NeuroDecision Technology and xAI for Credit Risk Management - Equifax Project in the Spanish Regulatory Sandbox

August 30, 2023

Credit Scoring and Credit Control XVIII

Outline

- Using Machine Learning algorithms for Credit Risk models
  - *Context and regulations*
- Equifax project in the Spanish Regulatory Sandbox
  - *Studying the impact of Machine Learning algorithms on Credit Risk models using real data*
- Main results from the Sandbox project
  - *Predictiveness and explainability*
- Conclusions

Why improve credit risk models?

Reduce overindebtedness and credit losses:

- **9.1%** of households in EU27 are in arrears (EU-SILC, 2021)
- **9.7%** of loans show significant credit risk (ECB, 2022)

Better access to affordable finance:

- **34%** of SMEs say access to finance is a significant problem (EIF, 2021)
- People who are young, rent their home, have lower incomes or who recently arrived in the EU all find it **harder to access** credit, or **pay more** for it.

Better compliance with the Mortgage Credit Directive, Consumer Credit Directive, GDPR and EBA guidelines

"Institutions need to maintain good credit risk management and monitoring standards, that is essential for supporting lending to the economy."

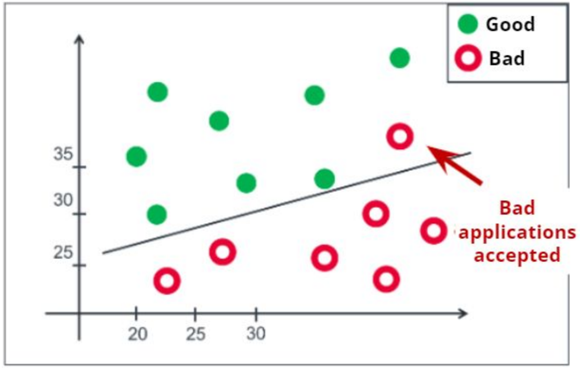
José Manuel Campa
EBA Chairperson

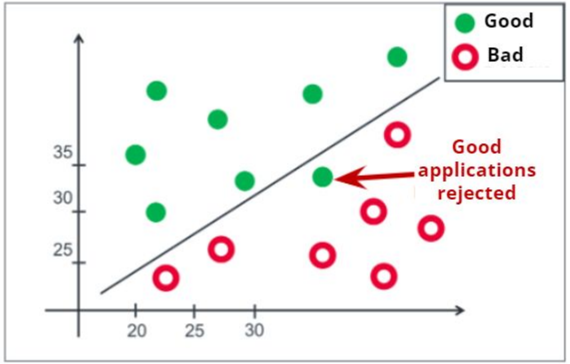
Using Machine Learning algorithms for Credit Risk models

Context and regulations

Logistic regression is effective and easy to explain, but there are false positive and false negative decisions

Can we improve the performance of the traditional statistical techniques used in Credit Risk models?



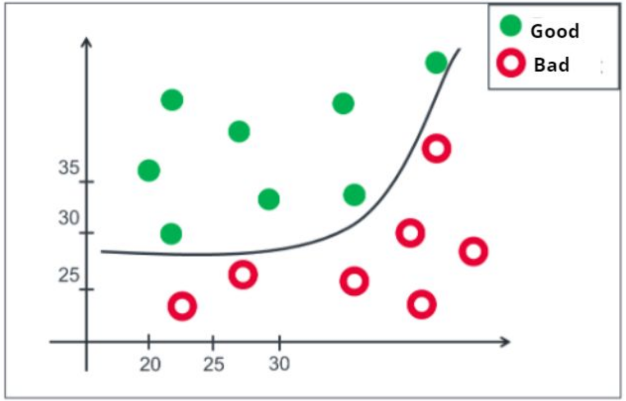


For decades linear algorithms such as **Logistic Regression** have been used successfully to build credit risk models to predict the probability of default.

Logistic Regression guarantees **full explainability**, but it can also miss nonlinear relationships between variables and lose precision

Using ML for Credit Risk management can mean fewer false positive and false negative decisions

Machine Learning (ML) algorithms can improve the performance of Logistic Regression (*)



ML algorithms such as XGBoost or Neural Networks can produce models which are **more effective** in detecting nonlinear relations between the input variables and the target, improving the **predictive power** of models.

So financial entities can maximize the number of loans they approve and consumers benefit from **financial inclusion**.

But when we **add complexity** to algorithms, we gain precision but we can **lose explainability.**

(*) *"Understanding the performance of machine learning models to predict credit default: A novel approach for Supervisory Evaluations"* Andrés Alonso and José Manuel Carbó (2021), Bank of Spain, Department of Financial Innovation.

How the AI black box effect impacts on credit risk management

Context

The **“Black Box”** effect in ML can cause mistrust among consumers and **problems** with risk management and supervision in financial entities

EBA(*) states that institutions must “**understand** the underlying models used” in credit granting. Explainability and interpretability are the key to using ML in Credit Risk

The main challenge is defining an **explainability framework** that allows the trustworthy use of ML

Risks

To add an explainability layer to unconstrained ML models, we can only use **approximations**: surrogate models or Shapley values.

The **global behaviour** of the average population does not always correspond to the **individual behaviour** of each consumer. That can make the model illogical, difficult to understand and difficult to explain.

The main risk for lenders when using approximations to explain ML models is giving the wrong explanation for a credit decision and losing control of credit risk

EQUIFAX®

(*) EBA (2020) *Guidelines on Loan origination and monitoring*, and (2020) *Report on Big Data and Advanced Analytics*

The European Artificial Intelligence Act: which will be the impact of the first worldwide law to regulate AI on Credit Risk?

European Parliament

The Artificial Intelligence Act(*) (AIA) is a law on Artificial intelligence currently under discussion in the European Parliament – if approved, it will be the first law in the world on AI by a major regulator.

Like the EU's General Data Protection Regulation (GDPR) in 2018, the EU AI Act could become a **global standard**.

- *In September 2022, Brazil's Congress passed a bill that creates a legal framework for artificial intelligence.*

Legislative process of the AIA:

- *The EU Commission proposed the first draft in April 2021*
- *The target of the EU Commission is approving the **final text** by the **end of 2023***

In the current version of the AIA, the use of AI for credit risk management is considered to be of **High Risk**, which means that using **Machine Learning** for **credit scoring** will occur on additional supervision and regulation.
In this scenario, the use of xAI can provide a trustworthy use of ML for credit scoring fulfilling with the European upcoming regulation

(*) https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206&from=EN

Using xAI techniques such as monotonicity restrictions can overcome the black box effect

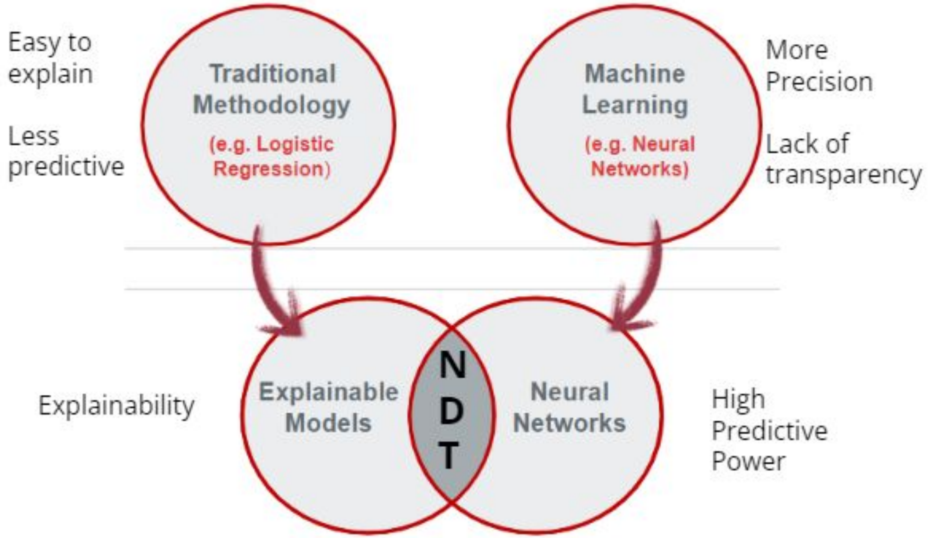
Explainable Artificial Intelligence

Explainable AI (xAI): we can explain every decision taken, for every observation - ML algorithms are explainable by design and no approximations are needed

Using **monotonic restrictions** we force the variables of a ML algorithm to have a strictly monotonic relationship with the target variable: this means the final is model is fully **explainable** and **interpretable** at global and local level

There are several open-source ML models available with monotonic restrictions, including a version of XGBoost or TensorFlow

Equifax developed **Neurodecision Technology (NDT)**: a neural network with monotonic restrictions



Equifax project in the Spanish Regulatory Sandbox

Studying the impact of Machine Learning algorithms on Credit Risk models using real data

What is a Regulatory Sandbox?

How does a regulatory sandbox work?

- A **Regulatory Sandbox** is a **virtual space controlled** and **supervised** by the **Government**, where tech companies working in the **Financial Sector can develop innovative projects** without the burden of following strict legal regulations.
- The Regulatory Sandbox provides innovators, both incumbents and new players, with access to regulatory expertise.
- The **projects** must **use real data** and are supervised by a public institution.
- The first European regulatory sandbox was launched by the **UK** in 2016 - many other european countries, including **Italy** and **Germany**, have recently launched their own

The regulatory Sandbox in Spain

- The **Spanish regulatory Sandbox** was approved in 2020 but finally launched in **2021** due to the pandemic emergency
- The Sandbox is controlled by the Spanish department of Treasury with the participation of three supervising entities, including the **Banco de España** (Central Bank of Spain)
- The Spanish Regulatory Sandbox is now at its fourth call for projects, and the Spanish government is organizing a specific **sandbox for Artificial Intelligence**, to test the possible effect of the AIA
- **Equifax Iberia** showed its interest to participate in 2020 and decided to apply to the first call of the Spanish regulatory sandbox in 2021

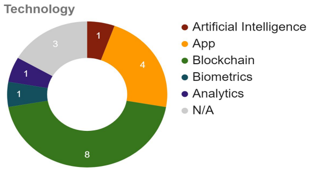
Project in the Bank of Spain's Sandbox: testing the predictive power and explainability of NeuroDecision Technology and other AI algorithms

Equifax proposal

- *Project:* NDT - showing the value of xAI for Credit Risk models
- *Partner and data provider:* Sabadell Bank (a Top 5 bank in Spain)
- *Supervisor:* Bank of Spain
- *Data used:* **Real data** on 3 years of historical data on pre-approved consumer loans from Sabadell Bank

The project had **2 main objectives,** to show that NDT and ML with monotonicity restrictions can:

1. Improve on the **predictive power** of Logistic Regression
2. **Mitigate the black box effect** and achieve an **explainability** level comparable with Logistic Regression



Of the 18 projects selected for the first round of Sandbox, only Equifax's related to Credit Risk and AI

Equifax Project

We developed a Credit Risk Model to predict the probability of default in 12 months, comparing the **performance** of several algorithms:

● NeuroDecision Technology ● Monotonic XGBoost	Monotonic ML
● Logistic Regression	Linear models
● Unconstrained Neural Networks ● XGBoost ● Random Forest	Unconstrained ML

After training the models, we compared the **explainability** and **interpretability** of each algorithm to assess if they provided logical explanations at at global and local levels.

Then we estimated the **financial inclusion** benefits of ML compared to Logistic Regression, simulating the improvement in the number of accepted loans according to the risk appetite of a financial institution.

EQUIFAX

Equifax project in the Spanish Sandbox: a data science project developed in six phases under the supervision of the central Bank of Spain

Equifax's sandbox project was developed in 6 months and completed in March 2022. The project was divided into 6 phases, consisting of the main stages of the development of a Machine Learning model for credit scoring, plus the study of the explainability framework of the algorithms and a study of the project impact on current EU current regulation and on financial inclusion.

	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6
	Data extraction and reception	**Data Ingestion and EDA**	**Preprocessing Data**	**Model training**	**Model comparison**	**Impacts and documentation**
	• Definition of the dataset and the target variable • Data extraction • Data Anonymization • Data secure transfer from Sabadell Bank to Equifax	• Data ingestion • Auditing and consistency analysis • Initial Exploratory Data Analysis	• Missings treatment • Variable creation • Outliers treatment • Normalization/ standardization • Bivariate analysis • Variable selection	• Logit, Probit, Decision Trees. • Unconstrained ML: XGBoost, Random Forest, … • NeuroDecision Technology	• Predictive performance • Stability and variable importance. • Explicability framework visualization. • xAI benefits over other techniques.	• Financial inclusion • Customer protection • Economic Impact • Alignment with the EBA BD&AA Report • Documentation
Objectives	Definition of the test case, dataset and target variable determination, extraction, anonymization and secure transfer from the Bank to Equifax.	Data Ingestion in the analytic environments, Initial Exploratory Analysis and data quality.	Modeling subsamples definition (train, test, etc.), variables creation, bivariate analysis and pre-selection of variables for training models.	Training and calibration of models using a set of algorithms that includes: traditional models, constrained and unconstrained Machine Learning models	Comparison of predictive power of models using KPI metrics such as GINI, KS, etc., study of local and global explainability for all the algorithms	Estimating the financial inclusion impact of using Machine Learning algorithms. Identifying the main risks and benefits of each algorithm. Conclusiones, recommendations, regulatory barriers identified and final documentation.

Main results from the Sandbox project

Predictiveness and explainability

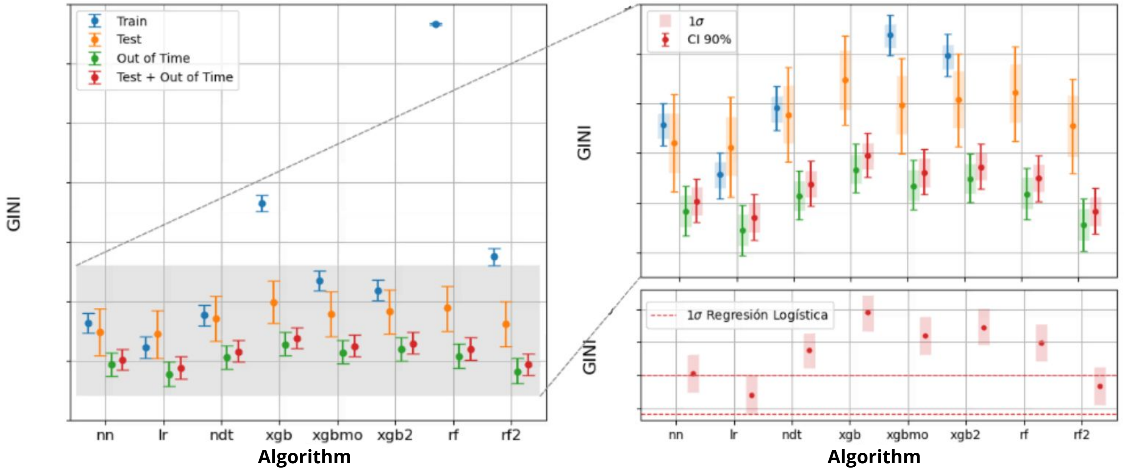
Predictiveness: improvement in predictive power with ML limited due to the characteristics of the dataset

Improvement for all ML algorithms compared to Logistic Regression was **limited**, and lower than what observed in other similar projects using real credit risk data. This was because:

- The baseline Logistic Regression model had already a very **good level of performance**, leaving less room for improvements.
- The most important variables in terms of predictive power showed a very **linear relationship** with the target variable.
- **Feature engineering** and **business logic** had an impact on the results - removing it significantly improved predictiveness.

But NDT still **reduced the rate of consumers declined for credit by 25%** and kept credit risk within the bank's risk appetite.

Improvement in out of time sample versus logistic regression	
XGBoost	3.5%
Monotonic XGBoost	2.6%
NDT	2.0%
Neural Networks	1.9%
Random Forest	1.3%



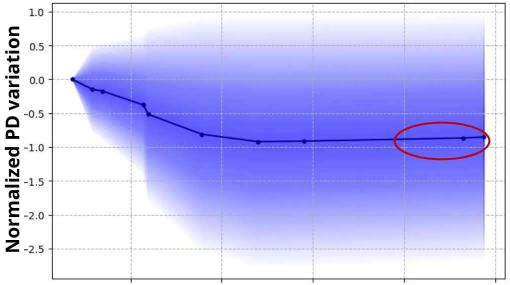
Explainability (1): unconstrained Machine Learning can produce illogical, unexplainable results

With **unconstrained ML**, we found several variables and observations did **not** show a **monotonic** relationship with the default probability, leading to the possibility of **illogical results.**

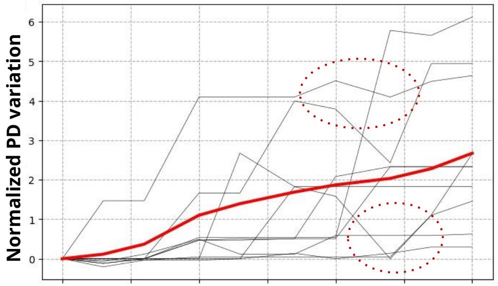
We found cases - on global or local scale - where negative credit behaviour was positively rewarded in the credit score (e.g. higher number of days in default increased the score). Even if we used using techniques like Shapley values. **Unconstrained ML models were not always fully explainable and interpretable.**

Explainability of XGBoost (unconstrained ML)

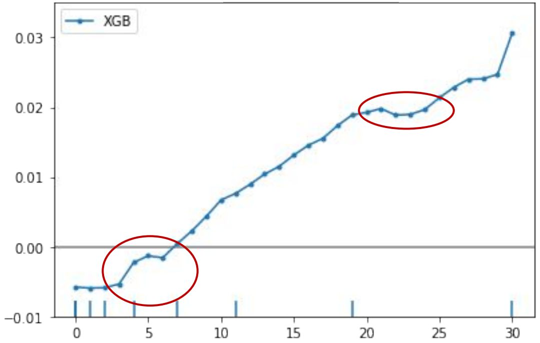
PDP



ICE



ALE



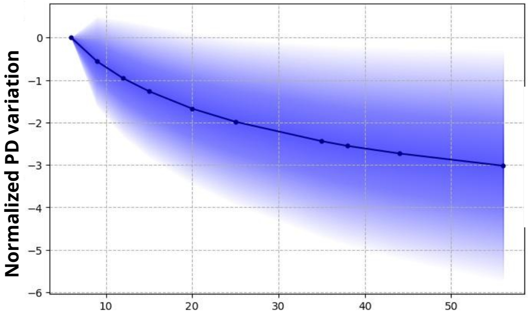
Explainability (2): NDT and ML with monotonicity restrictions can remove the “black box” effect and achieve full explainability and interpretability

In the case of **monotonic algorithms** like **NDT**, the relationship between variables and the default probability was always **strictly monotonic,** as with logistic regression. This was also valid for every individual observation of the dataset.

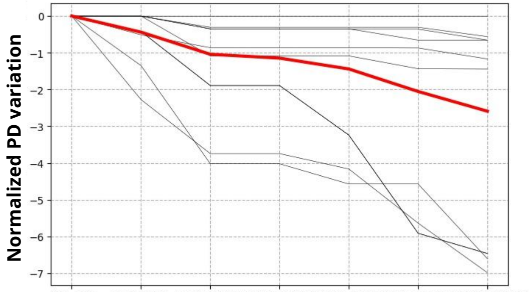
Monotonic ML models are always explainable and interpretable at global and local scale

Explainability of NDT (constrained ML)

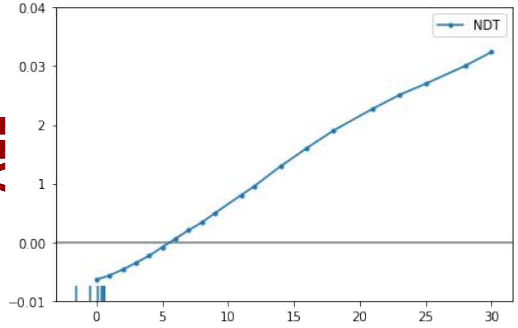
PDP



ICE



ALE



Shapley values: application to Credit Risk models and differences between constrained and unconstrained ML algorithms

- The sum of all Shapley values corresponds to the difference between the prediction of the model an a reference value.

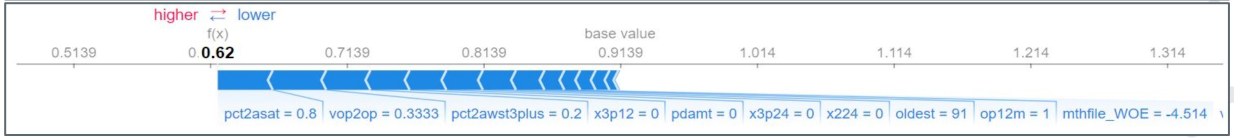
$$\sum_{i=1}^{F} \phi_i(x_j) = E_X[f(X)] - f(x_j)$$

- Shapley values can be expressed as the **marginal contribution** of each variable to a **credit score** : it is possible to select a **reference value** to provide explanations of each variable's contribution to a score using Shapley Values.
- The reference value can be selected as the **maximum score** of a model: in this case, the explanation of the score will be described by:

$$\sum_{i=1}^{F} \phi_i(x_j) = max[f(X)] - f(x_j)$$

- For monotonic models, using this reference value, all Shapley values will be **negative** when the value of a variable does not correspond to the **maximum** (or **minimum**) **-** the sorted Shapley values can be used to produce *Reason Codes(*)*
- For unconstrained models, reference values do not have the same meaning, and positive Shapley values can appear

Unconstrained Neural Network



Monotonically constrained Neural Network (*NeuroDecision Technology*)

EQUIFAX®

(*) [Comparative Analysis of Machine Learning Credit Risk Model Interpretability: Model Explanations, Reasons for Denial and Routes for Score Improvements], Michael McBurnett et al., Credit Scoring and Credit Control Conference, University of Edinburgh, 2021

SHapley Additive exPlanations

Applications to Credit Risk

- The sum of all Shapley values is equal to the difference between the prediction of the model an a reference value.

$$\sum_{i=1}^{F} \phi_i(x_j) = E_X[f(X)] - f(x_j)$$

- In this sense, Shapley values can be expressed as the **marginal contribution the the credit score.**
- It is possible to **select the reference value** to give explanations of a score with Shapley Values.
- The reference value can be selected as the maximum score of a model, in this case, the explanation of the score will be described by:

$$\sum_{i=1}^{F} \phi_i(x_j) = max[f(X)] - f(x_j)$$

- For monotonic models, using this reference value, **all Shapley values will be negative**
- The sorted Shapley values can be used to produce *Reason Codes(*)*
- For unconstrained models, reference values do not have the same meaning, and positive Shapley values can appear

(*) [Comparative Analysis of Machine Learning Credit Risk Model Interpretability: Model Explanations, Reasons for Denial and Routes for Score Improvements](), Michael McBurnett et al., Credit Scoring and Credit Control Conference, University of Edinburgh, 2021

Conclusions

In their report on the Sandbox project “NDT - IA explicable en la gestión de Riesgos”, the Bank of Spain concluded:

“The use of Machine Learning algorithms with monotonic constraints ***can represent an advantage for financial entities****, by giving them alternative tools to logistic models that have* ***an aligned explainability and interpretability framework.****"*

In the Sandbox project, and several other projects using real data provided by financial entities, Equifax observed:

- NDT and ML algorithms with monotonic restrictions always achieve **a predictive power higher** than Logistic Regression
- NDT always offers an **explainability** and **interpretability** framework comparable with that of Logistic Regression
- ML models with monotonic restrictions improve financial inclusion by helping to **maximize** the number of **accepted credit applications** while keeping the risk appetite of financial institutions **under control**.

powering the world with knowledge™

Federico Sembolini, PhD
Data Science & Innovation Manager
m: +34 652162893
federico.sembolini@equifax.com