

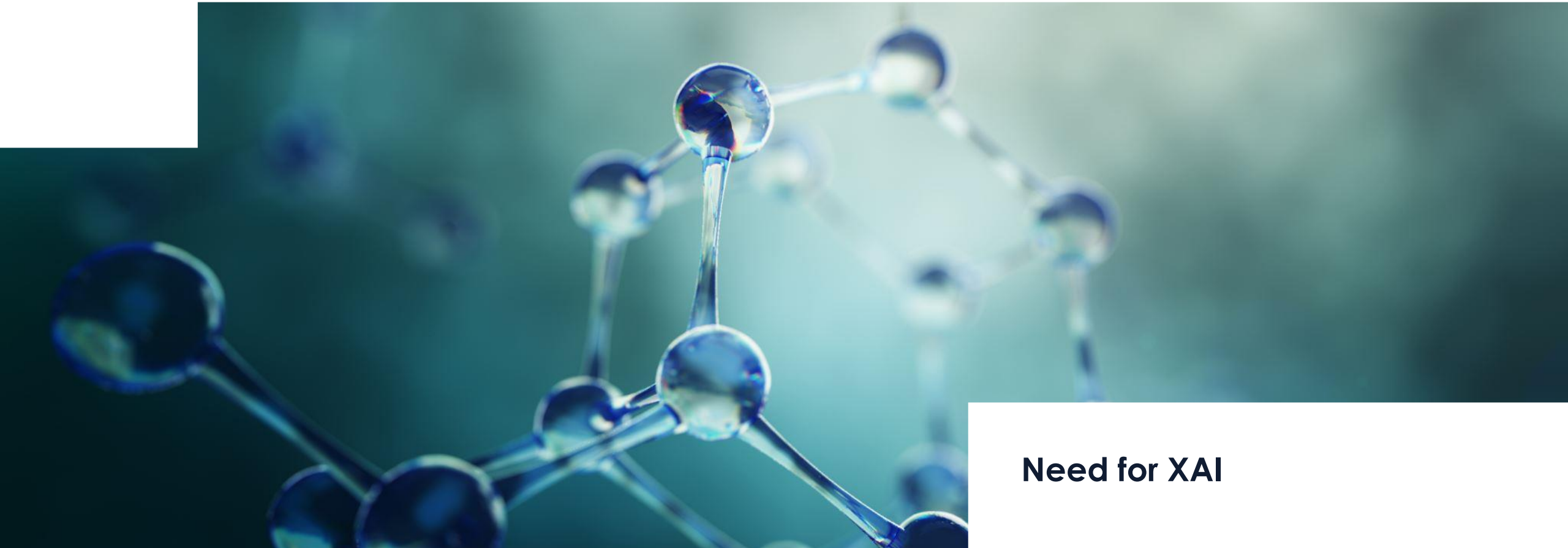
FinanceCom 2022 – Enterprise Applications, Markets & Services in the Finance Industry

University of Twente, 22-24 August 2022
The Netherlands

eXplainable AI for Finance

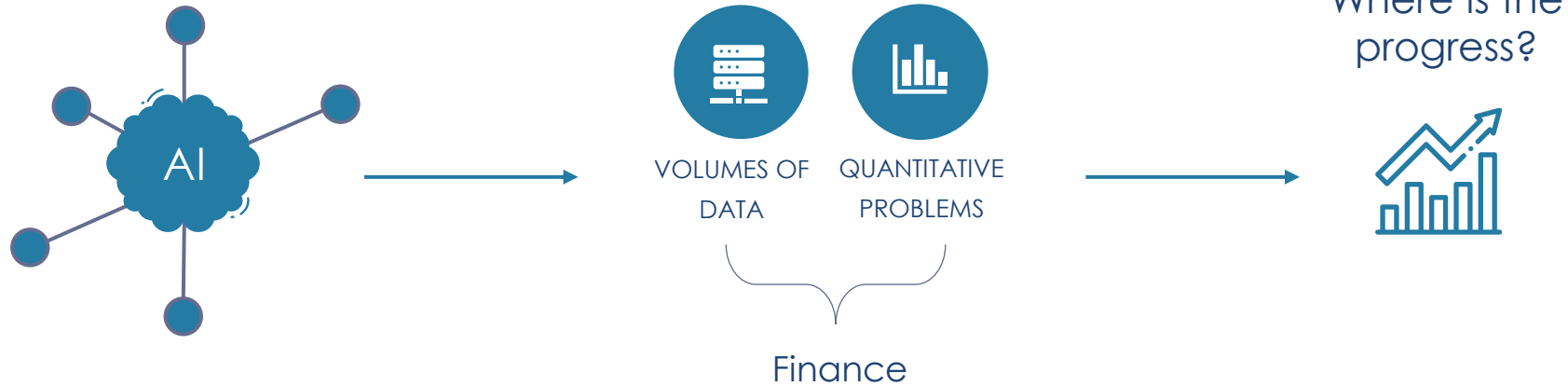
Prof. Dr. Branka Hadji Misheva
Bern University of Applied Sciences, Switzerland





Need for XAI

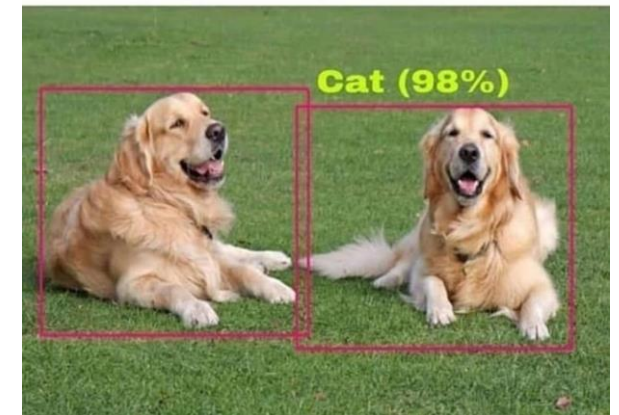
eXplainable AI for Finance



Well ...

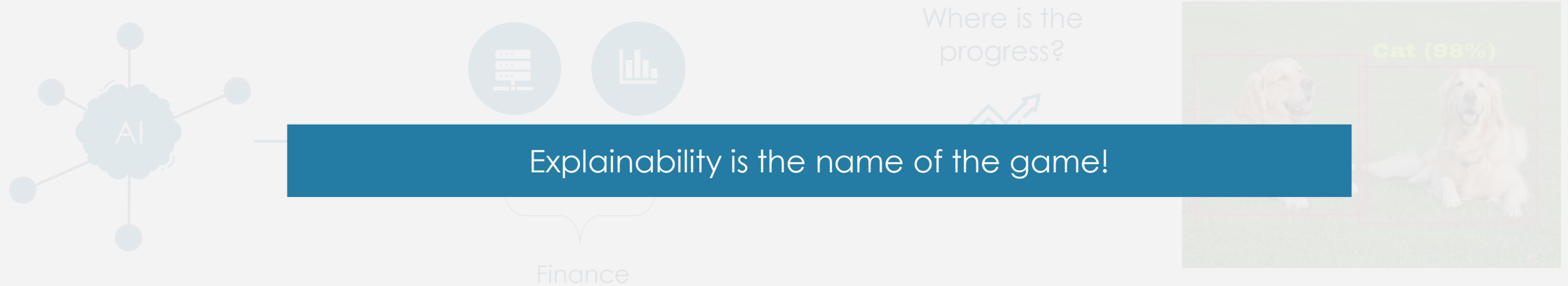
People: *fearing* AI takeover

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)



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)



Deploying eXplainability

POST-HOC Explainability

- There are many models that are highly interpretable.
- Regardless, for some ML models, **post-hoc explainability 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-specific vs. model-agnostic)
 - **the unit being explained** (if the method provides an explanation which is instance-specific 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).

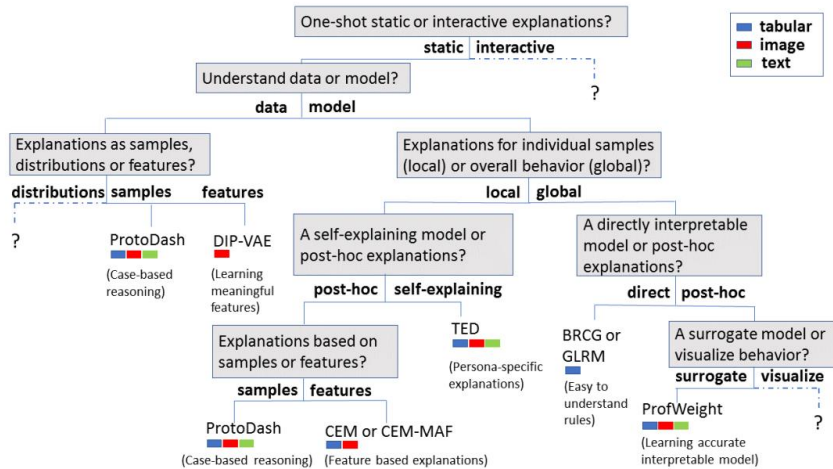


Figure 1. Arya et al. (2019) proposed taxonomy based on questions about what is explained, how it is explained and at what level

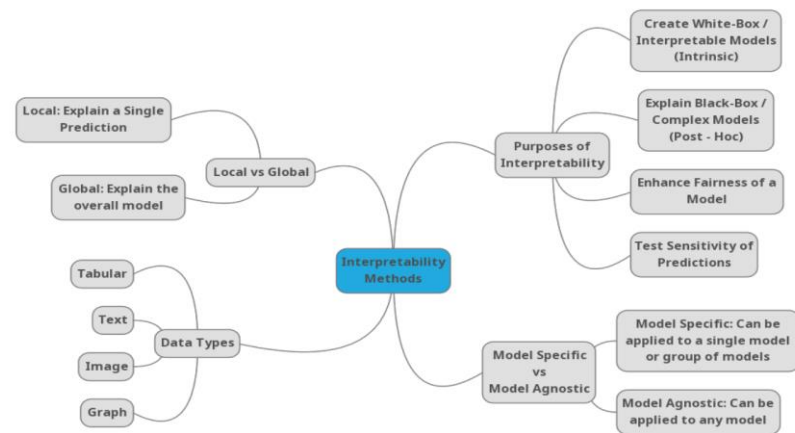


Figure 2. Linardatos et al. (2021) taxonomy mind-map of Machine Learning Interpretability Techniques.

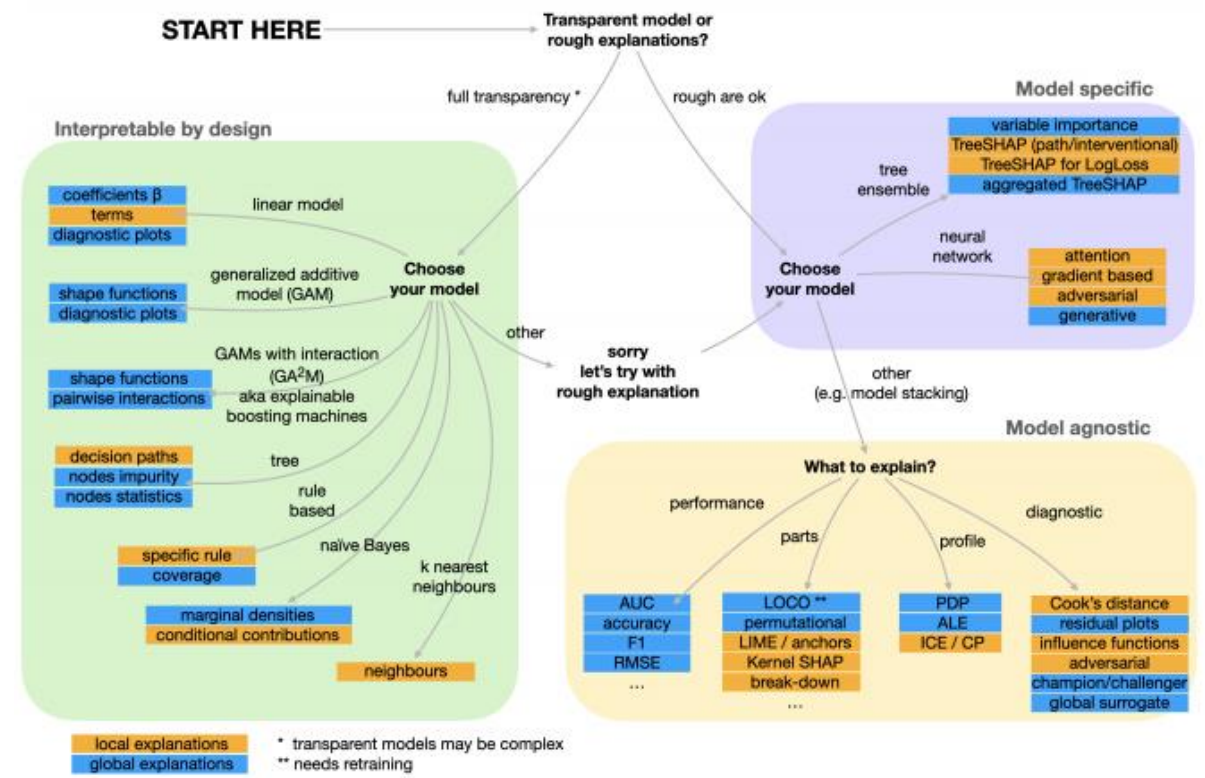


Figure 3. Maksymiuk et al. (2021) model-oriented taxonomy for XAI method

LIME & SHAPLEY: Details

LIME

- The explanation provided by LIME for each observation:

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

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 players 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 :

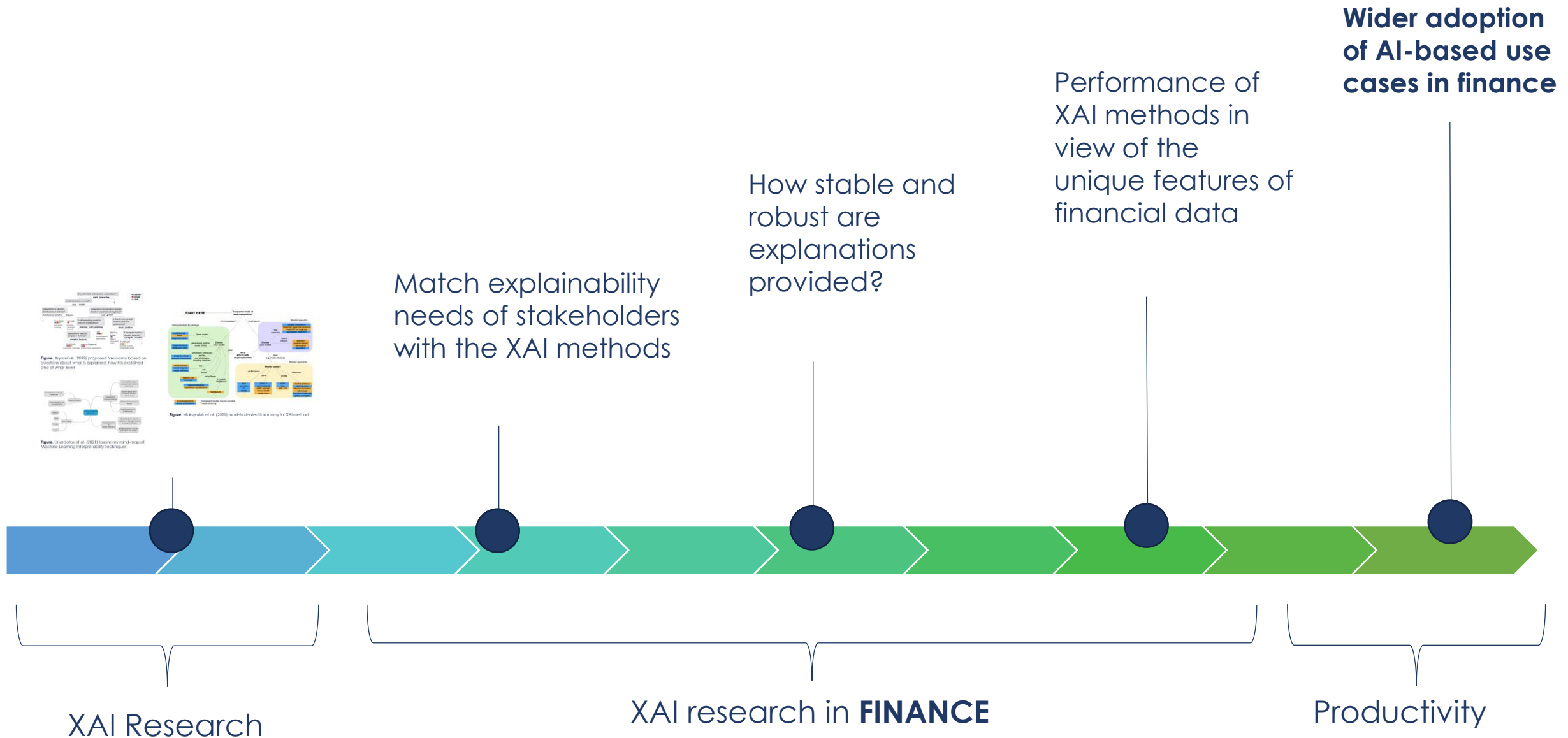
$$\Delta_v(i, S) = v(S \cup i) - v(S)$$

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

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



XAI for Finance



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.**



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

Branka Hadji Misheva^{1*}, David Jaggi², Jan-Alexander Posth², Thomas Gramespacher² and Joerg Osterrieder¹

¹ZHAW, School of Engineering, Institute of Data Analysis and Process Design, Winterthur, Switzerland, ²ZHAW, 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

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 organizations 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 (Gijbbers et al., 2019).

OPEN ACCESS

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

Reviewed by:
Bowei Chen,
University of Glasgow,
United Kingdom
Evadita Valciukynas,
Kaunas University of Technology,
Lithuania

***Correspondence:**
Branka Hadji Misheva
hadji@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

Explainability needs

Model developers

Model debugging

- Identify why the model is performing poorly
- Understand **time-dependent sensitivities**

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?

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?

Identified issues with classical XAI

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 provided

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**

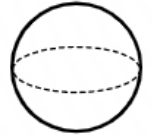
No framework for testing stability of explanations provide

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

eXplainable AI for Finance

“Trust opens up new and unimagined possibilities.”

Robert C. Solomon



EXPLAINABLE
AI FOR FINANCE

[Link to page](#)



| | | |
|--------------------------------|-------------------------------|-----------------------------------|
| 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 |

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 risks 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

Data Exploration

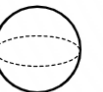
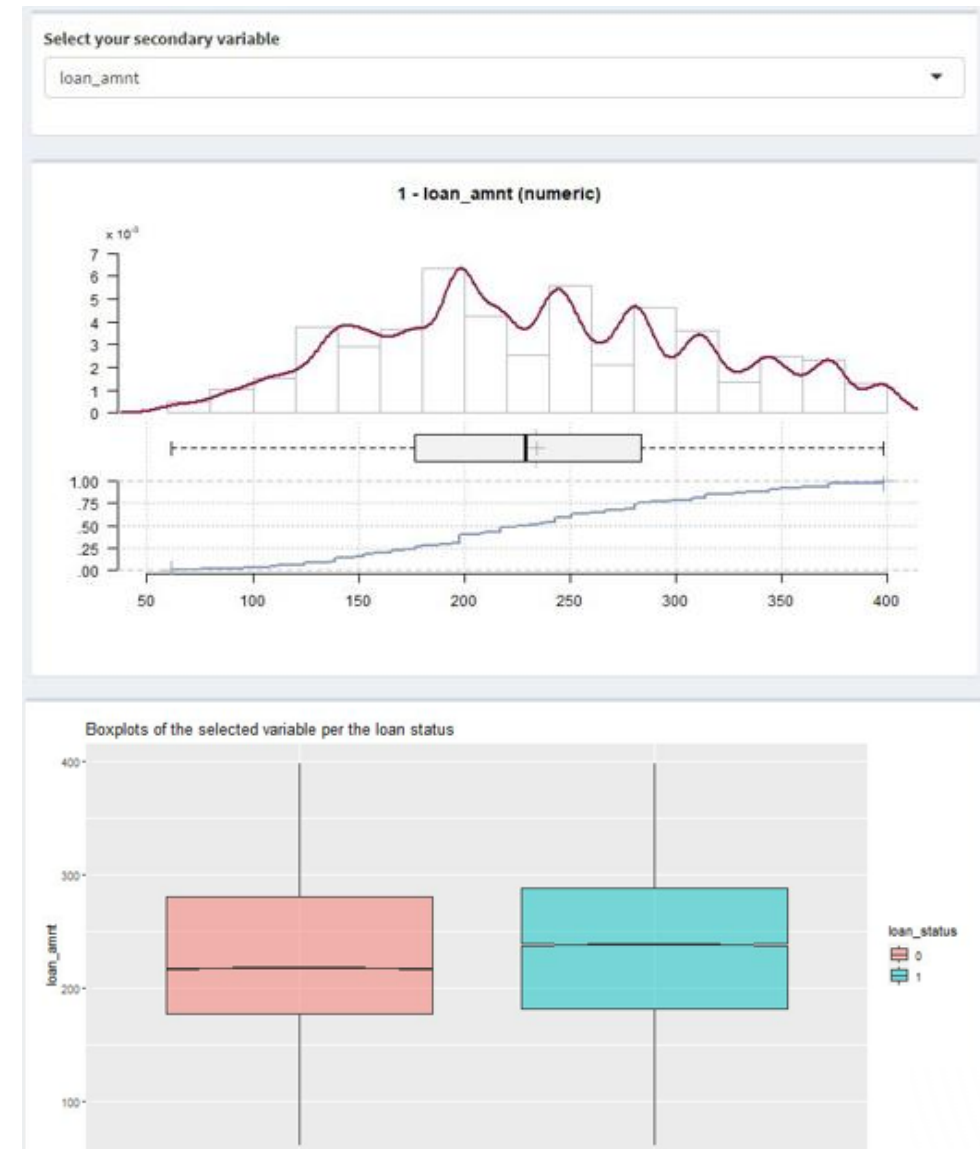
Numeric Variables

Show 10 entries

| | Variable | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 75 | Max |
|----|----------------|-------|------------|------------|--------|----------|----------|---------|
| 1 | loan_amnt | 50000 | 234.046 | 77.683 | 61.246 | 176.885 | 283.657 | 398 |
| 2 | annual_inc | 50000 | 77604.327 | 88950.967 | 191 | 46000 | 92000 | 802081 |
| 3 | dti | 50000 | 20.32 | 17.9 | 0 | 12.72 | 25.96 | 999 |
| 4 | fico_range_low | 50000 | 50.813 | 1.237 | 49.381 | 49.769 | 51.479 | 56.138 |
| 5 | inq_last_6mths | 50000 | 0.607 | 0.842 | 0 | 0 | 1 | 5 |
| 6 | revol_bal | 50000 | 15050.964 | 18621.312 | 0 | 5279 | 18735.25 | 526194 |
| 7 | revol_util | 50000 | 46.464 | 24.384 | 0 | 27.7 | 64.7 | 134.4 |
| 8 | total_acc | 50000 | 7.589 | 2.399 | 1.464 | 6 | 9.136 | 24.306 |
| 9 | tot_coll_amnt | 50000 | 254.749 | 1660.19 | 0 | 0 | 0 | 101936 |
| 10 | tot_cur_bal | 50000 | 141160.078 | 156561.554 | 0 | 30355.75 | 211601 | 3188187 |

Showing 1 to 10 of 48 entries

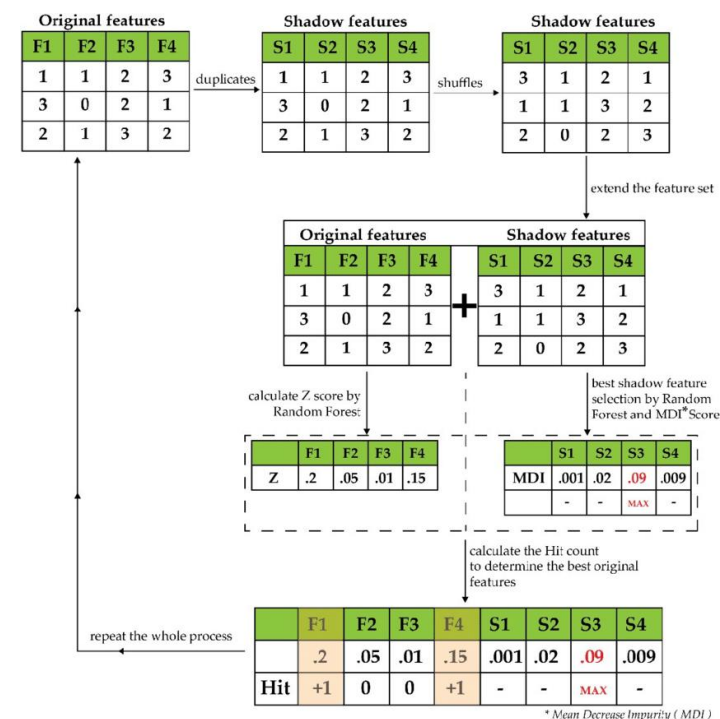
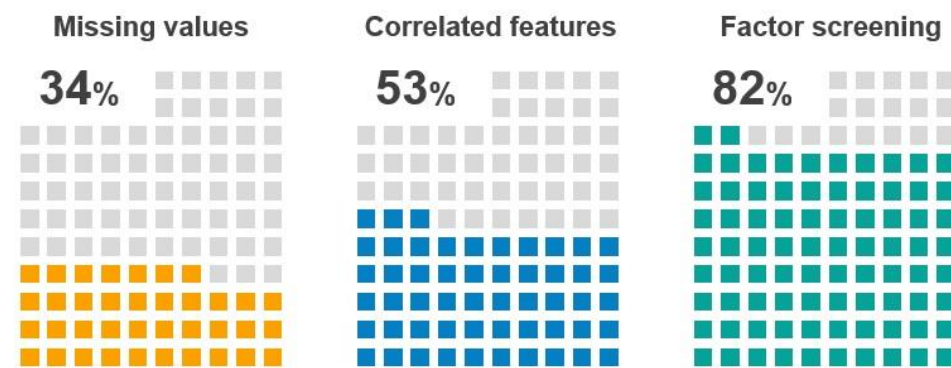
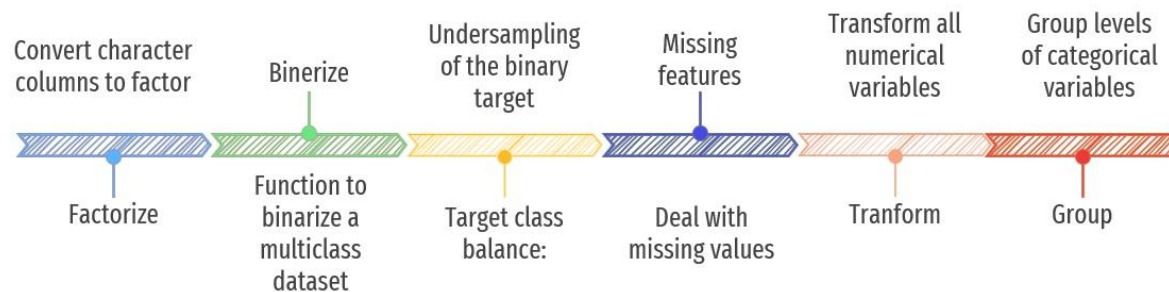
Previous 1 2 3 4 5 Next



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**



* Mean Decrease Impurity (MDI)

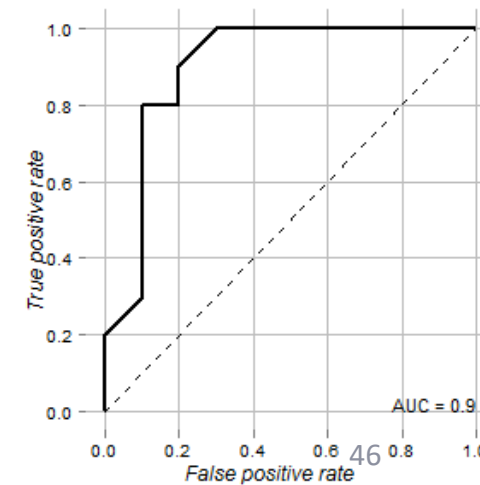
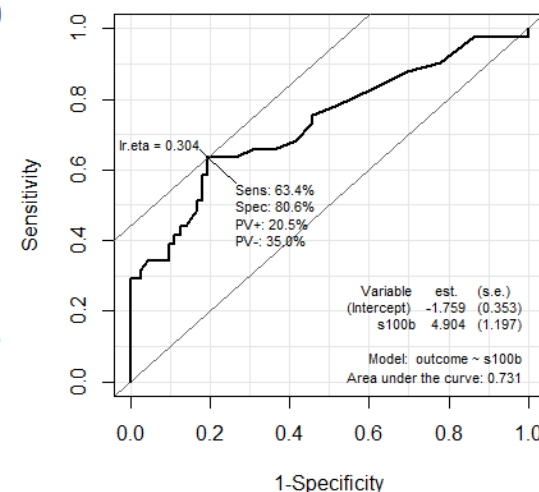
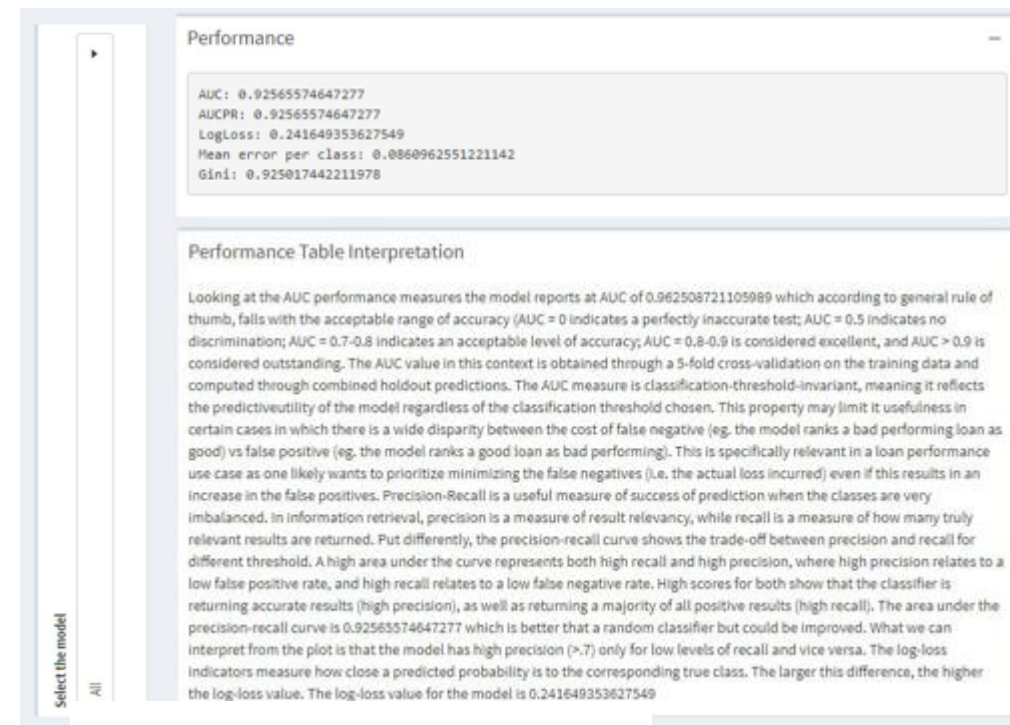


EXPLAINABLE
AI FOR FINANCE

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

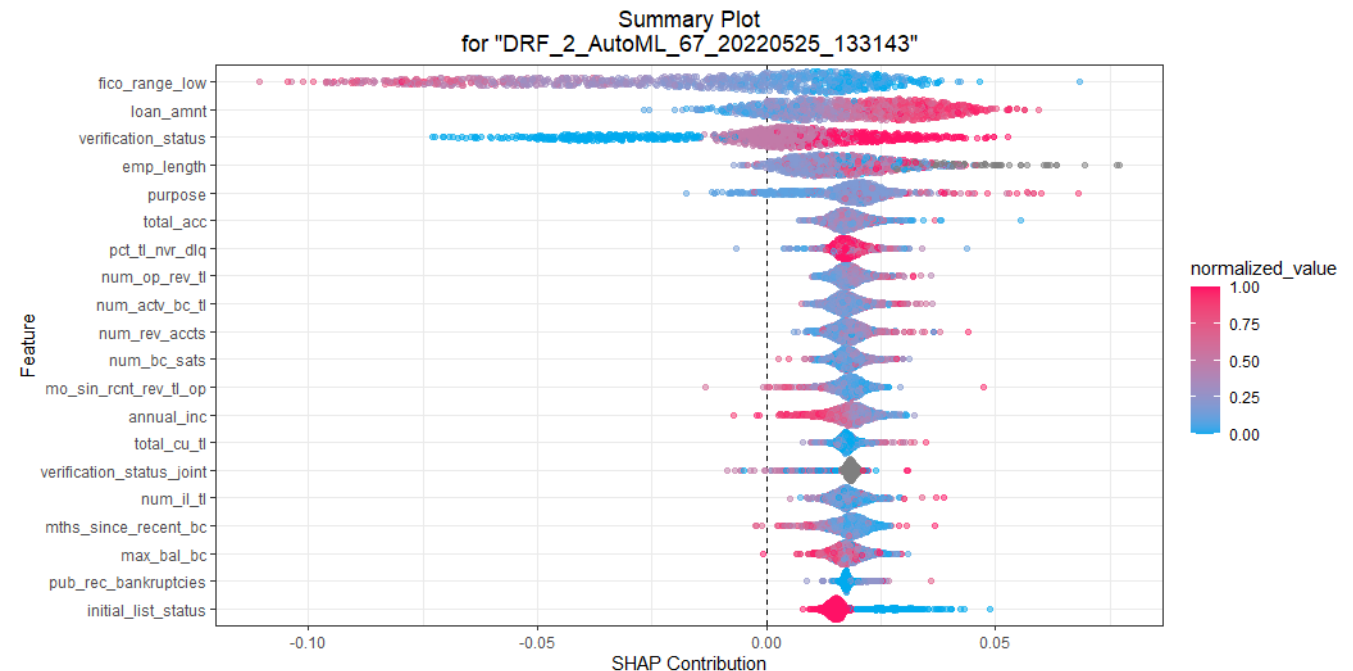
Explainability for ML-models applied to credit risk management

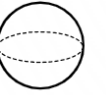
- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training
- **Explainability**



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

Global Explanations





XAI VA TOOL Developed

Explainability for ML-models applied to credit risk management

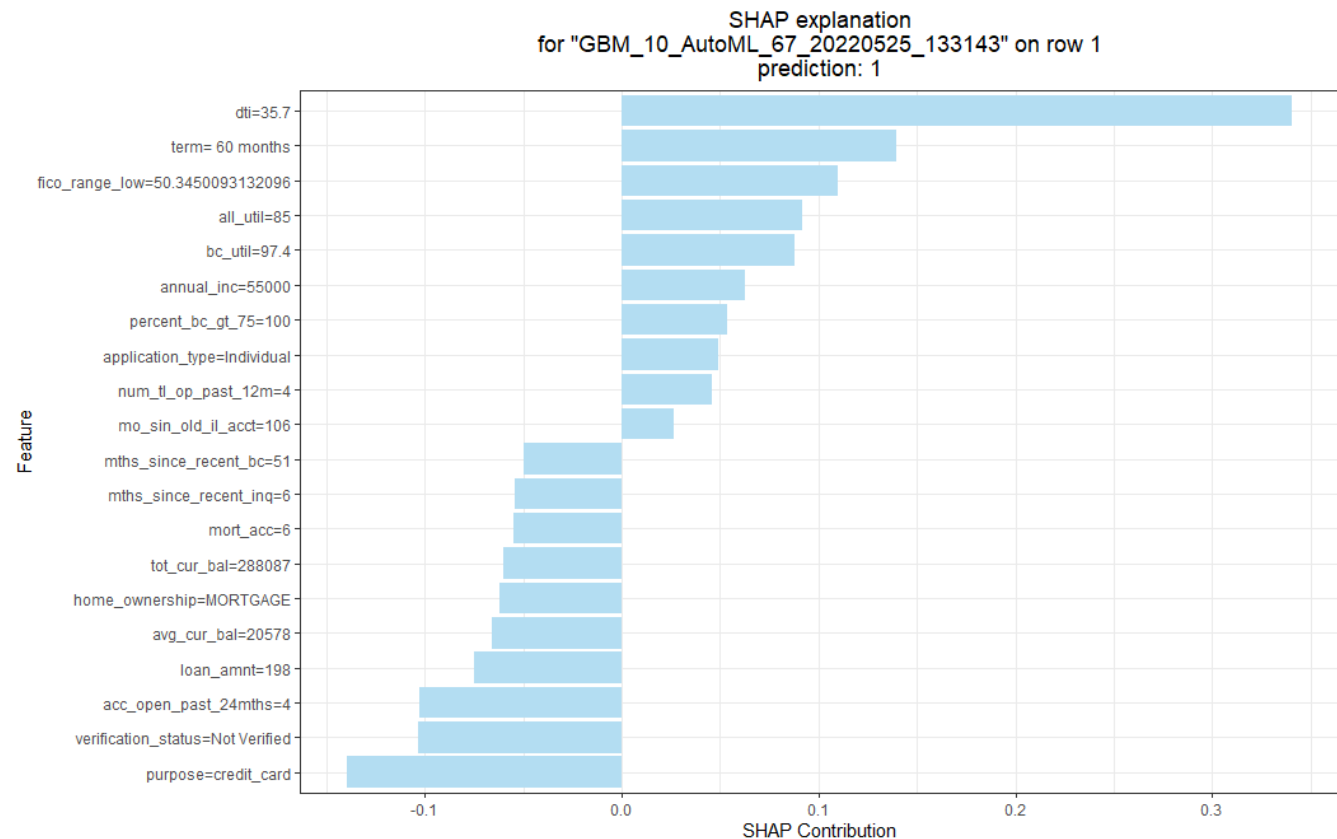
- **Toy data** (open source loan performance data –2GB of data containing information [160 features] on 2.2 million loan contracts)
- **Data pre-processing, feature selection & engineering**
- **ML model** testing & training
- **Explainability**



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

Local Explanations



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
- **Explainability**
- **Stability** of predictions



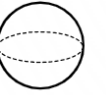
Select a perturbation approach

$Z = \max(x) - \min(x)$
 $X + \text{runif}(n, -a, +a)$

- $a = z/50$;
- $a = (1.1; 1.5; 2.5; 5) \times z/50$
- $d = \text{smallest difference between adjacent unique } x \text{ values, } a = d/5$
- $a = (1.1; 1.5; 2.5; 5) \times d/5$

Re-run ML model & get sensitivities

- 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

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 your manipulation | |
| Sensitivity 1 | |
| Select your feature | |
| loan_amnt | |
| Show 10 entries | Search: |
| Small jitter | |
| difference_in_AUC | -0.00116452711028459 |
| difference_in_AUCPR | -0.00151307714519694 |
| mean_PD_change | 0.03171684070948 |
| mean_change_in_variable | 0.0136055454110364 |
| min_PD_change | -0.513898926235879 |
| max_PD_change | 0.755349028970036 |
| class_change | 782 |
| correlation | 0.0168440846867815 |
| coef_lm | 0.000464054238593564 |
| p_value | 0.039117849089844 |

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:

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

- **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.

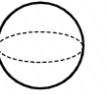
Stability & Robustness of Explanations: **Explanation Distance of Loans**

- 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**:

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

where,

- n are the top 10 features provided by XAI method
- x_i and x_j 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

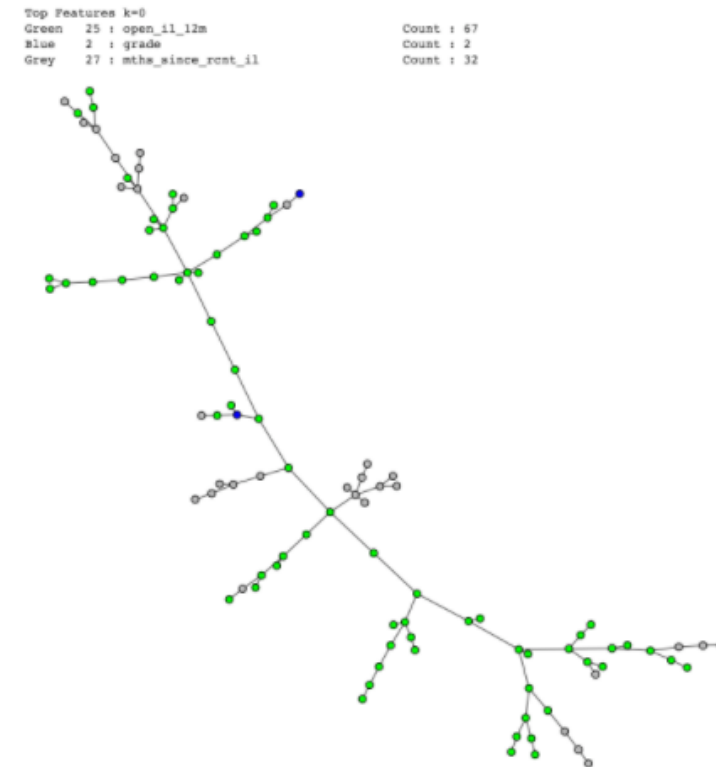
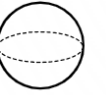
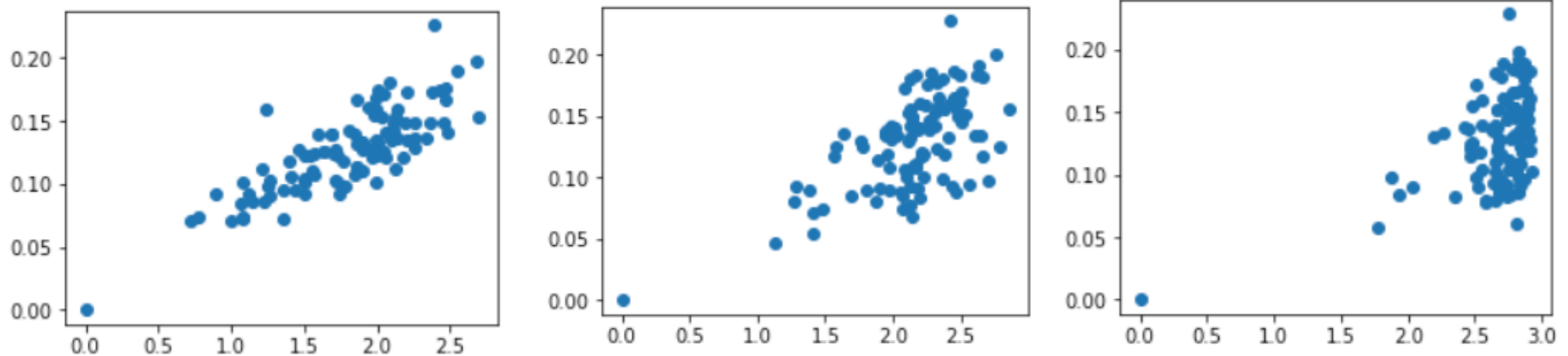


Figure 4. MST tree representation of 100 random data points.



Stability of Explanations though **GRAPH THEORY**

Explanation Distance



Similarity [Standardized Euclidean Distance]

Figure 5. Explanation Distance vs Spatial Distance for $\text{ref}_i = 1000$, $n = 100$ 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**

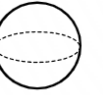
Available toolkits & visualization **not suited for a non-technical audience**

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**

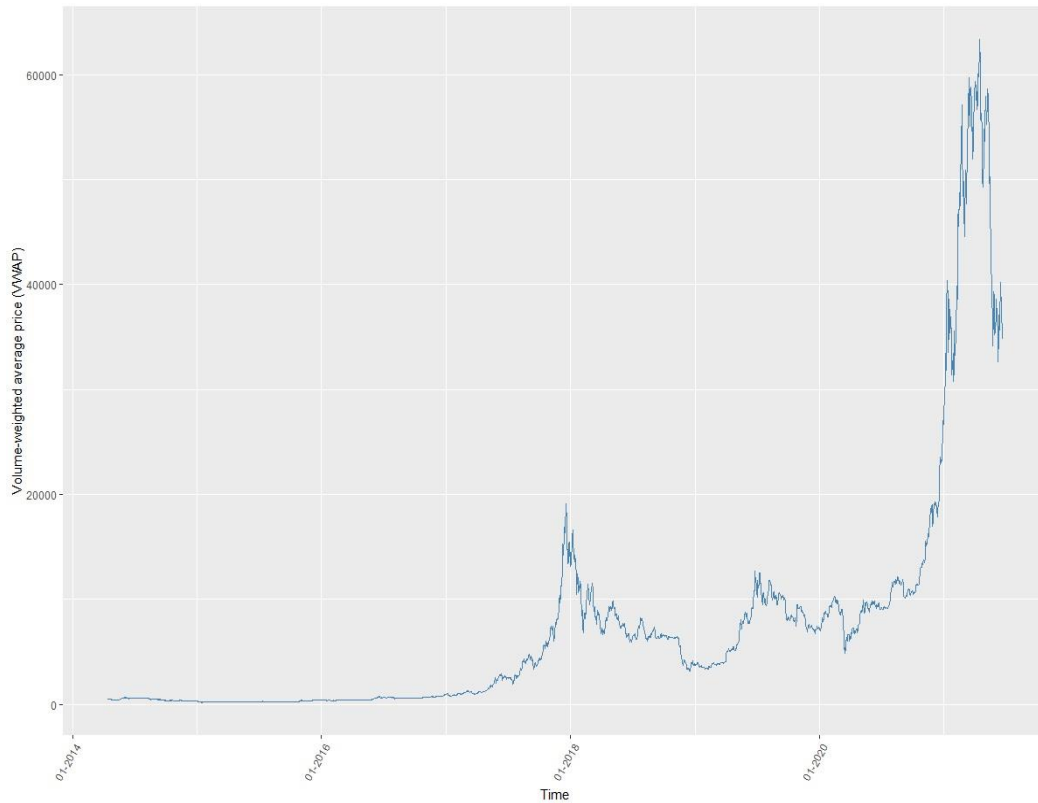


Figure 4. Bitcoin Prices (original time ordering)

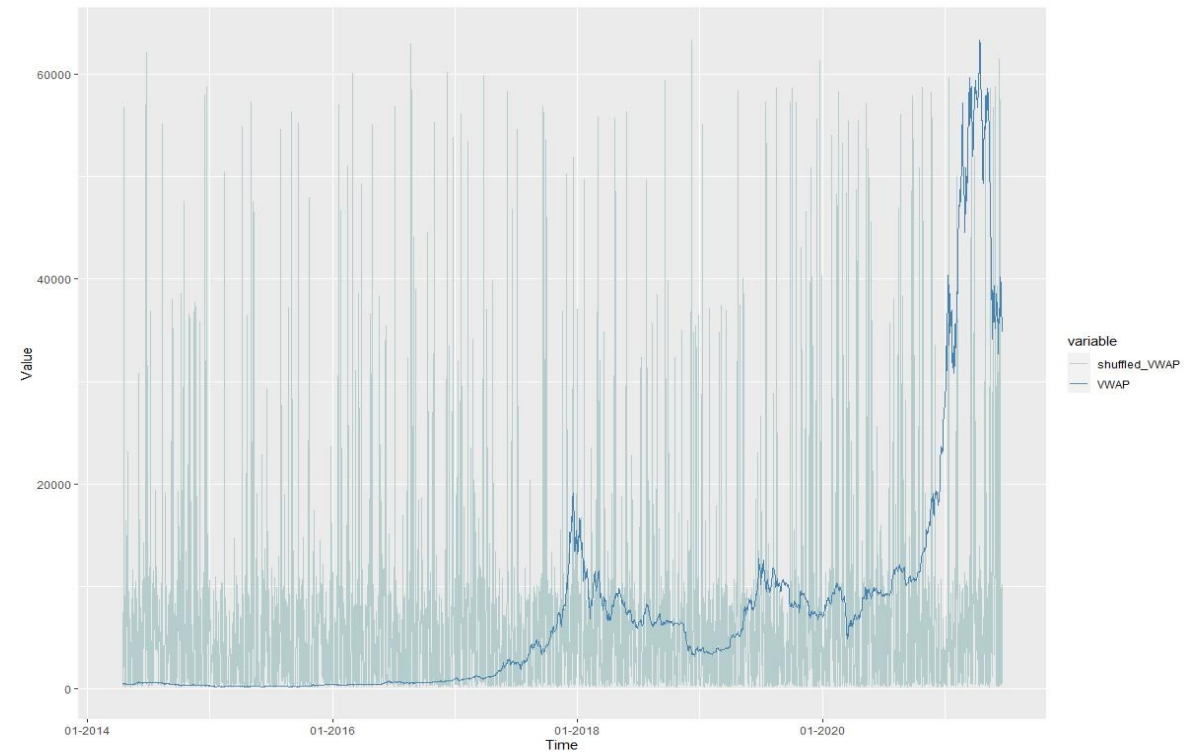
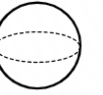


Figure 5. Bitcoin Prices (original time ordering and shuffled values)



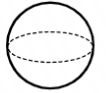
Simple X-Function: Identity and LPD

- We propose a family of Explainability (X-)functions $xf(.)$ for assigning meaning to the net's response or output o_t over time $t = 1, \dots, T$, where $o_t = (o_{1t}, \dots, o_{npt})$ 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 $w_{ijt} = \frac{\partial o_{jt}}{\partial x_{it}}, i = 1, \dots, n, j = 1, \dots, n_p$, 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:

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

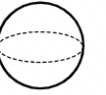
- For each output neuron o_{jt} , the resulting derivatives or 'explanations' $b_t, w_{1t}, w_{2t} \dots w_{pt}$ generate a new data-flow which is referred to as **Linear Parameter Data (LPDs)**
- The LPD is a matrix of dimension $T * (n + 1)$, irrespective of the complexity of the neural net, with t -th row denoted by $LPD_t := (b_t, w_{1t}, w_{2t} \dots w_{pt})$
- 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

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**

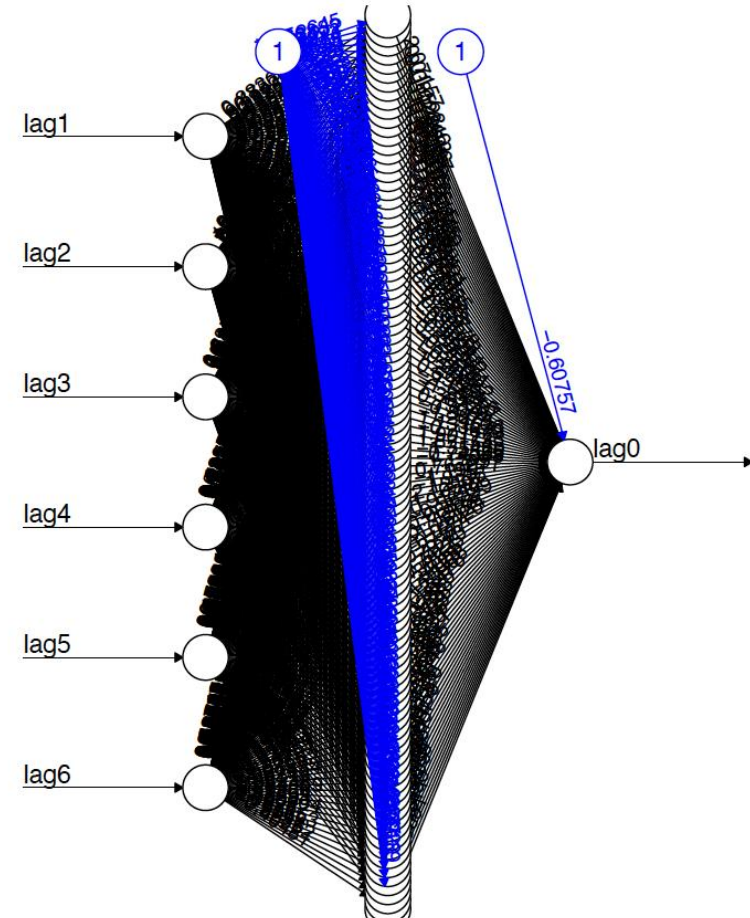
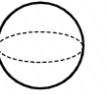


Figure 6. Neural net BTC: feedforward net with a single hidden-layer of dimension 100 and an input layer of dimension 6 comprising the last six lagged returns



Simple X-Function: Identity and LPD

- **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**

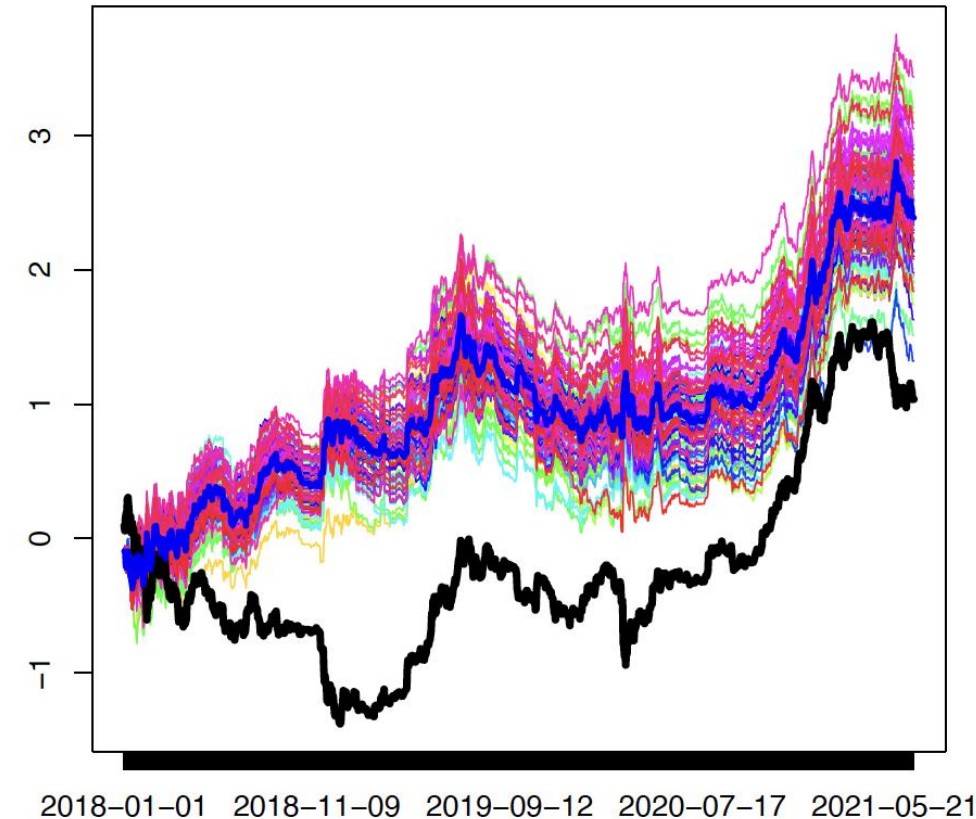
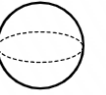


Figure 7. Cumulated log-performances out-of-sample based on sign-rule (buy or sell depending on sign of forecasted return): 'random' neural nets (colored) vs. buy-and-hold (bold black) and mean-net performance (bold blue)

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.**

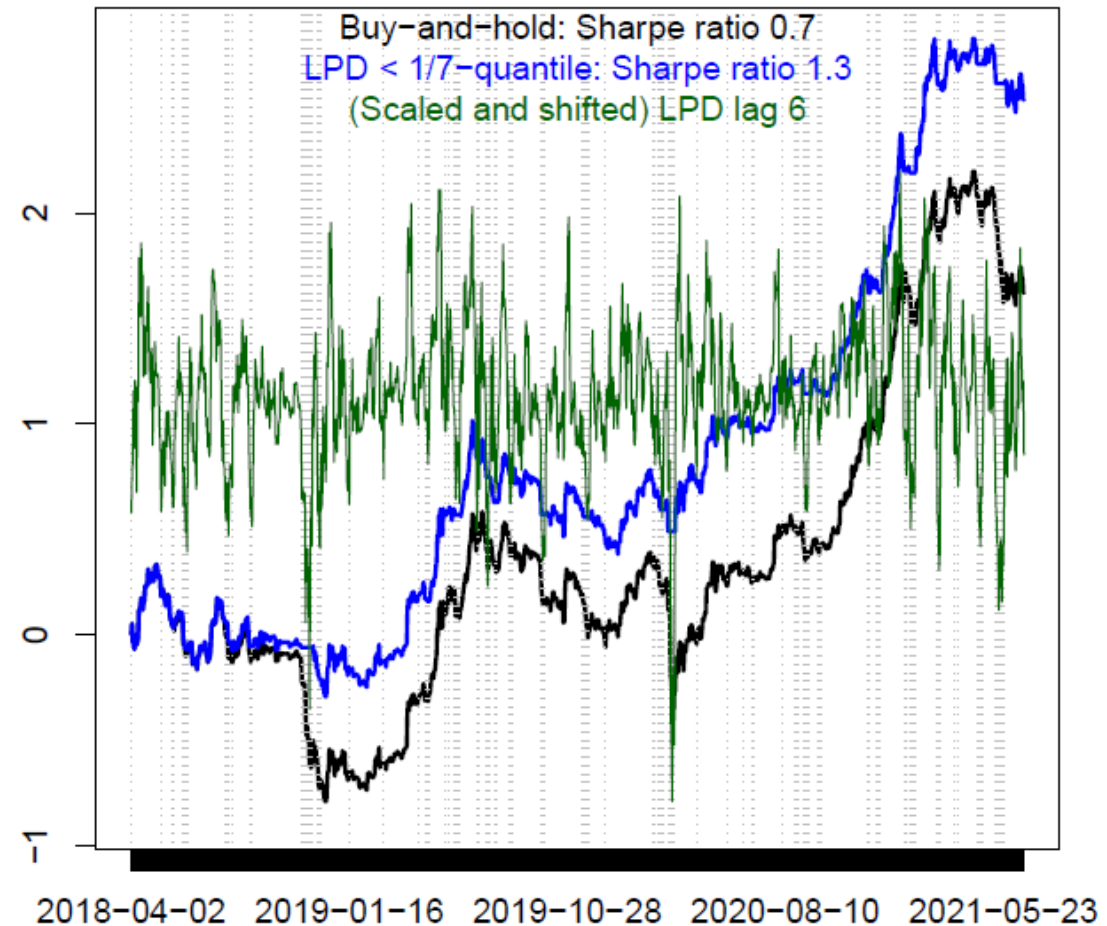


Figure 8. Buy-and-hold (black) vs. out-of-sample (mean-) LPD market-exit strategy (blue): exits (shaded in grey) occur if today's out-of-sample mean-LPD (green) drops below the 1/7-quantile based on a rolling-window of length one quarter of its own history. The LPD corresponding to the lag-6 BTC-value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

Concluding Remarks

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



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 through 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.



A Time Series Approach to Explainability for Neural Nets with Applications to Risk-Management and Fraud Detection

Marc Wildi and Branka Hadji Mishaeva
January 13, 2022

Abstract
Artificial intelligence (AI) is creating one of the biggest revolutions across technology-driven application fields. For the finance sector, it offers many opportunities for significant market innovation and yet broad adoption of AI systems heavily relies on our trust in their outputs. Trust in technology is enabled by understanding the rationale behind the predictions made. In other words, we need to make sure that values and domain knowledge are reflected in the algorithms' outcomes. To this end, the concept of explainable AI (XAI) emerged introducing a suite of techniques attempting to explain to users how complex models arrived at a certain decision. Even though many of the classical XAI approaches can lead to valuable insights about the models' inner workings, in most cases these techniques are not tailored for time series applications due to the presence of possibly complex and non-stationary dependence structure of the data. In this paper, we propose a generic XAI-technique for deep learning methods (DL) which preserves and exploits the natural time ordering of the data by introducing a family of so-called explainability (X)-functions. This concept bypasses severe identifiability issues, related among others to preclude numerical optimization problems, and it promotes transparency by means of intuitively appealing input-output relations, ordered by time. We illustrate the generic concept based on financial time series and we derive explicit expressions for two specific X-functions for tracking potential non-linearity of the model and, by extension, for tracking non-stationarity of the data generating process. Our examples suggest that this natural extension of the original XAI-approach, namely inferring a better understanding of the data from a better understanding of the model, might provide additional value in a broad range of application fields, including risk management and fraud detection.

1 Introduction

Developing accurate forecasting methodologies for financial time series remains one of the key research topics relevant from both a theoretical and applied viewpoints. Traditionally, researchers aimed at constructing a causal model, based on econometric modelling, that explains the variations in the specific time series as a function of other inputs. Yet, traditional approaches often struggle when it comes to modelling high-dimensional, non-linear landscapes often characterized with missing or sparse input space.

Recently, deep learning (DL) has become highly popularized in many aspects of data science and has become increasingly applied to forecasting financial and economic time series [1], [2], [3], [4]. Recurrent methods are related to time series modelling due to their memory state and their ability to learn relations through time; however, convolutional neural networks (CNN) are also able to build temporal relationships [5]. The literature offers various examples of the application of DL methods to stock and forex market forecasting, with results that significantly outperform

arXiv:2103.00949v1 [q-fin.RM] 1 Mar 2021

¹Financial support by the Swiss National Science Foundation within the project "Mathematics and Finance - the next revolution in the digital transformation of the finance industry" is gratefully acknowledged by the corresponding author. This research has also received funding from the European Union's Horizon 2020 research and innovation program (ERC-ETN - A Financial Supervision and Technology compliance training programme under the grant agreement No 825213 (Eupac - ERC-10-2018, Type of action: CSA). Moreover, this article is also based upon the work from the Innovation Project 41084 I 1P-SBM Towards Explainable Artificial Intelligence and Machine Learning in Credit Risk Management. Furthermore, this article is based upon work from COST Action 19130 Finance and Artificial Intelligence in Finance, supported by COST (European Cooperation in Science and Technology), www.cost.eu (Action Chair: Joerg Osterrieder). The authors are grateful to management committee members of the COST Action CA19130 Finance and Artificial Intelligence in Finance as well as speakers and participants of the 5th European COST Conference on Artificial Intelligence in Finance and Industry, which took place at Zurich University of Applied Sciences, Switzerland, in September 2020.

DOWNLOAD

DOWNLOAD