FinTech HO2020 Project

Deliverable 3.1

Grant agreement No. 825215 (Topic: ICT-35-2018 Type of action: CSA). 2021

Document Information

Deliverable information	
WP NO.	WP3
DEL. REL. NO.	D3.1
DEL. NO.	D7
TITLE	Repository of papers in artificial intelligence
DESCRIPTION	Repository of papers in artificial intelligence produced by consortium partners throughout the duration of the project.
NATURE	Document
EST. DEL. DATE	June 30 2021

Document information	
DATE	28/05/2021
WRITTEN BY	UCL and UNIPV
APPROVED BY	PAOLO GIUDICI (PC)

1

Repository of papers in artificial intelligence

The University of Pavia and the University College of London as WP3 Leader, with the support of all the Consortium, has developed the repository of papers related to artificial intelligence under the WP3. The repository of all research papers outputted regarding artificial intelligence. The material has been developed by individual partners and by collaborations from within the consortium. The papers have been presented at various academic conferences and have been published in Open Access Journals or have been archived by the authors to maintain an open access copy in an Open Access Repository e.g. Arxiv, SSRN.

Specifically, the final repository contains the following information about the papers.

- Title
- Authors
- Abstract
- Partners
- Journal
- Date
- Link (doi for open access articles and also SSRN for those not open access)

Highlights

Financial technology (FinTech) solutions that employ Artificial Intelligence solutions are growing steadily, both in investment lending, driven by peer to peer lending, and in asset management, where robot advisory is emerging as a new financial service. While the application of machine learning methods and AI may reduce transaction costs, improve financial inclusion and improve user experience and risk profile matching, it may also increase financial risks, coming from the strong interconnections between investments and assets that derive from a common asset management platform.

The measurement of the additional source of financial risk due to platform contagion is of key interest to regulators and supervisors. The EU-funded FIN-TECH has developed research aimed at measuring this risk.Twelve papers have been selected for inclusion in the project's AI research repository; three of them have been selected as use cases to be shared with regulators, supervisors, banks and fintechs. Among them, the paper "Explainable machine learning in credit risk management in peer to peer lending platforms" has received the best feedback.

The main contribution of the paper is the proposition of a methodology based on clustered Shapley values which, applied to the predictions that derive from a complex machine learning model, make it explainable, in terms of the features that mostly determine each single prediction. The use case and the feedback from the project's stakeholder reveal that the proposed method is, at the same time, highly predictively accurate and interpretable. It can thus be suggested as a standard risk measurement practice in the application of machine learning models to finance.

Following the EU proposed regulation of AI, much research activity has been dedicated to the topic of building trustworthy and explainable AI models. Besides the best paper, six other propose explainable AI models: two for credit lending, three for asset management, and one proposed an alternative to the mainstream XAI model, based on the combination between Shapley values and Lorenz zonoids, which produces an easy to interpret normalised measure of explainability. The remaining five papers concern the construction of robot advisory models for asset management: respectively for bond markets, currency markets, financial markets, venture capital investments and digital investments.

Title

Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments

Authors

Javier Arroyo; Francesco Corea; Guillermo Jimenez-Diaz; Juan A. Recio-Garcia

Abstract

The venture capital (VC) industry offers opportunities for investment in early-stage companies where uncertainty is very high. Unfortunately, the tools investors currently have available are not robust enough to reduce risk and help them managing uncertainty better. Machine learning data-driven approaches can bridge this gap, as they already do in the hedge fund industry. These approaches are now possible because data from thousands of companies over the world is available through platforms such as Crunchbase. Previous academic efforts have focused only on predicting two classes of exits, i.e., being acquired by other company or offering shares to the public, using only one or a few subsets of explanatory variables. These events are typically related to high returns, but also higher risk, making hard for a venture fund to get repeatable and sustainable returns. On the contrary, we will try to predict more possible outcomes including a subsequent funding round or the closure of the company using a large set of signals. In this way, our approach would provide VC investors with more information to set up a portfolio with lower risk that may eventually achieve higher returns than those based on finding unicorns (i.e., companies with a valuation higher than one billion dollars). We will analyze the performance of several machine learning methods in a dataset of over 120,000 early-stage companies in a realistic setting that tries to predict their progress in a 3-year time window. Results show that machine learning can support venture investors in their decision-making processes to find opportunities and better assessing the risk of potential investments.

Partners

- Universidad Complutense de Madrid

Journal

IEEE

Date of Publication

30 August 2019

Link

https://doi.org/10.1109/ACCESS.2019.2938659

Title

Sentiment Analysis of European Bonds 2016–2018

Authors

Peter Schwendner, Martin Schüle and Martin Hillebrand

Abstract

We revisit the discussion of market sentiment in European sovereign bonds using a correlation analysis toolkit based on influence networks and hierarchical clustering. We focus on three case studies of political interest. In the case of the 2016 Brexit referendum, the market showed negative correlations between core and periphery only in the week before the referendum. Before the French presidential elections in 2017, the French bond spread widened together with the estimated Le Pen election probability, but the position of French bonds in the correlation blocks did not weaken. In summer 2018, during the budget negotiations within the new Italian coalition, the Italian bonds reacted very sensitively to changing political messages but did not show contagion risk to Spain or Portugal for several months. The situation changed during the week from October 22 to 26, as a spillover pattern of negative sentiment also to the other peripheral countries emerged.

Partners

- Zurich University of Applied Sciences

Journal

Frontiers in Artificial Intelligence

Date of Publication

15 October 2019

Link

https://doi.org/10.3389/frai.2019.00020

Title

COVID-19 contagion and digital finance

Authors

Arianna Agosto & Paolo Giudici

Abstract

Digital finance is going to be heavily affected by the COVID-19 outbreak. We present a statistical model which can be employed to understand the contagion dynamics of the COVID-19, so that its impact on finance can possibly be anticipated, and digitally monitored. The model is a Poisson autoregression of the daily new observed cases, and considers both short-term and long-term dependence in the infections counts. Model results are presented for the observed time series of China, the first affected country, but can be easily reproduced for all countries.

Partners

- University of Pavia

Journal

Digital Finance

Date of Publication

11 May 2020

Link

https://doi.org/10.1007/s42521-020-00021-3

Title

Analysing Social Media Forums to Discover Potential Causes of Phasic Shifts in Cryptocurrency Price Series

Authors

Andrew Burnie, Emine Yilmaz and Tomaso Aste

Abstract

The recent extreme volatility in cryptocurrency prices occurred in the setting of popular social media forums devoted to the discussion of cryptocurrencies. We develop a framework that discovers potential causes of phasic shifts in the price movement captured by social media discussions. This draws on principles developed in healthcare epidemiology where, similarly, only observational data are available. Such causes may have a major, one-off effect, or recurring effects on the trend in the price series. We find a one-off effect of regulatory bans on bitcoin, the repeated effects of rival innovations on ether and the influence of technical traders, captured through discussion of market price, on both cryptocurrencies. The results for Bitcoin differ from Ethereum, which is consistent with the observed differences in the timing of the highest price and the price phases. This framework could be applied to a wide range of cryptocurrency price series where there exists a relevant social media text source. Identified causes with a recurring effect may have value in predictive modelling, whilst one-off causes may provide insight into unpredictable black swan events that can have a major impact on a system.

Partners

- University College London

Journal

Frontiers in Blockchain

Date of Publication

28 January 2020

Link

https://doi.org/10.3389/fbloc.2020.00001

Title

Network Models to Enhance Automated Cryptocurrency Portfolio Management

Authors

Paolo Giudici, Paolo Pagnottoni and Gloria Polinesi

Abstract

The usage of cryptocurrencies, together with that of financial automated consultancy, is widely spreading in the last few years. However, automated consultancy services are not yet exploiting the potentiality of this nascent market, which represents a class of innovative financial products that can be proposed by robo-advisors. For this reason, we propose a novel approach to build efficient portfolio allocation strategies involving volatile financial instruments, such as cryptocurrencies. In other words, we develop an extension of the traditional Markowitz model which combines Random Matrix Theory and network measures, in order to achieve portfolio weights enhancing portfolios' risk-return profiles. The results show that overall our model overperforms several competing alternatives, maintaining a relatively low level of risk.

Partners

- University of Pavia

Journal

Frontiers in Artificial Intelligence

Date of Publication

24 April 2020

Link

https://doi.org/10.3389/frai.2020.00022

Title

Shapley-Lorenz eXplainable Artificial Intelligence

Authors

Paolo Giudici and Emanuela Raffinetti

Abstract

Explainability of artificial intelligence methods has become a crucial issue, especially in the most regulated fields, such as health and finance. In this paper, we provide a global explainable AI method which is based on Lorenz decompositions, thus extending previous contributions based on variance decompositions. This allows the resulting Shapley-Lorenz decomposition to be more generally applicable, and provides a unifying variable importance criterion that combines predictive accuracy with explainability, using a normalised and easy to interpret metric. The proposed decomposition is illustrated within the context of a real financial problem: the prediction of bitcoin prices.

Partners

- University of Pavia

Journal

Expert Systems with Applications

Date of Publication

1 April 2021

Link

https://doi.org/10.1016/j.eswa.2020.114104

Title

Neural networks and arbitrage in the VIX

Authors

Joerg Osterrieder, Daniel Kucharczyk, Silas Rudolf and Daniel Wittwer

Abstract

The Chicago Board Options Exchange Volatility Index (VIX) is considered by many market participants as a common measure of market risk and investors' sentiment, representing the market's expectation of the 30-day-ahead looking implied volatility obtained from real-time prices of options on the S&P 500 index. While smaller deviations between implied and realized volatility are a well-known stylized fact of financial markets, large, time-varying differences are also frequently observed throughout the day. Furthermore, substantial deviations between the VIX and its futures might lead to arbitrage opportunities on the VIX market. Arbitrage is hard to exploit as the potential strategy to exploit it requires buying several hundred, mostly illiquid, out-of-the-money (put and call) options on the S&P 500 index. This paper discusses a novel approach to predicting the VIX on an intraday scale by using just a subset of the most liquid options. To the best of the authors' knowledge, this the first paper, that describes a new methodology on how to predict the VIX (to potentially exploit arbitrage opportunities using VIX futures) using most recently developed machine learning models to intraday data of S&P 500 options and the VIX. The presented results are supposed to shed more light on the underlying dynamics in the options markets, help other investors to better understand the market and support regulators to investigate market inefficiencies.

Partners

- Zurich University of Applied Sciences

Journal

Digital Finance

Date of Publication

13 August 2020

Link

https://doi.org/10.1007/s42521-020-00026-y

Title

Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending

Authors

Miller Janny Ariza-Garzón; Javier Arroyo; Antonio Caparrini; Maria-Jesus Segovia-Vargas

Abstract

Peer-to-peer (P2P) lending demands effective and explainable credit risk models. Typical machine learning algorithms offer high prediction performance, but most of them lack explanatory power. However, this deficiency can be solved with the help of the explainability tools proposed in the last few years, such as the SHAP values. In this work, we assess the well-known logistic regression model and several machine learning algorithms for granting scoring in P2P lending. The comparison reveals that the machine learning alternative is superior in terms of not only classification performance but also explainability. More precisely, the SHAP values reveal that machine learning algorithms can reflect dispersion, nonlinearity and structural breaks in the relationships between each feature and the target variable. Our results demonstrate that is possible to have machine learning credit scoring models be both accurate and transparent. Such models provide the trust that the industry, regulators and end-users demand in P2P lending and may lead to a wider adoption of machine learning in this and other risk assessment applications where explainability is required.

Partners

- Universidad Complutense de Madrid

Journal

IEEE

Date of Publication

30 March 2020

Link

https://doi.org/10.1109/ACCESS.2020.2984412

Title

Explainable Machine Learning in Credit Risk Management

Authors

Niklas Bussmann, Paolo Giudici, Dimitri Marinelli and Jochen Papenbrock

Abstract

The paper proposes an explainable Artificial Intelligence model that can be used in credit risk management and, in particular, in measuring the risks that arise when credit is borrowed employing peer to peer lending platforms. The model applies correlation networks to Shapley values so that Artificial Intelligence predictions are grouped according to the similarity in the underlying explanations. The empirical analysis of 15,000 small and medium companies asking for credit reveals that both risky and not risky borrowers can be grouped according to a set of similar financial characteristics, which can be employed to explain their credit score and, therefore, to predict their future behaviour.

Partners

- University of Pavia
- Firamis

Journal

Computational Economics

Date of Publication

25 September 2020

Link

https://doi.org/10.1007/s10614-020-10042-0

Title

Predictability and pricing efficiency in forward and spot, developed and emerging currency markets

Authors

Valerio Potì, Richard Levich, Thomas Conlon

Abstract

We study the predictability of forward and spot exchange rates of currencies of emerging and developed economies from 1994 to 2016. Our purpose is to shed light on the efficiency of currency markets and how and why it has evolved over this time. For the currencies of emerging economies, our analysis of rates of return on forward contracts finds some evidence of excess-predictability, especially in the earlier parts of the sample period, consistent with the view that this portion of the foreign exchange market has only become efficient in recent times. When we turn our attention to excess-returns computed from spot exchange rates and spot interest rates, however, we find much less predictability. In particular, over our full sample period, we find no evidence of excess-predictability, in contrast with the results reported by Hsu et al. (2016) but in agreement with Kuang et al. (2014). The different predictability of spot excess-returns and rates of return on forward contracts is a manifestation of the widespread violation of covered interest parity which emerged with the onset of the 2008 financial crisis.

Partners

- University College Dublin

Journal

Journal of International Money and Finance

Date of Publication

October 2020

Link

https://doi.org/10.1016/j.jimonfin.2020.102223

Title

Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective

Authors

Andreas G. F. Hoepner, David McMillan,Andrew Vivian & Chardin Wese Simen

Abstract

Although machine learning is frequently associated with neural networks, it also comprises econometric regression approaches and other statistical techniques whose accuracy enhances with increasing observation. What constitutes high quality machine learning is yet unclear though. Proponents of deep learning (i.e. neural networks) value computational efficiency over human interpretability and tolerate the 'black box' appeal of their algorithms, whereas proponents of explainable artificial intelligence (xai) employ traceable 'white box' methods (e.g. regressions) to enhance explainability to human decision makers. We extend Brooks et al.'s [2019. 'Financial Data Science: The Birth of a New Financial Research Paradigm Complementing Econometrics?' European Journal of Finance 25 (17): 1627–36.] work on significance and relevance as assessment critieria in econometrics and financial data science to contribute to this debate. Specifically, we identify explainability as the Achilles heel of classic machine learning approaches such as neural networks, which are not fully replicable, lack transparency and traceability and therefore do not permit any attempts to establish causal inference. We conclude by suggesting routes for future research to advance the design and efficiency of 'white box' algorithms.

Partners

- University College Dublin

Journal

The European Journal of Finance

Date of Publication

03 Dec 2020

Link

https://doi.org/10.1080/1351847X.2020.1847725

Title

Interpretable Machine Learning for Diversified Portfolio Construction

Authors

Markus Jaeger, Stephan Krügel, Dimitri Marinelli, Jochen Papenbrock and Peter Schwendner

Abstract

In this article, the authors construct a pipeline to benchmark hierarchical risk parity (HRP) relative to equal risk contribution (ERC) as examples of diversification strategies allocating to liquid multi-asset futures markets with dynamic leverage (volatility target). The authors use interpretable machine learning concepts (explainable AI) to compare the robustness of the strategies and to back out implicit rules for decision-making. The empirical dataset consists of 17 equity index, government bond, and commodity futures markets across 20 years. The two strategies are back tested for the empirical dataset and for about 100,000 bootstrapped datasets. XGBoost is used to regress the Calmar ratio spread between the two strategies against features of the bootstrapped datasets. Compared to ERC, HRP shows higher Calmar ratios and better matches the volatility target. Using Shapley values, the Calmar ratio spread can be attributed especially to univariate drawdown measures of the asset classes.

Partners

- Firamis
- ZHAW

Journal

Journal of Financial Data Science

Date of Publication

August 2021

Link

https://doi.org/10.3905/jfds.2021.1.066

Title

Matrix Evolutions: Synthetic Correlations and Explainable Machine Learning for Constructing Robust Investment Portfolios

Authors

Jochen Papenbrock, Peter Schwendner, Markus Jaeger and Stephan Krügel

Abstract

In this article, the authors present a novel and highly flexible concept to simulate correlation matrixes of financial markets. It produces realistic outcomes regarding stylized facts of empirical correlation matrixes and requires no asset return input data. The matrix generation is based on a multiobjective evolutionary algorithm, so the authors call the approach matrix evolutions. It is suitable for parallel implementation and can be accelerated by graphics processing units and quantum-inspired algorithms. The approach is useful for backtesting, pricing, and hedging correlation-dependent investment strategies and financial products. Its potential is demonstrated in a machine learning case study for robust portfolio construction in a multi-asset universe: An explainable machine learning program links the synthetic matrixes to the portfolio volatility spread of hierarchical risk parity versus equal risk contribution.

Partners

- Firamis
- ZHAW

Journal

Journal of Financial Data Science

Date of Publication

April 2021

Link

https://doi.org/10.3905/jfds.2021.1.056

Title

Evaluation of multi-asset investment strategies with digital assets

Authors

Alla Petukhina Erin Sprünken

Abstract

The drastic growth of the cryptocurrencies market capitalization boosts investigation of their diversification benefits in portfolio construction. In this paper with a set of classical and modern measurement tools, we assess the out-of-sample performance of eight portfolio allocation strategies relative to the naive 1/N rule applied to traditional and crypto-assets investment universe. Evaluated strategies include a range from classical Markowitz rule to the recently introduced LIBRO approach (Trimborn et al. in Journal of Financial Econometrics 1–27, 2019). Furthermore, we also compare three extensions for strategies with respect to input estimators applied. The results show that in the presence of alternative assets, such as cryptocurrencies, mean–variance strategies underperform the benchmark portfolio. In contrast, CVaR optimization tends to outperform the benchmark as well as geometric optimization, although we find a strong dependence of the former's success on trading costs. Furthermore, we find evidence that liquidity-bounded strategies tend to perform very well. Thus, our findings underscore the non-normal distribution of returns and the necessity to control for liquidity constraints at alternative asset markets.

Partners

- Humboldt University Berlin

Journal

Digital Finance

Date of Publication

4 April 2021

Link

https://doi.org/10.1007/s42521-021-00031-9

Annex List of publications

Title	Authors	Journal	Keywords
Sentiment Analysis of European Bonds 2016–2018	Peter Schwendner, Martin Schüle, Martin Hillebrand	Frontiers in Artificial Intelligence	Bond market models, Correlation network models, Machine learning econometrics
Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending	Miller Janny Ariza-Garzon, Javier Arroyo, Antonio Caparrini, Maria-Jesus Segovia-Vargas	IEEE Access	Explainable AI, credit scoring, lending decisions
Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments	Javier Arroyo, Francesco Corea, Guillermo Jimenez-Diaz, Juan A. Recio-Garcia	IEEE Access	Investment models, machine learning models, predictive accuracy
Explainable Machine Learning in Credit Risk Management	Niklas Bussmann, Paolo Giudici, Dimitri Marinelli, Jochen Papenbrock	Computational Economics	Explainable AI, peer to peer lending, clustering models
COVID-19 contagion and digital finance	Arianna Agosto, Paolo Giudici	Digital Finance	Financial market models, covid-19 contagion, predictive models
Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective	Andreas G. F. Hoepner, David McMillan, Andrew Vivian, Chardin Wese Simen	The European Journal of Finance	Financial market models, Machine learning models, econometric models
Neural networks and arbitrage in the VIX	Joerg Osterrieder, Daniel Kucharczyk, Silas Rudolf, Daniel Wittwer	Digital Finance	Financial market models, Neural networks, volatility models
Shapley-Lorenz eXplainable Artificial Intelligence	Paolo Giudici, Emanuela Raffinetti	Expert Systems with Applications	Explainable AI, Shapley method, Rank-based performance
Interpretable Machine Learning for Diversified Portfolio Construction	Markus Jaeger, Stephan Krügel, Dimitri Marinelli, Jochen Papenbrock, Peter Schwendner	The Journal of Financial Data Science	Explainable AI, portfolio allocation, machine learning models

Matrix Evolutions: Synthetic Correlations and Explainable Machine Learning for Constructing Robust Investment Portfolios	Jochen Papenbrock, Peter Schwendner, Markus Jaeger, Stephan Krügel	The Journal of Financial Data Science	Explainable AI, portfolio robustness, machine learning models
Evaluation of multi-asset investment strategies with digital assets	Alla Petukhina, Erin Sprünken	Digital Finance	Financial Market imodels, Digital investments, portfolio allocation.
Predictability and pricing efficiency in forward and spot, developed and emerging currency markets	Valerio Potì, Richard Levich, Thomas Conlon	Journal of International Money and Finance	Currency markets models, Machine learning models, predictive models