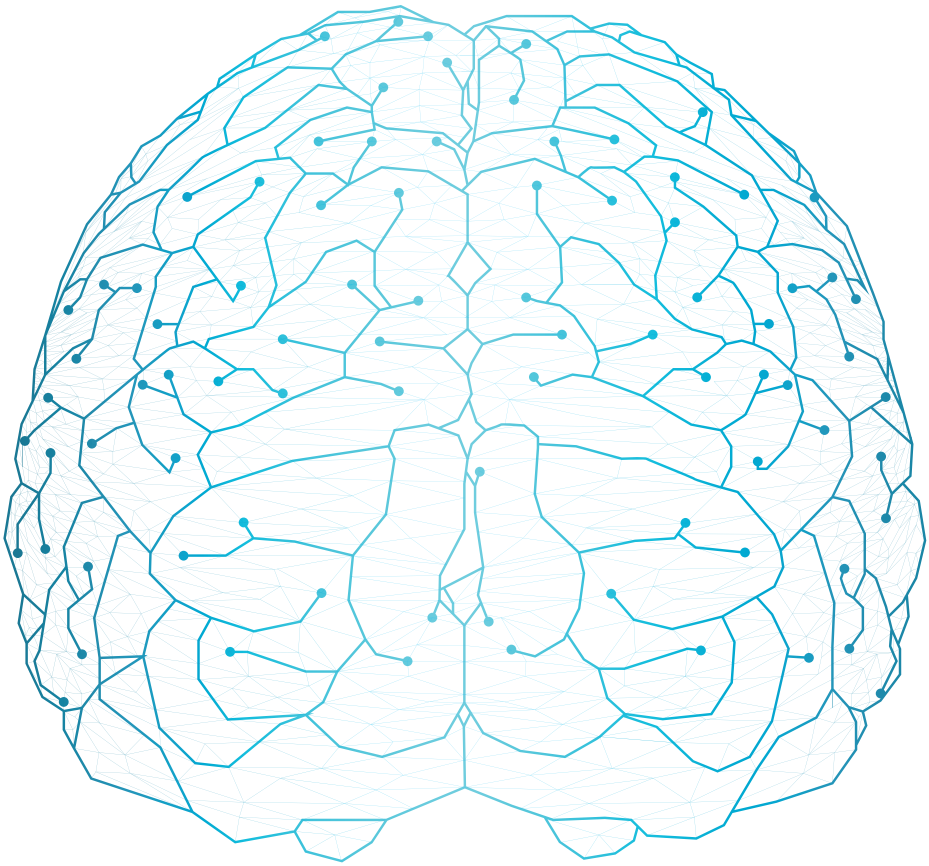
Deloitte.



Explain Artificial Intelligence for Credit Risk Management

Post 2008 crisis, Paul Willmott and Emmanuel Derman were already pointing out one of the major challenges of financial institutions in the Modeler's Hippocratic Oath: "Nor will I give the people who use my model false comfort about its accuracy. Instead, **I will make explicit its assumptions and oversights.**" In addition to this needed transparency, this emphasizes the ethical point related to model use and understanding.
With the development of computational methods, **understanding the output provided by complex models is becoming stronger than ever.**

After the financial crisis, regulators have put a great focus on risk management supervision and expect financial institutions to have transparent, auditable risk measurement frameworks, depending on portfolios characteristics for regulatory, financial or business decision-making purposes. In addition, start-ups and fintechs are developing AI components very quickly, powered by digital growth and the increasing amount of available data.

To align with those new agents, banks must develop more reliable models in order to reduce the decision time and develop better business.

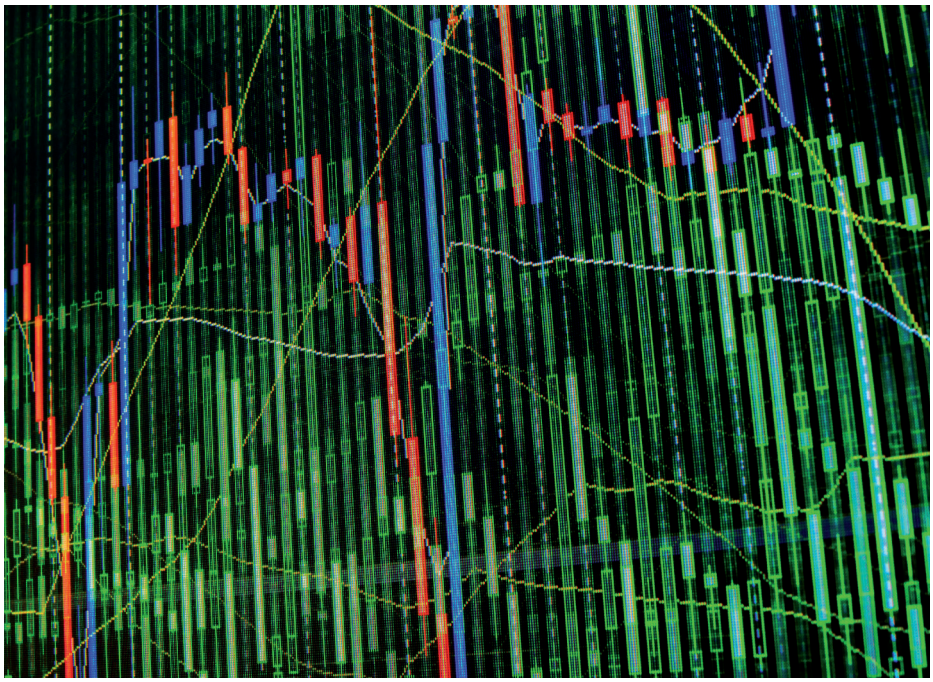
Quantitative modeling techniques are used to get more insights from data, reduce cost and increase overall profitability. Every model contains inherent deficiencies and it is important to keep the focus on reducing model errors.

At the same time, so-called machine-learning method development is skyrocketting. The progresses in computers' processing allows the use of methods such as Deep Learning, tree-based algorithms (Decision Trees, Random Forest and Gradient-Boosting Machines such as the most recent and powerful XGBoost and Light GBM), and ensembling techniques that combine the ouputs of machine-learning models (such as Stacking).
In the credit risk industry, the usage of Machine Learning techniques for model development faces skepticism, notably for regulatory purposes because of the lack of transparency and the known "black box" effect of these techniques.

Although Artificial Intelligence can help model developers to reduce model risk and improve general model predictive power, a wide part of the financial industry remains careful regarding the explainability barrier faced by machine learning techniques. Indeed, the progress observed in the accuracy of models, are often made at the cost of their explainability. Moreover, this lack of explanation constitutes both a practical and an ethical issue for credit professionals, as said by Guidotti and al. (2018) [1]

As pointed out by the latest reports produced by the World Economic Forum and the French prudential authority (ACPR), Artificial Intelligence as a topic has reached on inflection point. In the short term, it seems important that the development of Artificial Intelligence in the banking and insurance sectors satisfy minimum criteria for governance and control. This reflection should cover the proof of the reliability of the algorithms used (with a view to their internal auditability as well as external), models' explainability and interactions between humans and intelligent algorithms.

Artificial Intelligence is already transforming the financial ecosystem, offering a wide range of opportunities and challenges, across different sectors (deposit and lending, insurance, investment management, etc.), therefore the definition of AI model governance is becoming a key concern. As a consequence, **understanding and explaining** the output of machine learning is becoming a top priority for banks and regulators.



The toolkit opening the black-box

Deloitte has designed the Zen Risk platform, which enable its users access to the most advanced and modern tools at each modelling and validation step. Its objective is to provide pre-approval with tools enabling them to automate a part of their work, to compare their model with different approaches and, finally, to give them the keys to integrate transparent rules identified by Artificial Intelligence into existing models.

Originally designed for challenging internal models, this approach could be adapted to underwriting and preapproval credit processes, building eligibility scores or underwriting scorecards. This black-box opening approach can also be extended to collections and recoveries, loss management, including litigation recoveries models, recovery forecasting models or restructuring and discount models.

The Zen Risk toolbox aims to explain both an isolated observation and the overall decisions taken by the algorithms. These techniques are instrumental in the era of increasingly precise models, to the detriment of their interpretability.

The approach flows smoothly and gradually, through many dimensions (model explanation, important features, outcome global explanation, individual forecast explanation).

Machine learning modelling

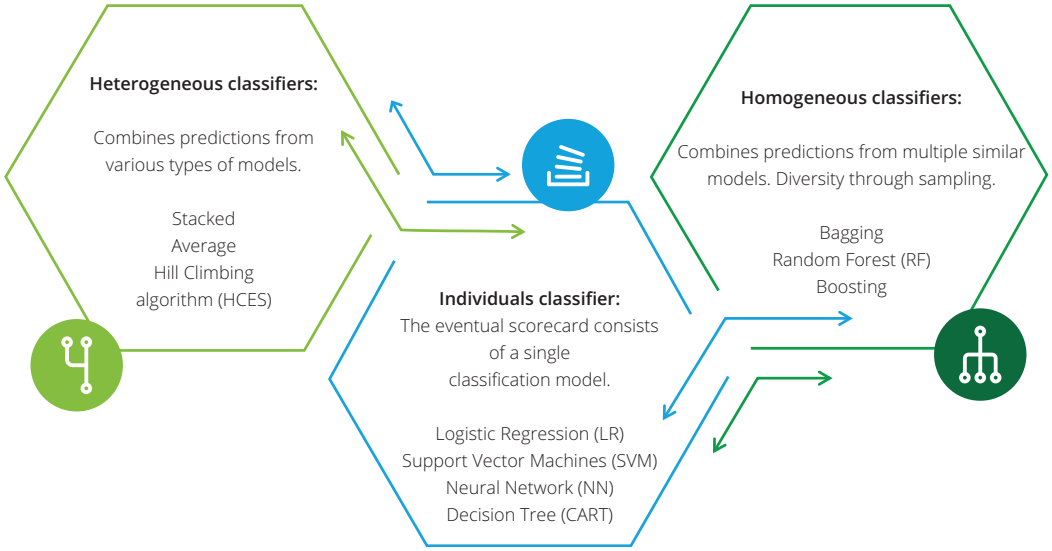
The process starts with an advanced data pre-processing step, where a wide range of machine learning tools are used to improve data quality. The process includes data imputation, when relevant, data filtering for very sparse data, and detection and management of outliers. The choice of the techniques used is left to the modeller, with the possibility to visualize the real time effect of the methods used on the dataset.

[1] A Survey of Methods for Explaining Black Box Models, Guidotti and al. (2018)

The solution then explores many models, from the well-known logistic regression to the most recent Boosting (Light GBM, XGBoost) and Neural Networks.

The table below sums-up some of the most widely used methods:

Algorithms considered:



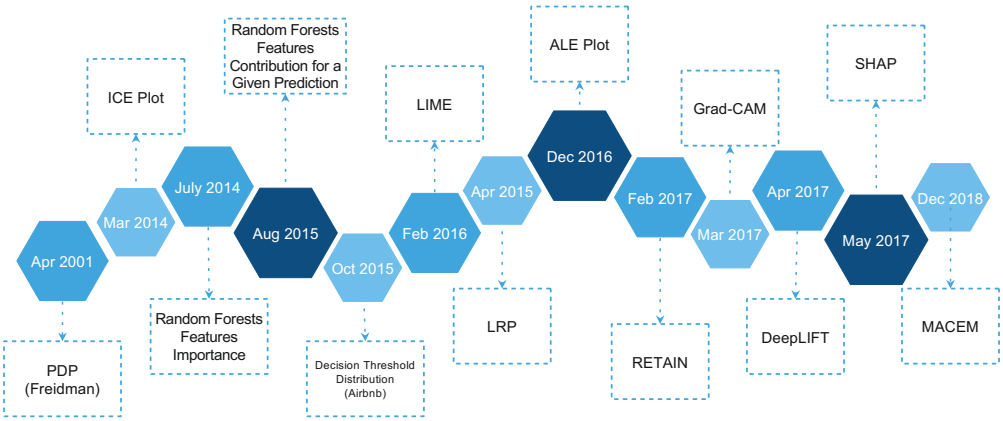
For these complex models that need a high degree of parameter tuning, the use of **hyperparameter optimization algorithms such as genetic algorithms are necessary to reach the best performance.**

Indeed, the number of hyperparameters is too important to test all set of possible combinations. Borrowed from biostatistics, genetic algorithms are inspired by the concept of natural selection and help to fight back this issue.

Generally, genetic algorithms consist in keeping the most resilient parents models. Crossing over parents and allowing genetic mutation creates further generations of models. The process is then iterated until reaching satisfactory results.
At the cost of its complex implementation, using the right parameters leads to an increase of performance metrics (ROC AUC, F1 Scores, ...) in a reasonable amount of time. Indeed, in a case for a PD model development, optimizing hyperparameters could lead to a 10% increase in AUC compared to default parameters.

But if a machine-learning model performs well, why do not we just trust the model and ignore why it made a certain decision? [2]
Are we ready to believe the outputs with high degree of confidence, without taking into account ethical consideration? Of course, the answer is no, and it becomes necessary to open the black-box!

The development of methods for opening the black box has increased considerably in recent years.



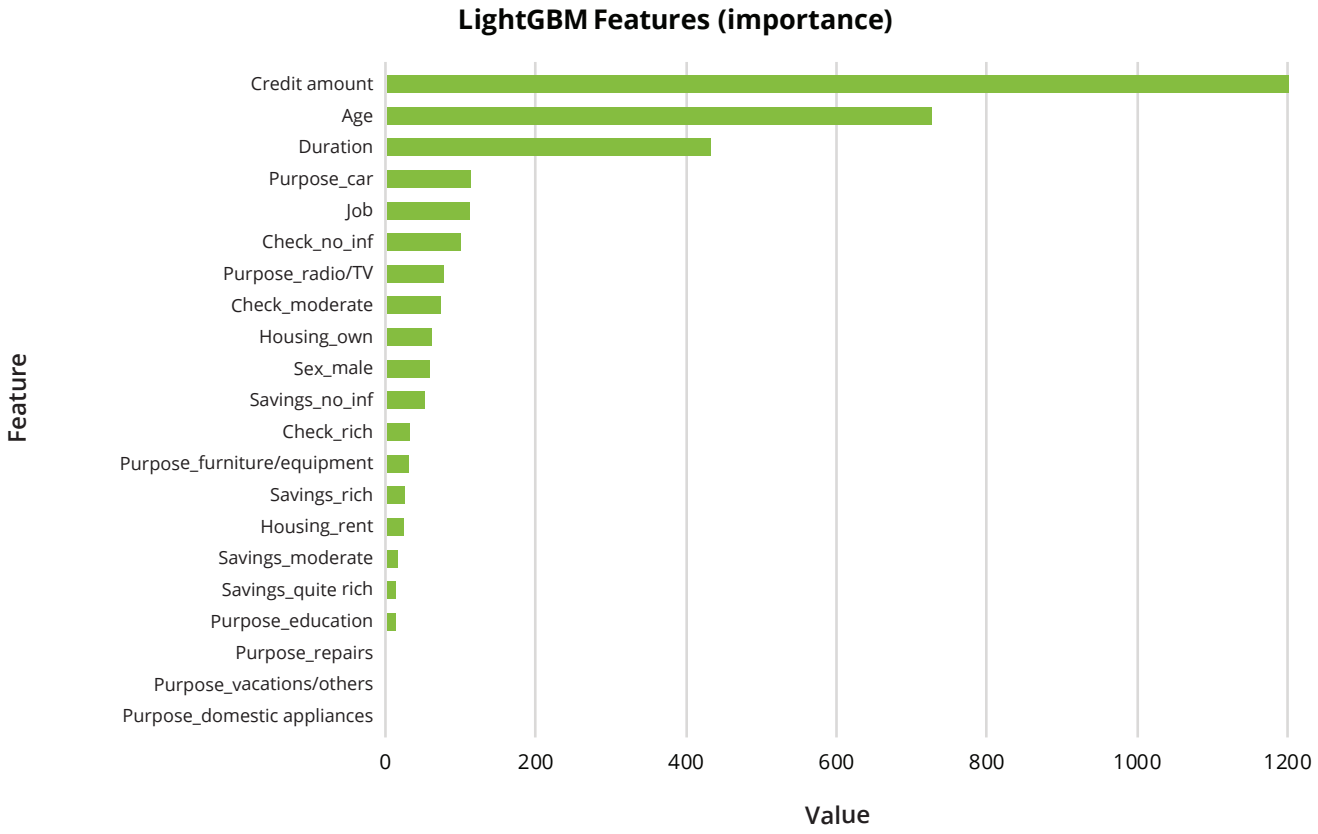
[2] Christoph Molnar, A guide for Making Black Box Models Explainable

Model explainability

The first model explainability tool is often provided by the model itself. Indeed, and this is especially true for tree-based methods, the algorithms can assess the importance of each variables, giving a hierarchy of features importance. To do so, it computes the impact of changing a variable in a tree (by another random one) on the model evaluation metric. The more the model quality decreases in average, the more the variable is important.

For example, in the graph below, we can observe that the credit amount, the borrower's age or the maturity (duration) are the most important variables in the dataset. On the contrary, some variables have a limited impact and bring almost no information.

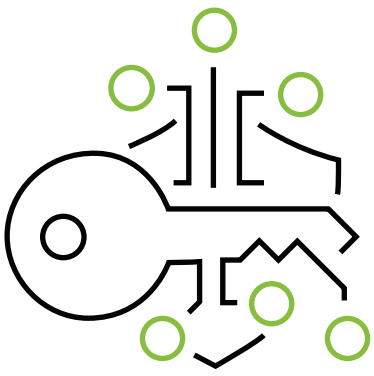
This method could be used for variable selection, before the application of a most usual kind of model such as a logistic regression.



Model-Agnostic methods

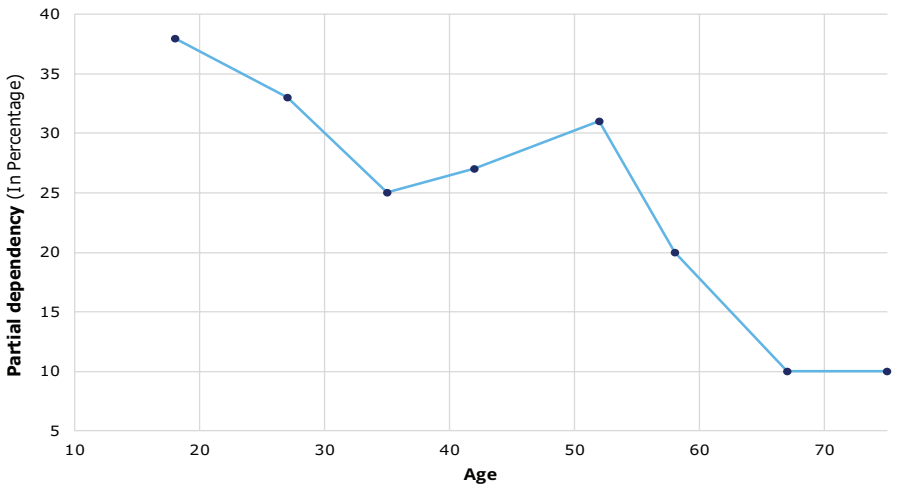
In the recent literature, the research about models explainability has increased. Some approaches are remarkable:

- Partial Dependencies Plot (PDP) aims to introduce variations in input variables and plot the output of the model along these variations. It is one of the most used techniques, and it gives users a good review of the response of the model to a feature globally.
- Local Models (LIME) focus on using interpretable models to explain locally the model's decisions.
- Shapley value, the most sophisticated available approach.



Model sensitivity to variables, the Partial Dependecy Plot example

Partial Depency Plots highlight the sensitivity of a model output to variation of a feature. It exhibits the sense of the relationship (positive or negative effect) and quantify the impact of a variable through a response function.



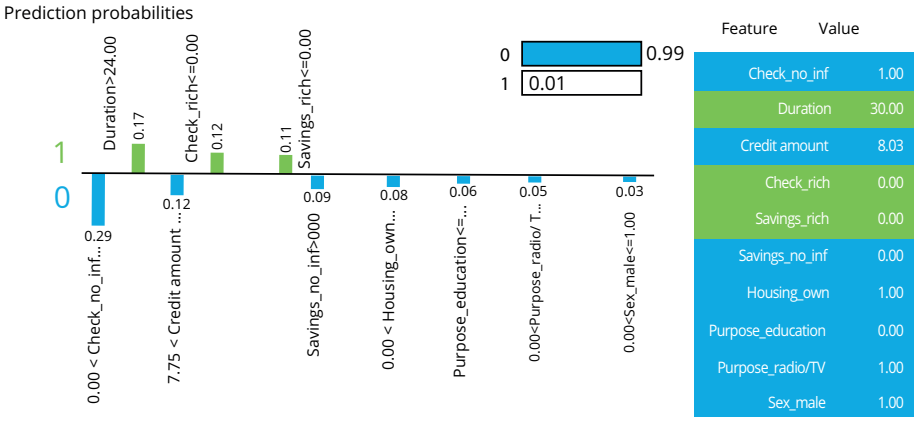
In the example above, the blue line summarizes the average effect of the variable Age on the Default rate. We can see that the default rate is more important for young adults than retired people (that have a fix revenue).
It underlines the non-linear effect of the Age variable and can be used to practice either segmentation, variable selection (if the average effect is linear, the variable is thus useless) or optimization of the binning of numerical variables.
Likewise, it is possible to combine the effect of two variables by using 3Dimensionnal PDP.

Local MODELS: LIME

PDP give a global and a local view in term of **features**. However, it is difficult to plot effect on more than two dimensions (crossed effect of two variables) or to explain the relationship between variables at a global and local (for a given observation) level.

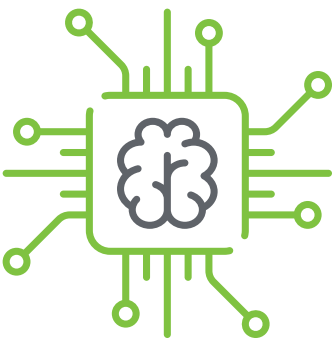
LIME belongs to local models' family, which is a set of models used to explain individual predictions of black box machine learning models. It gives a good approximation of the machine learning output, **locally**.

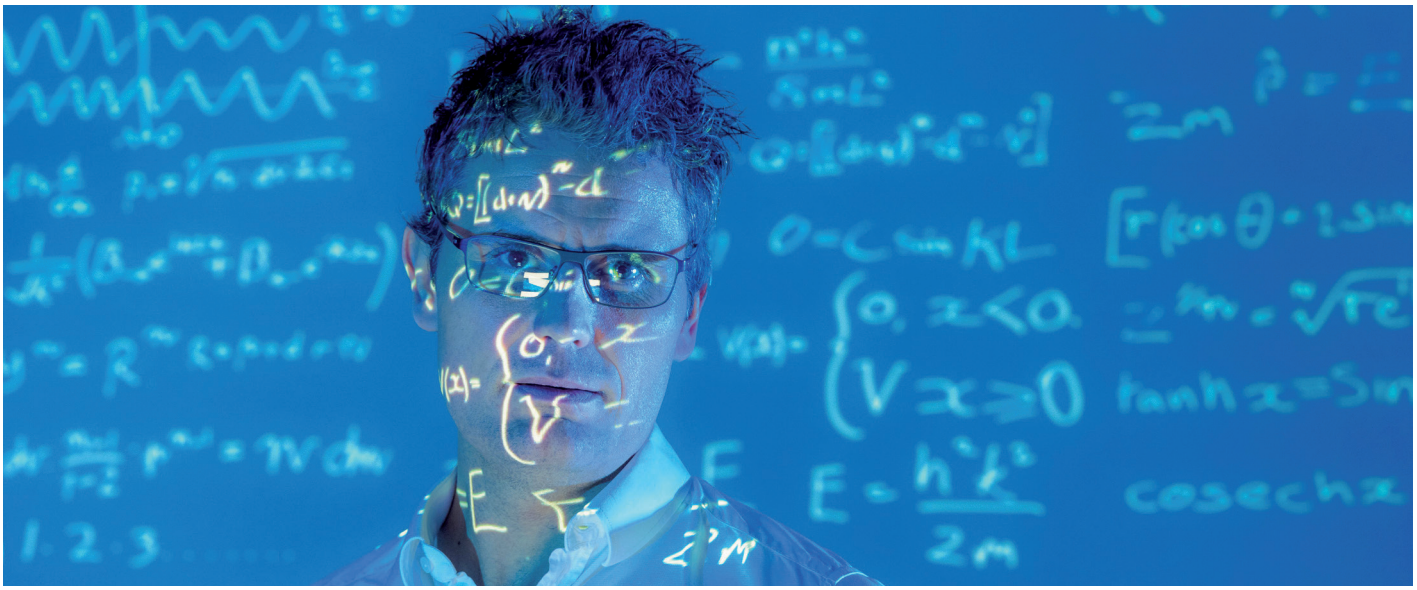
To do so, LIME generates depending on a kernel a new dataset containing permuted samples and the predictions of the black-box non-interpretable model. Then, the principle is to fit an interpretable model (linear regression, decision tree, ...) on machine learning outputs, to explain why a chosen borrower is classified as default or not for example.



The figure above explains why the individual has been classified as default at a local level, and what are the most important characteristics behind this decision. For instance, the duration remaining equals 30 months, which is considered to be high in the studied portfolio. Indeed, having a residual maturity above 24 months increases the probability of default for this specific individual. Samewise, owning a house (House_own variable=1) gives some guarantees and decreases the probability of default.

One can notice that this model is made to be used locally, for borrowers sharing similarities with the example above. Moreover, LIME relies on sampling techniques so the consecutive use of LIME for the same observation will give different results.





The Shapley value analysis

In most recent research works, the Shapley value approach inspired by the game theory essay of Lloyd Shapley seems to be the most promising.
Assuming a Probability of Default model as previously, where the objective is to estimate the probability of default of a counterparty, the Shapley value could help to understand:

- How much has each feature contributed to the average prediction?
- How much has each feature contributed to an individual targeted prediction?

Answering these questions is easy with linear models, however, it is much harder with complex algorithms.

The shapley values aswers this specific issue. It aims to find each variable marginal contribution, averaged over every possible sequence in which the variable could have been added to the set of explanatory variables.

$$\phi_i = \sum_S \frac{|S|!\,(|F| - |S| - 1)!}{|F|!}\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right]$$

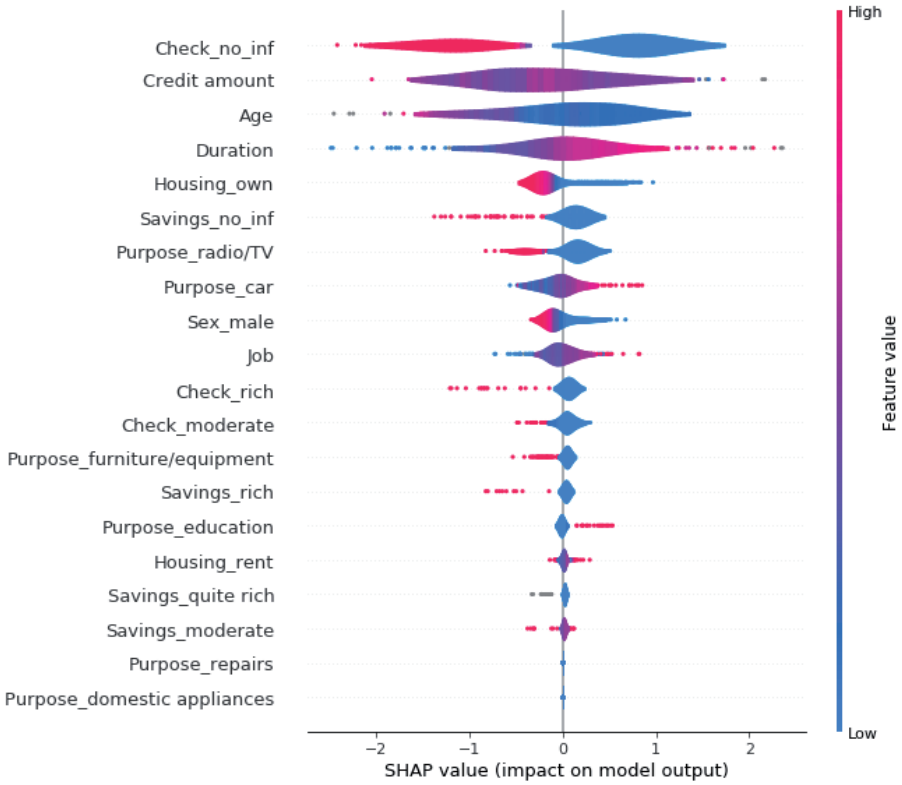
Where S is a vector with a subset of features, F the full number of features, f() the output of a model and i the added feature. Contrary to LIME, the shapley value is unique.

Finally, it gives both a global picture and a downscale effect on a specific individual. Solutions to the feature importance attribution problem are proven unique thanks to their properties of local accuracy, missingness, and consistency.

In the example above, an individual borrower is studied with a Shapley value analysis. His profile seems to be risky (unskilled and non-resident, he wants to buy a car, but does not have a lot of money in his bank account).

Besides, in a global manner, the graphs below exhibit the fact that younger people and small duration loans are riskier, while people making loans for their education are more willing to repay than others. Besides, it quantifies the risk with the Shapley value impact on model output.

However, the main drawback of this method is its computational time, proportionate to the number of features, observations, and the complexity of the model.



The Hybrid approach

Deloitte hybrid's approach is a two stages analysis crossing over the outputs of machine learning models and usual logistic regressions. Indeed, it consists in looking at population that have been misclassified by logistic regression but well classified by an advanced model (e.g Random Forest, Neural Network, ...). Once the population identified, the extraction of business rules takes place in order to override logistic regression model outputs.

Therefore, by using simple logistic regression and additional rules, we succeed in reaching an accuracy level way better than the traditional model and picturing the reality in a more comprehensive way, with the possibility for the modeller to adjust different choices of business rules. Another value of this approach is that expert business can validate the new rules extracted from advanced algorithms and decide whether it makes sense or not.

Conclusion

Looking at algorithm explainability and transparence may also be an enabler to quantify model risk. Indeed, addressing problematics such as inputs and methodology, financial institutions will be able to better quantify model risk arising from these types of techniques.
By analogy, this understanding of models' sensitivity to variables could be useful for risk management and stress testing purposes.

Moreover, recent progresses on the academic field and discussions around the governance of Artificial Intelligence, emphasize the premises for future changes in the model arena in the years to come.

About the authors

Hervé PHAURE
Partner, Risk Advisory
hphaure@deloitte.fr

Hervé is Partner in the Risk Advisory department, in charge Credit Risk Advisory services. Hervé has been involved in risk management areas since more than 25 years. His expertise relates with statistical models in finance, Risk Management, IFRS9, credit processes, valuation of credit portfolios. He coordinates Deloitte Credit Risk Community at European level.

Erwan ROBIN
Senior Consultant, Risk Advisory
erobin@deloitte.fr

Erwan Robin is a senior consultant and works as a data scientist within Deloitte France. His work consists in applying innovative solutions to credit risk management, notably for credit risk modelling. He is involved in research and development topics regarding machine learning.

Deloitte
6, place de la Pyramide - 92908 Paris-La Défense Cedex
Tél. : 33 (0)1 40 88 28 00