



Machine Learning in Credit Monitoring: an example of Early Warning System

ABI Supervision, Risks & Profitability 2020

September 22, 2020

INTRODUCTION

Executive summary

GOALS



Banco BPM Group has launched a specific project with the aim of **re-developing the Early Warning group model**, which leads to the **improvement of the Early Warning System efficiency** through the integration of **underlying statistical model**.

The goal is to obtain an **advanced Early Warning System** that **accurately intercepts the significant increase in credit risk**, to allow **more reactive and efficient management actions**.

ACTIVITIES



To achieve this goal, the **Credit department** of Banco BPM, **with the support of CRIF**, has been involved in the **following activities**:

1. **Develop a new Early Warning scoring model**, even through the use of alternative techniques (e.g. Machine Learning), for estimating the probability of occurrence of the target event
2. **Evolution of actions and strategies** according to anomalies detected by EW scoring
3. **Integration of the EW model in the processes used for credit monitoring**
4. The new **Early Warning System will be implemented by a specific engine** integrated with bank procedures.

INTRODUCTION


Executive summary

Banco BPM Group is actively working to improve the ability to monitor and manage underperforming exposure.

In order to do so, **Banco BPM has recently deployed the new PMG** and has started to evolve to Early Warning System.


The new Early Warning System will be data-driven, easy-to-use and integrated in the process & procedures of the Group

STATISTICAL APPROACH




The model development will be based on statistical techniques and machine learning in order to exploit the information available

SUPERVISORY




The JST has expressed interest in the quantitative monitoring of performing exposures. Therefore the action plan proposed by Banco BPM contains the integration of the new Early Warning System

AN EYE ON THE FUTURE



The model will be in place in time for the «new normal» after the Covid 19 pandemic: the teachings of the crisis will be taken into account

COLLABORATION

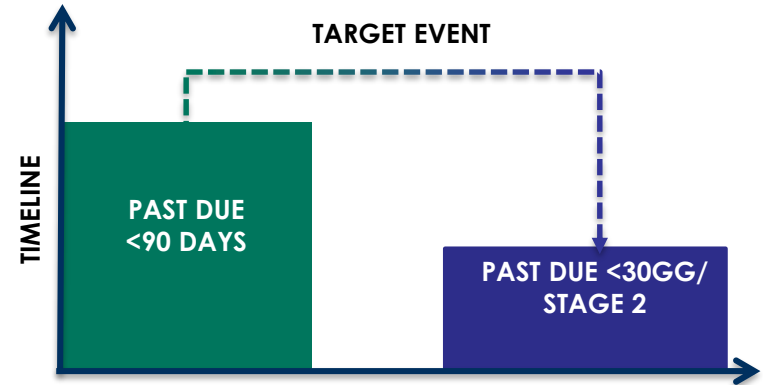
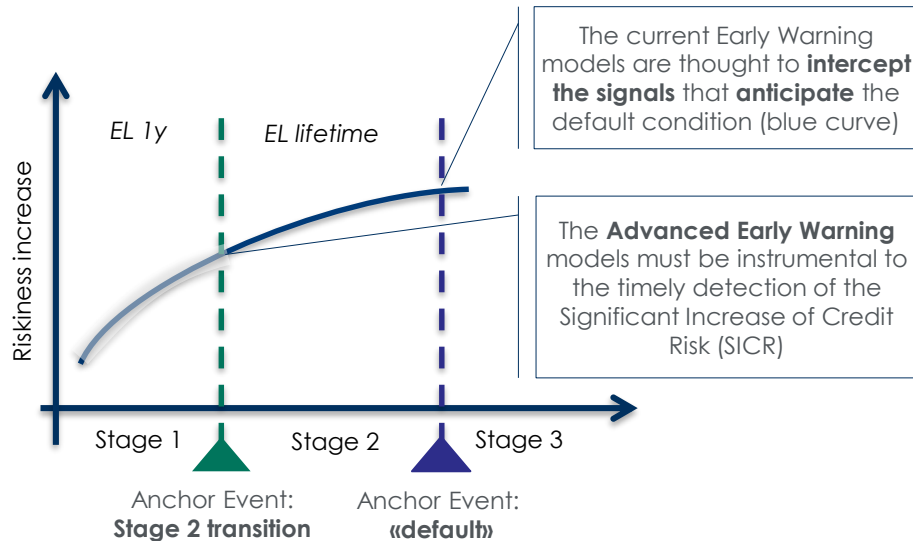


Close interaction between the model development team, the IT team in order to speed up the implementation and the «Monitoraggio Crediti» team to choose the best triggers and to set up the thresholds

THE EVOLUTION OF THE MODEL

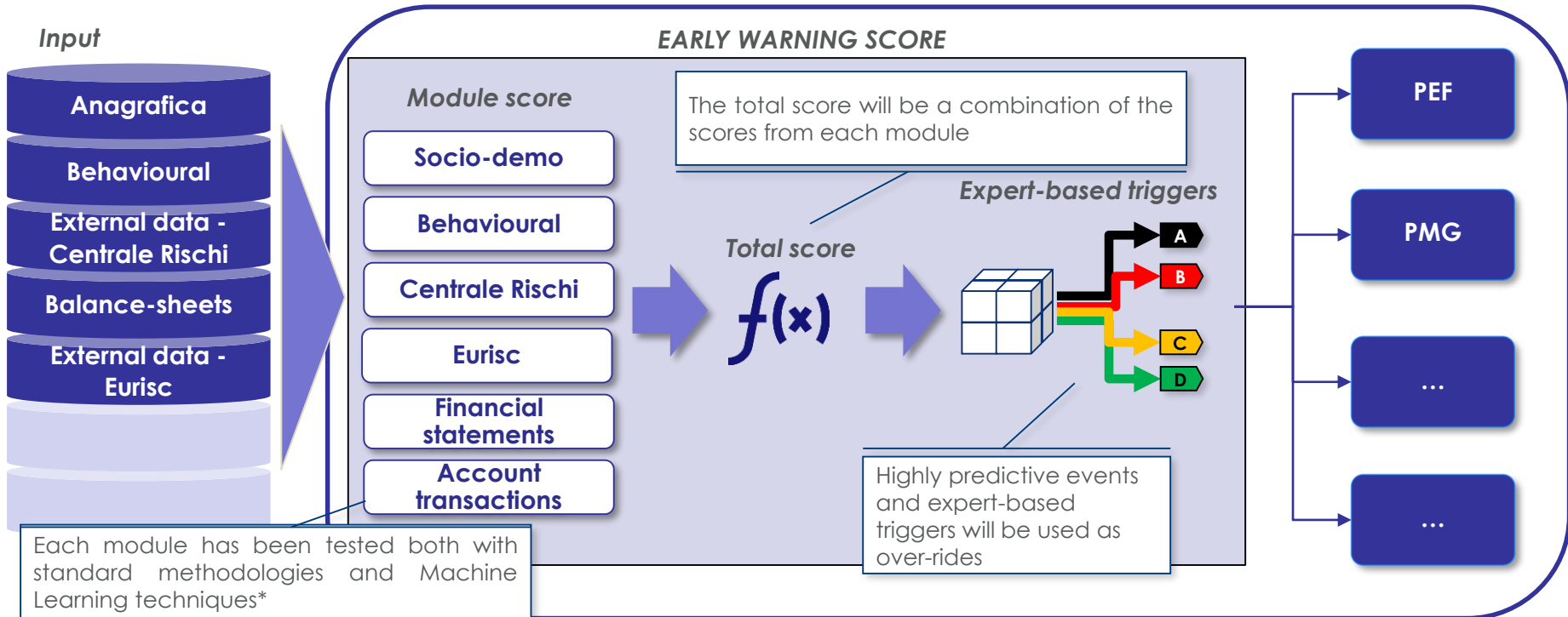
The target event

In line with market best practices and with the evolution of accounting regulations (IFRS 9), the shift of the target event from the “actual default” (the so-called “past-due” to 90 days) to the “significant increase of risk” (for example, the past-due at 30 days), in order to prevent and minimize the transition of credits in Stage 2 (and the related impact on provisions).



THE EVOLUTION OF THE MODEL

High-level description of the «to be» model



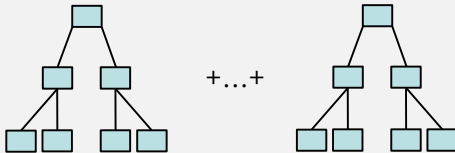
THE MACHINE LEARNING COMPONENT

The model

The chosen Machine Learning approach is the **Random Forest**

DECISION TREES

Random forest algorithm creates many decision trees with bad as target variable.
Every split includes a random set of features (variables)



HOW TO ESTIMATE THE PROBABILITY OF THE EVENT

Each tree «cast a vote» for the target variable. The probability of the target event is the ratio between the number of tree voting «bad» and the total number of trees



THE MODEL

Every feature entering decision trees is a variable of the model.
The feature importance is assigned to every variable and represents the weight in the model.
It depends on the number of trees including the feature and on the position in the trees



THE MACHINE LEARNING COMPONENT



The results of the modules

Random Forest models out-perform Logit models on most of the Segments/Modules: the working group has preferred Logit for Centrale Rischi and Financial statements modules for interpretability issues; Eurisc has been synthesized in the Perform 2 score (proprietary score)

CORPORATE		SMALL BUSINESS		PRIVATE INDIVIDUALS	
Module	RF AR vs Logit AR*	Module	RF AR vs Logit AR*	Module	RF AR vs Logit AR*
Account transactions	==	Account transactions	↑	Account transactions	↑
Socio-demo	==	Socio-demo	↑	Socio-demo	↑↑
Centrale Rischi	↑	Centrale Rischi	↑	Centrale Rischi**	↑↑
Financial statements (Industrial)	↓	Financial statements (Industrial)	↑↑		
Financial statements (Real Estate)	↑↑	Financial statements (Real Estate)	==		

LEGEND

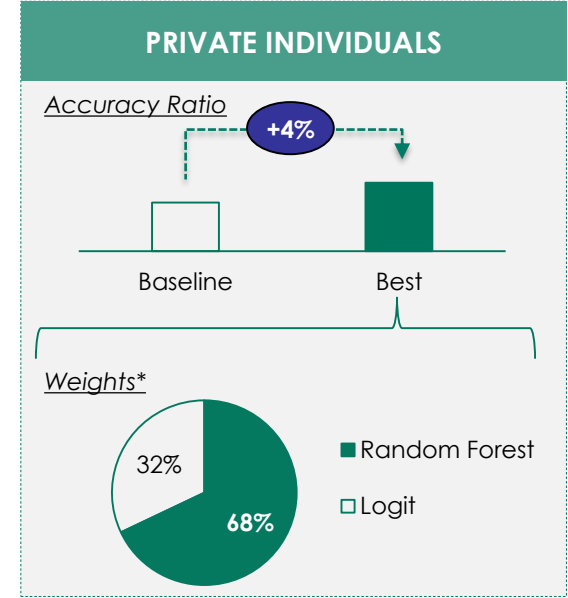
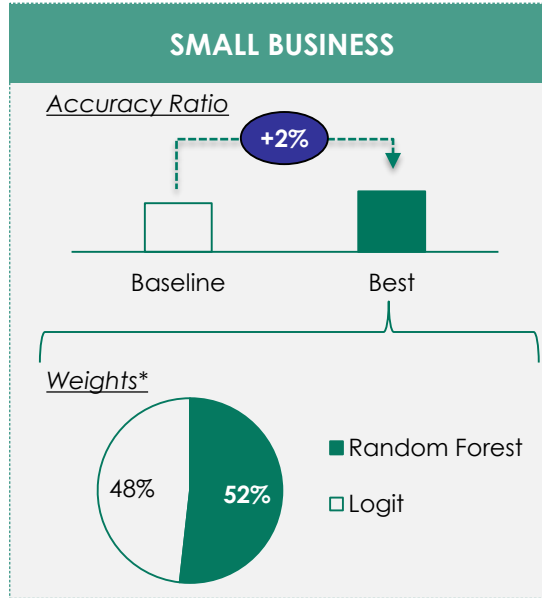
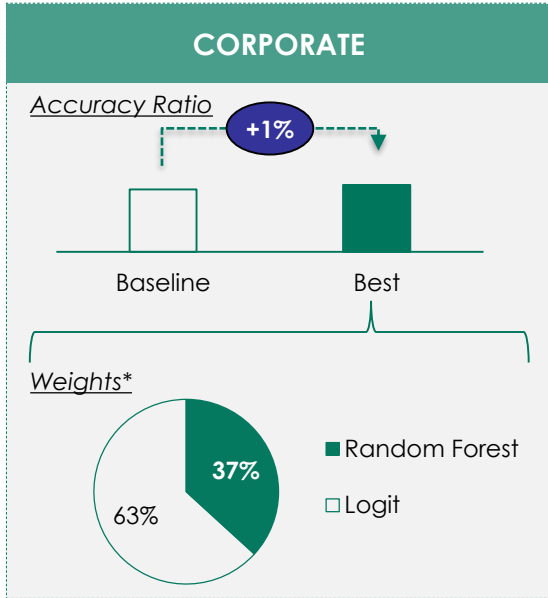
 >8%
  >4%

 Within 4% range
  < -4%

THE OVERALL RESULTS

Results show the highest marginal value for private obligors

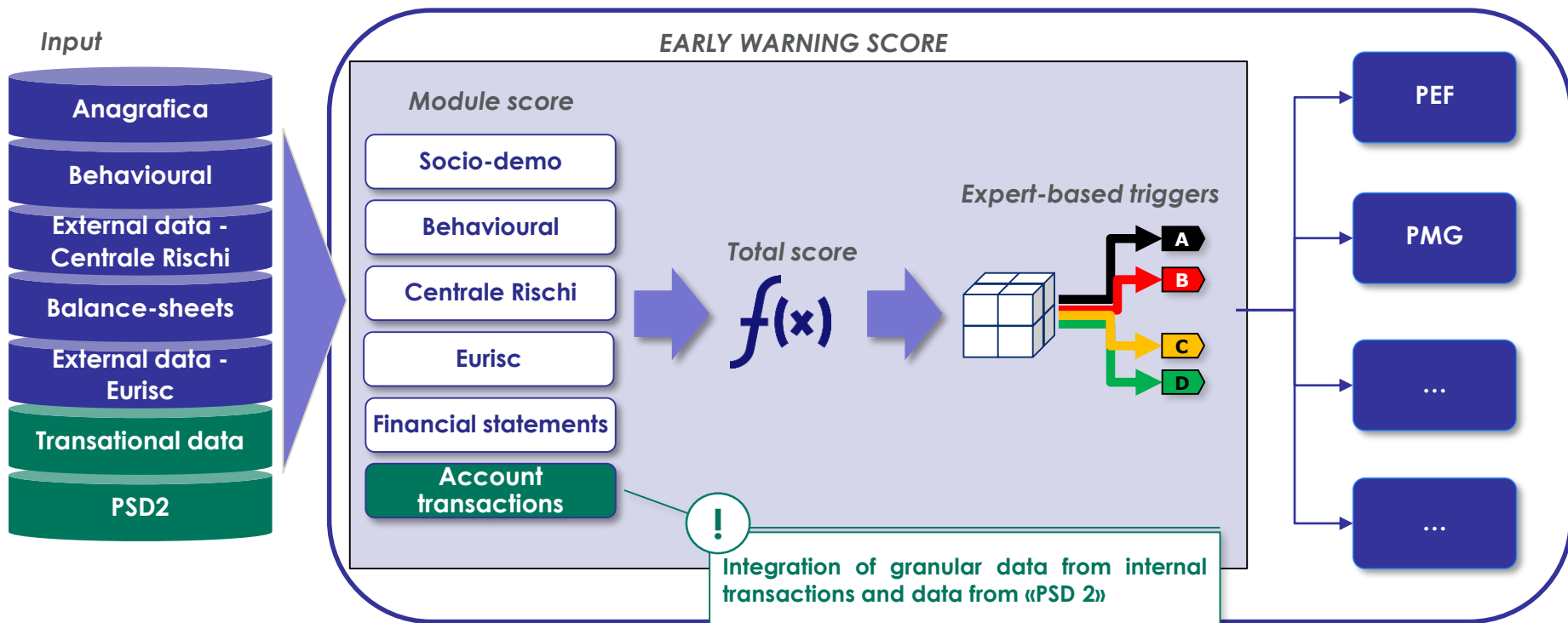
The Machine Learning components proved to be more effective even when integrated with the modules based on logistic regression (*Best*). *Baseline* considers only modules developed with logistic regression



*Weights are normalized excluding the Eurisc component because the inclusion of the Perform 2 is still to be evaluated

NEXT STEPS

Integration of transational data and PSD 2



CONTACTS



Virginia Tirri

Banco BPM – Responsabile Performance Management Bonis – Governo del credito

Mail: virginia.tirri@bancobpm.it

Matteo Camelia

CRIF Transformation Services – Project Leader

Mail: m.camelia@crif.com – Mobile: +39 340 106 2526

