Explainable AI in finance

ASEN PANTOV, University of Twente, The Netherlands

The demand for models that are interpretable and transparent is on the rise as Artificial Intelligence (AI) is getting more widely used. When discussing AI in the financial sector, making decisions is supposed to be effective, but most importantly, it must be done openly and responsibly. The need for models to be interpretable is especially crucial in this industry for various reasons, including regulatory compliance, trust and reliability and informed decision-making. The central theme of this paper is an examination of the common applications and methodologies of Explainable AI (XAI) in the financial industry. Different aspects of finance are explored, and various commonly used XAI techniques are identified. This study analyzes relevant research articles through a systematic literature review focusing on the application and types of XAI methods. The results show that XAI is widely applied in fields such as credit risk assessment, fraud detection, stock market forecasting and customer profiling. Another finding from the examined literature is that SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) are solidified as the de facto industry standard for XAI currently. Thus, such research assists to introduce more developments within finance for XAI while highlighting major trends, gaps and future directions that need to be addressed in future studies.

Additional Key Words and Phrases: Explainable AI (XAI), Interpretable Machine Learning, AI Transparency, AI Accountability, Finance Industry

1 INTRODUCTION

At present, innovation in most areas involves Artificial Intelligence (AI) as a driving factor with great potential. This technological wave of advanced models is getting increasingly integrated into different industries and areas of study [Crompton and Burke 2023; Miotto et al. 2017; S Nair 2022; Szczepaniuk and Szczepaniuk 2022; Waltersmann et al. 2021]. All businesses are pushing in that direction and looking for a way to improve through the use of AI. With the desire for such innovation, there is also a need for explainability, especially in high-stakes environments like finance, where decisions impact economic outcomes and human lives.

The decision-making process may be hidden by the complexity and sometimes vague nature of AI systems, which can undermine confidence and accountability. This is especially important for the financial industry, as regulatory issues[1] and stakeholder trust loss might arise from an inability to comprehend and justify AI-driven choices. Thus, the goal of this research is to address key questions that are vital for understanding and improving the implementation of explainable AI in finance. The exact research questions can be found in Section 1.1. The first related research question concerns the techniques and methods used for achieving XAI. Assimilating them and their uses is essential as it can lead the way for future technological development in the area and further help financial enterprises understand what is appropriate for their purposes. Another crucial research question of this paper is about understanding the common applications of Explainable AI (XAI). This question explores the key areas within finance, such as credit risk assessment and fraud detection, where XAI is actively applied. Based on the answer, the extent of present implementations is established, and areas that may require more innovation are understood. This paper answers important questions that have great effect and point to future research directions. This is achieved through the means of a systematic literature review conducted by identifying and evaluating all relevant literature linked to XAI in the context of finance.

[1] https://artificialintelligenceact.eu/

TScIT 41, July 5, 2024, Enschede, The Netherlands

1.1 Problem Statement & Research Question

With the rapid advancement of AI, their integration within finance has increased. From that, a need for explainability emerges in the way that decisions made by those systems are derived. The financial industry is inherently complex, with high regulation standards and requiring a high level of trust from its customers [Parra Rodriguez 2022]. The issue is how traceable a decision made by an AI algorithm is. Not knowing why a person had their credit risk assessed in a particular way raises bias and unfair treatment concerns, potentially violating regulatory compliance that asks for clear explanations. Despite the growing integration of AI in finance, there is insufficient discussion of Explainable AI (XAI), which could fix the 'black box' problem with AI algorithms by helping humans understand the AI decision-making process. The primary issue, though, is not only integrating XAI to improve explainability but also comprehending the precise applications and uses of these explainable models. There is a need to structure the knowledge that is out there on the topic of Explainable AI implementation in finance and assess how and with what purpose it has been used. This investigation is critical as it will benefit the discussed problem and major point of concern. Through this approach, a path should form on how advancements in the application of AI can be made in a way that should pose no concerns to any stakeholders or customers involved in a financial service. Taking this into account, the main research question of this paper is:

How is Explainable AI (XAI) utilized in the financial industry to enhance explainability and interpretability?

The following subquestions have been derived to help answer the main research question:

(1) Which techniques and models are most commonly used for providing explainability in AI systems in finance?
(2) What are the application areas of Explainable AI (XAI) in the financial industry?

The 1st research subquestion is concerned with keeping track of the most frequent methodologies in XAI. The 2nd research subquestion revolves around key areas within finance, such as risk management, credit scoring, investment decisions, and fraud detection, where XAI is and has the potential to be actively applied.

The main research question and related subresearch questions will be explored through the means of a systematic literature review. This methodology has proved successful in answering research questions related to those laid out in this work [Avila-Garzon 2020]. The practical outcome of the employed methodology is a curated table of the examined research, summarizing the articles by mentioning the important aspects of every single one individually. From this process, the answers to the posed questions will be found and discussed extensively.

This work's major contributions lie in its examination of the current landscape of Explainable Artificial Intelligence (XAI) in finance. In doing this, the research reviewed a wide range of literature that helped to identify key areas where XAI is applied and how it is being used most frequently, thereby enabling insights into the current intersection of XAI and the field of finance. The review classified studies by objectives, methodologies, explanations techniques, applications area, and main contribution, thus providing a blueprint for how XAI is employed to facilitate decision-making processes in finance. The findings add to the existing knowledge base by pointing out trends, gaps, and future directions for research, hence setting an agenda for further development in the field.

The paper is further divided into sections, with the following one being Section 2, which elaborates on the research methodology taken. After it is explained how the systematic literature review was conducted, Section 3 reveals the landscape of collected literature. Consequently, Section 4 delves into an extensive discussion of the examined research, answering the posed research questions. The final part of this work is Section 5, which notes the conclusions of this paper, its limitations, and future research agenda.

2 RESEARCH METHODOLOGY

A systematic literature review should guarantee unbiased and replicable outcomes [Amato et al. 2024]. Therefore, for the purpose of this research, the review is based on the work of S et al. [2024]. Following this approach, one can be certain that the research is done properly. Scopus[2] is the database utilized for the search because of its extensive library and high standards [Baas et al. 2020]. This will ensure the quality of the explored papers.

A relatively straightforward query was initially utilized to explore the database. The query is the following:

- (TITLE-ABS-KEY ("explainable ai") AND TITLE-ABS-KEY (financ*))

The simplicity of the query ensures a wide range of results, guaranteeing that most papers with the potential to be relevant are indeed part of the search. This serves as a good starting point, yielding 243 search results. However, it is too generic for the purpose of a systematic literature review; thus, based on the research questions, it was refined. Advancing our query with the application of filters provides a more curated list of related papers. The query with filters that was used is the following:

- (TITLE-ABS-KEY ("explainable ai") AND TITLE-ABS-KEY (financ*)) AND (LIMIT-TO (EXACTKEYWORD , "Finance") OR LIMIT-TO (EXACTKEYWORD , "Forecasting") OR LIMIT-TO (EXACTKEYWORD , "Financial Markets") OR LIMIT-TO (EXACTKEYWORD , "Investments") OR LIMIT-TO (EXACTKEYWORD , "Risk Assessment") OR LIMIT-TO (EXACTKEYWORD , "Risk Management") OR LIMIT-TO (EXACTKEYWORD , "Financial Service") OR LIMIT-TO (EXACTKEYWORD , "Decision Making")) AND (LIMIT-TO (LANGUAGE , "English")) AND (EXCLUDE (DOCTYPE , "tb"))

The use of keywords to focus on particular aspects of finance limits the results to those that are more relevant. Furthermore, the article should have been written in English. Then, it should not be retracted as that would imply there is a major problem with the article [COPE Council 2019]. Citation count was also considered as a filtering metric; however, due to how recent the articles are, it was found inappropriate in this case. Furthermore, papers which were not open access and could not be accessed with the University of Twente institution rights were removed by hand. The updated version of the search request narrows down the list to 93 papers. There is no room for appropriate further filtering of the works through the advanced query; thus, they are manually reviewed based on their relevancy. Whether a work is relevant is judged by the Title, Keywords, and Abstract sections of the set article in a meticulous manner by taking into account the posed research questions. An overview of the whole selection and filtering process with respective numbers can be viewed in Fig. 1.

After the papers are individually assessed on their relevancy, those which are deemed suitable are further summarized in a table (Appendix) that includes important information about them. More specifically, the categories in the table are Objective, Methodology, XAI Techniques, Application Area and Main Contributions. Those particular categories are chosen with their connection to the research questions in mind. Based on the results from that activity, the landscape of literature, meaningful analysis, conclusions, and recommendations for future research are drawn.

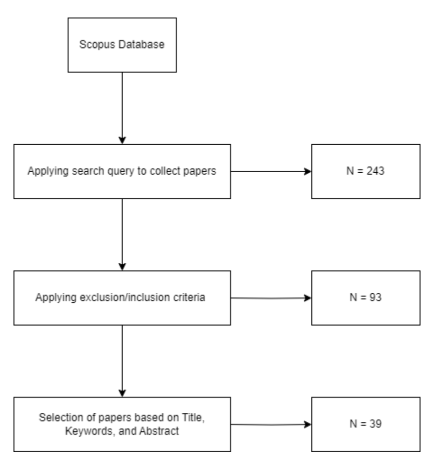

Fig. 1. Papers selection process

[2] https://www.scopus.com/

3 LANDSCAPE OF COLLECTED LITERATURE

This section presents an overview of the papers resulting from the systematic literature review. There are different ways to deconstruct the research landscape, and it has been found appropriate to structure this part into three subsections, each offering a different perspective to analyze the academic work related to Explainable Artificial Intelligence in the field of finance.

The temporal distribution of articles is presented in the first subsection. The chronological order of relevant literature is explored through a visualisation chart. The potential to distinguish trends, patterns or shifts throughout time is possible because of the provided visual.

The second subsection relates the papers and their source distribution. The distinction here is made on the basis of whether an article was published in an academic journal or presented at a conference. To visually represent this distribution, a pie chart is utilized, in order to provide a clear comparison of the number of papers from conferences compared to journals.

In the final part of this section, we explore the keywords within the reviewed studies. The main ideas and themes, as well as emerging areas of interest that guide research in the field, are exemplified through word clouds. For this purpose, two separate word clouds are made with the author and index keywords to showcase interesting deviations.

Besides this, as previously mentioned, an all-encompassing table (Table 1) that summarizes each paper into concise categories relevant to the research questions of this work is present in the Appendix. Discussions based on the findings of it are present in the Discussion section of this work. Due of its widespread application for our goal, citation-based distribution was also taken into consideration. However, given how recent the publications are, it was determined that this strategy was inappropriate in this particular instance.

The combination of all these segments can truly decipher the research environment and set the stage for an insightful interpretation of the current landscape and the evolving trends within the scope of Explainable AI in finance. This extended perspective helps distinguish between areas that have been more thoroughly examined and potentially those that would benefit from further research.

3.1 Temporal Distribution of Literature

Investigating the temporal distribution is the first step in exploring the landscape of the examined research. The bar chart in Figure 2 clearly highlights a trend of growing interest and research activity over recent years. Starting with a modest number of publications in 2019, where only one paper is part of our list, there is a prominent increase almost every year. There is only a decline from 2023 to 2024 because 2024 has not concluded at the point of writing this research. The decline is from 12 to 8 articles, which does not indicate a downward trend in the momentum of XAI, given that 2024 has not passed. The biggest observed surge is from 2021 to 2022, from 4 to 11 scholarly works in that period, showcasing a possible correlation with increased reliance on AI in the financial sector as confirmed by the work of El Hajj and Hammoud [2023]. The upward trend of research in the area reflects the ongoing developments in AI and the need of the financial industry to address the 'black box' nature of many advanced AI models.

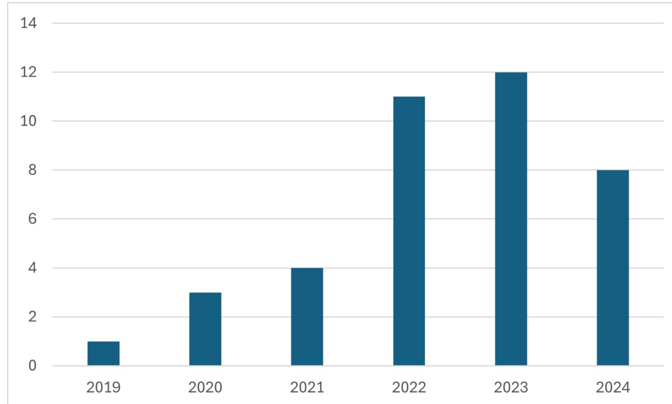

Fig. 2. Temporal distribution of filtered research papers on the topic of XAI in finance

3.2 Source Distribution of Literature

The source distribution of literature refers to the split between conference papers and articles from the examined research. As it can be seen in Figure 3, the proportion is quite even. The compiled data from our systematic literature review reveals that there are exactly 21 articles published in academic journals and 18 papers presented at conferences. The relatively balanced distribution is a testament to how dynamic a field it is. The journal articles are subjects of a peer-review process, serving as a benchmark for quality in academic contributions. On the other hand, the rapid evolution that characterizes the application of XAI in financial contexts is reflected in conference papers, which capture discoveries and promote real-time exchange of ideas and collaboration possibilities. This balance implies a healthy exchange of fundamental research and novel applications within the community, demonstrating that the discipline is simultaneously expanding on existing knowledge and adapting fast to new problems and technologies.

3.3 Keywords Distribution of Literature

The analysis of keywords within the works of the systematic literature review carries insights into the main themes and trends within the field of XAI in finance. Keywords serve a vital part in making the research process smoother for both search engines and academics. In this work, the exploration is visualised through two distinct word clouds illustrating the author (Figure 4) and index keywords (Figure 5), respectively.

The author keyword word cloud emphasizes terms such as "Explainable AI," "Machine Learning," "Deep Learning," and "Credit Scoring." These keywords point to the core technologies and applications driving current research efforts. The prominence of the terms "Deep Learning" and "Machine Learning" makes them appear as the foundation upon which much of the XAI techniques in finance are built. This is a reflection of the goal to accelerate algorithmic complexity while retaining explainability. There are many application areas that

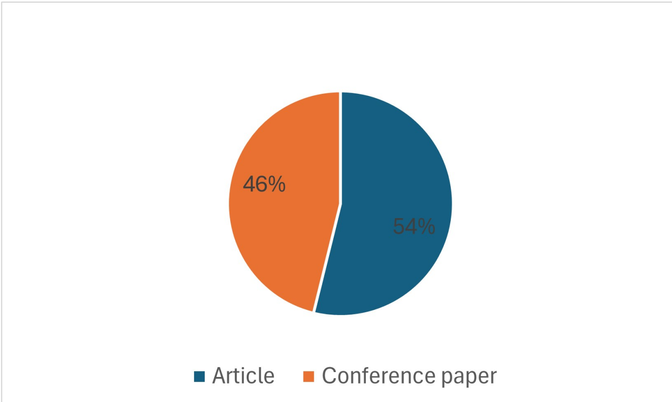

Fig. 3. Source distribution of filtered research papers on the topic of XAI in finance

can be seen, like "Credit Scoring," "Blockchain," and "Stock Market Prediction," with all of them slightly faded in the background. What is interesting is that on the XAI techniques side, we have the general term and "Shap" as words that really stand out. From there on, it is more difficult to find particular XAI methods in the illustration, thus signalling possibly how diverse the field is besides the most prominent technique with SHAP (SHapley Additive exPlanations).

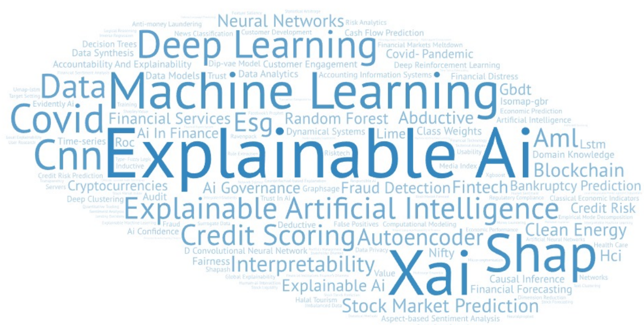

Fig. 4. Author keywords word cloud of reviewed papers

In contrast, the index keywords word cloud shifts slightly the emphasis on terms like "Forecasting," "Financial Markets," and "Investments." The adjusted focus suggests a broader thematic context. Usually allocated by database indexers or created algorithmically, these keywords help to place a study in more accessible academic and practical fields.

The disparity between these two sets of words highlights the double intention of sharing research outcomes with audiences, that is, being precise and being easily accessible. Author keywords tend to target specific interests in a particular domain, whereas index keywords expand the reach, linking the themes to general ones. In our case, the keywords word clouds have shown common XAI techniques (SHAP, LIME), AI algorithms upon which XAI techniques are utilized (Deep Learning, Neural Networks, Random Forest) and different application domains which were not part of the query

Fig. 5. Index keywords word cloud of reviewed papers

(Stock Market Prediction, Credit Scoring, Fraud Detection) among other insights. All of those are directly related to the posed research questions in this paper (Section 1.1). The analysis of these word clouds has been useful for delving into how the intersection of XAI and finance is framed within the academic landscape.

4 DISCUSSION

The systematic literature review conducted for this thesis provides an exhaustive examination of the current scope and depth of Explainable Artificial Intelligence (XAI). Following a meticulous research methodology approach, the culmination of this systematic literature review is the carefully curated table which can be found in the Appendix of this paper. As previously mentioned, it contains five categories that identify each reviewed article, namely Objective, Methodology, XAI Techniques, Application Area and Main Contributions. These particular categories were selected to correspond with the research question closely and, thus, with the subquestions concerning the types of XAI approaches used and how in the financial domain.

Elaborating on the columns from the curated Table 1, the objective outlines the specific purpose or problem characterizing each study, thereby echoing the unique challenges and opportunities in the context of XAI in finance. A clear understanding of what is sought after helps highlight the main driving forces for conducting research as well as the issues it seeks to resolve, such as the ambiguous nature of high-performing AI models that are currently used or could be used in practice. Understanding the process and assessing the scientific validity and relevance of the findings requires knowledge of the implemented methodology. It provides much necessary context for each scholarly work. The XAI techniques category highlights the specific tools and algorithms that can make AI systems more explainable. This column is clearly crucial and directly related to the 1st research subquestion. The application area category showcases where the discussed project is implemented or could be. It paints a picture of the financial segments associated with XAI and directly answers the 2nd research subquestion. Lastly, the main contributions column is critical for presenting the outputs, focusing on the novelty of each individual study. The intersection of XAI with various fields of finance can be recognized in this category.

Building on the findings provided by the systematic literature review, the discussion below is segmented into two elaborate subsections. Section 4.1 and Section 4.2 answer the 1st and 2nd research question, respectively.

4.1 XAI methods and tools

There are many different types of AI models, but a very general distinction which can be made is based on whether they're a white box or a black box model. White box models are clear and transparent, enabling their decision-making process to be understood. Those types of models are explainable by definition. However, a lot of advanced models are black box, meaning that their decisions cannot be tracked back and assimilated. Those models are not inherently explainable, but their explainability can be improved with the help of XAI techniques. The paragraphs hereafter delve into XAI discerned through a methodical review of contemporary scholarly works. These methods encompass a spectrum from model-agnostic techniques to particular algorithms devised to offer comprehension of the decision-making mechanics of AI systems.

Based on the observed XAI techniques in Table 1, the most prominent one used throughout the literature is SHapley Additive explanAtions (SHAP). As described in the work of Dwivedi et al. [2023], *"SHAP is a game-theoretic approach to explain the output of any ML model."* The way that it works for clarifying individual predictions is by ascertaining the contributory extent of each feature to the prediction. SHAP values deliver a fair attribution of the outcome to every input feature. This method is predicated on Shapley values stemming from cooperative game theory, assuring an equal distribution of all features' contributions. It is capable of generating global explanations, which show the overall impact of each feature across all predictions. In addition to this, it also has the capability of producing local explanations that show how each unique feature leads to specific decisions., which can be useful for understanding why a particular decision was made by the model for a given instance. The consistency and ability to explain complex models like deep learning networks and ensemble methods is what makes SHAP a de facto industry standard. That can be easily seen in the examined research as well, with such a large portion having implemented SHAP. Interestingly, there is an article that proposes a different technique, which uses SHAP but builds upon it. It is the conference paper by Müller et al. [2022] introducing Reconstruction Error SHapley Additive exPlanations Extension (RESHAPE) for the purpose of financial statement audits.

A descried combination of two XAI techniques joint in a paper [Sathe and Mahalle 2023] is that of SHAP and Local Interpretable Model-agnostic Explanations, known as LIME. Interpretability of local models is enabled via LIME as well. What the method does to understand the model is to vary the input data and examine the changes in predictions. Some feature values are altered for a single data point, and the influence on the output is measured as a consequence. The local surrogate model thereby offers insights into which features hold the most importance for a given prediction. LIME's model-agnostic nature means it can be utilized with any variant of a machine learning model.

Other less-sighted but still relevant techniques are Permutation Importance (PI) and Mean Decrease Impurity (MDI). They are used to interpret the contribution of each feature in a model. PI measures the decrease in a model's performance after randomly changing the value of each feature, thus indicating how important a feature is. Then, MDI is mainly used in relation to a Random Forest model and is a means to gauge how much each feature contributes to the organization of our data. The combination of the two can be seen in the paper from Carta et al. [2022b] where PI, MDI and LIME are utilized for feature selection and model explanation for a Random Forest model.

Additionally, another noteworthy publication is that from Dessain et al. [2023], which describes the innovative models for Explainable Boosting Machine (EBM) and GAMI-Net (GAMI). EBM is an interpretable model that combines the strong sides of Generalized Additive Models (GAMs) with advanced machine learning methods and constructs a model where each feature's relationship to outcome is learnt using boosting, which makes the model easy to understand. On the other hand, GAMI-Net includes additional interaction terms but still remains understandable making it possible to make more precise claims about interactions between features. Because of how they are constructed, these models balance well accuracy with explainability. A very unique part of the study goes into the costs associated with explainable models and concludes that enhancing model transparency can sometimes lead to trade-offs with model complexity and performance. Moreover, they view EBM and GAMI as promising alternatives to traditional models.

Those are among the notable and intriguing techniques found throughout the explored papers. Besides those, there are more unique ones present; in fact, apart from the unrivalled two (SHAP and LIME), there is almost no technique that can be seen more than once. This is an indication of the different kinds of approaches and interpretations of XAI. The range of methods used in the field highlights its continuous evolution and the targeted efforts to adapt explainability approaches to meet the needs of specific applications and model types. With that being the case, it is infeasible to discuss each and every single technique at length. Their variety points out the innovative environment in XAI, where novel means are consistently being actualized.

4.2 Usage of XAI in finance

In the realm of finance, the application of XAI is explored in a variety of scenarios, each trying to solve a particular need or challenge of the industry. These uses span from assessing and managing risk to optimizing stock market strategies, enhancing banking services, and beyond. A word cloud hinting at how diverse and distinct the applications are can be seen in Figure 6.

Considering how varied the applications are, it is beneficial to categorize them into broader groups that reflect their primary focus and operational context. This grouping allows for a clearer analysis of how XAI is being leveraged to address specific challenges and opportunities. This kind of classification makes the conversation flow effortlessly and emphasizes the variety of ways that XAI may be used to increase trust, efficiency, and transparency across many financial sectors. In order for a cluster to be formed, at least

Fig. 6. Application areas word cloud of reviewed papers

three papers should relate to a particular broader application area. The specified categories are credit and risk management, fraud and bankruptcy detection, stock and market analysis, customer profiling and varied. The number of papers connected to each broader classification and how they compare can be observed in Figure 7.

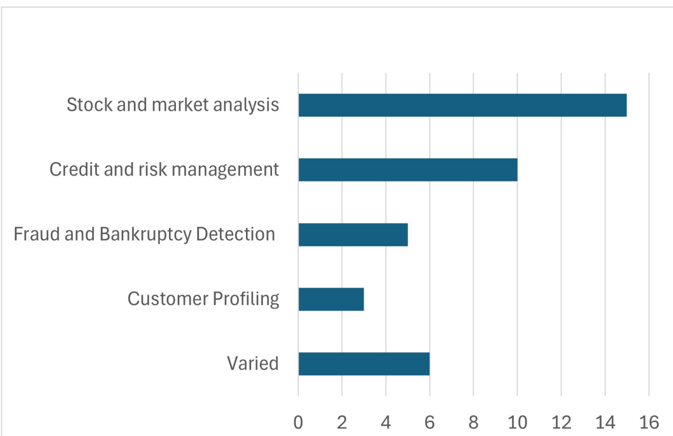

Fig. 7. Broader application categories and their respective number of papers

Significant traction can be seen in the papers discussing the utilization of XAI for stock market forecasting, market analysis and related specific applications. That is the first big cluster of articles that can be recognized. Studies like the one by Ghosh and Jana [2024] where Facebook's Prophet and NeuralProphet are combined with SHAP and Partial Dependence Plots (PDP) have demonstrated how technical indicators such as closing prices and moving averages play a critical role in forecasting clean energy investments. In parallel, the utilization of Gradient Boosting Decision Trees (GBDT) with SHAP values for predicting financial market crises underscores the superior accuracy and clarity that these models bring to identifying key predictors of large S&P 500 price drops [Ohana et al. 2021]. Moreover, through the utilization of deep learning approaches for stock price forecasting with the integration of SHAP for explaining the influence of sentiment indices, it is shown the impact textual analysis can have on financial predictions [Abdullah et al. 2024]. The diverse ways in which XAI can be employed prove the ability of those techniques to enable people to understand those complex financial models, which is crucial for stakeholders in that domain. The consistent theme throughout these studies is the improvement in model interpretability, and stakeholder trust achieved through the application of various XAI methods.

The broader classification containing the second most papers is that of credit and risk management. The literature inside this grouping mainly focuses on credit scoring and credit risk. An interesting approach which has been examined is the integration of SHAP and Multi-Objective Game-based Counterfactual Explanation (MOGCE) to create a decision support framework for credit scoring, which transparently explains rejection reasons and suggests actionable improvements for customers [Onari et al. 2024]. Another intriguing work demonstrates how the rarely sighted model of Type-2 Fuzzy Logic can be combined with evolutionary optimization for superior explainability in banking risk management [Adams and Hagras 2020]. Additionally, the investigation of domain knowledge in peer-to-peer lending, showing how it reduces reliance on incorrect AI predictions [Dikmen and Burns 2022], and the use of the AIX360 tool to analyze case studies for stakeholder-specific explanations [Karpagam et al. 2022], further illustrate the broad applicability and critical role of XAI in enhancing decision-making processes in finance.

In the group of fraud and bankruptcy detection, five scholarly articles employ XAI techniques to enhance explainability, accuracy, and reliability. All of them except one apply SHAP for their explainability goals. The paper in question is by Yohei [2021] and argues for the use of Heterogeneous Mixture Learning in the domain of fraud and risk detection over the proven SHAP model. The justification inside the work is that *"LIME and SHAP are methods that provide explainability locally to specific input data after an AI model is generated while Heterogeneous Mixture Learning provides explainability to the point where an AI model is generated."* The issue that should be pointed out with the work of Yohei [2021] is that it is incredibly concise and has no experiment or demonstration showcasing the model and its results, so the findings of that study should be taken with reservations. From the other works, a compelling work is that of Cho and Shin [2023], which introduces a novel counterfactual-based explanation method for bankruptcy prediction utilizing a Genetic Algorithm (GA) combined with SHAP to generate feature-weighted counterfactuals, providing insights into how changes in features could prevent bankruptcy.

The smallest out of the recognized broader groups is that of customer profiling, with three articles related to it. While the finance industry is not the immediate association when thinking about customer profiling, in this case, all of the works relate to finance. A notable study is that of Murindanyi et al. [2023], which predicted customer engagement in retail banking by utilizing feature engineering, deep clustering, and ensemble learning models. These models were evaluated using Evidently AI and interpreted with LIME, resulting in a predictive model that effectively leverages ensemble learning and XAI techniques. Another work creates a three-phase clustering approach for segmenting and profiling financial customers through the use of complex models and SHAP values for explanations [Choi et al. 2024]. These diverse applications of XAI in customer profiling are fascinating because they show how explainability helps us better understand and predict customer behaviour, leading to more accurate and reliable financial services.

The last cluster of papers that deserves discussion is that of the varied type of application areas. It contains five articles which did not really align with any of the other groupings and were not enough to form a new one. Two of them are related to transaction classification with the one by Kotios et al. [2022] combining CatBoost and DeepAR with SHAP and LIME in order to illustrate how rule-based and machine learning techniques can work together to deliver precise and interpretable banking services. Another notable study comes from Sun et al. [2024] and introduces a GraphSAGE and DRL coupled model (GRL) for financial portfolio optimization where SHAP is used for feature selection. The outcome of the work is an advanced model that can both optimize portfolios and ensure the interpretability of their decisions.

In summary, there are so many diverse usages of XAI in finance that are meant for specific purposes. Even though we were successful in grouping them into larger areas of applications, they are still varied in nature. The ability to combine XAI techniques with so many intriguing and different models exemplifies how endless the possibilities are. The massively diverse application areas are only another testament to how young the field is and how much more potential it contains.

5 CONCLUSIONS

The exploration of Explainable Artificial Intelligence (XAI) in the financial sector covered in this work emphasizes the significant development and various uses of XAI. The techniques remain indispensable to improving transparency and responsibility and assuring trust within AI-based financial systems. The results indicate that XAI methods are not simply beneficial but necessary in a field like finance, where high-stakes decisions have massive impact.

Various finance domains have extensively applied XAI, such as credit risk assessment, fraud detection, stock market forecasting, and customer profiling, which were reported by this research. XAI has the potential to be significant in credit risk assessment for providing clear explanations and rationalizations of credit decisions, fostering trust, and compliance with regulations. XAI can help improve fraud detection through transparency into the determinants of various fraudulent activities, enhancing the reliability and efficacy of measures to prevent fraud. XAI techniques in stock market forecasting explain predictions produced by complex models and enable stakeholders to understand what drives market trends as well as make sound investment choices. Customer profiling utilizes XAI to explain customer segmentation in detail and forecast their behaviour so that financial services providers can adjust their operations accordingly and enhance satisfaction among clients.

The transparency and trust brought by XAI enhance each application area's outcomes by addressing problems that are unique to the financial industry. SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were revealed as the most commonly used XAI techniques because of their capacity to give global and local explanations. These methodologies improve the explainability of most complex models like deep learning networks and ensemble methods, thus making them more understandable.

Moreover, the study has shed light on some new techniques that fuse multiple XAI technologies to increase explainability. These include approaches that demonstrate how dynamic and developing XAI methodologies are by now being tailored to particular tasks in order to satisfy various requirements of the finance domain. An example of this is the combination of SHAP with Multi-Objective Game-based Counterfactual Explanation (MOGCE) for feature selection and model explanation [Onari et al. 2024]. Hybridization in XAI depicts a deepened understanding of AI structures and signifies a resolve to move XAI forward in response to the ever-changing financial environments.

This research emphasizes the importance of XAI in changing the financial sector by creating more explainable and responsible AI systems. The findings from the conducted systematic literature review build a solid basis for future research that will address current limitations and further develop XAI in finance. The ongoing evolution and application of XAI approaches are important for guaranteeing the efficiency, as well as responsibility and fairness of AI systems in finance.

5.1 Limitations

Despite the significant progress, there are several limitations and areas for future research that must be addressed. An aspect of this research which can be considered a confining factor is restricting the use of multiple databases of scholarly works and utilizing only Scopus. While this was done for a reason, and that was to ensure the quality of all papers explored in the systematic literature review, it could have proved beneficial to explore articles available elsewhere as well. Another limitation of this work is the inclusion and exclusion criteria that were applied. They were constructed in such a way as to narrow down the results of papers, and for reasons similar to the previous point, it could have been advantageous to incorporate more works from the initial query. Additionally, a limiting factor that was encountered during the course of the research was the paywall behind some articles. Even though they were present in Scopus, it was not possible to download them from there. It seems that the University of Twente institution access did not give permission to download those works. A final limitation was not exploring how the two keywords' word clouds would have formed if the keywords used in the search query were removed from the visualisations.

5.2 Future Research

A challenge that was noticed during the examination of all the relevant literature is the scalability and integration of XAI techniques into large-scale real-time financial systems. There is much room for further exploration in this area, which can certainly make XAI advance to a higher stage. Part of future research should concentrate on coming up with scalable XAI solutions that can be integrated with existing financial infrastructures. This area warrants further investigation.

There exist many opportunities for further research into the impact of XAI on decision-making processes in finance. It will be beneficial to see more development related to the stakeholders for whom explanations are meant. Understanding how explanations influence user trust, reliance, and overall task performance in AI-assisted financial decision-making contexts is essential, but it is underexplored. From the examined research, there are only two articles [Golbin et al. 2019; Karpagam et al. 2022] investigating this area. With this in mind, projects such as the one on Audience-dependent explanations[3] should be developed and serve beneficially. This information will help refine XAI approaches to better cater for users' requirements.

REFERENCES

Mohammad Abdullah, Zunaidah Sulong, and Mohammad Ashraful Ferdous Chowdhury. 2024. Explainable deep learning model for stock price forecasting using textual analysis. *Expert Systems with Applications* 249 (2024). https://doi.org/10.1016/j.eswa.2024.123740 Cited by: 2.

Janet Adams and Hani Hagras. 2020. A type-2 fuzzy logic approach to explainable ai for regulatory compliance, fair customer outcomes and market stability in the global financial sector. *IEEE International Conference on Fuzzy Systems* 2020-July (2020). https://doi.org/10.1109/FUZZ48607.2020.9177542 Cited by: 19.

Alessandra Amato, Joerg R. Osterrieder, and Marcos R. Machado. 2024. How can artificial intelligence help customer intelligence for credit portfolio management? A systematic literature review. *International Journal of Information Management Data Insights* 4, 2 (2024), 100234. https://doi.org/10.1016/j.jjimei.2024.100234

Cecilia Avila-Garzon. 2020. Applications, Methodologies, and Technologies for Linked Open Data: A Systematic Literature Review. *International Journal on Semantic Web and Information Systems (IJSWIS)* 16, 3 (2020), 53–69. https://doi.org/10.4018/IJSWIS.2020070104

Tomisin Awosika, Raj Mani Shukla, and Bernardi Pranggono. 2024. Transparency and Privacy: The Role of Explainable AI and Federated Learning in Financial Fraud Detection. *IEEE Access* 12 (2024), 64551 – 64560. https://doi.org/10.1109/ACCESS.2024.3394528 Cited by: 0.

Jeroen Baas, Michiel Schotten, Andrew Plume, Grégoire Côté, and Reza Karimi. 2020. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies* 1, 1 (02 2020), 377–386. https://doi.org/10.1162/qss_a_00019 arXiv:https://direct.mit.edu/qss/article-pdf/1/1/377/1760882/qss_a_00019.pdf

Harit Bandi, Suyash Joshi, Siddhant Bhagat, and Dayanand Ambawade. 2021. Integrated Technical and Sentiment Analysis Tool for Market Index Movement Prediction, comprehensible using XAI. *Proceedings - International Conference on Communication, Information and Computing Technology, ICCICT 2021* (2021). https://doi.org/10.1109/ICCICT50803.2021.9510124 Cited by: 5.

Jonathan Boardman, Md Shafiul Alam, Xiao Huang, and Ying Xie. 2022. Integrated Gradients is a Nonlinear Generalization of the Industry Standard Approach to Variable Attribution for Credit Risk Models. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022* (2022), 5012 – 5023. https://doi.org/10.1109/BigData55660.2022.10020687 Cited by: 1.

Salvatore Carta, Sergio Consoli, Alessandro Sebastian Podda, Diego Reforgiato Recupero, and Maria Madalina Stanciu. 2022a. Statistical arbitrage powered by Explainable Artificial Intelligence. *Expert Systems with Applications* 206 (2022). https://doi.org/10.1016/j.eswa.2022.117763 Cited by: 6; All Open Access, Hybrid Gold Open Access.

S. Carta, A.S. Podda, D. Reforgiato Recupero, and M.M. Stanciu. 2022b. Explainable AI for Financial Forecasting. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 13164 LNCS (2022), 51–69. https://doi.org/10.1007/978-3-030-95470-3_5 cited By 5.

Federico Maria Cau, Hanna Hauptmann, Lucio Davide Spano, and Nava Tintarev. 2023. Supporting High-Uncertainty Decisions through AI and Logic-Style Explanations. *International Conference on Intelligent User Interfaces, Proceedings IUI* (2023), 251 – 263. https://doi.org/10.1145/3581641.3584080 Cited by: 5; All Open Access, Bronze Open Access, Green Open Access.

Soo Hyun Cho and Kyung-shik Shin. 2023. Feature-Weighted Counterfactual-Based Explanation for Bankruptcy Prediction. *Expert Systems with Applications* 216 (2023). https://doi.org/10.1016/j.eswa.2022.119390 Cited by: 4.

Insu Choi, Woosung Koh, Bonwoo Koo, and Woo Chang Kim. 2024. Network-based exploratory data analysis and explainable three-stage deep clustering for financial customer profiling. *Engineering Applications of Artificial Intelligence* 128 (2024). https://doi.org/10.1016/j.engappai.2023.107378 Cited by: 1.

COPE Council. 2019. COPE Guidelines: Retraction Guidelines. https://doi.org/10.24318/cope.2019.1.4 Accessed: 2024-05-18.

[3] https://www.digital-finance-msca.com/audience-dependent-explanations

Helen Crompton and Diane Burke. 2023. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education* 20 (04 2023). https://doi.org/10.1186/s41239-023-00392-8

Yiqi Deng, Yuzhi Liang, and Siu-Ming Yiu. 2024. Towards interpretable stock trend prediction through causal inference. *Expert Systems with Applications* 238 (2024). https://doi.org/10.1016/j.eswa.2023.121654 Cited by: 1.

Sahil Deo and Neha Sontakke. 2021. User-Centric Explainability in Fintech Applications. *Communications in Computer and Information Science* 1420 (2021), 481 – 488. https://doi.org/10.1007/978-3-030-78642-7_64 Cited by: 2.

Jean Dessain, Nora Bentaleb, and Fabien Vinas. 2023. Cost of Explainability in AI: An Example with Credit Scoring Models. *Communications in Computer and Information Science* 1901 CCIS (2023), 498 – 516. https://doi.org/10.1007/978-3-031-44064-9_26 Cited by: 2; All Open Access, Hybrid Gold Open Access.

Murat Dikmen and Catherine Burns. 2022. The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human Computer Studies* 162 (2022). https://doi.org/10.1016/j.ijhcs.2022.102792 Cited by: 27; All Open Access, Hybrid Gold Open Access.

Rudresh Dwivedi, Devam Dave, Het Naik, Smiti Singhal, Rana Omer, Pankesh Patel, Bin Qian, Zhenyu Wen, Tejal Shah, Graham Morgan, and Rajiv Ranjan. 2023. Explainable AI (XAI): Core Ideas, Techniques, and Solutions. *ACM Comput. Surv.* 55, 9, Article 194 (jan 2023), 33 pages. https://doi.org/10.1145/3561048

Mohammad El Hajj and Jamil Hammoud. 2023. Unveiling the Influence of Artificial Intelligence and Machine Learning on Financial Markets: A Comprehensive Analysis of AI Applications in Trading, Risk Management, and Financial Operations. *Journal of Risk and Financial Management* 16, 10 (2023). https://doi.org/10.3390/jrfm16100434

Jacopo Fior, Luca Cagliero, and Paolo Garza. 2022. Leveraging Explainable AI to Support Cryptocurrency Investors. *Future Internet* 14, 9 (2022). https://doi.org/10.3390/fi14090251 Cited by: 2; All Open Access, Gold Open Access.

Masaru Fuji, Katsuhito Nakazawa, and Hiroaki Yoshida. 2020. "Trustworthy and explainable AI" achieved through knowledge graphs and social implementation. *Fujitsu Scientific and Technical Journal* 56, 1 (2020), 39 – 45. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086072973&partnerID=40&md5=1a012d8912d5becf51265279e6481214 Cited by: 5.

Indranil Ghosh, Esteban Alfaro-Cortés, Matías Gámez, and Noelia García-Rubio. 2023. Role of proliferation COVID-19 media chatter in predicting Indian stock market: Integrated framework of nonlinear feature transformation and advanced AI. *Expert Systems with Applications* 219 (2023). https://doi.org/10.1016/j.eswa.2023.119695 Cited by: 8; All Open Access, Hybrid Gold Open Access.

Indranil Ghosh and Rabin K. Jana. 2024. Clean energy stock price forecasting and response to macroeconomic variables: A novel framework using Facebook's Prophet, NeuralProphet and explainable AI. *Technological Forecasting and Social Change* 200 (2024). https://doi.org/10.1016/j.techfore.2023.123148 Cited by: 2.

Ilana Golbin, Kyungha Kay Lim, and Divyanshi Galla. 2019. Curating explanations of machine learning models for business stakeholders. *Proceedings - 2019 2nd International Conference on Artificial Intelligence for Industries, AI4I 2019* (2019), 44 – 49. https://doi.org/10.1109/AI4I46381.2019.00019 Cited by: 2.

G.R. Karpagam, Aditya Varma, M. Samrddhi, and V. Shri Shivathmika. 2022. Understanding, Visualizing and Explaining XAI Through Case Studies. *8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022* (2022), 647 – 654. https://doi.org/10.1109/ICACCS54159.2022.9785199 Cited by: 0.

Dimitrios Kotios, Georgios Makridis, Georgios Fatouros, and Dimosthenis Kyriazis. 2022. Deep learning enhancing banking services: a hybrid transaction classification and cash flow prediction approach. *Journal of Big Data* 9, 1 (2022). https://doi.org/10.1186/s40537-022-00651-x Cited by: 10; All Open Access, Gold Open Access.

Akshat Mahajan and Kaushal Kumar Shukla. 2023. Analyzing False Positives in Bankruptcy Prediction with Explainable AI. *2023 International Conference on Artificial Intelligence and Applications, ICAIA 2023 and Alliance Technology Conference, ATCON-1 2023 - Proceeding* (2023). https://doi.org/10.1109/ICAIA57370.2023.10169390 Cited by: 0.

Charl Maree, Jan Erik Modal, and Christian W. Omlin. 2020. Towards Responsible AI for Financial Transactions. *2020 IEEE Symposium Series on Computational Intelligence, SSCI 2020* (2020), 16 – 21. https://doi.org/10.1109/SSCI47803.2020.9308456 Cited by: 12; All Open Access, Green Open Access.

Charl Maree and Christian W. Omlin. 2022. Understanding Spending Behavior: Recurrent Neural Network Explanation and Interpretation. *2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics, CIFEr 2022 - Proceedings* (2022). https://doi.org/10.1109/CIFEr52523.2022.9776210 Cited by: 1; All Open Access, Green Open Access.

Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. 2017. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics* 19, 6 (05 2017), 1236–1246. https://doi.org/10.1093/bib/bbx044 arXiv:https://academic.oup.com/bib/article-pdf/19/6/1236/27119191/bbx044.pdf

Sudi Murindanyi, Margaret Nagwovuma, Barbara Nansamba, and Ggaliwango Marvin. 2023. Explainable Ensemble Learning and Trustworthy Open AI for Customer Engagement Prediction in Retail Banking. *ACM International Conference Proceeding Series* (2023), 198 – 206. https://doi.org/10.1145/3607947.3607983 Cited by: 0; All

Open Access, Hybrid Gold Open Access.
Ricardo Müller, Marco Schreyer, Timur Sattarov, and Damian Borth. 2022. RESHAPE: Explaining Accounting Anomalies in Financial Statement Audits by enhancing SHapley Additive exPlanations. *Proceedings of the 3rd ACM International Conference on AI in Finance, ICAIF 2022* (2022), 174 – 182. https://doi.org/10.1145/3533271.3561667 Cited by: 3; All Open Access, Green Open Access, Hybrid Gold Open Access.
Jean Jacques Ohana, Steve Ohana, Eric Benhamou, David Saltiel, and Beatrice Guez. 2021. Explainable AI (XAI) Models Applied to the Multi-agent Environment of Financial Markets. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12688 LNAI (2021), 189 – 207. https://doi.org/10.1007/978-3-030-82017-6_12 Cited by: 27.
Mohsen Abbaspour Onari, Mustafa Jahangoshai Rezaee, Morteza Saberi, and Marco S. Nobile. 2024. An explainable data-driven decision support framework for strategic customer development. *Knowledge-Based Systems* 295 (2024). https://doi.org/10.1016/j.knosys.2024.111761 Cited by: 0.
Keane Ong, Wihan Van Der Heever, Ranjan Satapathy, Erik Cambria, and Gianmarco Mengaldo. 2023. FinXABSA: Explainable Finance through Aspect-Based Sentiment Analysis. *IEEE International Conference on Data Mining Workshops, ICDMW* (2023), 773 – 782. https://doi.org/10.1109/ICDMW60847.2023.00105 Cited by: 1; All Open Access, Green Open Access.
Sangjin Park and Jae-Suk Yang. 2022. Interpretable deep learning LSTM model for intelligent economic decision-making. *Knowledge-Based Systems* 248 (2022). https://doi.org/10.1016/j.knosys.2022.108907 Cited by: 26.
Carmen Parra Rodriguez. 2022. Ethical principles in the use of Artificial Intelligence in the financial sector from a European perspective. *Studia Prawnicze KUL* 1 (mar. 2022), 199–221. https://doi.org/10.31743/sp.13029
N. Pavitha and Shounak Sugave. 2024. Explainable Multistage Ensemble 1D Convolutional Neural Network for Trust Worthy Credit Decision. *International Journal of Advanced Computer Science and Applications* 15, 2 (2024), 351 – 358. https://doi.org/10.14569/IJACSA.2024.0150237 Cited by: 0; All Open Access, Gold Open Access.
Varsha P S, Amrita Chakraborty, and Arpan Kumar Kar. 2024. How to Undertake an Impactful Literature Review: Understanding Review Approaches and Guidelines for High-impact Systematic Literature Reviews. *South Asian Journal of Business and Management Cases* 13, 1 (2024), 18–35. https://doi.org/10.1177/22779779241227654 arXiv:https://doi.org/10.1177/22779779241227654
Jayakrishnan S Nair. 2022. Artificial Intelligence (AI) in Retailing- A Systematic Review and Research Agenda. *SSRN Electronic Journal* (01 2022). https://doi.org/10.2139/ssrn.4134959
S.S. Sathe and P. Mahalle. 2023. Predictive Analytics in Financial Services Using Explainable AI. *Lecture Notes in Networks and Systems* 641 LNNS (2023), 431–444. https://doi.org/10.1007/978-981-99-0483-9_35 cited By 1.
Qiguo Sun, Xueying Wei, and Xibei Yang. 2024. GraphSAGE with deep reinforcement learning for financial portfolio optimization. *Expert Systems with Applications* 238 (2024). https://doi.org/10.1016/j.eswa.2023.122027 Cited by: 3.
Hubert Szczepaniuk and Edyta Szczepaniuk. 2022. Applications of Artificial Intelligence Algorithms in the Energy Sector. *Energies* 16 (12 2022), 347. https://doi.org/10.3390/en16010347
Xue Wen Tan and Stanley Kok. 2023. Explainable Risk Classification in Financial Reports. *International Conference on Information Systems, ICIS 2023: "Rising like a Phoenix: Emerging from the Pandemic and Reshaping Human Endeavors with Digital Technologies"* (2023). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192547934&partnerID=40&md5=e16c68d2b5a998aa52943560aa482465 Cited by: 0.
Tamara Teplova, Tatiana Sokolova, and David Kissa. 2023. Revealing stock liquidity determinants by means of explainable AI: The role of ESG before and during the COVID-19 pandemic. *Resources Policy* 86 (2023). https://doi.org/10.1016/j.resourpol.2023.104253 Cited by: 1.
Ana Todorovska, Hristijan Peshov, Ivan Rusevski, Irena Vodenska, Lubomir T. Chitkushev, and Dimitar Trajanov. 2023. Using ML and Explainable AI to understand the interdependency networks between classical economic indicators and crypto-markets. *Physica A: Statistical Mechanics and its Applications* 624 (2023). https://doi.org/10.1016/j.physa.2023.128900 Cited by: 3.
Kim Long Tran, Hoang Anh Le, Thanh Hien Nguyen, and Duc Trung Nguyen. 2022. Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam. *Data* 7, 11 (2022). https://doi.org/10.3390/data7110160 Cited by: 16; All Open Access, Gold Open Access.
Lara Waltersmann, Steffen Kiemel, Julian Stuhlsatz, Alexander Sauer, and Robert Miehe. 2021. Artificial Intelligence Applications for Increasing Resource Efficiency in Manufacturing Companies—A Comprehensive Review. *Sustainability* 13 (06 2021), 6689. https://doi.org/10.3390/su13126689
Sugiyama Yohei. 2021. AI-Based Fraud and Risk Detection Service Ensures Transparency While Improving Operational Efficiency and Performance. *NEC Technical Journal* 15, 1 (2021), 128 – 131. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85168117581&partnerID=40&md5=c29916f2b7d5943696545bc8f1266020 Cited by: 0.
Taha Buğra Çelik, Özgür İcan, and Elif Bulut. 2023. Extending machine learning prediction capabilities by explainable AI in financial time series prediction[Formula presented]. *Applied Soft Computing* 132 (2023). https://doi.org/10.1016/j.asoc.2022.109876 Cited by: 12.

APPENDIX

Table 1. Table reporting all the articles that have been examined to conduct the research

Study	Objective	Methodology	XAI Techniques	Application Area	Main Contributions
[Mahajan and Shukla 2023]	Analyze false positives in bankruptcy prediction using explainable AI techniques	Utilize SHAP and TreeSHAP for post-hoc analysis on bankruptcy datasets, employing various models; feature importance	SHAP and TreeSHAP	Bankruptcy prediction	Provided in-depth analysis of false positives using SHAP, highlighting differences in feature importance and demonstrating class weights for handling imbalanced datasets.
[Karpagam et al. 2022]	Achieve Responsible AI through XAI and provide explainability from various stakeholder viewpoints	AIX360 tool to analyze two case studies	Protodash Explainer, Contrastive Explanations, Class Activation Mapping (CAM) within AIX360 framework	Credit Scoring	Conceptual model for XAI-based predictive analysis, demonstrated with case studies in finance and healthcare using AIX360
[Ong et al. 2023]	To provide explainability in financial analysis	Combined Sentic-GCN model for ABSA with statistical techniques on social media data and stock prices	Sentic-GCN with Granger causality, Uncertainty coefficient, Pearson correlation	Financial analysis, Stock price prediction	Novel framework to derive explainable relationships between sentiment and stock prices.
[Abdullah et al. 2024]	Forecast stock prices using text analysis and deep learning; explain the models using explainable AI	Developed World Halal Tourism Composite Sentiment Index (WHTCSI) using text analysis. Utilized ARIMAX, SVR, MLP, CNN, and LSTM for forecasting and SHAP for explainability	SHAP	Stock price forecasting	Provided explainability using SHAP, showing significant contribution of sentiment index to the model
[Tan and Kok 2023]	Assess post-event return volatility risk using an explainable deep-learning model on 10-K financial reports	Developed FinBERT-XRC, used attention mechanisms for explanations at word, sentence, and corpus levels	Attention mechanism in FinBERT-XRC	Risk classification in financial reports	FinBERT-XRC for multi-level explanations
[Pavitha and Sugave 2024]	Develop an explainable deep learning model for credit risk prediction that balances high accuracy and interpretability	Utilized multistage ensemble technique with 1D CNN, incorporating ARSLRP for explanations	Adaptive Relevance Scaling for Layer-wise Relevance Propagation (ARSLRP)	Credit risk prediction	Achieved superior performance metrics compared to state-of-the-art models and provided multi-level explanations.
[Yohei 2021]	Develop an AI-based fraud and risk detection service that enhances transparency, efficiency, and performance	Utilized NEC's Heterogeneous Mixture Learning technology, incorporating visualization techniques for explainability	Heterogeneous Mixture Learning	Fraud and risk detection	Developed a fraud detection service that provides detailed explanations.
[Murindanyi et al. 2023]	Predict customer engagement in retail banking	Used a dataset with feature engineering, deep clustering, and ensemble learning models, evaluated using Evidently AI and interpreted with LIME.	LIME	Customer engagement prediction in retail banking	Developed a predictive model using ensemble learning and XAI techniques
[Onari et al. 2024]	Propose a framework that helps rejected customers improve their applications	Integrates SHAP and Multi-Objective Game-based Counterfactual Explanation (MOGCE) to provide a transparent and actionable framework for improving credit application outcomes	SHAP, Multi-Objective Game-based Counterfactual Explanation (MOGCE)	Credit scoring and customer development	Developed a decision support framework combining SHAP and MOGCE to explain rejection reasons and provide actionable modifications
[Awosika et al. 2024]	Develop a fraud detection system using Federated Learning (FL) combined with XAI	Utilized Federated Learning to train models on local devices with aggregated updates, integrated SHAP for explainability	SHAP	Financial fraud detection	Developed a privacy-preserving fraud detection system combining FL and XAI
[Teplova et al. 2023]	Analyze the impact of ESG factors on stock liquidity in the Russian market before and during the COVID-19 pandemic	Applied a three-stage approach: principal component analysis for liquidity indices, neural networks with dense layers, and Shapley values for interpretation	SHAP	Stock liquidity	Identified key ESG determinants affecting stock liquidity
[Maree and Omlin 2022]	Explain and interpret segmentation of customers	Trained RNNs on financial transactions, used linear regression for explanation, and inverse regression for interpretation	Linear Regression, Inverse Regression	Customer segmentation	Developed a method for symbolic explanation and interpretation of RNNs, segmenting customers based on spending behaviour and personality traits

[Boardman et al. 2022]	Introduce Integrated Gradients as a method for variable attribution in nonlinear models	Compared logistic regression with Integrated Gradients for neural networks	Integrated Gradients	Credit risk modeling	Demonstrated that Integrated Gradients can generalize the industry standard variable attribution for linear models to nonlinear neural networks, satisfying regulatory requirements for explainability
[Deng et al. 2024]	Enhance the interpretability of stock trend prediction models	Combined causal inference with a hierarchical attention network (HAN) for stock trend prediction. Introduced DWF-KST for keyword extraction	Causal inference, Hierarchical Attention Network (HAN)	Stock trend prediction	Proposed a novel framework that integrates causal graphs with deep learning models, providing explanation for stock trends
[Sathe and Mahalle 2023]	Enhance predictive analytics in financial services using XAI	Applied machine learning models, integrated with explainability techniques	SHAP, LIME	Financial services	Developed a framework combining predictive models with XAI techniques
[Choi et al. 2024]	Develop an explainable, three-stage clustering framework for effective financial customer segmentation and profiling	Utilized deep neural network-based autoencoders for dimension reduction, followed by non-neural network-based algorithms for further reduction, and K-means clustering. SHAP values used for explanation	SHAP	Financial customer profiling	Novel network-based visualization for high-dimensional data, demonstrating the practical application.
[Golbin et al. 2019]	Create a framework for curating explanations of machine learning models tailored for different business stakeholders	Used FICO HELOC dataset to develop and evaluate explanation methods for various stakeholders	SHAP, Bayesian Rule Lists (BRL), Logistic Regression	Financial risk prediction	Proposed a stakeholder-specific framework for model explainability, combining local and global explanations to meet diverse needs, demonstrating application on financial risk prediction
[Fior et al. 2022]	Develop a visual analytics tool to explain AI-based cryptocurrency trading systems	Combined SHAP with a visual analytics tool (CryptoMLE) to explain the impact of various features on AI predictions for cryptocurrency prices	SHAP	Cryptocurrency trading	Developed CryptoMLE, an interactive dashboard to visualize and monitor the influence of features on cryptocurrency price predictions
[Ghosh and Jana 2024]	Investigate the predictability of clean energy investment in the US market	Utilized Facebook's Prophet and NeuralProphet for forecasting, with SHAP and Partial Dependence Plots (PDP) for explainability.	SHAP, Partial Dependence Plots (PDP)	Stock forecasting	Developed a framework combining Prophet and NeuralProphet for accurate predictions and enhancing model interpretability through SHAP and PDP
[Deo and Sontakke 2021]	Analyze the effectiveness of user-centric explanations in conveying the decision-making logic of complex algorithmic systems in fintech applications	Conducted a user study using a custom-built Robo Advisor (RA) equipped with explanation strategies	SHAP, LIME	Fintech applications	Demonstrated that explanations increase user trust and comprehension
[Todorovska et al. 2023]	Analyze the relationships between cryptocurrencies and traditional financial markets	Developed an Explainable ML model combining XGBoost and SHAP to analyze dependencies between cryptocurrencies and classical economic indicators	SHAP	Financial markets, Cryptocurrency markets	Created a methodology that combines multimodal data from traditional financial markets and cryptocurrencies
[Dessain et al. 2023]	Analyze the cost implications of implementing explainability in AI models, specifically in credit scoring	Conducted a comparative analysis of interpretable models and black-box models using credit scoring datasets	Logistic Regression (LR), Explainable Boosting machine (EBM), Gami-net (GAMI)	Credit scoring	Demonstrated that while interpretable models incur higher computational costs, they ensure regulatory compliance and enhance trust, providing a cost-benefit analysis
[Sun et al. 2024]	Improve Proximal Policy Optimization (PPO) in portfolio optimization by integrating GraphSAGE-based feature extraction and SHAP for feature selection	Combined GraphSAGE with PPO to create a GRL model, using SHAP for feature selection	SHAP	Financial portfolio optimization	Developed a novel GRL model incorporating GraphSAGE for feature extraction and SHAP for feature selection
[Müller et al. 2022]	Develop a method for generating attribute-level explanations for autoencoder neural networks (AENNs) used in detecting accounting anomalies	Introduced RESHAPE, which builds upon SHAP	Reconstruction Error SHapley Additive exPlanations Extension (RESHAPE)	Financial statement audits	Developed RESHAPE for attribute-level explanations of AENN outputs, providing an evaluation framework for benchmarking XAI methods in auditing
[Cau et al. 2023]	Explore how different types of logic-style explanations and AI confidence levels affect decision-making quality and user trust in financial trading	Conducted user experiments to compare the effectiveness of abductive, deductive, and counterfactual explanations across various confidence levels	k-NN algorithm, SHAP, Collection of High Importance Random Path Snippets (CHIRPS)	Financial trading	Investigated how AI confidence and explanation styles affect information reliance, task performance, and agreement with AI, finding that high confidence and abductive/deductive explanations improve outcomes

[Bandi et al. 2021]	Develop a tool for predicting market index movements using technical and sentiment analysis, enhanced by explainable AI	Combined technical analysis (MACD, RSI, EMA) with sentiment analysis of news, applying LIME for sentiment prediction	LIME	Market index prediction	Developed a novel system and provided a dashboard with explainable recommendations
[Cho and Shin 2023]	Develop a counterfactual-based explanation method for bankruptcy prediction	Utilized a genetic algorithm (GA) to generate counterfactual explanations, incorporating feature importance derived from SHAP values	SHAP, Genetic Algorithm (GA)	Bankruptcy prediction	Introduced a GA-based method for generating feature-weighted counterfactuals
[Fuji et al. 2020]	Develop trustworthy and explainable AI to provide the basis for AI inferences	Employed knowledge graphs to represent expert knowledge, combined with AI inferences	Knowledge Graphs	Loan application	Introduced an approach using knowledge graphs to enhance the trustworthiness and explainability of AI, demonstrating application in finance and chemistry.
[Carta et al. 2022a]	Enhance statistical arbitrage trading strategies by integrating machine learning with XAI	Developed an ML model for stock returns prediction, integrated with feature selection methods using permutation importance scores to identify relevant features	Permutation Importance (PI)	Stock market trading	Integrated ML and XAI techniques to improve prediction performance and financial returns, demonstrating enhanced performance
[Carta et al. 2022b]	Improve the accuracy and interpretability of financial forecasting models by applying explainable AI techniques to feature selection	Used Random Forest models with lagged returns as features for stock return prediction, employing PI, MDI, and LIME for feature selection and model explanation	Permutation Importance (PI), Mean Decrease Impurity (MDI), LIME	Stock market trading	Developed a feature selection strategy using PI, MDI, and LIME to enhance prediction accuracy and interpretability by removing uninformative features to improve the model's performance and computational efficiency
[Ghosh et al. 2023]	Assess the impact of COVID-19 media chatter on predicting stock prices using AI	Developed two hybrid predictive models and XAI techniques to interpret results.	SHAP, Partial Dependence Plots (PDP)	Stock market prediction	Demonstrated significant influence of media chatter on stock prices, providing interpretable insights
[Kotios et al. 2022]	Develop a hybrid model for transaction classification and cash flow prediction for SMEs	Utilized CatBoost for transaction classification and DeepAR for cash flow prediction, applying SHAP and LIME for model interpretability	SHAP, LIME	Banking services, Transaction classification	Developed a hybrid model combining rule-based and machine learning approaches for accurate transaction classification, ensuring explainability with SHAP and LIME
[Çelik et al. 2023]	Improve the reliability and accuracy of financial time series predictions	Developed an ensemble model integrating EMD with ANNR and ANNC, followed by random forest (RF) and XGBoost classifiers, applying LIME for model interpretability	LIME	Financial time series prediction, stock market prediction	Proposed a hybrid model EMD-ANN-RF, integrating LIME
[Maree et al. 2020]	Develop an explainable AI model for financial transaction classification	Used a deep neural network with word2vec encoding for transaction text, combined with SHAP for feature importance and DBSCAN for text clustering, evaluated robustness with adversarial attacks	SHAP, DBSCAN, Decision Trees	Financial transaction classification	Developed a model integrating SHAP for feature importance and DBSCAN for clustering
[Tran et al. 2022]	Enhance the predictability of financial distress prediction models	Applied machine learning models (XGBoost, Random Forest) to financial data, utilized SMOTE for data balancing, and employed SHAP values for explainability	SHAP	Financial distress prediction	Showcased the performance of XGBoost and Random Forest, and identified key financial ratios impacting predictions
[Adams and Hagras 2020]	Propose an AI risk management banking solution, ensuring explainability and regulatory compliance	Evaluated Type-2 Fuzzy Logic with evolutionary optimization against Neural Networks and Logistic Regression across nine banking use cases	Type-2 Fuzzy Logic	Banking, Risk Management	Demonstrated the superior explainability of Type-2 Fuzzy Logic compared to Neural Networks and Logistic Regression, while maintaining competitive accuracy
[Ohana et al. 2021]	Improve the prediction of financial market crises and enhance the interpretability of machine learning models in financial markets	Used Gradient Boosting Decision Trees (GBDT) to predict large S&P 500 price drops from 150 features, applying SHAP values to identify important variables and provide local explanations	SHAP	Financial market prediction, crisis detection	Demonstrated the superior accuracy of GBDT over other ML methods in predicting market crises and highlighted the use of SHAP for identifying key predictors and providing local explanations
[Park and Yang 2022]	Improve economic prediction and decision-making using LSTM and SHAP for better interpretability	Developed LSTM model to predict GDP growth rates and crises for G20 countries, utilizing SHAP for interpretability	SHAP	Economic prediction, Financial forecasting	Introduced an interpretable LSTM model, predicting economic outcomes and identified key economic factors impacting growth and crises
[Dikmen and Burns 2022]	Investigate how domain knowledge affects trust, reliance, and task performance in AI-assisted decision-making	Developed a P2P lending simulator with a gradient boosting classifier (CatBoost) to predict default risk and used SHAP for local explanations and domain knowledge-based decision aids for non-expert participants	SHAP	Peer-to-peer lending, financial investment	Demonstrated that providing domain knowledge reduces reliance on incorrect AI predictions