while fitting the model
- Inherently interpretable models provide accurate explanation but sacrifice model performance to an extent[1]. US POP model reflects this scenario with series of business adjustments after prediction
- To combat, we developed a high-performing intrinsically interpretable POP model (i-pop) using an explainable neural network (xNN) based on Generalized Additive Models with structured Interactions [2] [3]
- ROC (train -95.81 % & test -81.4 %) show that i-pop model has high discriminatory power than the current models
- DeLong's test show that the AUCs of the current and previous models are statistically different

* Email: ybabu@ford.com & rshyaamp@ford.com

Motivation

- POP Models are used by the Risk team for the Credit Approval Process to Ford credit portfolio
- Current POP models use XGBoost and can be extended to Explainable XGBoost [EBM], where main effect or pairwise interaction can be estimated via gradient boosted shallow trees
- But, gradient boosted shallow trees are shown to have strong approximation ability
- EBM outputs shape functions with unexpected jumps and may become worse when there exist outliers
- EBM can be extremely complex for high dimensional data : Lacks regularization when both parent and child pairwise interactions explain the target variable

Research Objective

- To build an xNN that pursue a good balance between POP prediction accuracy and model interpretability
- Ensure that Neural network architecture follow the below interpretability constraints:

a)Sparsity : *Select the most significant effects for parsimonious representations*

b)Heredity [4]*: Pairwise interaction could only be included when at least one of its parent main effects exists*

c)Marginal clarity : *To make main effects and pairwise interactions mutually distinguishable*

Methodology

- Generalized additive index model (GAIM) is used in POP prediction. The relationship between raw features x ε $\mathbb{R}^p$ and the response y is represented by

$$g[E[y/x]] = u + \sum_{j=1}^{M} h_j(w_j{}^T x) \qquad \text{--------- (1)}$$

g is a pre-specified link function, u is the intercept, and M is the number of additive functional components.

- GAIM is estimated using back fitting algorithm, iteratively estimates a pair of $\{w_j, h_j\}$ at a time, with other pairs fixed
- Nonparametric regression (ex: smoothing splines) is used to fit the shape functions in (1)
- GAIM includes both main effects and interaction effects between individual features for performance improvement
- In addition to neural network parametrization, the interpretability of (1) is enhanced with below three constraints:

Sparsity : Prune the trivial main/interaction effects

$$D(h_j) = \frac{1}{n-1} \sum_{j \varepsilon s_1} h^2{}_j(x_j)$$

$$D(f_{j,k}) = \frac{1}{n-1} \sum_{j \varepsilon s_1} f^2{}_{jk}(x_j x_k)$$

Main effects (h(x))

Interaction effect (f(x_j, x_k))

Heredity : Atleast one main effect is significant

$\forall$(j; k) ϵ S_2 : j ϵ S_1 or k ϵ S_1

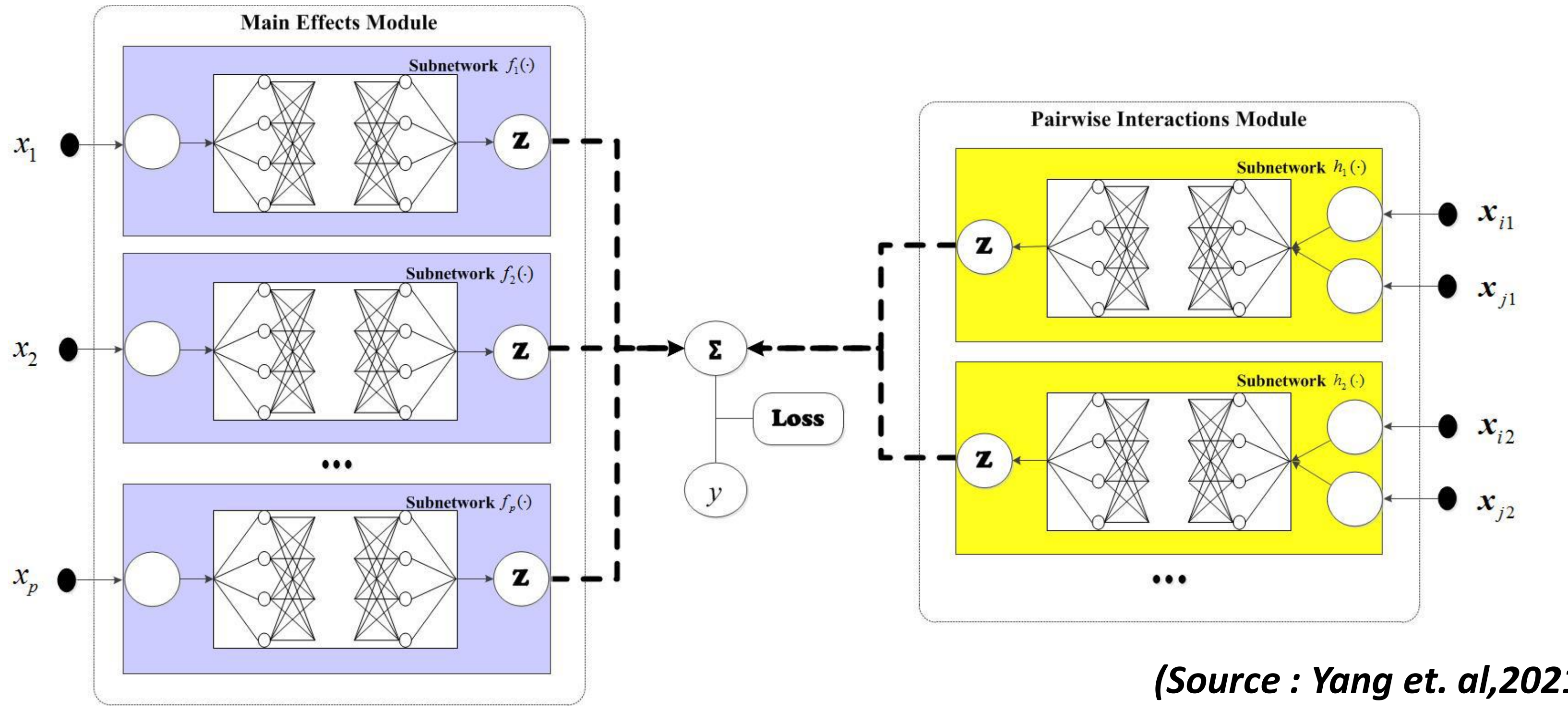
S_1,S_2 – List of main & interaction effects

Marginal : Separate main and interaction effect

$$\Phi(h_j, f_{j,k}) = \left| \frac{1}{n} \sum_{j \varepsilon s_1} h_j(x_j)\, f_{j,k}(x_j, x_k) \right|$$

Smaller the value of orthogonality $\Phi(h_j, f_{j,k})$, clearly marginal effect hj is separated from child interaction fjk

Network Architecture



(Source : Yang et. al,2021)

Proposed xNN is formulated as follows:

$$g\left[E\left[y/x\right]\right] = u + \sum_{j\varepsilon s_1} h_j\left(x_j\right) + \sum_{(j,k)\varepsilon s_2} f_{j,k}\left(x_j, x_k\right) \quad \text{---------} \ (2)$$

- The main effects (h(x)) are first fitted
- Top-K ranked pairwise interactions ($f(x_j, x_k)$) are selected & fitted to the residuals, subject to heredity constraint
- The dashed arrows to Σ nodes denote the sparsity constraints, the trivial subnetworks are pruned
- Finally, the marginal clarity is imposed for regularizing pairwise interactions

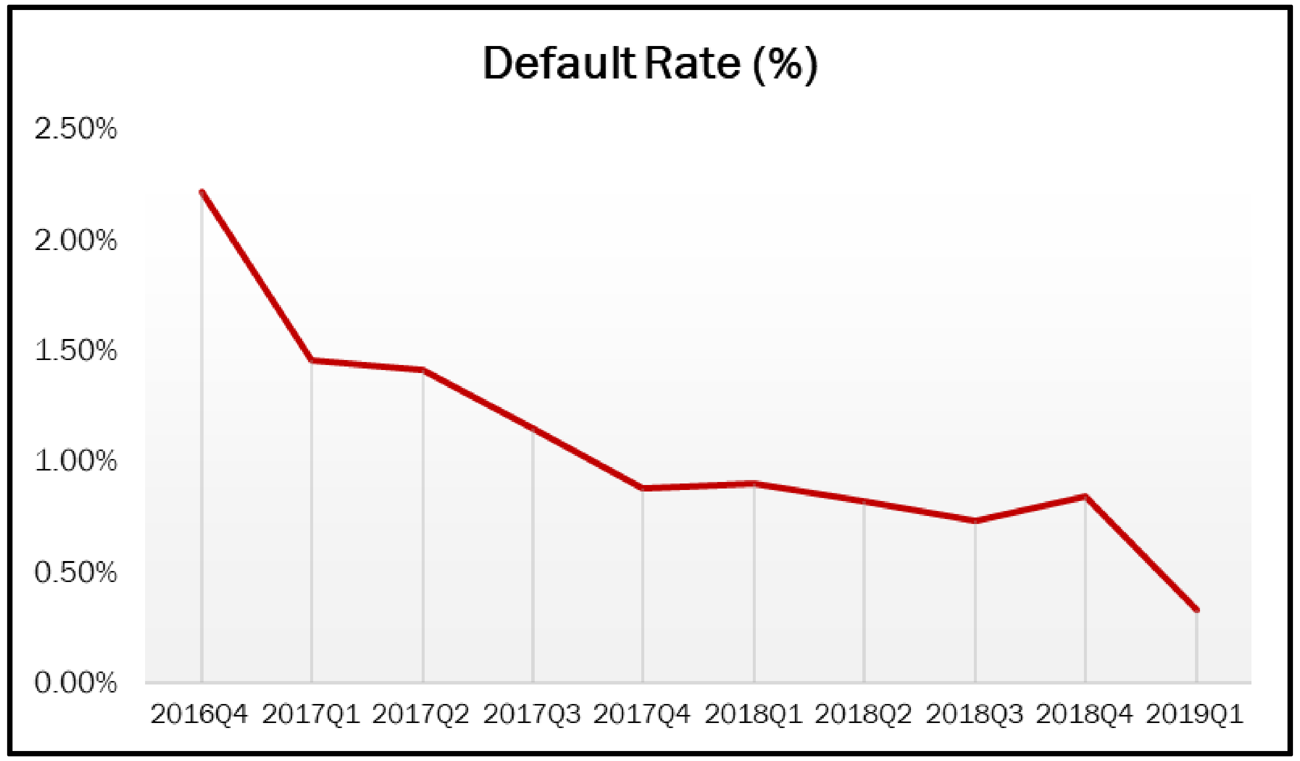
Dataset

- **Data sources:**

  1. SCOPE (**Credit bureau UK portfolio** - File name: uk_current - Location: prodfauk)
  2. Baseline dataset: default_uk_commercial_2020.sas7bdat

- **UK contracts** Originations timeframe – Oct,2016 to July,2019

Defaults distribution over the sample period (UK)



- Contract defaults (Target variable) are defined below

	Default definition
UK	All accounts open or closed in loss and recovery (goodzbad_indic in(7,8))

- The following table shows the distribution of the dataset

UK data time frame	**Jan,2017 - Feb,2021**
Total contracts	20,980
Default volume	139

- The following table shows the volume distribution by product

Purchase Quarter	Non-Default	Default	Total	Default Rate	Percent Population	Non-Default/ Default ratio
Balloon	16803	86	16889	0.51%	80.50%	195.38
Non-Balloon	4038	53	4091	1.30%	19.50%	76.19

Hyperparameters and Interpretability

- Maximal number of pairwise interactions is set to K = 30
- Subnetwork is configured with 5 ReLU hidden layers each with 40 nodes
- Subnetwork weights are initialized using the Gaussian orthogonal initializer
- Initial learning rate of the Adam optimizer is set to 0.0001
- Mini-batch sample size is determined according to the sample sizes of different datasets
- A 20% validation set is split for early stopping, and the early stopping threshold is set to be 50 epochs
- The tolerance threshold is set to be 1% of the minimal validation loss.
- The marginal clarity regularization strength can be empirically selected from 0.0001 to 1

Importance Ratio (IR) :

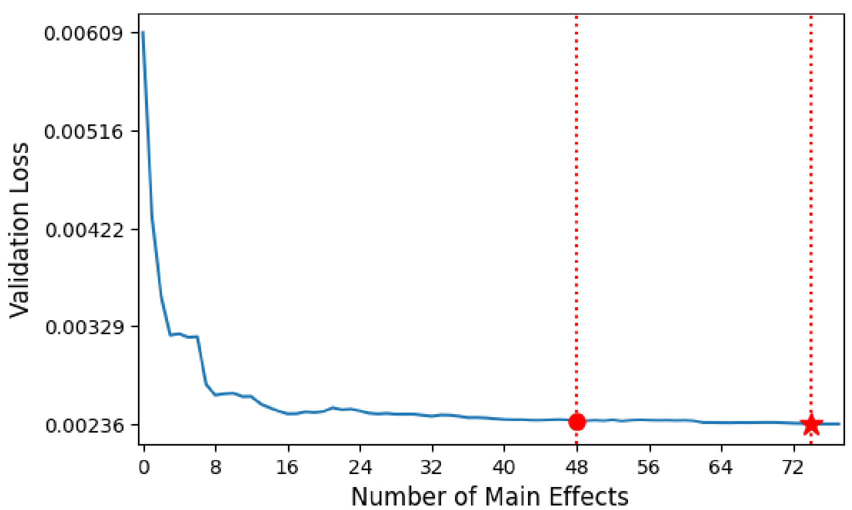
Contribution of each individual variable to the overall prediction is measured by following :

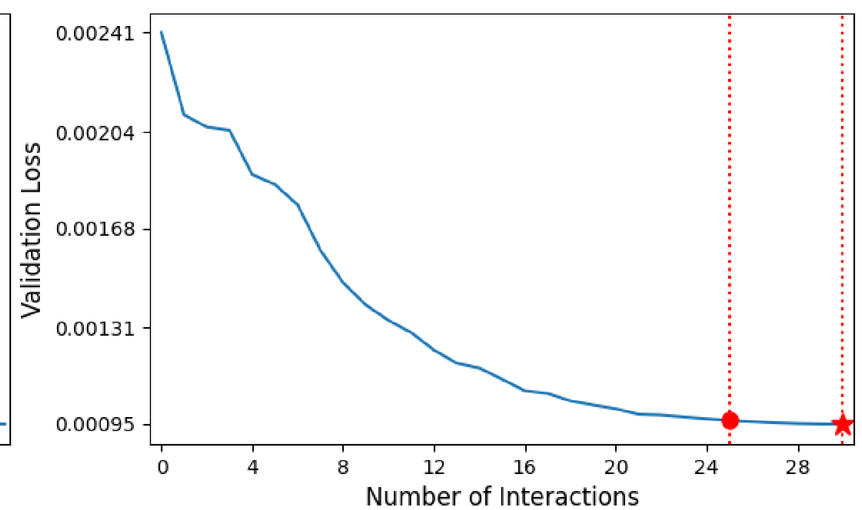
Main effects : $IR[j] = \frac{D(h_j)}{T}$

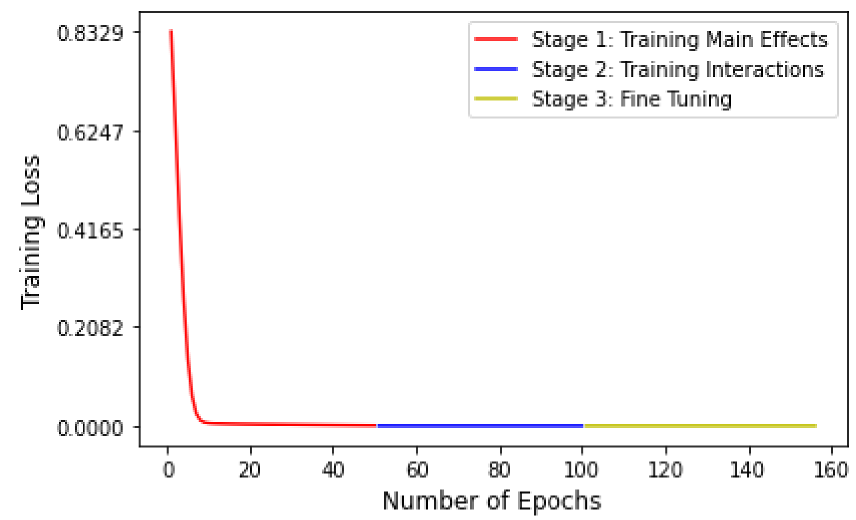
Interaction effects : $f_{j,k}[j,k] = \frac{D(f_{j,k})}{T}$

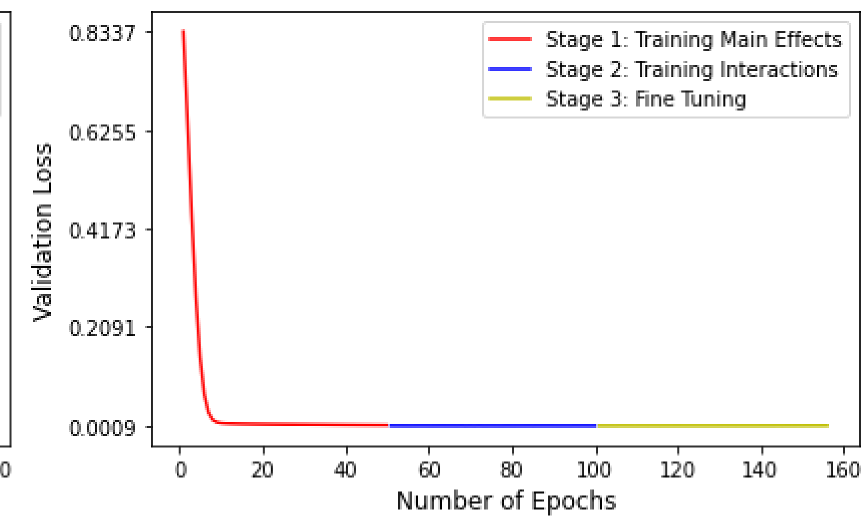
$where \quad T = \sum_{j \varepsilon s_1} D(h_j) + \sum_{(j,k) \varepsilon s_2} D(f_{j,k})$

Results
UK Contracts POP prediction

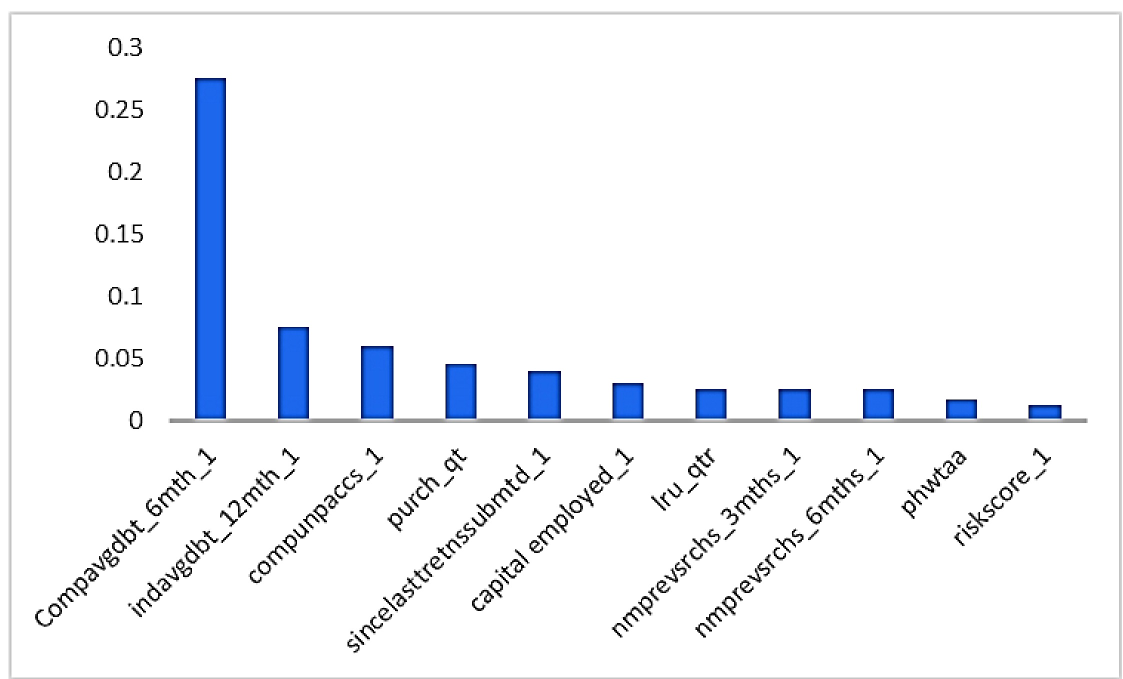




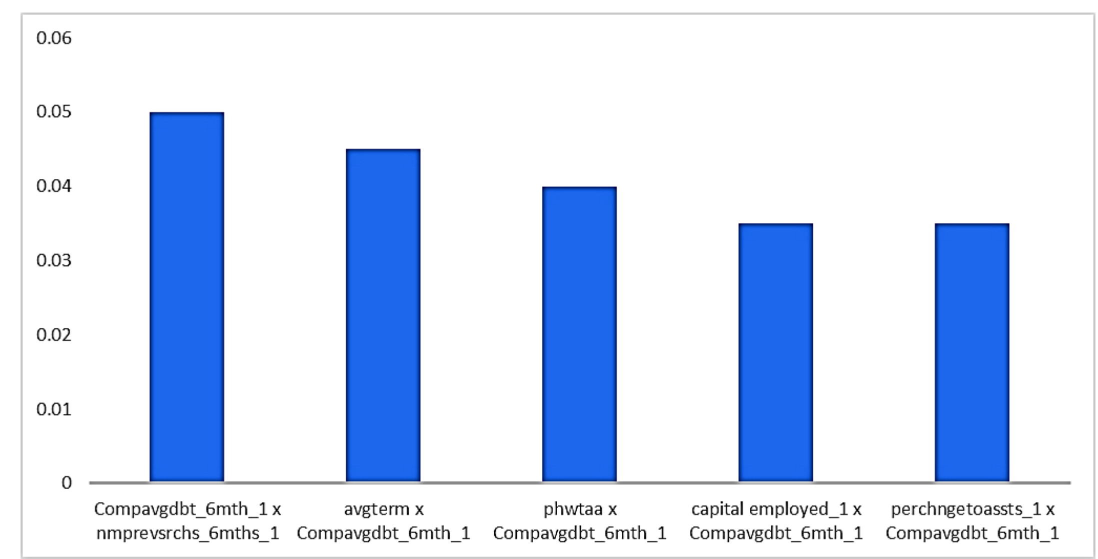




Top 10 Main effects



Top 5 Interaction effects



- Training Sample : 14,686
- Test Sample : 6,294
- No. of Features : 77
- No of main effects: 48
- No of interaction effects:25
- No. of GPUs :1
- ROC (%) train data : 95.81 %
- ROC (%) test data : 81.4 %

Conclusion

- Current POP models use interpretable XGBoost to predict raw POP Scores
- Empirically, gradient boosted shallow trees used by XGBoost are shown to have strong approximation ability and estimated shape functions by boosted trees are all piecewise constant.
- To combat, we built an intrinsically explainable **i- POP** model using GAIM architecture to predict POP scores
- Our xNN approximates the functional relationship using subnetwork-represented main effects and pairwise interactions
- We include three different constraints (Sparsity, Heredity and Marginal clarity) to enhance the model interpretability
- Experimental results from UK Contracts show that proposed model has higher discriminatory power

Next Steps

- Evaluate the robustness of the model using US Contracts data
- To extend the i-POP model with higher-order interactions
- Additional shape constraints for each component function, e.g., monotonic increasing/decreasing, convex or concave

i- POP model is highly interpretable and has competitive predictive performance