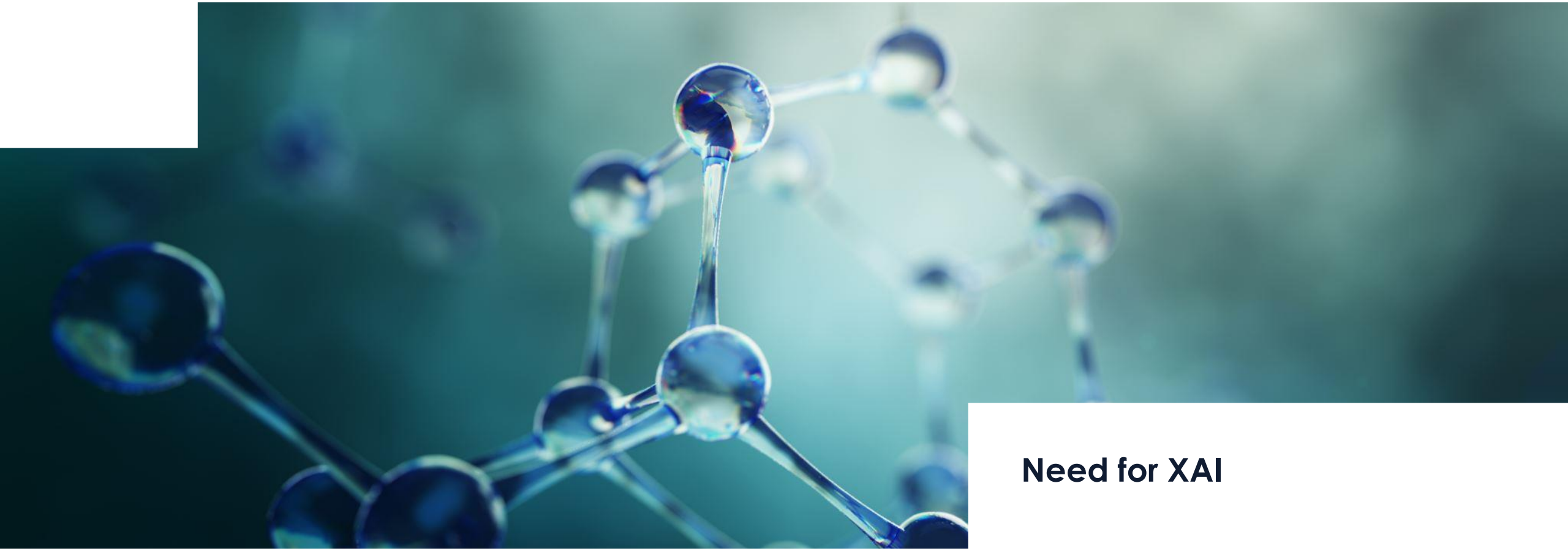
FinanceCom 2022 – Enterprise Applications, Markets & Services in the Finance Industry

University of Twente, 22-24 August 2022
The Netherlands

eXplainable AI for Finance

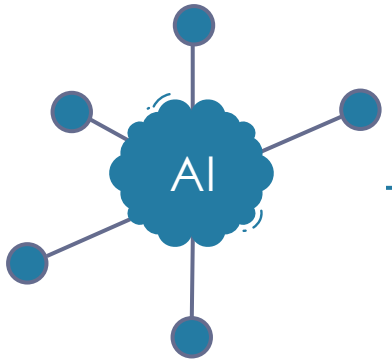
Prof. Dr. Branka Hadji Misheva
Bern University of Applied Sciences, Switzerland

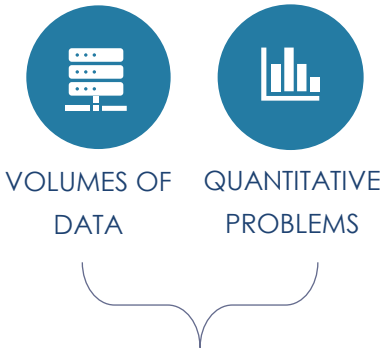
Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences



Need for XAI

eXplainable AI for Finance



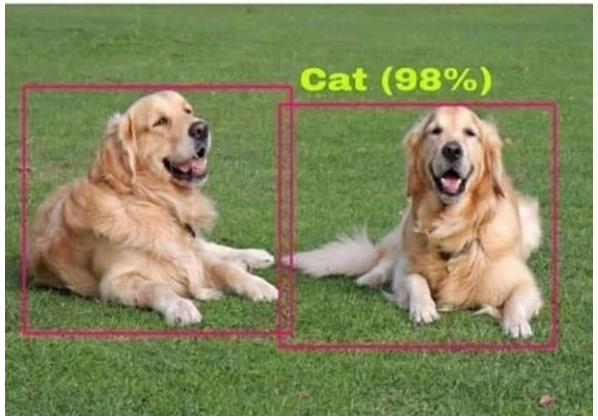


Where is the progress?

Well …

People: *fearing*AI takeover

AI:



No black box excuses – explainability/traceability of models is necessary and can improve the analysis process | It is the responsibility of supervised firms to ensure that BDAI-based decisions can be explained and are understood by third-party experts. Supervisory authorities take a critical view of models that are categorised purely as black boxes. New approaches allow firms using such models to at least gain some insight into how these models work and identify the reasons behind decisions. In addition, a better understanding of models provides an opportunity to improve the analysis process – allowing, for instance, the responsible units in the supervised firm to identify statistical problems.

Bafin (2020)

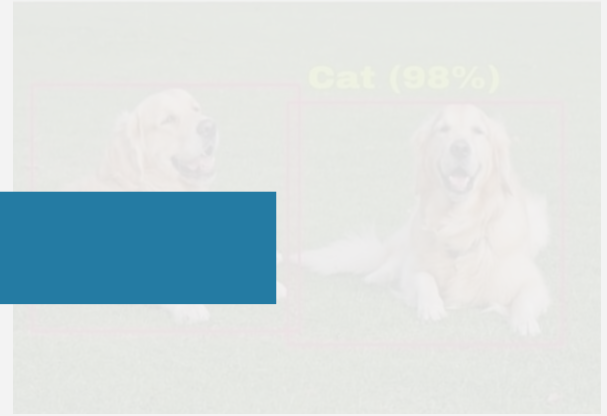
eXplainable AI for Finance

Well …

People: *fearing*AI takeover

AI:

Where is the progress?



Finance

Explainability is the name of the game!

No black box excuses – explainability/traceability of models is necessary and can improve the analysis process | It is the responsibility of supervised firms to ensure that BDAI-based decisions can be explained and are understood by third-party experts. Supervisory authorities take a critical view of models that are categorised purely as black boxes. New approaches allow firms using such models to at least gain some insight into how these models work and identify the reasons behind decisions. In addition, a better understanding of models provides an opportunity to improve the analysis process – allowing, for instance, the responsible units in the supervised firm to identify statistical problems.

Bafin (2020)

Deploying eXplainability

POST-HOC Explainability

- There are many models that are highly interpretable.
- Regardless, for some ML models, **post-hoc explaianbility is required**!
- Post-hoc explainability techniques → understandable information about how an already developed model produces its predictions for any given input!
- Methods are considered in view of two main criteria (Linardatos et al. 2021):
  - the **type of algorithm** on which they can be applied (model-specic vs. model-agnostic)
  - **the unit being explained** (if the method provides an explanation which is instance-specic then this is a local explainability technique and if the method attempt to explain the behavior of the entire model, then this is a global explainability technique).

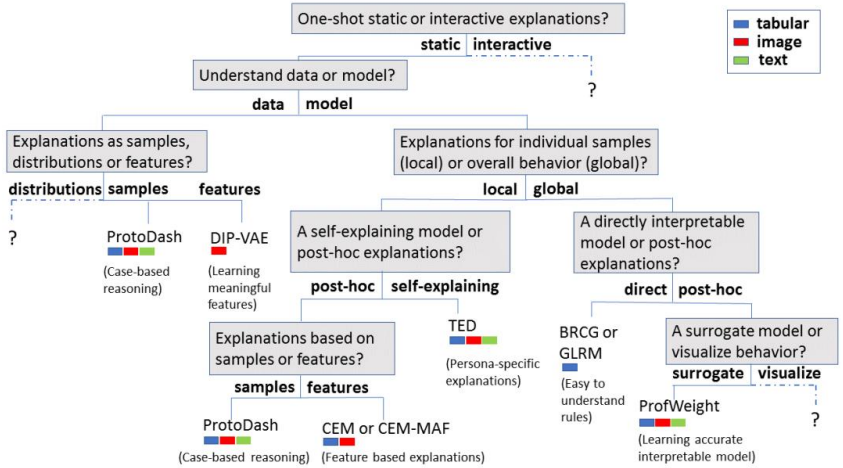


Figure 1. Arya et al. (2019) proposed taxonomy based on questions about what is explained, how it is explained and at what level

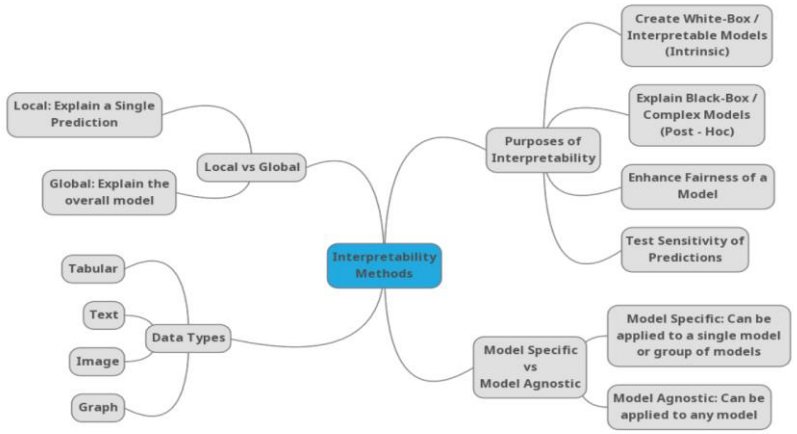


Figure 2. Linardatos et al. (2021) taxonomy mind-map of Machine Learning Interpretability Techniques.

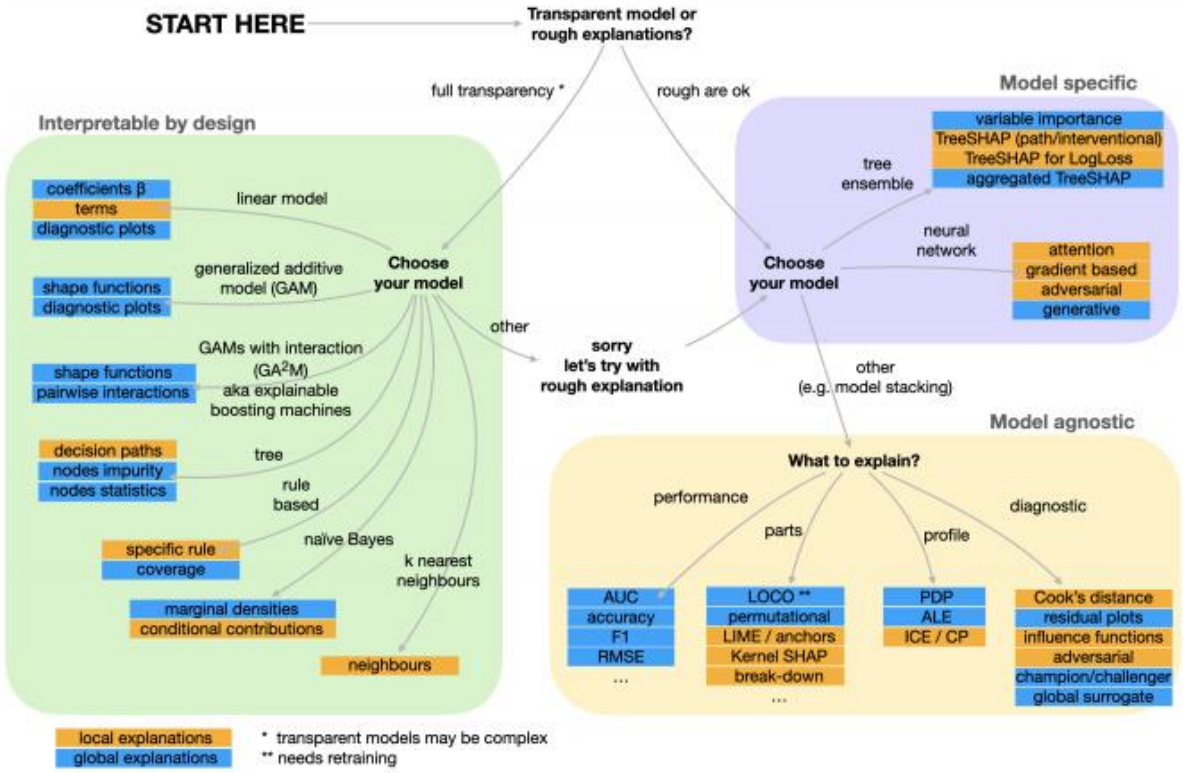


Figure 3. Maksymiuk et al. (2021) model-oriented taxonomy for XAI method

LIME & SHAPLEY: Details

LIME

- The explanation provided by LIME for each observation:

$$\xi(x) = argmin_{g \in G} L(f, g, \pi_x) + \Omega(g)$$

where G is the class of potentially interpretable models (i.e. linear models)

$g \in G$: An explanation considered as a model

$f: \mathbb{R}^d \rightarrow \mathbb{R}$: The main classifier being explained

$\pi_x(z)$: The proximity measure of an instance z from x

- The goal is to **minimize the locality aware loss** L without making any assumptions about f, since a key property of LIME is that it is model agnostic.
- L is the measure of how unfaithful g is in approximating f in the locality defined by π_x.

SHAPLEY

- Given a model

$$f(x_1, x_2, x_3 \dots x_n)$$

with feature 1 to n being payers in a game in which the payoff v is the measure of importance of the subset.

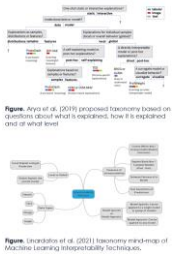
- Marginal contribution $\Delta_v(i, S)$ of a feature i:

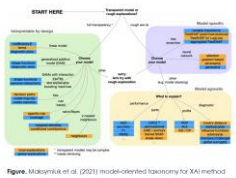
$$\Delta_v(i, S) = \mathrm{v}(\mathrm{S} \cup i) - v(S)$$

- Let $\prod$ be the set of permutations of the integers up to N, and given $\pi \in \prod$ let $S_{i,\pi} = \{j: \pi(j) < \pi(i)\}$ are the players preceding payer i in π, then:

$$\Phi_v(i) = \frac{1}{N!} \sum_{\pi \in \prod} \Delta_v(i, S_{i,\pi})$$

XAI for Finance





Match explainability needs of stakeholders with the XAI methods

How stable and robust are explanations provided?

Performance of XAI methods in view of the unique features of financial data

Wider adoption of AI-based use cases in finance

XAI Research

XAI research in **FINANCE**

Productivity

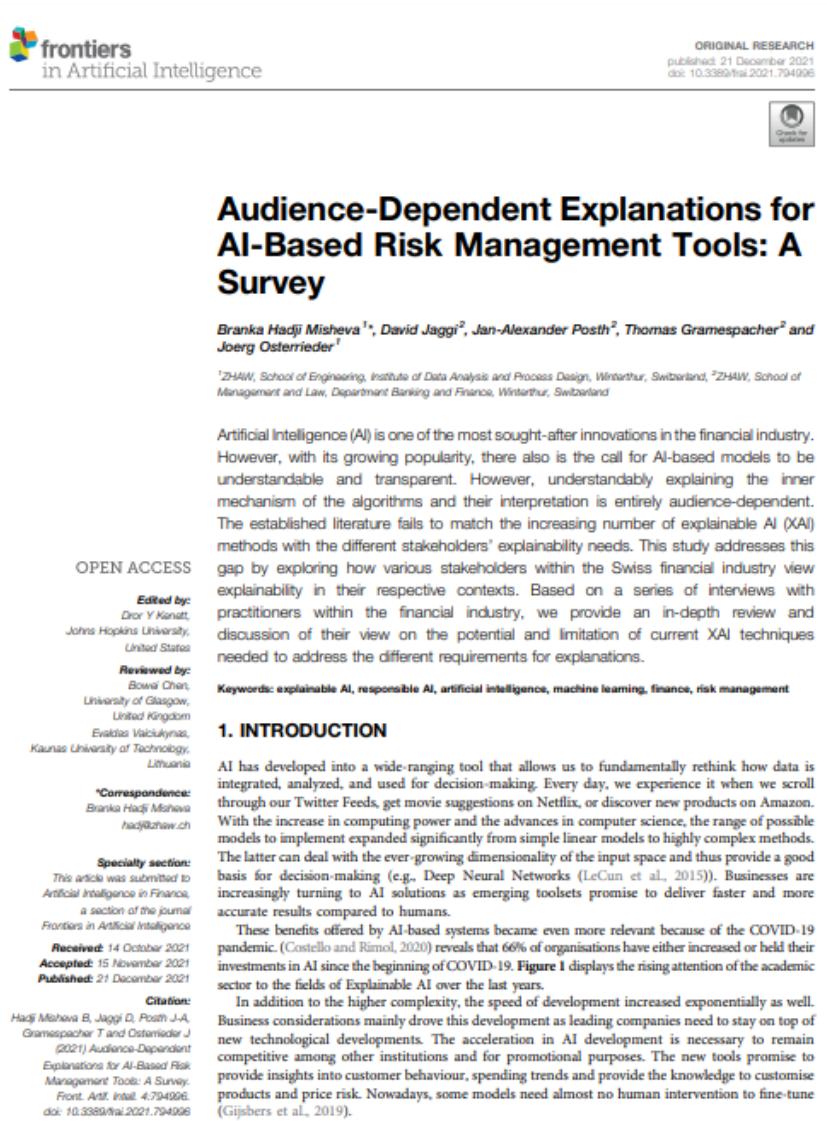
How do we start?

- **We ask the industry!**

- Following the example of (Holstein et al., 2019) and (Bhatt et al., 2020), we ran interviews so to **study how the industry looks at the recent developments in ML and the need for deploying explainable AI solutions.**

- The interviewees were representatives from several **FinTech companies, banks, large banking associations and insurance companies** all of which operate in Switzerland.

- The interviewees were selected to represent the various stakeholders within the financial sector: **model developers/ML engineers, risk and legal experts, higher management & end users.**



frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 21 December 2021
doi: 10.3389/frai.2021.794996

Audience-Dependent Explanations for AI-Based Risk Management Tools: A Survey

Branka Hadji Misheva[1], David Jaggi[2], Jan-Alexander Posth[2], Thomas Gramespacher[2] and Joerg Osterrieder[1]*

[1]ZHAW, School of Engineering, Institute of Data Analysis and Process Design, Winterthur, Switzerland, [2]ZHAW, School of Management and Law, Department Banking and Finance, Winterthur, Switzerland

Artificial Intelligence (AI) is one of the most sought-after innovations in the financial industry. However, with its growing popularity, there also is the call for AI-based models to be understandable and transparent. However, understandably explaining the inner mechanism of the algorithms and their interpretation is entirely audience-dependent. The established literature fails to match the increasing number of explainable AI (XAI) methods with the different stakeholders' explainability needs. This study addresses this gap by exploring how various stakeholders within the Swiss financial industry view explainability in their respective contexts. Based on a series of interviews with practitioners within the financial industry, we provide an in-depth review and discussion of their view on the potential and limitation of current XAI techniques needed to address the different requirements for explanations.

Keywords: explainable AI, responsible AI, artificial intelligence, machine learning, finance, risk management

OPEN ACCESS

Edited by:
Dror Y Kenett,
Johns Hopkins University,
United States

Reviewed by:
Bowei Chen,
University of Glasgow,
United Kingdom
Evaldas Vaicilukynas,
Kaunas University of Technology,
Lithuania

****Correspondence:***
Branka Hadji Misheva
hadj@zhaw.ch

Specialty section:
This article was submitted to Artificial Intelligence in Finance, a section of the journal Frontiers in Artificial Intelligence

Received: *14 October 2021*
Accepted: *15 November 2021*
Published: *21 December 2021*

Citation:
Hadji Misheva B, Jaggi D, Posth J-A, Gramespacher T and Osterrieder J (2021) Audience-Dependent Explanations for AI-Based Risk Management Tools: A Survey. Front. Artif. Intell. 4:794996. doi: 10.3389/frai.2021.794996

1. INTRODUCTION

AI has developed into a wide-ranging tool that allows us to fundamentally rethink how data is integrated, analyzed, and used for decision-making. Every day, we experience it when we scroll through our Twitter Feeds, get movie suggestions on Netflix, or discover new products on Amazon. With the increase in computing power and the advances in computer science, the range of possible models to implement expanded significantly from simple linear models to highly complex methods. The latter can deal with the ever-growing dimensionality of the input space and thus provide a good basis for decision-making (e.g., Deep Neural Networks (LeCun et al., 2015)). Businesses are increasingly turning to AI solutions as emerging toolsets promise to deliver faster and more accurate results compared to humans.

These benefits offered by AI-based systems became even more relevant because of the COVID-19 pandemic. (Costello and Rimol, 2020) reveals that 66% of organisations have either increased or held their investments in AI since the beginning of COVID-19. **Figure 1** displays the rising attention of the academic sector to the fields of Explainable AI over the last years.

In addition to the higher complexity, the speed of development increased exponentially as well. Business considerations mainly drove this development as leading companies need to stay on top of new technological developments. The acceleration in AI development is necessary to remain competitive among other institutions and for promotional purposes. The new tools promise to provide insights into customer behaviour, spending trends and provide the knowledge to customise products and price risk. Nowadays, some models need almost no human intervention to fine-tune (Gijsbers et al., 2019).

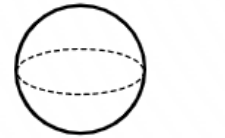
	Explainability needs	Identified issues with classical XAI
Model developers	**Model debugging** • Identify why the model is performing poorly • Understand **time-dependent sensitivities**	Classic XAI methods are deemed useful only for **model developers**
Higher management	**Model transparency** • Are the predictions understandable to a non-technical audience (**show your work approach**)? • Are they in line with current (and maybe forthcoming) regulation?	Available toolkits & visualization **not suited for a non-technical audience**
		No framework for testing stability of explanations provided
Risk experts	**Model audit** • Is the model picking features that an risk expert would pick? • Are **explanations provided stable**? Would similar units receive similar explanations?	No XAI-technique for deep learning methods (DL) which preserves and exploits the natural time ordering of the data

Classic XAI methods are deemed useful only for **model developers**

Available toolkits & visualization **not suited for a non-technical audience**

No framework for testing stability of explanations provide

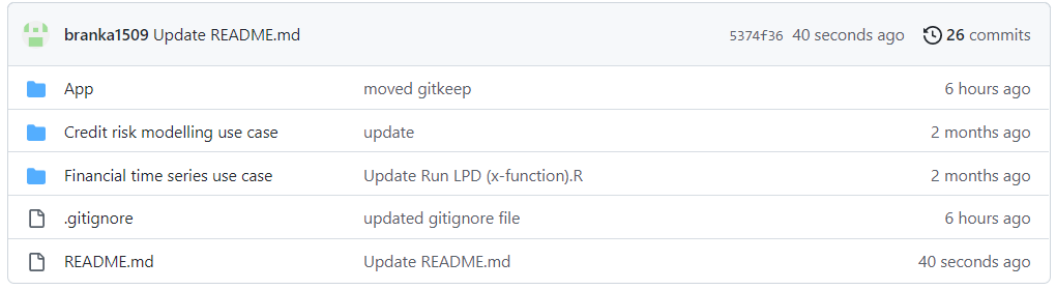
No XAI-technique for deep learning methods (DL) which preserves and exploits the natural time ordering of the data



EXPLAINABLE AI FOR FINANCE

Link to page

git GitHub



branka1509 Update README.md	5374f36 40 seconds ago	26 commits
App	moved gitkeep	6 hours ago
Credit risk modelling use case	update	2 months ago
Financial time series use case	Update Run LPD (x-function).R	2 months ago
.gitignore	updated gitignore file	6 hours ago
README.md	Update README.md	40 seconds ago

Shiny

README.md

Towards eXplainable AI in Financial Applications

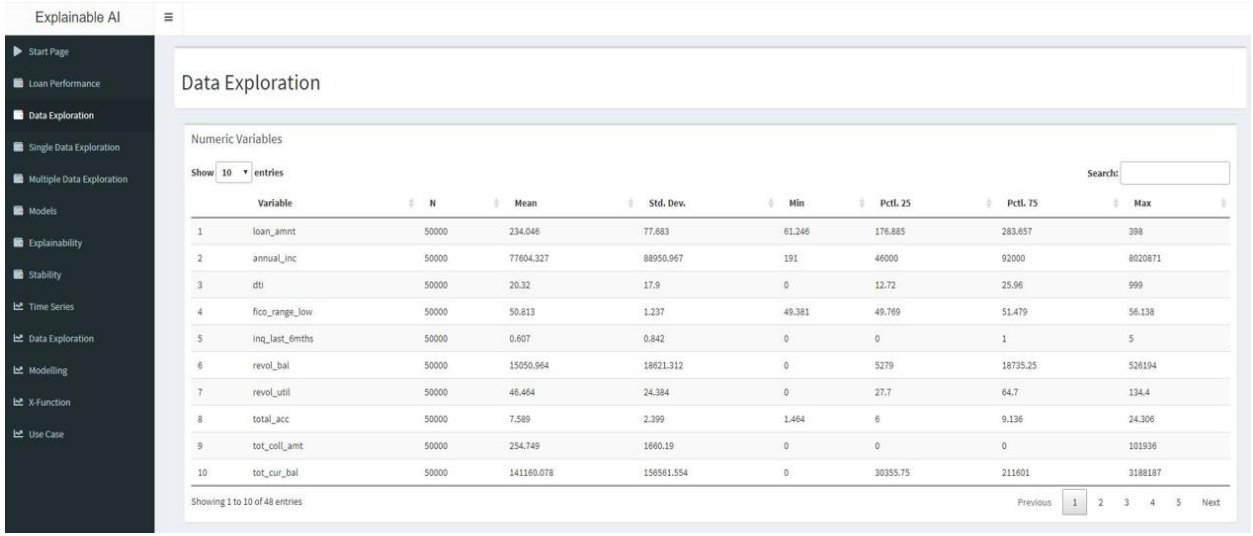
This repository contains the code to the explainable AI project funded by Innosuisse xxx project. In this project, we aim to build a VA framework which will enable users to get some insights into the innerworkings of ML models as applied to financial problem sets. Specifically, we consider two different use cases:

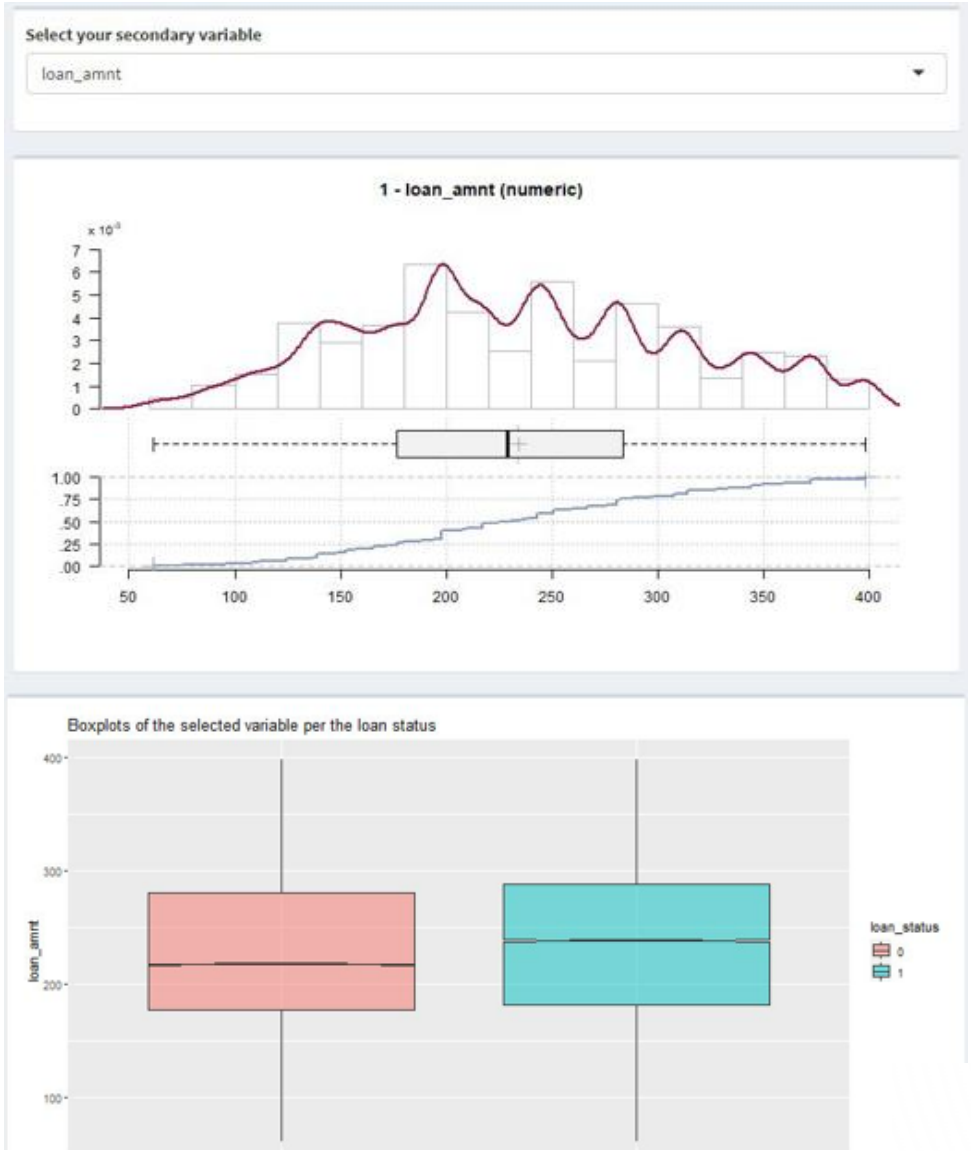
- credit riks modelling
- financial time series forecasting

XAI VA TOOL Developed

Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)



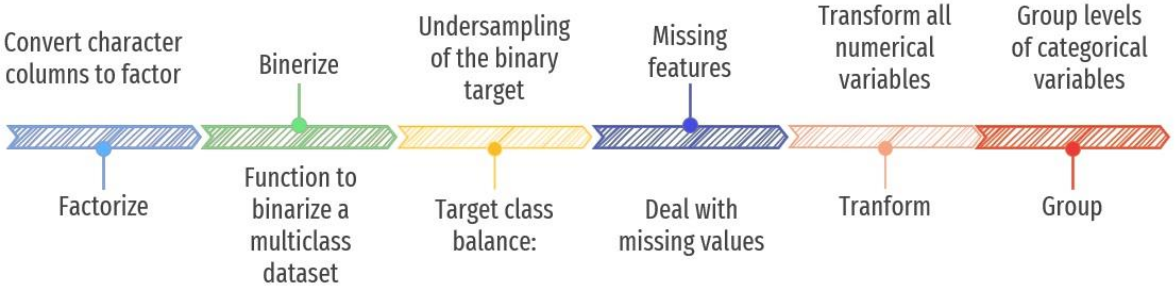


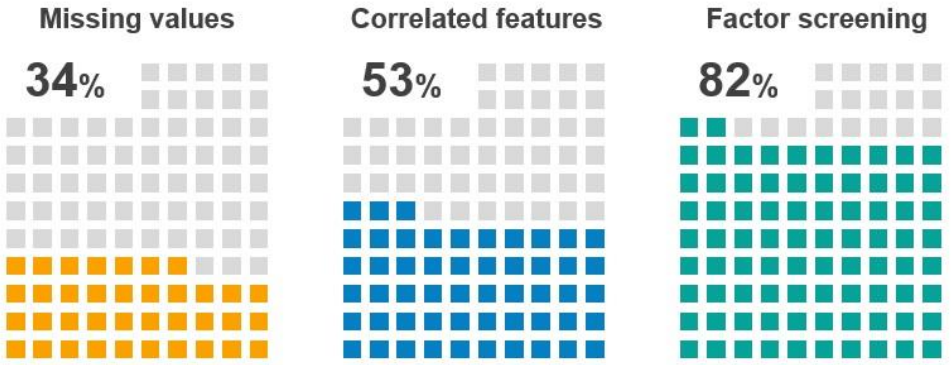
EXPLAINABLE AI FOR FINANCE

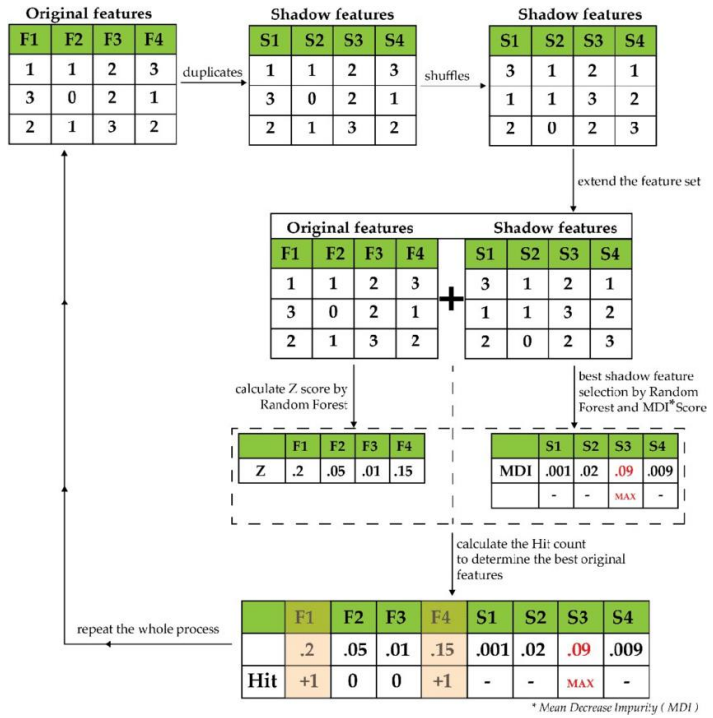
XAI VA TOOL Developed

Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**





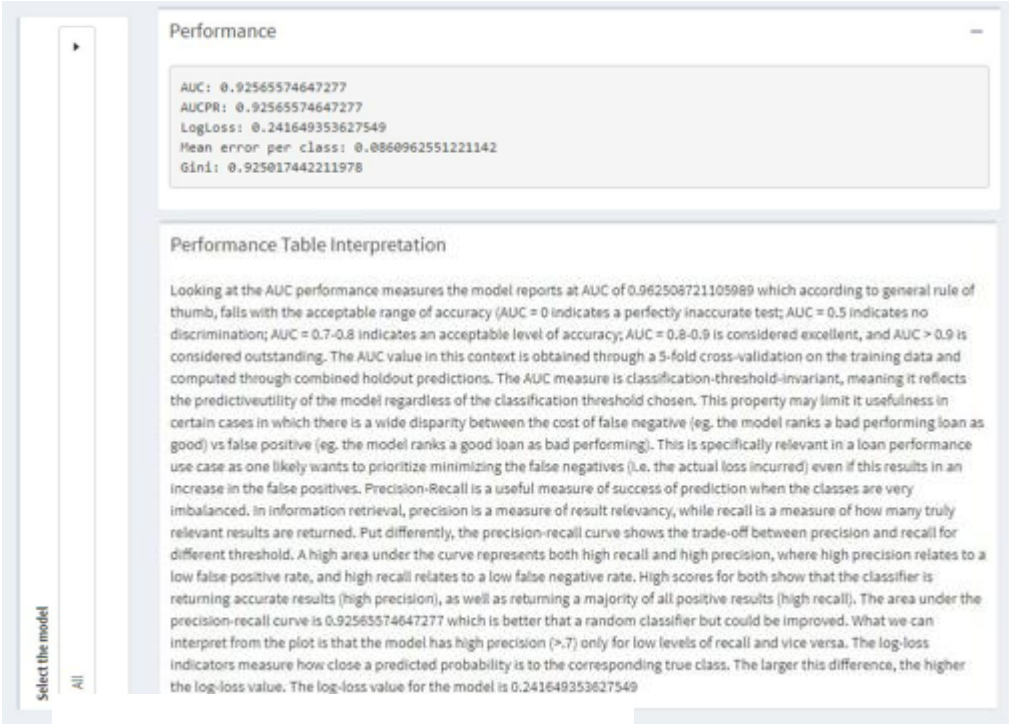


EXPLAINABLE AI FOR FINANCE

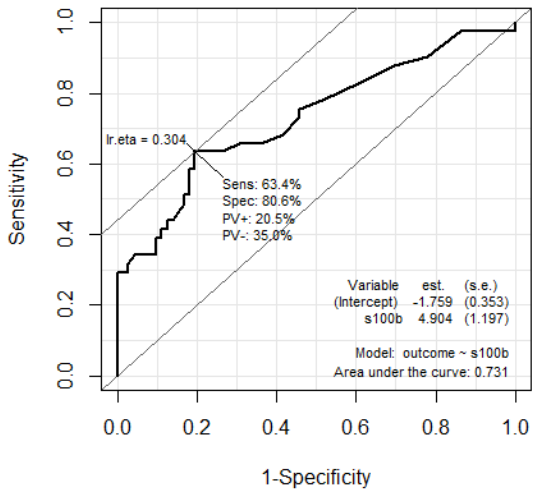
XAI VA TOOL Developed

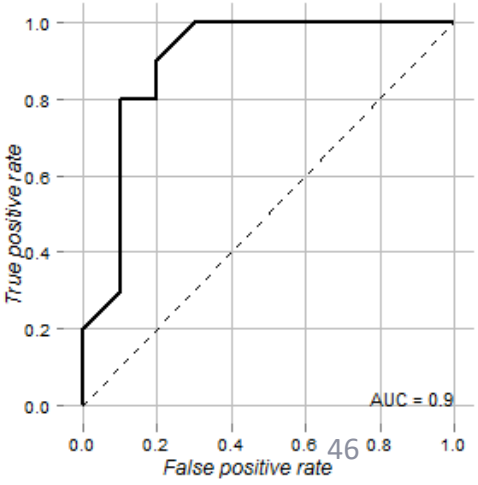
Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training



- **DRF** (This includes both the Distributed Random Forest (DRF) and Extremely Randomised Trees (XRT) models.)
- **GLM** (Generalised Linear Model with regularisation)
- **XGBoost** (XGBoost GBM)
- **GBM** (H2O GBM)
- **DeepLearning** (Fully-connected multi-layer artificial neural network)
- **StackedEnsemble** (Stacked Ensembles, includes an ensemble of all the base models and ensembles using subsets of the base models)





XAI VA TOOL Developed

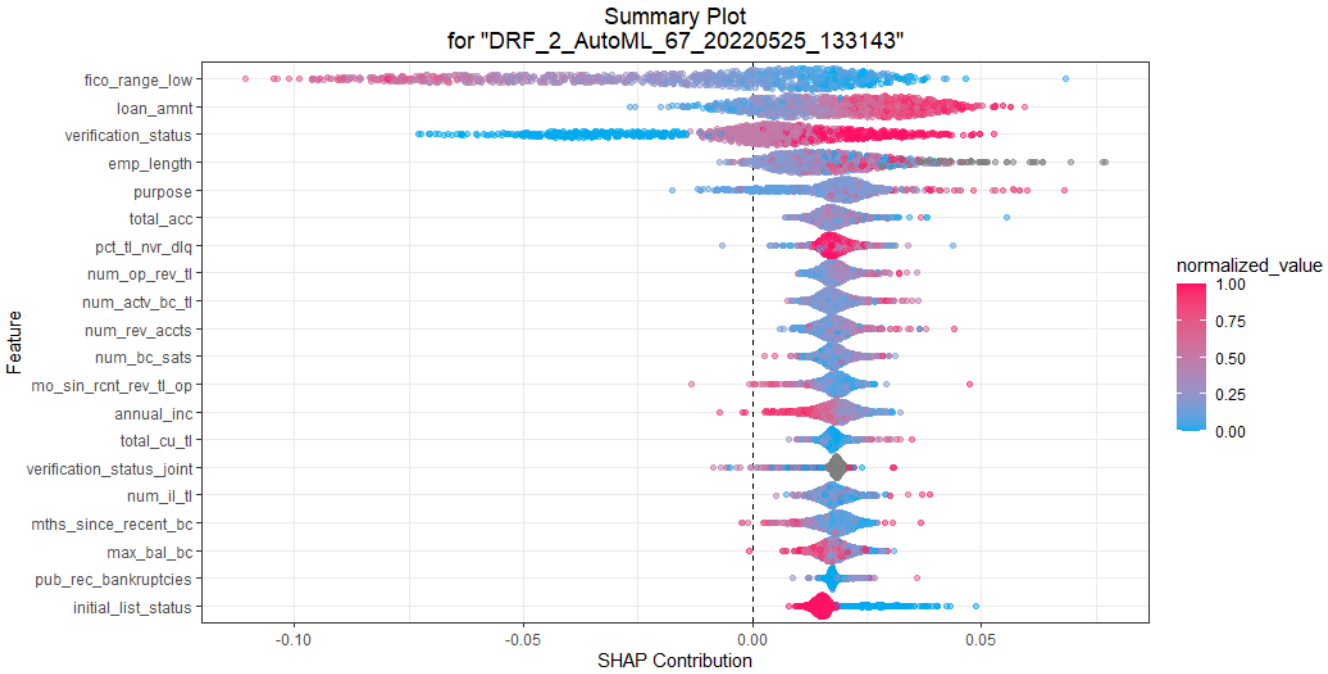
EXPLAINABLE AI FOR FINANCE

Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training
- **Explainability**

All accompanied by detailed **interpretations & scores** concerning the overall alignment with financial logic

Global Explanations



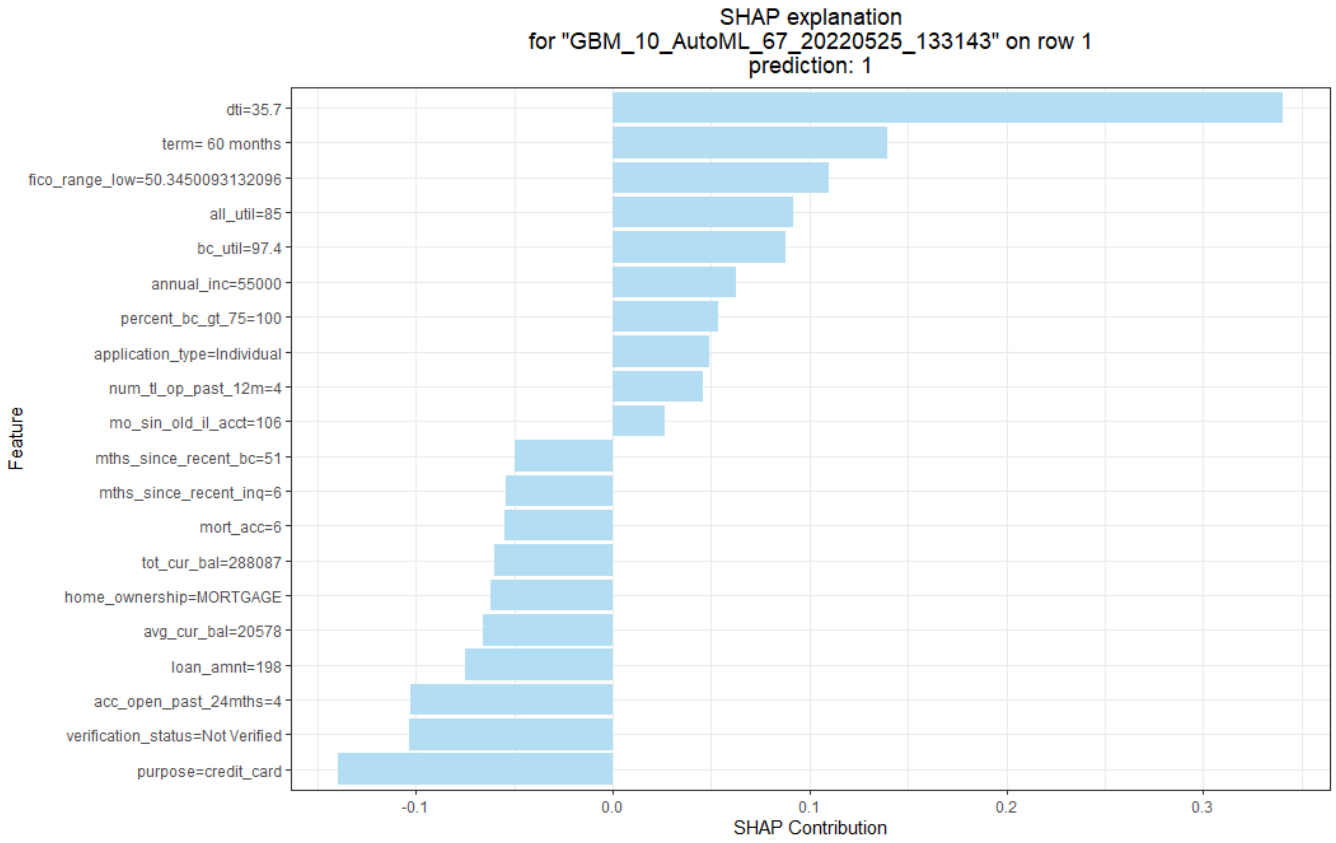
XAI VA TOOL Developed

EXPLAINABLE AI FOR FINANCE

Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training
- **Explainability**

Local Explanations



Users would be able to input information about themselves. We then find the loan contract included in the training sample that is closes in profile to that of the user and display the local explanations

XAI VA TOOL Developed

EXPLAINABLE AI FOR FINANCE

Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training
- **Explainability**
- **Stability** of predictions

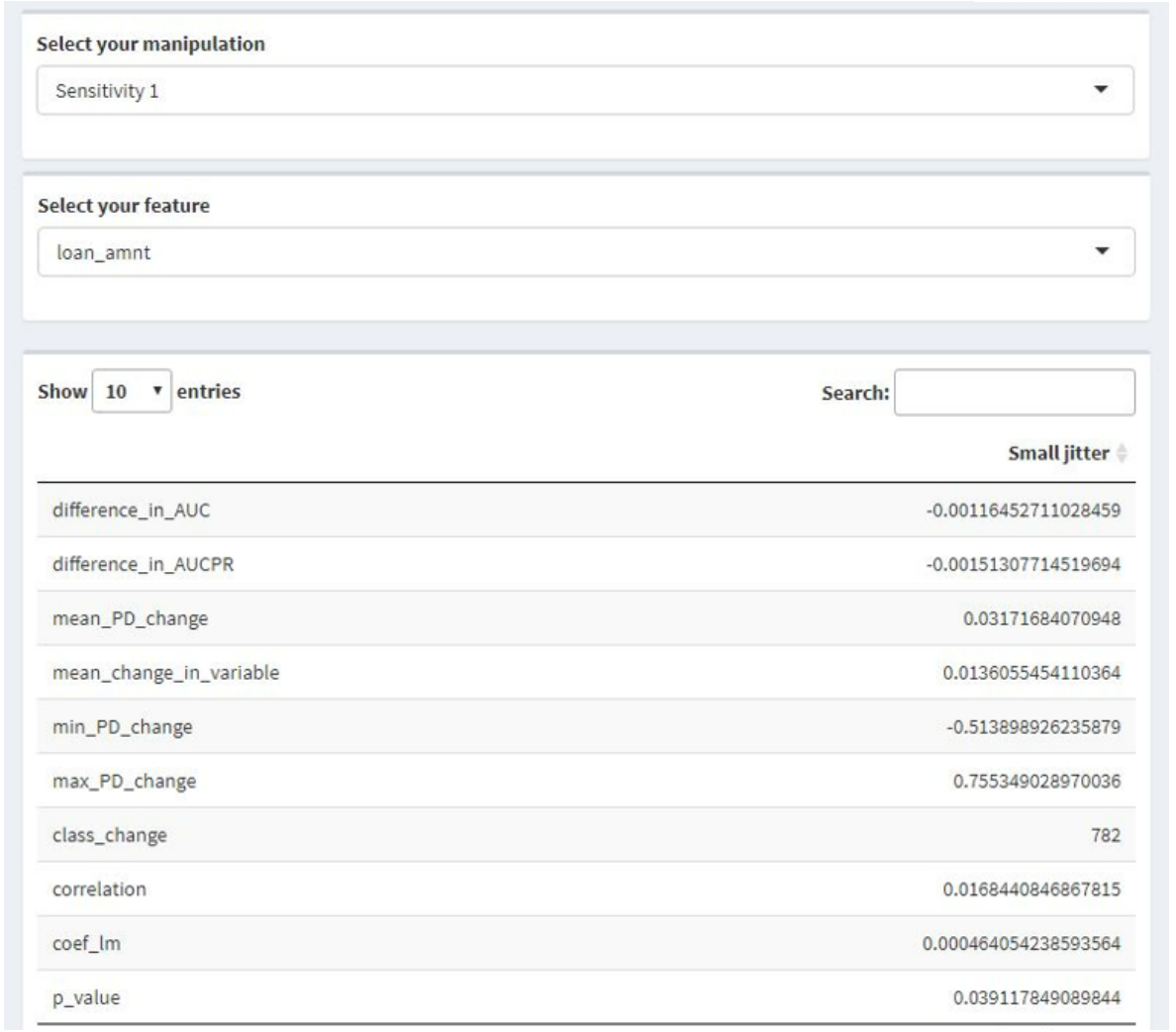
Select a perturbation approach	Re-run ML model & get sensitivities
Z = max(x) – min (x) X + runif(n, -a, +a) • a = z/50; • a = (1.1; 1.5; 2.5; 5) x z/50 • d = smallest difference between adjacent unique x values, a = d/5 • a = (1.1; 1.5; 2.5; 5) x d/5	• Change in AUC and AUCPR • The mean change in the variable • The mean, min and max change in PD • The number of class changes • The correlation coefficient • The regression coefficient • The scatter plot

XAI VA TOOL Developed

EXPLAINABLE AI FOR FINANCE

Explainability for ML-models applied to credit risk management

- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training
- **Explainability**
- **Stability** of predictions



Classic XAI methods are deemed useful only for **model developers**

Available toolkits & visualization **not suited for a non-technical audience**

No framework for testing stability of explanations provide

No XAI-technique for deep learning methods (DL) which preserves and exploits the natural time ordering of the data

Stability & Robustness of Explanations

- Papers have suggested that **post hoc explanation methods are unstable** (i.e., small perturbations to the input can substantially change the constructed explanations), as well as not robust to distribution shift (Ghorbani et al., 2019; Lakkaraju & Bastani, 2020).

- **Graph theory** applicability:
  - **Visualization**
  - **Modelling:** we can define a **special distance** between loan contracts using different distance measures (e.g. Euclidian, HEOM distance, etc.) and we can estimate the relationship between the spatial and **explanation distance**.

Stability & Robustness of Explanations: **Spatial Distance of Loans**

- We exploit information derived from the numerical features collected in a vector x_n representing the different loan contacts n.

- We define a metric **D - standardized Euclidean distance** between each pair $(x_j; y_j)$ loan feature vectors:

$$D_{x,y} = \sqrt{\sum \left(\frac{x_i}{s_i} - \frac{x_j}{s_j} \right)^2}$$

- **Visualization**: For a **Graph** G, the goal is to find a tree T which is a spanning subgraph of G, i.e. every node is included to at least one edge of T and has minimum total weight.
  - Pick some arbitrary start node u. Initialize $T = u$
  - At each step add the lowest-weight edge to T (the lowest-weight edge that has exactly one node in T and one node not in T);
  - Stop when T spans all the nodes.

Stability & Robustness of Explanations: **Explanation Distance of Loans**

- We further allow users of the VA tool to investigate the dependence between the special distance and the explanation difference between similar contracts.
- For this purpose, we define an **explanatory distance measure**:

$$expDis = \sqrt{\sum_{z=1}^{10}(SHAP_{n_{zx_i}} - SHAP_{n_{zx_j}})^2 + \sum_{z=1}^{n}(SHAP_{n_{1x_i}} + \ldots SHAP_{n_{zx_j}})^2}$$

where,

- n are the top 10 features provided by XAI method
- x_i and x_i are the different pairs of loan feature vectors
- SHAP are the specific SHAP contributions

Visualization: Stability of Explanations

- Users would be able to select **a sample of loans**
- For each pair of loan contracts $D_{x,y}$ is calculated and the MST is found. Then, for each of the nodes, the top explanatory feature is extracted.
- The visualization displays the MST of the sample of companies selected. Coloring based on the top explanatory feature [green = "Number of instalment accounts opened in past 12 months"; grey = "Months since most recent instalment accounts opened" ; blue = "Grade"]
- Users can get first glimpse as to whether similar loans have the same top explanatory feature

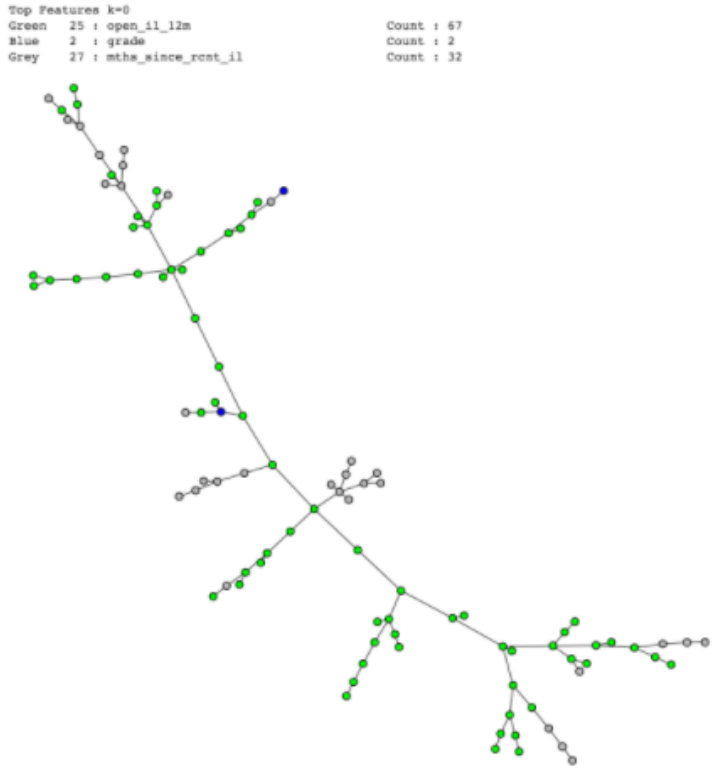


Figure 4. MST tree representation of 100 random data points.

Stability of Explanations though **GRAPH THEORY**

EXPLAINABLE AI FOR FINANCE

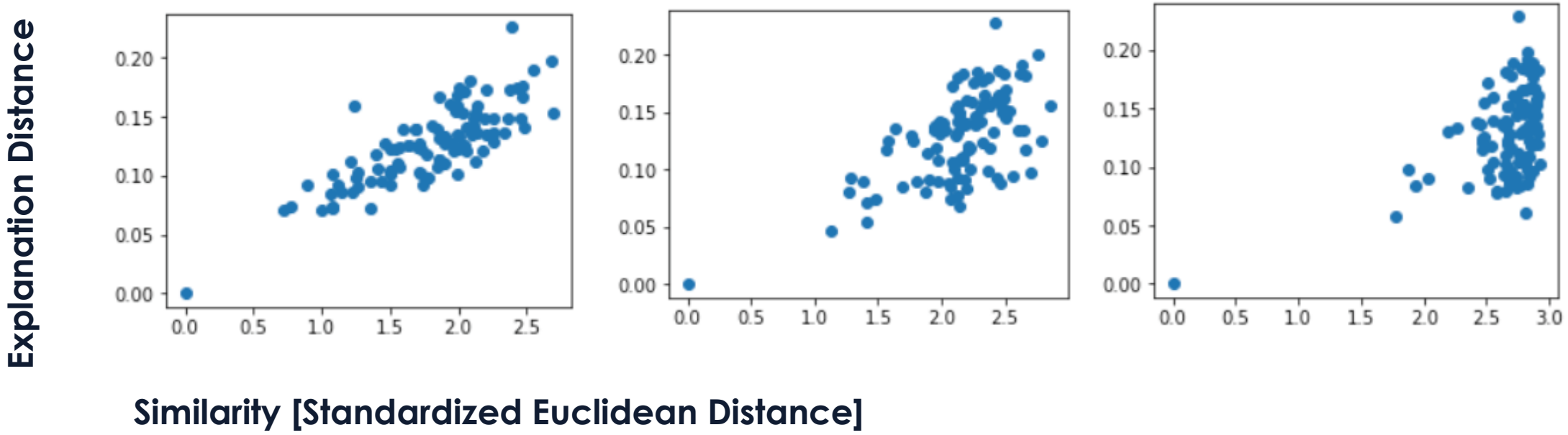


Figure 5. Explanation Distance vs Spatial Distance for ref_i = 1000, n = 100 for 5, 10, and 20 Features.

*Remember: the explanation difference formula takes the top n features of two points, adds up the squared difference of the contributions of each feature in common, and for each feature that is not common, adds up the square of each contribution then finally take the square root of the sum.

Classic XAI methods are deemed useful only for **model developers**

Available toolkits & visualization **not suited for a non-technical audience**

No framework for testing stability of explanations provide

No XAI-technique for deep learning methods (DL) which preserves and exploits the natural time ordering of the data

XAI VA TOOL Developed – Financial Time Series Use Case

- Classical approaches and their current implementation are not tailored for **financial time series** (subject to trends, vola-clusters,...).

- Specifically, **perturbation-based methods are fully dependent on the ability to perturb samples in a meaningful way**. In the context of financial data:

  - if features are correlated, the artificial coalitions created will lie outside of the multivariate joint distribution of the data,

  - if the data are independent, coalitions/generated data points can still be meaningless;

  - generating artificial data points through random replacement disregards the time sequence hence producing unrealistic values for the feature of interest.

Applicability for **Time Series Data**

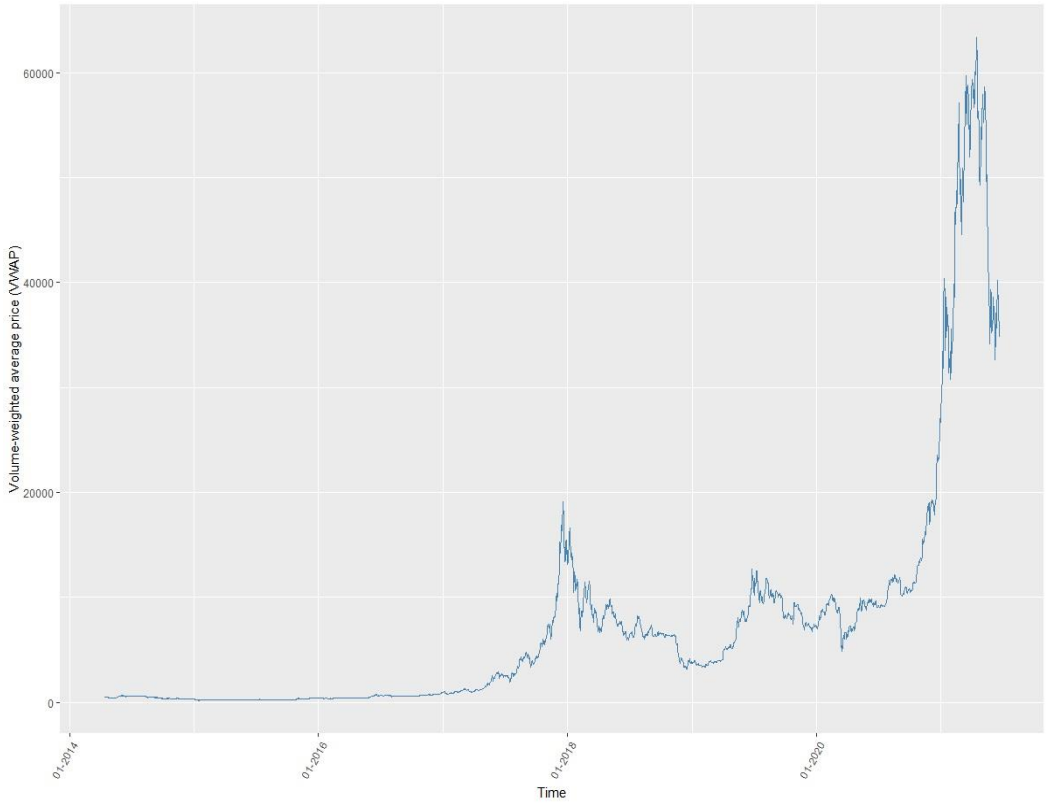


Figure 4. Bitcoin Prices (original time ordering)

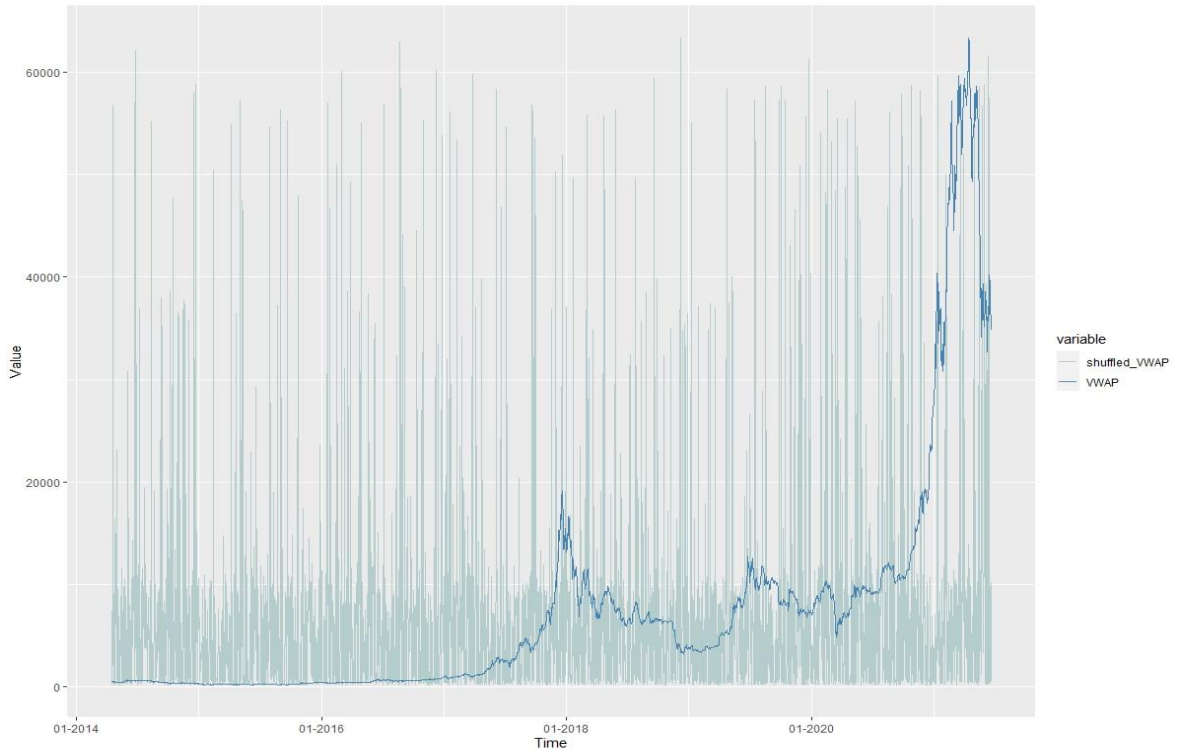


Figure 5. Bitcoin Prices (original time ordering and shuffled values)

Simple X-Function: Identity and LPD

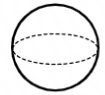
EXPLAINABLE AI FOR FINANCE

- We propose a family of Explainability (X-)functions xf(.) for assigning meaning to the net's response or output o_t over time t = 1, …, T, where $o_t = (o_{1t}, \ldots, o_{npt})$ is a n_p dimensional vector of possibly multiple output neurons.
- By selecting the identity $xf(o_t) = o_t$ we can mark preference for the sensitivities or partial derivatives $w_{ijt} = \frac{\partial o_{jt}}{\partial x_{it}}, i = 1, \ldots, n, j = 1, \ldots, n_p$, for each explanatory variable x_{it} of the net.
- In order to complete the 'explanation' derived from the identity one can add a synthetic intercept to each output neuron o_{jt} defined according to:

$$b_{jt} := o_{jt} - \sum_{i=1}^{n} w_{ijt} x_{it}$$

- For each output neuron o_{jt}, the resulting derivatives or 'explanations' $b_t, w_{1t}, w_{2t} \ldots w_{pt}$ generate a new data-flow which is referred to as **Linear Parameter Data (LPDs)**
- The LPD is a matrix of dimension T $*$ (n + 1), irrespective of the complexity of the neural net, with t−th row denoted by $LPD_t := (b_t, w_{1t}, w_{2t} \ldots w_{pt})$
- The LPD can be interpreted in terms of **exact replication of the net by a linear model at each time point t** and the natural time-ordering of LPDs subsequently allows to examine changes of the linear replication as a function of time.

Simple X-Function: Identity and LPD



EXPLAINABLE
AI FOR FINANCE

In order to give an intuition as to the sensitivities/explanations we want to obtain let's imagine the following brute approach:

1. We start by training a neural network (NN) on the specified inputs and response and store the results
2. Next, we perturb a selected input slightly
3. We use the trained NN and make the predictions for the changed inputs
4. For each changed variable, we collect the perturbed data and the corresponding NN-output
5. We fit a linear model and obtain the weights
6. We train the net for 100 different random initialization and **we observe the dependency of the LPDs across the different random nets**.

Simple X-Function: Identity and LPD

- **Data:** Bitcoin returns 15-04-2014 to 30-06-2021

- **Model:** simple feedforward net with a single hidden layer and an input layer collecting the last six lagged (daily) returns: the net is then trained to predict next day's return based on the MSE-criterion. The number of estimated parameters then amounts to a total of $6 * 100 + 100 = 700$ weights and $100 + 1 = 101$ biases

- We optimize the net 100-times, based on different random initializations of its parameters, and we compute **trading performances of each random-net based on the simple sign-rule**

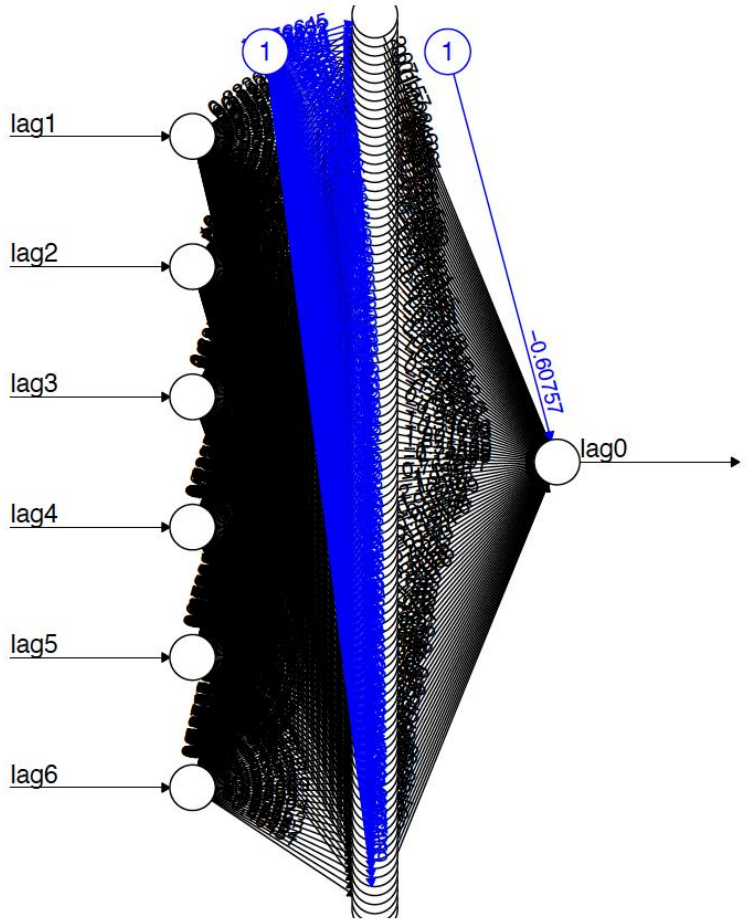


Figure 6. Neural net BTC: feedforward net with a single hidden-layer of dimension 100 and an input layer of dimension 6 comprising the last six lagged returns

Simple X-Function: Identity and LPD

EXPLAINABLE AI FOR FINANCE

- **Data:** Bitcoin returns 15-04-2014 to 30-06-2021

- **Model:** simple feedforward net with a single hidden layer and an input layer collecting the last six lagged (daily) returns: the net is then trained to predict next day's return based on the MSE-criterion. The number of estimated parameters then amounts to a total of $6 * 100 + 100 = 700$ weights and $100 + 1 = 101$ biases

- We optimize the net 100-times, based on different random initializations of its parameters, and we compute **trading performances of each random-net based on the simple sign-rule**

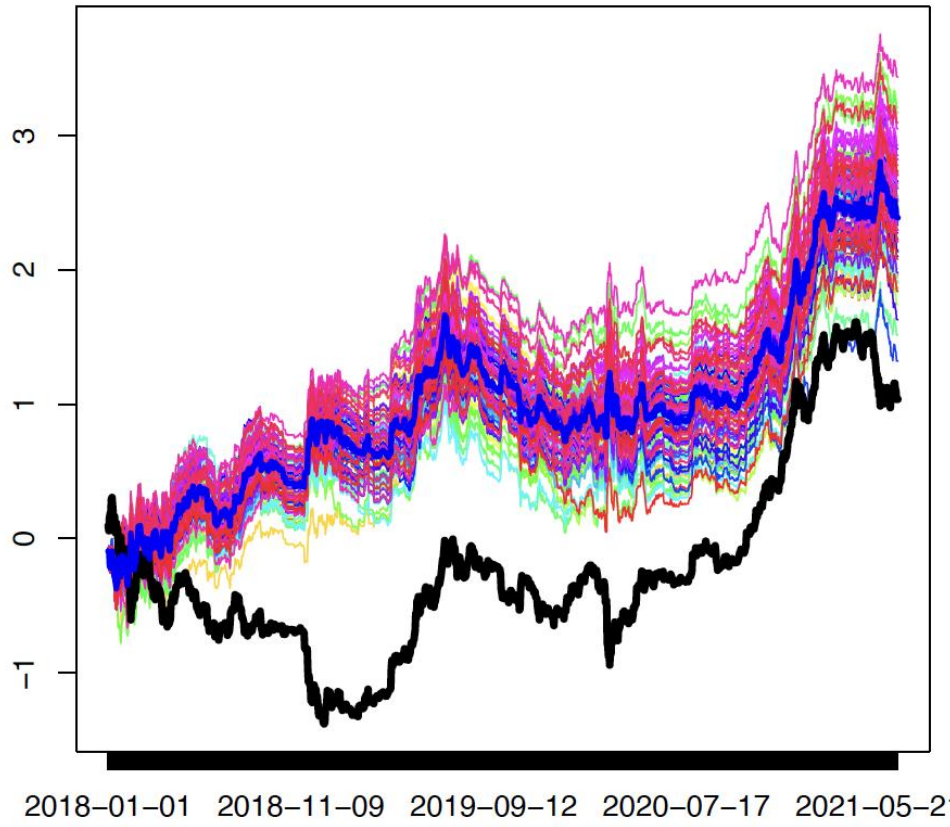


Figure 7. Cumulated log-performances out-of-sample based on sign-rule (buy or sell depending on sign of forecasted return): 'random' neural nets (colored) vs. buy-and-hold (bold black) and mean-net performance (bold blue)

LPDs & **RISK MANAGEMENT**

EXPLAINABLE AI FOR FINANCE

- Our initial insights suggest that **the time-varying dependency of the data measured by the LPDs is indicative of different states of the market.**

- In particular **weak dependency** (small absolute LPD) **is an indicator of randomness or 'chaos'.**

- We therefore propose a simple rule for managing risks: **exit markets at times tagged as chaotic by the LPD.**

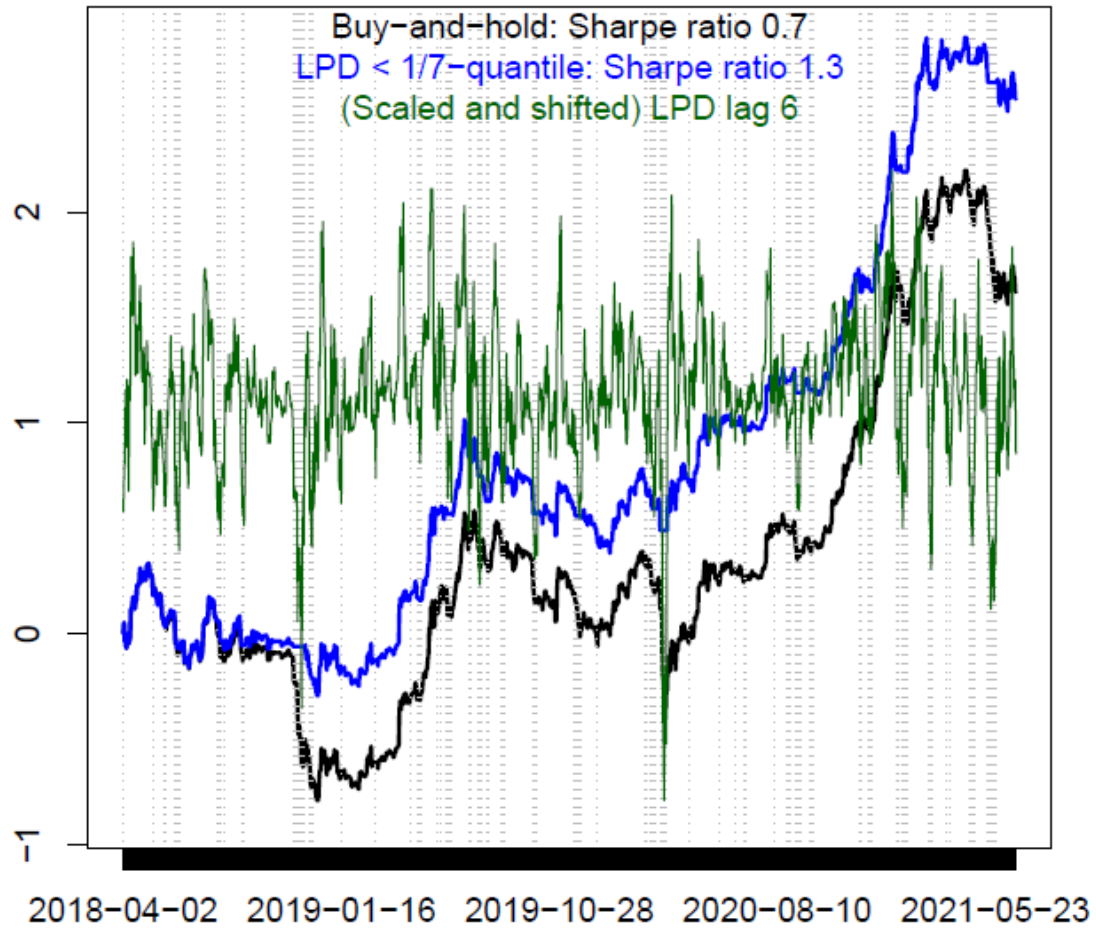


Figure 8. Buy-and-hold (black) vs. out-of-sample (mean-) LPD market-exit strategy (blue): exits (shaded in grey) occur if today's out-of-sample mean-LPD (green) drops below the 1/7-quantile based on a rolling-window of length one quarter of its own history. The LPD corresponding to the lag-6 BTC-value is used.

Concluding Remarks

- Check out our app & code https://www.explainableaiforfinance.com/
- Check out our papers https://www.explainableaiforfinance.com/repository-of-papers

EXPLAINABLE AI FOR FINANCE

Code & App

All the material produced under this project is open source. The data used as well as the scripts, functions and models developed are available in a GitHub repository whereas the visual analytics tool (i.e. the app) is created though R Shiny and it is made available for all model developers and model evaluators to check and understand the inner-workings of ML models applied to financial problem sets.