Extending application of explainable artificial intelligence for managers in financial organizations

Sabharwal, Renu, Miah, Shah Jahan, Wamba, Samuel Fosso and Cook, Peter (2024) Extending application of explainable artificial intelligence for managers in financial organizations. Annals of Operations Research. ISSN 0254-5330

Annals of Operations Research
https://doi.org/10.1007/s10479-024-05825-9

ORIGINAL RESEARCH

Extending application of explainable artificial intelligence for managers in financial organizations

Renu Sabharwal[1] · Shah J. Miah[1] · Samuel Fosso Wamba[2] · Peter Cook[3]

Received: 27 March 2023 / Accepted: 1 January 2024

Abstract
Anomalies are a significant challenge for businesses in all industries. Artificial intelligence (AI) based machine learning (ML) detection models can help find aberrant customer transaction behaviour in financial datasets. However, the output responses provided by these AI-based ML models lack transparency and interpretability, making it challenging for financial managers to comprehend the reasoning underlying the AI detections. Suppose managers cannot comprehend how and why AI models develop responses based on the input information. In such cases, AI is unlikely to enhance data-driven decision-making and add value to organizations. This article's primary objective is to illustrate the capacity of the SHapley Additive exPlanations (SHAP) technique to give finance managers an intuitive explanation of the anomaly detections AI-based ML models generate for a specific customer transaction dataset. Theoretically, we contribute to the literature on international finance by offering a conceptual review of AI algorithmic explainability. We discuss its implications for sustaining a competitive advantage using the concepts of action design research methodology following the research onion framework. We also suggest an explainable AI implementation methodology based on SHAP as a valuable guide for finance managers seeking to boost the transparency of AI-based ML models and to alleviate trust difficulties in data-driven decision-making.

Keywords Artificial intelligence · Machine learning · Financial organization · Supervised learning · Unsupervised learning

1 Introduction

In recent years Artificial Intelligence (AI) applications have been widely studied in the operational research domain. AI is currently the most common disruptive technology capable of transforming business operations. Previous AI research has established various related

✉ Shah J. Miah
shah.miah@newcastle.edu.au

[1] Newcastle Business School, University of Newcastle, Callaghan, NSW, Australia

[2] TBS Business School, Toulouse, France

[3] Novatti Group, Melbourne, Australia

Published online: 30 January 2024

schools of thought on such disruption, such as using AI for enhancing business innovation (Wamba-Taguimdje et al., 2020), re-engineering of business processes (Al-Anqoudi et al., 2021), analysing customer requirements (Zhou et al., 2020), and using data analytics for improved and precise managerial decision support (Gupta et al., 2022). While many studies have indicated significant as well as potential benefits of AI applications in business operations, the impact of clearly explainable and understandable AI in transforming both businesses and customers has yet to be fully developed. This could not only provide simplification in managerial decision-making, but also address an apparent lack of concern.

Transparent or explainable AI offers end users suitable understandability regarding the decisions or predictions derived by the system (Arrieta et al., 2020). The idea of explainable AI articulates a position which has the opposite role of AI as a "black box", as the machine of which "its designers cannot explain why an AI arrived at a specific decision" (Castelvecchi, 2016; Sample, 2017). Explainable AI intends to address managers' concerns by establishing positive attitudes and enhancing trust in the AI output through a context of transparency. This positivity can help create a sense of facilitating understandable insights on how the machine learning (ML) algorithm treats and transforms data to generate new insights in a hitherto hidden, obscure space.

In transparency context, explainable AI attempts to answer a basic question regarding the decision-making process, also considering human and machine-level explainable AI for better decision outcomes (Yang et al., 2022). Previous literature has discussed technical specifications regarding AI application validity; however, not many studies refer to issues of value and productivity improvement in businesses. The low transparency and explainability of AI output has emerged as a fundamental obstacle to achieving the anticipated benefits that would confidently transform data-centric decisions into practical strategies (Chowdhury et al., 2022a, 2022b; Makarius et al., 2020; Shin & Park, 2019). While AI-based models have become increasingly critical for data-driven decision-making in many management sectors, the complex nature of these models often creates a barrier to comprehending and interpreting them. Put differently, building trust and ownership of AI processed is paramount in ensuring their use. Despite spending time, effort, and resources on AI, many organizations cannot reap the envisaged benefits, primarily due to a lack of digital skills, cognitive skills in dealing with AI complexity, and information-processing expertise (Makarius et al., 2020). The ability to provide explanations for AI results will most likely eliminate bias in organizational operations, processes, and decision-making, thereby improving fairness (Satell & Sutton, 2019). For example, the lack of transparency in AI systems that set credit card borrowing limitations has led to companies' and their customers' mistrust (BBC, 2020).

Further, Chowdhury et al. (2022c) have explained the value of AI transparency in identifying and resolving flaws in ML models, thereby enhancing their utility for business organizations. Jabeur et al. (2021) exemplify this in their empirical analysis, which applied ML models in forecasting gold price fluctuations. Building on these insights, this above study specifically concentrates on utilizing data from January 1986 to December 2019. Although the research context is different, it demonstrates that the XGBoost algorithm has superior prediction accuracy compared to other algorithms, as various statistical tests also supported. They also provided a sensitivity analysis and comparisons with other established approaches (Jabeur et al., 2021).

Furthermore, the above-mentioned study utilizes SHAP to enhance the interpretability of the model, thereby assisting policymakers in comprehending the crucial elements that influence gold prices. Using XGBoost with SHAP demonstrates its efficacy in reliability prediction and interpretability. The results have significant consequences for policymakers, investors, and researchers.

This paper aims to introduce a newly developed AI-based SHAP ML model designed to assist managers in understanding what the AI system is doing, how it generates particular output responses, and why a given response is generated in the context of a financial organization. The positive operational view is to help managers confidently understand and assess the accuracy of the responses they get from AI, based on their expertise. This, we assume, will enhance their trust in using ML. The ability to consume explanations for the AI output can also potentially reduce biases in business processes, operations, and decision-making, thus enhancing accountability and transparentability.

In addressing the complexities surrounding the application of AI in financial organizations, this paper aims to bridge two main gaps in the existing literature: (1) the transparency and explainability of AI-driven systems, and (2) the identification of organizational resources required to leverage AI transparency (Makarius et al., 2020; Chowdhury et al., 2022c). The research presented in this paper, therefore, intends to fill the above-mentioned knowledge gaps, as they are graphically depicted in Fig. 1. Our study is inspired by action design research (ADR), which is a "method for developing and evaluating ensemble IT artifacts in an organizational environment to produce prescriptive design knowledge" (Sein et al., 2011, p.40) within a structure that uses the research onion framework. It addresses two issues that appear unrelated at first glance: Firstly, it considers intervening and evaluating a problem situation observed in a specific organizational setting, and secondly, it reflects on "constructing and evaluating an IT artifact that addresses the problem category exemplified by the observed situation" (Sein et al., 2011, p.40). ADR identifies the stages and concepts embedded in problem formation for the ensemble artefacts on which it focuses (Sein et al., 2011). This research procedure is depicted in Fig. 2 below. Inspired by the ADR methodology (Sein et al., 2011), our study addresses this challenge taking a theoretical and practical perspective (Bromiley & Rau, 2016; Mikalef & Gupta, 2021). ADR serves as the theoretical lens that allows us to engage with the problem through both organizational intervention and IT artefact construction. It aims to develop prescriptive design knowledge that enhances AI transparency. Hence, we propose two research questions:

RQ1: How can ML algorithms, based on AI, effectively identify anomalies in financial organizations?

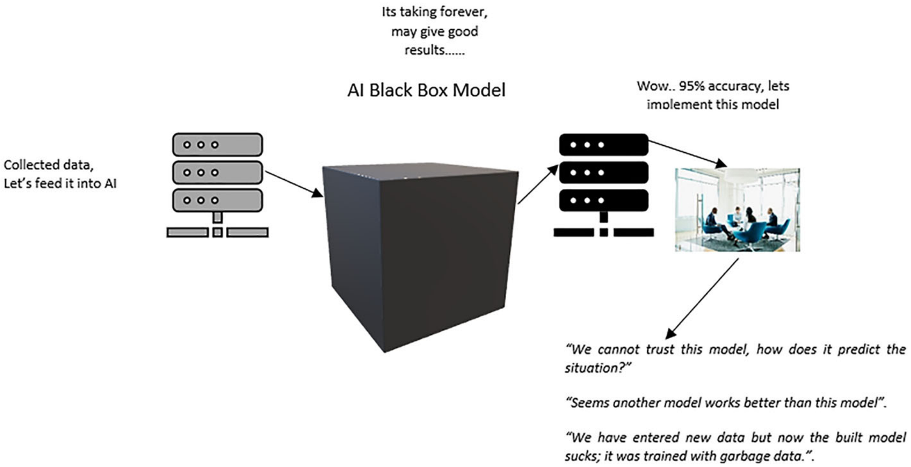

Its taking forever, may give good results......
AI Black Box Model
Wow.. 95% accuracy, lets imolement this model
Collected data, Let's feed it into AI
"We cannot trust this model, how does it predict the situation?"
"Seems another model works better than this model".
"We have entered new data but now the built model sucks; it was trained with garbage data.".

Fig. 1 The AI transparency gaps

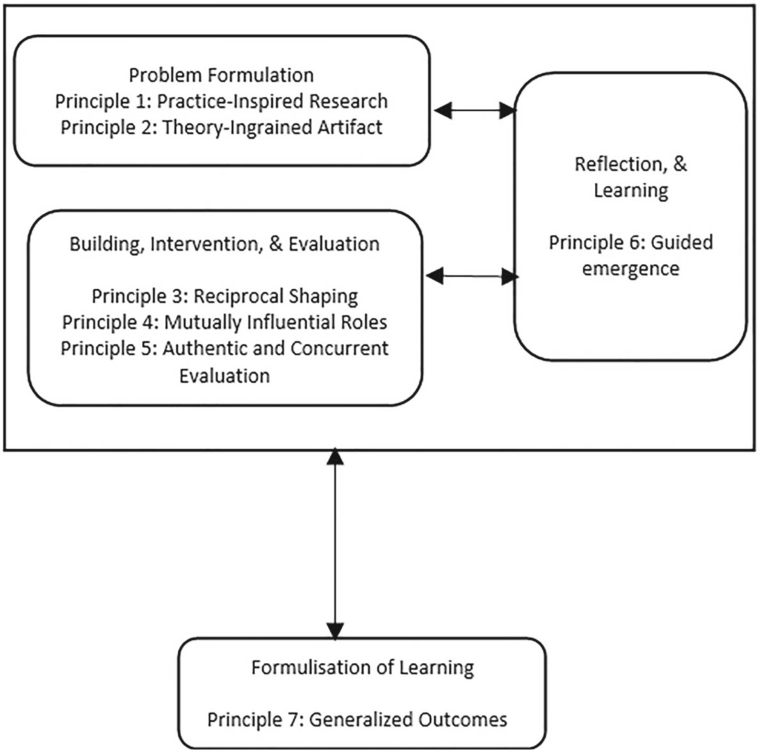

Problem Formulation
Principle 1: Practice-Inspired Research
Principle 2: Theory-Ingrained Artifact
Building, Intervention, & Evaluation
Principle 3: Reciprocal Shaping
Principle 4: Mutually Influential Roles
Principle 5: Authentic and Concurrent Evaluation
Reflection, & Learning
Principle 6: Guided emergence
Formulisation of Learning
Principle 7: Generalized Outcomes

Fig. 2 The ADR model (Source: Sein et al., 2011)

RQ2: What mechanism can be incorporated into ML algorithms to ensure transparency and justify the output developed by these algorithms in detecting anomalies?

The answers to these research questions could improve organizational executives' and managers' capacity to confidently use AI-based decision support systems utilizing ML algorithms (RQ1) and could assist these organizational executives and managers in comprehending how these algorithms detect anomalies in developing strategies and initiatives towards combating fraud and money laundering (RQ2).

Answers to these concerns are crucial since organizational researchers have noted and acknowledged the growing use and influence of AI in decision-making processes to obtain a competitive advantage (Kaplan & Haenlein, 2020; Shrestha et al., 2019, 2021).

The importance of introducing explainable AI-based decision support systems, or of explaining how the outputs of AI algorithms are formed, is widely acknowledged and apparent (Chowdhury et al., 2022c). Such explanations are not well-established in the academic literature on basic management and financial issues. In the light of technological and algorithmic advancements, the challenge now is for scholars to clarify how AI algorithms produce the conclusions they present rather than being obliged to collect huge amounts of data which they can turn into knowledge and conclusions (Kersting & Meyer, 2018; Belizón & Kieran, 2022; Gunasekaran et al., 2017). This will make it easier to establish trust in the management and to transform AI conclusions into useful information.

Our study utilizes ADR as methodology and additionally employs the research onion framework, as given in Fig. 3, to enhance methodological rigour. Using this dual framework technique offers a methodical strategy for clarifying the intricacies of our research questions (Mardiana, 2020). The research onion allows us to elucidate our pragmatic philosophical perspective while supporting inductive reasoning, which integrates effectively with the ADR's naturally iterative and problem-solving nature. Utilizing this layered approach enables us to delineate pragmatic problem-solving approaches targeted at tangible issues in the real world while establishing a strong theoretical framework. Taking the ADR's iterative approach, our research employs a longitudinal methodology to encompass the necessary temporal scope for numerous cycles of implementation and evaluation. To achieve our research aims, we have chosen to utilize quantitative data collection approaches and employ rigorous statistical analysis. Using this multifaceted methodology ensures a comprehensive and steadfast approach that can effectively address the research questions and objectives. We frame the implications of this work and its contributions to the financial sector in the following ways by combining AI literature with theoretical tenets of ADR within the research onion framework. Considering the ADR methodology (Boxall, 1996) in anomaly detection, we have investigated the explicability and transparency of AI-based ML models as a strategic resource. Technology is frequently one of the organization's fundamental strategic resources, crucial to achieving and maintaining a competitive edge (Alalie et al., 2018). Adopting and utilising a technology resource can significantly impact the organization's effectiveness (Wernerfelt, 1984). Technology adoption and value creation will suffer if there is a lack of trust in what the technology can achieve (Andriopoulos & Lewis, 2009). For AI-based ML models to be a successful strategic resource, they must include transparency (Raisch & Krakowski, 2021), allowing decision-makers to use AI's analytical and computational capabilities to support and drive data-centric decision-making.

By creating an implementation framework which incorporates explainability components in AI-based systems, we can advance the field of AI. Such a framework in the financial sector will make it easier for managers to comprehend the reasoning behind decisions the AI systems suggest. AI applications must be transparent and understandable now and in the future to

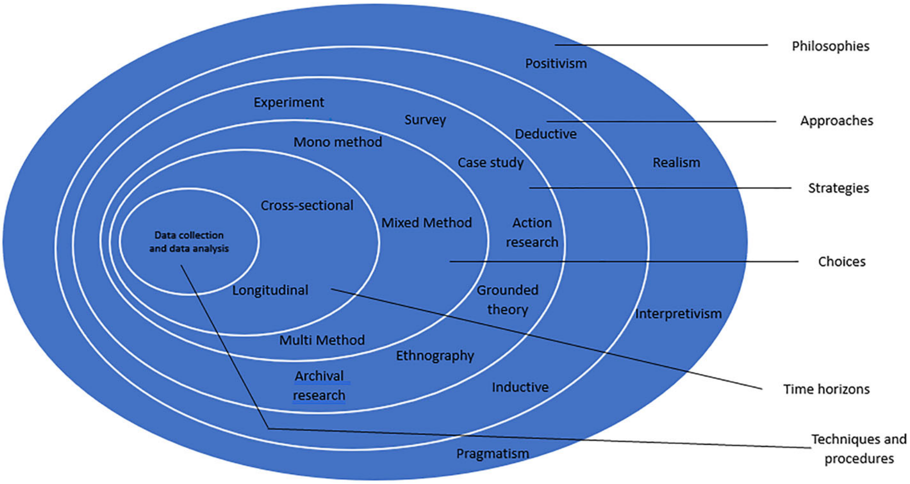


Fig. 3 The research onion (Source: Saunders et al., 2007)

supplement and support human intellect in taking decisions. This will promote ethical and responsible AI use in businesses (Brock & Von Wangenheim, 2019; Shrestha et al., 2021).

Various AI applications have been used in other areas of practice, such as in healthcare (Lasaga & Santhana, 2018; Šabić et al., 2021), stock market forecasting (Bao et al., 2017), customer retention (Motiwalla et al., 2019), and even social media (Grandinetti, 2021; Munk et al., 2022). In the financial service industries, the science of detecting anomalies (Ahmed et al., 2016) has been developed by applying different ML algorithms for credit card fraud, mobile phone fraud, insurance claim fraud, and insider trading fraud (Sariannidis et al., 2020). The ML approaches were used to distinguish various kinds of fraud using synthetic data. Our research design, therefore, proposes a new transparent AI artefact to identify insights with a high degree of precision for the decision-making process using a synthetic and real-world dataset which we collected in a financial organization. We investigated how applying AI technologies can produce more accurate results to address these issues in the financial domain.

There is a growing realization that a responsible approach to AI technologies is necessary to ensure the beneficial, transparent and explainable use of AI technologies and to comprehend the AI concept of machines taking moral decisions. Several efforts try to propose standards and principles for responsibly developing and using AI technologies. Rapid advances in autonomy and ML enable AI technologies to make decisions and take actions without direct human control. Greater autonomy must be accompanied by greater responsibility. However, compared to dealing with human subjects, these concepts necessarily have a different meaning when applied to computers. Ensuring that systems are constructed responsibly contributes to our faith in their behaviour. This also involves accountability, i.e., the capacity to explain and justify decisions and improve transparency, Thus, technologies should support the human ability to comprehend how systems make decisions and which data they use (Felzmann et al., 2020).

Developments in individual autonomy and knowledge have significantly empowered AI technologies to determine necessary actions and operate independently (Calvo et al., 2020). To date, algorithm development has been driven by the desire to improve performance, but such development has resulted in opaque black boxes. Nonetheless, demands now are that increased machine autonomy must be accompanied by increased levels of company responsibility and the ability to explain decisions. Transparency is likely to be the foundation of trust in AI since if humans cannot connect to how machines operate, trust falters. Putting human values at the centre of AI technologies requires a mental shift on the part of researchers and engineers. The objective should be to enhance transparency rather than performance, which could imply the development of unique and fascinating algorithms (Müller, 2020). This is the central focus of our paper.

The remainder of this paper is organized as follows: Firstly, we present relevant background by giving a brief exposition of previous literature covering ML used in finance. Secondly, we describe the research methodology, including how an ML model is developed to detect anomalies, and thirdly, we evaluate and discuss the results using SHAP techniques from the ML algorithms proposed for this research also considering how as they indicate future research directions.

2 Literature review

This section provides a comprehensive review of the existing literature pertaining to three crucial areas: anomaly detection, ML in the realm of AI, and explainable AI. This study aims to illuminate the current knowledge gaps, particularly regarding the application of AI for improving data-driven decision-making and ensuring the explainability of AI algorithms.

2.1 Anomaly detection

The increasing supply of big data in the contemporary digital environment presents organizations with opportunities and challenges. Despite the expanding availability of data, many organizations still have difficulties with transforming raw data into practical and useful insights. Utilizing ML algorithms in anomaly detection plays a pivotal role in bridging the gap between data analytics and informed decision-making in the business context. Bishop and Nasrabadi (2007) find that this technology enables organizations to efficiently identify outliers or anomalies in extensive and intricate datasets, thereby facilitating the optimization of the decision-making process.

ML technologies have become a crucial instrument in addressing these intricacies. Anomaly detection catalyses enhanced organizational efficiency by repairing the disparity between data measurements and business operations. By employing ML techniques and specialized algorithms designed for anomaly detection, organizations can identify abnormal patterns in massive data sets, particularly when they are confronted with non-analogous variables. In the context of the information age, ML brings a transformative paradigm change that can significantly improve the operational efficiency of diverse businesses. ML is a subset of AI that provides a robust framework for facilitating automated learning and enhancement through experience, which eliminates the need for manual updates (Bishop & Nasrabadi, 2006).

2.2 Machine learning

Numerous scientific disciplines rely heavily on ML and its applications permeate our everyday lives. For example, among other applications, it is used for email spam filtering, weather forecasting, clinical findings, product suggestions, facial recognition, and fraud detection. ML research investigates learning defined as obtaining information through experience. The ML cycle often entails gathering information and formulating a hypothesis that enables users to make predictions or, more generally, conduct analytical actions. For PCs, experience, or the ability to learn, is provided by data. Hence, we can define ML as the process of extracting knowledge from data. This data-driven learning is crucial in numerous sectors, such as finance, healthcare and education. Specifically, ML methods are broadly categorized into supervised, unsupervised, and reinforcement learning. Despite the ML advantages, in the current context of multiple organizations, anomalies have occurred that could result in money laundering, identity theft and credit card fraud in the financial domain, medical errors and a decreased patient survival rate in the clinical domain, as well as a decline in student enrolment and a higher teacher attrition rate in the educational domain.

This paper uses explainable AI technologies to approach the particular problem of predicting and detecting anomalies using a certain technology, namely an ML-based model, using a dataset of customers' financial transactions. We developed an automated process for performing this task in future transactions and improving the decision-making processes.

We, therefore, studied recent literature that presents a wide range of highlighted datasets, AI technologies used to detect anomalies, and the relevant supplemental material. ML uses data and algorithms to gradually imitate how humans interact to improve AI programs' accuracy (Chen et al., 2021; Jarrin et al., 2019; Sarker et al., 2021).

2.3 Explainable AI

Explainable AI is a set of techniques and strategies that enable human users to interpret and therefore also rely on the output and results ML algorithms produce, as illustrated in Fig. 4. The term "explainable AI" characterises a given kind of AI model, its anticipated impact, and any potential biases. It assists in identifying model accuracy, fairness, transparency and outcomes in AI-powered decision-making. Explainable AI is essential for a company to create trust and confidence before deploying AI models, and it facilitates adopting a responsible approach to AI development in an organization. Organization needs to thoroughly understand the AI decision-making processes, including model monitoring and AI responsibility, rather than blindly trust them. For instance, organizations sacrifice rust if a model's creators cannot describe how, it determines credit results, nor identify which elements impacted the results the most. The emerging field of explainable AI can aid financial organizations in navigating trust and transparency concerns and provide a better understanding of their AI governance. The field aims to make AI models more explicable, intuitive, and intelligible to human users without losing performance or forecast accuracy (Gunning & Aha, 2019). Explainable AI is also a growing concern for financial organizations who want to ensure that their financial personnel "reasonably comprehend" AI procedures and outputs. Explainable AI techniques assist organizations in implementing explainability, transparency and accuracy.

Further, explainable AI ensures that organizations conduct non-biased evaluations of AI systems. The explainable AI techniques explain and interpret AI models. Also, the system and its techniques synchronize methods and strategies to enable AI technology. These methods and techniques can be explained through data visualization, the logistic regression ML model, decision tree ML models, the neural network learning model, SHAP, and LIME, each contributing to human users comprehending the model outcomes.

Bussmann et al. (2021) offer an empirical assessment of this concept in their study on credit risk management in the context of peer-to-peer lending. The approach incorporates SHAP values and correlation networks to group comparable risk profiles using a dataset including 15,000 small and medium-sized firms (SMEs). The empirical investigation demonstrates that borrowers, regardless of their risk level, can be classified based on common financial characteristics, which improves the accuracy and dependability of credit score forecasts.

In a different context, i.e., in solar energy forecasting, Khan et al. (2022) illustrates the adaptability and efficacy of advanced ML techniques. Utilizing the DSE-XGB stacked ensemble approach, they integrated artificial neural network (ANN) and LSTM base models. The

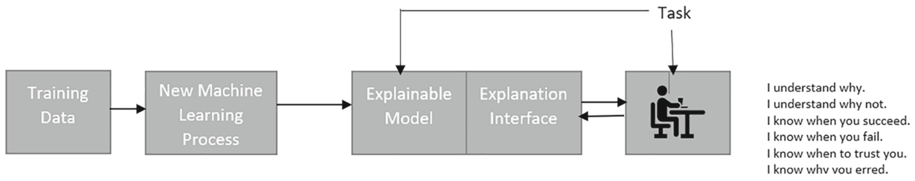


Fig. 4 Explainable AI process

technique was tested and verified using four separate datasets and a range of meteorological circumstances. Compared to standard models such as ANN, LSTM and Bagging, the approach demonstrates an improvement in R^2 values ranging from 10 to 12%. To enhance the outcome's interpretability, these researchers implemented the SHAP framework, which provided insight into the model's uncertainties and the elements that had a significant impact. The challenge of typically high computing demands associated with SHAP, especially when dealing with large datasets and deep learning models, could be mitigated by incorporating it with XGBoost. Using XGBoost's rapid convergence, it was possible to develop an efficient model informed by the data.

The idea of transparency encompasses the specifics of service rationale and other sorts of data management involving sensible data and/or the potential repercussions of the system indirectly gaining user knowledge (Ananny & Crawford, 2018). The significance of algorithmic transparency became apparent in resolving the matter of Facebook's role in Russian interference in the 2016 U.S. presidential election. The concepts of explainability that have become topical relate to whether stakeholders can interpret and comprehend the system's operation and results. Users are not required to have comprehensive and transparent access to a dataset and the underlying algorithms if a qualified, trusted professional or entity provides them with easily accessible information on the system. When people understand how a system operates, they are more likely to use it correctly and to have faith in its designers and developers (Lee & Boynton, 2017).

In recent years, financial services have rapidly started to utilize AI technologies. This research paper is the first to propose transparent AI technologies that emphasize the researchers' obligation to explain not only the algorithm's inputs and outputs but also which decisions the machines make, and why. Since the objective is to comprehend how an algorithmic system arrives at decisions, we ensure that the model can be described, can detect anomalies, and can perform a meta-analysis on ML application in a financial context (Nelson, 2019).

2.4 Knowledge gap

Although the finance literature offers conceptual frameworks on AI applications, we do not find demonstrations of these frameworks being actually implemented in empirical studies employing AI-based algorithms to comprehend the limitations presented by these algorithms' lacking transparency for decision-makers. A few reports on AI frameworks incorporate explainability in the financial literature, as well as in the health and educational domains. This calls for additional research in the field of explainable AI to enhance confidence in the automated analytic capabilities AI technologies provide in the digital era and to supplement such AI with managers' perceptive and social intelligence, both centred on AI's intangible expertise than can expedite data-driven strategic decision-making. Enhancing the transparency and explainability of AI algorithm responses by providing useful insight into the algorithms' accuracy, relevance and methodology can help to increase managers' trust in these models. Therefore, globally, and locally, AI transparency can facilitate strategic changes in financial processes, decision-making, and policies.

We eventually propose an ML-based solution artefact that can help other researchers create ML models that will assist in detecting anomalies. We validate this research by comparing two ML models of which the codes are provided in GitHub (https://github.com/MachineLearning-UON/ML-Financial_transaction.git).

2.5 Existing solutions and limitations

Several studies have investigated anomaly detection across multiple domains they often lack code transparency or applicability. For instance, Chang and Chang (2010) use X-Means clustering approaches to identify changes in the data that indicate fraudulent behaviour in online auction fraud. However, this experimental work does not provide a coding strategy to help other researchers and it does not explain how the results were generated. Le Khac and Kechadi's (2010) study employ K-Means ML algorithms to identify money laundering in CE banks. However, the research is limited to money laundering at CE banks without covering model validation. In a similar context, Glancy and Yadav's (2011) study presents a computational model to detect fraud in financial reporting; however, elements such as transactions, credit cards, and the absence of algorithmic openness, are not addressed. Ngai et al.'s (2011) literature analysis indicates that logistic models, neural networks, Bayesian belief networks, and decision trees are the most often employed strategies for detecting fraud. The research examines four categories of fraud, namely those in banking, insurance, securities, and commodities, but it provides neither a comprehensive understanding of fraud nor explains how the models could detect the fraud.

Chang and Chang (2012) present a cost-effective method for early identification of online auction fraud. This study's focus is limited to online auction fraud, yet the proposed model does not demonstrate that it can also be utilized to detect other unknown facts. Another study, Ahmed et al. (2016), provides a comprehensive overview of several clustering approaches for detecting anomalies and comparing methodologies. Their analysis compares clustering algorithms but provides no findings. Further, it lacks an experimental study to determine the optimal strategy. Similarly, Abdallah et al. (2016) did a survey leading to a systematic and exhaustive review of the obstacles that hamper the operation of fraud detection and prevention systems. The study focuses only on the factors that impede fraud detection and prevention systems' performance and delivers conclusions based on previous research. There is no experimental study validating the outcomes.

West and Bhattacharya (2016) show the results of a survey on using classification algorithms to detect financial fraud while also analysing the algorithm's weaknesses and strengths. The study focuses solely on classifying various kinds of ML, examining the literature from 2004 to 2014 that could have influenced the survey results. Yaram's (2016) research proposes employing document clustering and classification algorithms to detect insurance claim fraud; however, the research focuses on fraud detection without describing the type of fraud. Additionally, the study compares ML algorithms without recommending any specific algorithm for future implementation. Ahmed et al. (2017) analyses the underlying assumptions of finding unusual untold facts using partition-based and hierarchical-based clustering algorithms on historical Australian Security Exchange (ASX) data. The research describes the assumptions; however, it provides neither the local outlier factor (LOF) nor the clustering-based multi-variate Gaussian outlier factor (CMGOS) codes. Therefore, the study could not confirm whether these strategies are effective. Bao et al. (2017) present a deep learning framework that predicts stock prices by combining wavelet transformations (WT), stacked auto-encoders (SAEs), and long and short-term memory (LSTM). The research is limited to the time-series component of finance.

Huang et al. (2018) introduces the CoDetect framework for detecting financial fraud by analysing network and feature data. With a focus on financial fraud, numerous other types of fraud that can occur in a financial institution are not identified, which could deliver ambiguous findings. A review by Ryman-Tubb et al. (2018) presents a survey on credit card fraud detection, identifying eight strategies that could be used to detect credit card fraud in the financial industry. The research focuses on detecting fraudulent card payments but ignores other types of fraud, such as money laundering, identity theft, account takeover, and risk modelling fraud. Further, Chalapathy and Chawla (2019) investigate credit card transactions using deep learning. This instrument demonstrates that the absence of consistent patterns presents the largest obstacle in combating credit card fraud. The paper discusses twelve distinct models without endorsing any particular model for detecting anomalies. Magomedov et al.'s (2018) analysis provides an anomaly detection solution for fraud control based on ML and a graph database. The article gives experimental results utilizing Random Forest and MinMaxScaler but cannot support these algorithms' ability to detect anomalies in fraud management. Pourhabibi et al.'s (2020) research outlines the graph-based anomaly detection method. They examine and evaluate articles published between 2007 and 2018, which we deem too wide to yield excellent findings.

3 Research methodology

The methodology we employ in this paper is based on the information system literature used to develop a predictive analytics data science resource for qualitative researchers (Ciechanowski et al., 2020), the fundamentals of using machine and deep learning algorithms (Shrestha et al., 2021), and ML models used for predicting (Hwang et al., 2020). We provide a framework for deploying ML that incorporates explainable AI to detect anomalies.

Many methods to detect anomalies are in use, but the demand to process them in real-time poses one of the greatest challenges (Chalapathy & Chawla, 2019). The existing approaches' accuracy in detecting anomalies is generally lower than that of classification or regression ML-based models. Existing approaches are also time-consuming and slow in identifying the patterns in financial transactions. Another challenge is that the profiles of common and abnormal transactions depend on consistent change, whereas existing information about abnormal behaviour is regularly skewed and is, therefore, unreliable. Although many ML approaches have been introduced over the past few years, we discuss only those that fit our requirements for building an artefact. The term 'solution artefact' refers to "*a thing that has been or can be transformed into material existence as an artificially made object (e.g., model, instantiation) or process (e.g., method, software). Many IT artifacts have some degree of abstraction but can be readily converted to material existence; for example, an algorithm converted to operational software*" (Gregor & Hevner, 2013, p.341). Kafai (1996) developed an artefact by establishing a link between learning and designing, whereas we want an artefact with the potential to create operational software in the financial industry to support their regular business.

AI refers to a collection of algorithms and methods that can automatically incorporate, process, and learn from data and then utilize such learnt knowledge to accomplish particular goals and activities (Haenlein & Kaplan, 2019). AI-based applications can aid in anticipating organization development, i.e., predicting anomalous financial transactions or increases in sales. However, these applications lack transparency and explainability, making it challenging for organizational managers to accept the AI results (Bieda, 2020; Chowdhury et al., 2022a; Rai, 2020). From a transparency perspective, it is regarded as ideal that AI should promote the disclosure of how data is incorporated in an algorithm, analysed by the algorithm, and used to gain knowledge (Cheng & Hackett, 2019). The problem with explainability is that organizational managers do not grasp how the AI-based ML algorithm processes input data to produce outputs, either because the algorithm is confidential or because the mathematical computational models are difficult to comprehend (Shin & Park, 2019).

3.1 Proposed ML-based solution artefact

The solution we propose consists of the steps set out in Fig. 5 below, which should help prepare the ML-based model to detect anomalies. The artefact model is constructed by following the steps as illustrated.

3.2 Data acquisition

This study utilizes a systematic methodology for collecting and preparing data, acknowledging its crucial significance in constructing a customized model suitable for predictive or classification purposes. We used two datasets to conduct a full evaluation of the efficacy of the model we built. The initial dataset was acquired from Kaggle, a popular repository for ML resources, while the second real-world dataset was procured through an actual financial organization. We recognized that, as elsewhere, our raw data necessitated significant pre-processing from databases, files, or other sources, regardless of origin, because factors such as missing values, extreme data points, and disorganized textual or noisy data could be present and confounding. Therefore, we implemented meticulous data preparation protocols to guarantee the utmost quality and dependability of the datasets included in our research.

3.3 Data pre-processing—extraction, wrangling and visualization

Data pre-processing is one of ML's most important steps because it helps build more accurate ML models. There is an 80/20 rule in ML, according to which 80% of the time is spent on pre-processing data and 20% on analysis. Pre-processing a dataset means cleaning the raw data, carrying out data wrangling and visualization using data conversion, sometimes ignoring the missing values, at other times filling in the missing values, and detecting outliers to convert raw data into a clean dataset, which can then be used to train the model. In real life, raw data is often messy due to missing, noisy and inconsistent data.

Data extraction consists of collecting or retrieving disparate types of data from various sources, many of which might be poorly organized or completely unstructured. Systematic extraction makes it possible to consolidate, process and refine data to be stored in a central location for transformation. Data wrangling consists of cleaning and converting 'raw data' into a format that enables convenient consumption. Data analysis involves selecting and filtering the data needed to prepare the model. Data visualization consists of translating huge datasets into charts or graphs for presentation. Data visualization enables users to identify data trends, recognize outliers, and impart new insights on the information represented in the dataset.

3.3.1 Training and testing algorithms

The training algorithm is prepared in the training dataset to understand the patterns and rules that govern the data, whereas the testing dataset determines the model's accuracy. In other words, a split training and testing algorithm is used to estimate the ML algorithm's performance when it is used to make predictions on the dataset, not to train the model. This

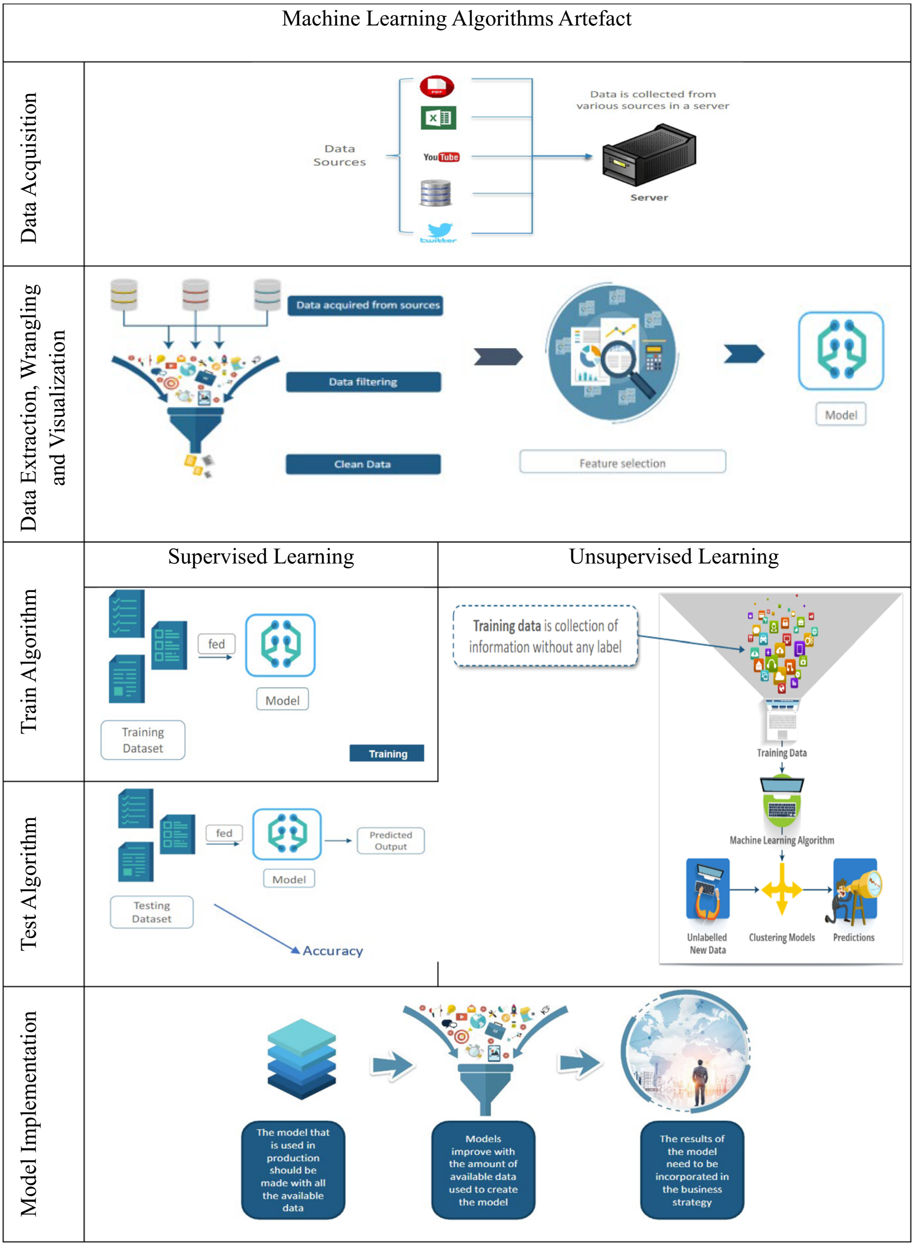

Machine Learning Algorithms Artefact
Data Acquisition
Data Sources
You Tube
twitter
Data is collected from various sources in a server
Server
Data Extraction, Wrangling and Visualization
Data acquired from sources
Data filtering
Clean Data
Feature selection
Model
Train Algorithm
Supervised Learning
Unsupervised Learning
fed
Model
Training Dataset
Training
Training data is collection of information without any label
Training Data
Machine Learning Algorithm
Unlabelled New Data
Clustering Models
Predictions
Test Algorithm
fed
Model
Predicted Output
Testing Dataset
Accuracy
Model Implementation
The model that is used in production should be made with all the available data
Models improve with the amount of available data used to create the model
The results of the model need to be incorporated in the business strategy

Fig. 5 Proposed ML framework

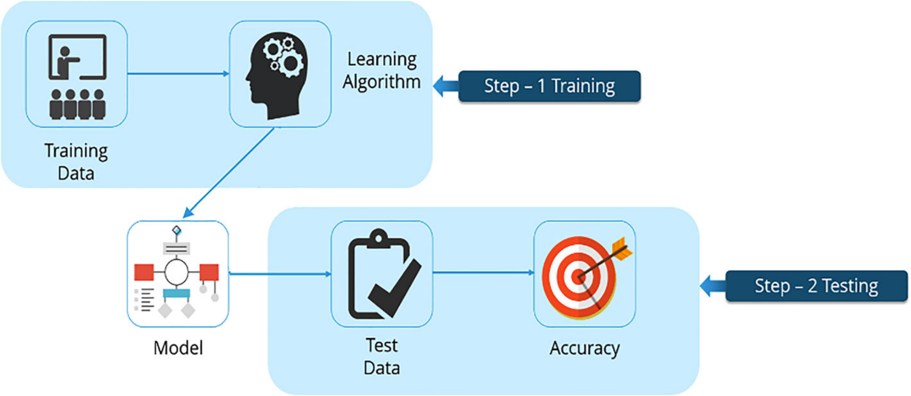


Fig. 6 Training and testing techniques with a 'sufficiently large' dataset

fast and simple technique enables the results to compute an ML algorithm's performance for a predictive modelling problem. This technique can also be used in classification or regression problems by dividing the data into two subsets, as shown in Fig. 5. The first aim is to fit the model. Therefore, we refer to the first subset as the training dataset. The second subset is not used to train the model; rather, the input element of the dataset is fed into the model, and then the resulting predictions are contrasted with the expected values.

The training dataset is used to fit the ML model. The testing dataset is used to evaluate the fitted ML model. The training and the testing algorithm's objective is to gauge the ML model's performance on a new dataset, not to train the model. The possibility of a dataset being 'sufficiently large' is explicit in every predictive modelling problem. It implies enough data to split the dataset into separate training and testing datasets, where every dataset appropriately portrays the problem domain. An appropriate representation of the problem domain implies enough records to cover all the common and uncommon causes of patterns. Notably, the train-test technique is not appropriate if the dataset is small since, then, dividing it into training and testing datasets is problematic. A small dataset does not offer enough information in the preparation dataset for the model to learn an effective mapping of input and output, and it does not provide enough data in a testing dataset to evaluate the model's performance (Fig. 6).

The train-test technique has one fundamental configuration parameter: the size of the training and testing datasets. This is most commonly expressed as a percentage between 0 and 1 for the training and testing datasets. For example, a training set with a size of 0.67 (67%) means that the remaining percentage of 0.33 (33%) is assigned to the testing set. The split rate depends on the computational case of the training model or the computational case when the model of the project's objectives has to be evaluated. The most commonly used split percentages are:

Training: 80%, Testing: 20%
Training: 67%, Testing: 33%
Training: 50%, Testing: 50%

3.3.2 Model implementation

Suppose that the speed and accuracy of the ML model are acceptable. In such a case, the model should be implemented in the real system in the target operational environment of the financial organization. This can be done using the following:

A trained model ready to deploy — save the model into a file to be further loaded and used by the web service.
A web service that gives a purpose for a model to be used in practice. The common platforms used to deploy trained models include Flask, Docker, Django, and Flask GitHub.
A cloud service provider — a special cloud server required to deploy the application. The services that can be used for simplicity include Heroku, AWS, and GCP.

3.3.3 Artefact evaluation

One of the greatest challenges in research involving financial organizations is the absence of publicly accessible datasets. This is mostly due to privacy issues because the existing datasets can contain sensitive and personal information on the clients. Below we list publicly accessible datasets that, to the best of our knowledge, are not obsolete. Figure 7 shows the synthetic dataset processes with two pipelines, taking Gregor and Hevner's (2013) experimental evaluation as guideline. The first pipeline obtains the data, analyses the data using data wrangling and visualization, prepares the model using the train-test technique and then carries out an evaluation.

This model can be prepared accurately and speedily, after which we move to the second pipeline, which helps the prepared ML model to detect anomalies in the new dataset. Various organizations can develop this to automate anomaly detection. Our study is limited to implementing the prepared model because there are no publicly available sources that will deploy a fully equipped ML model.

The freely accessible data sourced from the Kaggle website contains 6,362,620 data instances built from 11 features to present financial transactions identified as follows: step, type, amount, nameOrig, oldbalanceOrg, newbalanceOrig, nameDest, oldbalanceDest, newbalanceDest, isFraud, isFlaggedFraud. The financial transaction dataset 'isFraud' feature is annotated such that '0' represents a 'normal transaction' and '1' an 'anomalous transaction' (Fig. 8).

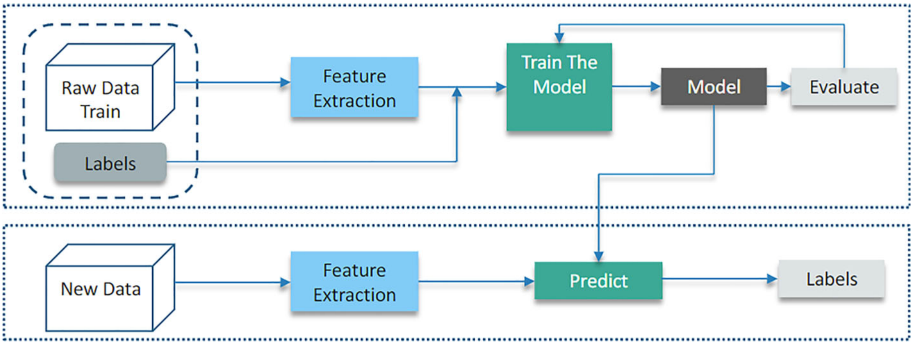
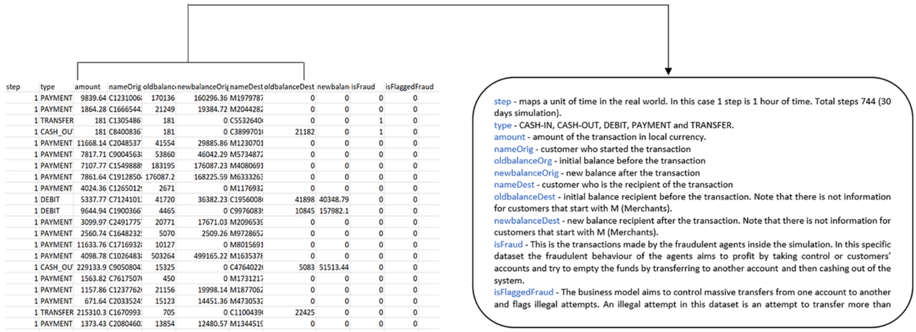


Fig. 7 Process flow of the proposed ML modelling

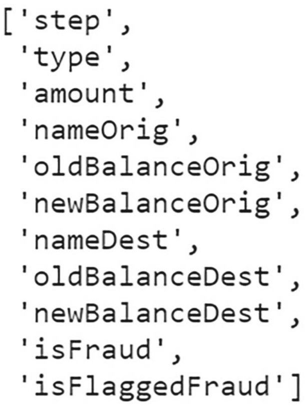


step	type	amount	nameOrig	oldbalanc	newbalanceOrig	nameDest	oldbalanceDest	newbalan	isFraud	isFlaggedFraud
1	PAYMENT	9839.64	C1231006	170136	160296.36	M1979787	0	0	0	0
1	PAYMENT	1864.28	C1666544	21249	19384.72	M2044282	0	0	0	0
1	TRANSFER	181	C1305486	181	0	C5532640	0	0	1	0
1	CASH_OU	181	C8400836	181	0	C3899701	21182	0	1	0
1	PAYMENT	11668.14	C2048537	41554	29885.86	M1230701	0	0	0	0
1	PAYMENT	7817.71	C9004563	53860	46042.29	M5734872	0	0	0	0
1	PAYMENT	7107.77	C1549888	183195	176087.23	M4080691	0	0	0	0
1	PAYMENT	7861.64	C1912850	176087.2	168225.59	M6333263	0	0	0	0
1	PAYMENT	4024.36	C1265012	2671	0	M1176932	0	0	0	0
1	DEBIT	5337.77	C7124101	41720	36382.23	C1956008	41898	40348.79	0	0
1	DEBIT	9644.94	C1900366	4465	0	C9976083	10845	157982.1	0	0
1	PAYMENT	3099.97	C2491775	20771	17671.03	M2096539	0	0	0	0
1	PAYMENT	2560.74	C1648232	5070	2509.26	M9728652	0	0	0	0
1	PAYMENT	11633.76	C1716932	10127	0	M8015691	0	0	0	0
1	PAYMENT	4098.78	C1026483	503264	499165.22	M1635378	0	0	0	0
1	CASH_OU	229133.9	C9050804	15325	0	C4764022	5083	51513.44	0	0
1	PAYMENT	1563.82	C7617507	450	0	M1731217	0	0	0	0
1	PAYMENT	1157.86	C1237762	21156	19998.14	M1877062	0	0	0	0
1	PAYMENT	671.64	C2033524	15123	14451.36	M4730532	0	0	0	0
1	TRANSFER	215310.3	C1670993	705	0	C1100439	22425	0	0	0
1	PAYMENT	1373.43	C2080460	13854	12480.57	M1344519	0	0	0	0

step - maps a unit of time in the real world. In this case 1 step is 1 hour of time. Total steps 744 (30 days simulation).
type - CASH-IN, CASH-OUT, DEBIT, PAYMENT and TRANSFER.
amount - amount of the transaction in local currency.
nameOrig - customer who started the transaction
oldbalanceOrg - initial balance before the transaction
newbalanceOrig - new balance after the transaction
nameDest - customer who is the recipient of the transaction
oldbalanceDest - initial balance recipient before the transaction. Note that there is not information for customers that start with M (Merchants).
newbalanceDest - new balance recipient after the transaction. Note that there is not information for customers that start with M (Merchants).
isFraud - This is the transactions made by the fraudulent agents inside the simulation. In this specific dataset the fraudulent behaviour of the agents aims to profit by taking control or customers' accounts and try to empty the funds by transferring to another account and then cashing out of the system.
isFlaggedFraud - The business model aims to control massive transfers from one account to another and flags illegal attempts. An illegal attempt in this dataset is an attempt to transfer more than

Fig. 8 Synthetic data dictionaries

Fig. 9 Process flow of the ML modelling



```
['step',
 'type',
 'amount',
 'nameOrig',
 'oldBalanceOrig',
 'newBalanceOrig',
 'nameDest',
 'oldBalanceDest',
 'newBalanceDest',
 'isFraud',
 'isFlaggedFraud']
```

We have renamed some columns for clarification and to show the new features, as in Fig. 9:

We use common libraries shown in Fig. 10 and some additional libraries as required for the outcome.

We have plotted some graphs to visualize the dataset. If the correlation values in the heatmap need to be seen, we note that these features are highly correlated: newBalanceDest and oldBalanceDest, newBalanceOrig and oldBalanceOrig. Hence, we can remove one of the two features from dataset to balance the corelation. In future, we could remove oldBalanceDest and oldBalanceOrig. Figure 11 presents the visual explanation of the heatmap using some features, whereas Fig. 12 displays the heatmap plotting the amount of money and isFraud features.

The graph in Fig. 13 presents the relationship between the number of transactions and the days, hours and weeks.

The graph in Fig. 14 below shows the transaction when a fraudulent incident happened (yellow graph) and when no fraudulent incident happened (blue graph), marking the time as during which hours of the day, then which day of the month, and then which weeks of the month.

Besides adding additional features using One-hot encoding to simplify the features provided, this dataset does not need any pre-processing because it is already in a clean format. The 'step' feature extended to three directly encoded features, such as 'hour', presents the

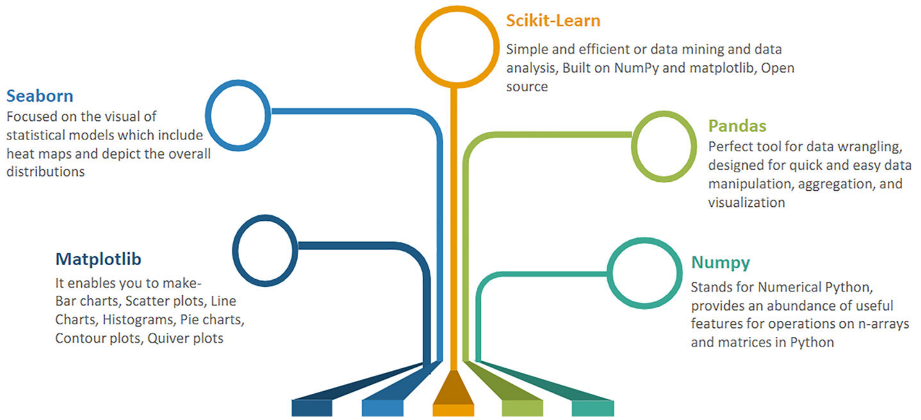

Scikit-Learn
Simple and efficient or data mining and data analysis, Built on NumPy and matplotlib, Open source
Seaborn
Focused on the visual of statistical models which include heat maps and depict the overall distributions
Pandas
Perfect tool for data wrangling, designed for quick and easy data manipulation, aggregation, and visualization
Matplotlib
It enables you to make- Bar charts, Scatter plots, Line Charts, Histograms, Pie charts, Contour plots, Quiver plots
Numpy
Stands for Numerical Python, provides an abundance of useful features for operations on n-arrays and matrices in Python

Fig. 10 ML common basic libraries

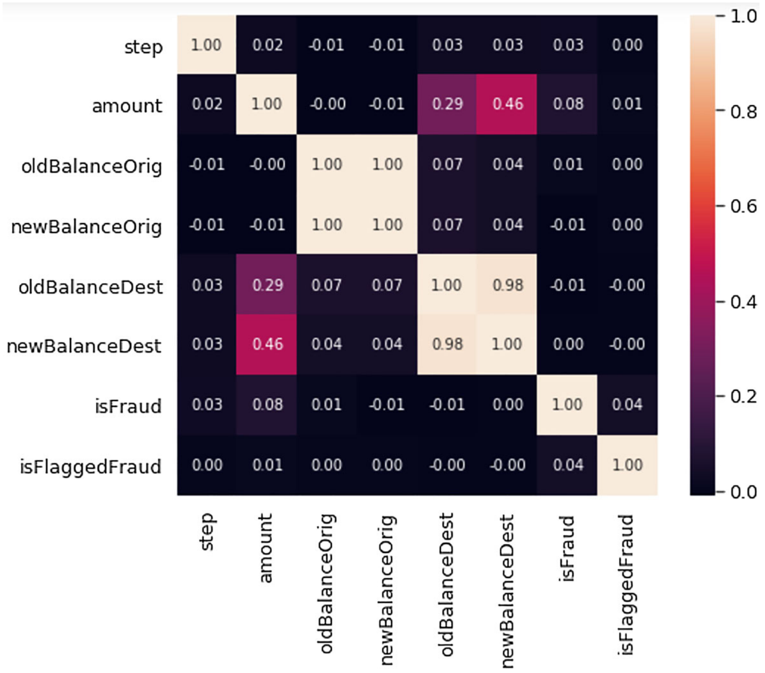
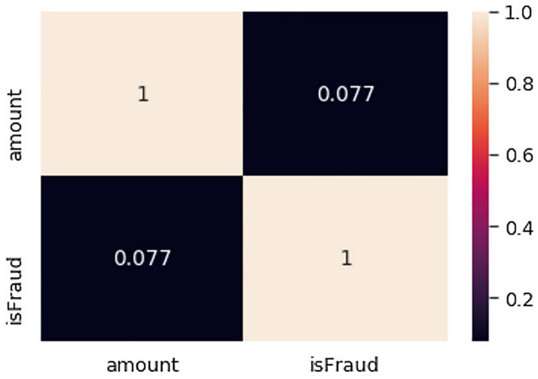


Fig. 11 Heatmap using a synthetic dataset for anomaly detection

number of hours from the first transaction in a 24-h time cycle with a range of 0–23. The ‘step_day’ feature represents the day in a month with the feature range 1–30, and ‘step_week’ represents the day in a seven-day week cycle with a feature range of 1–7. Another ‘hot’ encoding is carried out on the ’type’ feature, which creates another five columns: type_CASH_IN, type_CASH_OUT, type_DEBIT, type_PAYMENT, and type_TRANSFER.

Fig. 12 Heatmap using amount and is Fraud



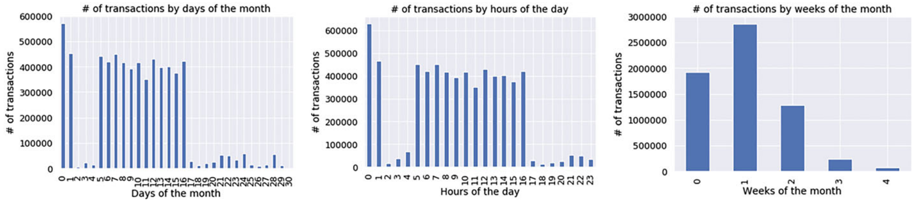


Fig. 13 Graph presenting the number of transactions according to days, hours and weeks

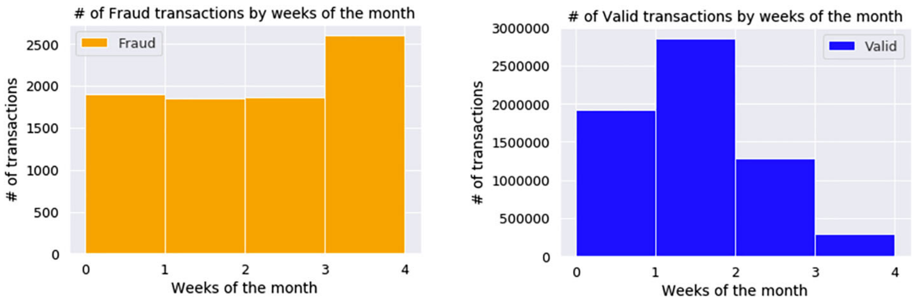
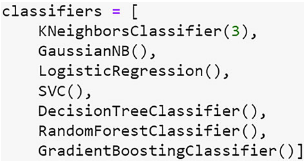


Fig. 14 Fraud and no fraud during hours, days and weeks

The synthetic dataset has been extended, and the new dataset consists of 6,362,620 and 23 instances. The box plot has been created in the data visualization process, as presented in Fig. 14 below. It shows that the number of outliers for all features is high, and therefore, the contamination factor used in later methods should not be too low. The box plots also show that none of the dataset features are distributed uniformly, while the box plot indicates that the features amount, oldbalanceOrg, newbalanceOrig, oldbalanceDest, and newbalanceDest are highly skewed and should be further processed before use in training ML models.

We performed the ML algorithms (see Fig. 15) below to detect an anomaly and then presented the results. The list of supervised ML algorithms includes K-Neighbors Classifier, GaussianNB (Gaussian Naïve Bayes), Logistic Regression, SVC (Support Vector Classifier), Decision Tree Classifier, Random Forest Classifier, and Gradient Boosting Classifier.



```
classifiers = [
    KNeighborsClassifier(3),
    GaussianNB(),
    LogisticRegression(),
    SVC(),
    DecisionTreeClassifier(),
    RandomForestClassifier(),
    GradientBoostingClassifier()]
```

Fig. 15 ML algorithms used in the model

Figure 16 below presents the accuracy, precision and recall scores in the training dataset (LHS) and accuracy, precision and recall in the testing dataset (RHS).

To evaluate the validity of the proposed ML framework given in Fig. 17, we used the ML algorithm to detect anomalies in real-world datasets and shared the results on GitHub (https://github.com/MachineLearning-UON/ML-Financial_transaction.git). The dataset consists of 8963 instances and 26 features which are defined as: ’id’, ’trans_type’, ’trans_ref’, ’status’, ’ppan’, ’cts’, ’bsb’, ’account_no’, ’indicator’, ’trans_code’, ’currency’, ’amount’, ’description’, ’lodgement_ref’, ’remitter_bsb’, remitter_account_no’, remitter_name’, withhold_tax_amount’, ’trans_data_id’, ’stan’, ‘post_date’, ’approval_code’, ’hold_number’, ‘exchange_rate’, ’ext_trans_no’, ’error_msg’.

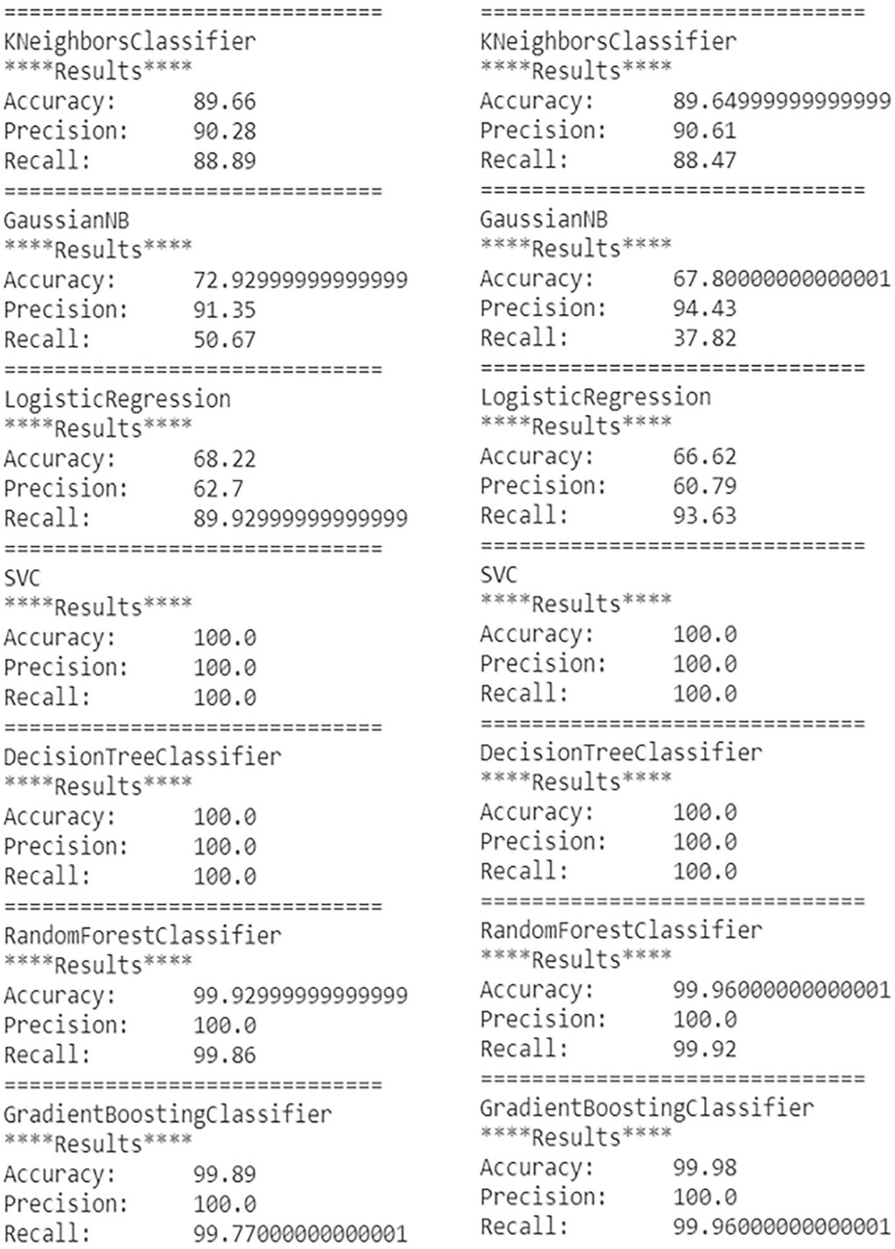
The data are in a raw format and need time to be cleansed and analysed. We carried out One-hot encoding on the ’trans_type’ feature, which added more features to the dataset and removed others that did not contribute to the outcome. We also carried out data visualization and created the heatmap shown in Fig. 18.

Figure 19 below shows the outliers in the amount feature since no outcome was provided. Then, it divided the amount into five parts ranging from 0 to 2000, 2000 to 4000, 4000 to 6000, 6000 to 8000, and then 8000 to 10,000. Next, we predicted that instances in the amount 0–2000 category are normal transactions. Using different algorithms in the more-than-2000 categories would be defined as an anomaly, as shown in the three Figs. 19, 20 and 21 below. Figure 19 shows the box plot graph using the seaborn library.

We carried out an unsupervised K-means ML because no outcome is given in the dataset. Figure 20 below shows that after 2000 transactions, one was detected as an anomaly.

We used another isolation Forest unsupervised algorithm, shown in Fig. 21 to compare the results. The figure shows the inliers and outliers in the amount of data.

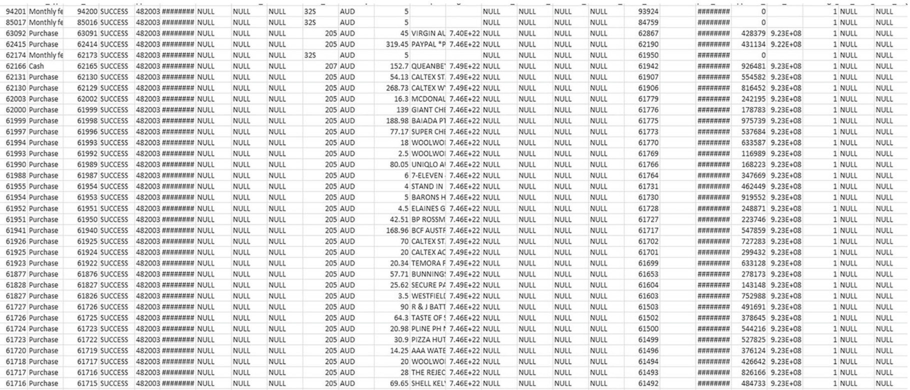
In this research, we used the SHAP framework for explainable AI, specifically in the context of anomaly detection within financial organizations. Based on the principles of cooperative game theory, SHAP ensures a consistent and equitable attribution of feature importance. In contrast to alternative global interpretability techniques such as feature importance and partial dependence plots and to local methods like LIME, SHAP offers both local and global interpretability. Its simultaneous consideration of transactional and systemic aspects provides valuable insights for comprehensive anomaly identification. The SHAP method effectively captures complex feature interactions while maintaining a model-agnostic approach. Whereas global approaches typically have limited local interpretative capabilities and LIME may produce inconsistent outcomes, SHAP emerges as a comprehensive and computationally efficient approach that is particularly well-suited to addressing the intricate complexities associated with financial anomaly identification (Elshawi et al., 2019; Khan et al., 2022). SHAP enables a unified approach to predictions regardless of the ML model employed, as Fig. 22 below shows in demonstrating the outcome. Here, the diagonals represent the SHAP values for the main effects, while the off diagonals represent the interaction effects. Similar to the beeswarm plot, the colours are determined by the feature value for



```
================================
KNeighborsClassifier
****Results****
Accuracy:       89.66
Precision:      90.28
Recall:         88.89
================================
GaussianNB
****Results****
Accuracy:       72.92999999999999
Precision:      91.35
Recall:         50.67
================================
LogisticRegression
****Results****
Accuracy:       68.22
Precision:      62.7
Recall:         89.92999999999999
================================
SVC
****Results****
Accuracy:       100.0
Precision:      100.0
Recall:         100.0
================================
DecisionTreeClassifier
****Results****
Accuracy:       100.0
Precision:      100.0
Recall:         100.0
================================
RandomForestClassifier
****Results****
Accuracy:       99.92999999999999
Precision:      100.0
Recall:         99.86
================================
GradientBoostingClassifier
****Results****
Accuracy:       99.89
Precision:      100.0
Recall:         99.77000000000001
```

```
================================
KNeighborsClassifier
****Results****
Accuracy:       89.64999999999999
Precision:      90.61
Recall:         88.47
================================
GaussianNB
****Results****
Accuracy:       67.80000000000001
Precision:      94.43
Recall:         37.82
================================
LogisticRegression
****Results****
Accuracy:       66.62
Precision:      60.79
Recall:         93.63
================================
SVC
****Results****
Accuracy:       100.0
Precision:      100.0
Recall:         100.0
================================
DecisionTreeClassifier
****Results****
Accuracy:       100.0
Precision:      100.0
Recall:         100.0
================================
RandomForestClassifier
****Results****
Accuracy:       99.96000000000001
Precision:      100.0
Recall:         99.92
================================
GradientBoostingClassifier
****Results****
Accuracy:       99.98
Precision:      100.0
Recall:         99.96000000000001
```

Fig. 16 Machine Learning algorithms score



94201	Monthly fe	94200	SUCCESS	482003	########	NULL	NULL	NULL	32S	AUD	5			NULL	NULL	NULL	NULL	93924		########	0		1	NULL	NULL
85017	Monthly fe	85016	SUCCESS	482003	########	NULL	NULL	NULL	32S	AUD	5			NULL	NULL	NULL	NULL	84759		########	0		1	NULL	NULL
63092	Purchase	63091	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	45	VIRGIN AU	7.40E+22	NULL	NULL	NULL	NULL	62867		########	428379	9.23E+08	1	NULL	NULL
62415	Purchase	62414	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	319.45	PAYPAL *P	7.46E+22	NULL	NULL	NULL	NULL	62190		########	431134	9.22E+08	1	NULL	NULL
62174	Monthly fe	62173	SUCCESS	482003	########	NULL	NULL	NULL	32S	AUD	5			NULL	NULL	NULL	NULL	61950		########	0		1	NULL	NULL
62166	Cash	62165	SUCCESS	482003	########	NULL	NULL	NULL	207	AUD	152.7	QUEANBE	7.49E+22	NULL	NULL	NULL	NULL	61942		########	926481	9.23E+08	1	NULL	NULL
62131	Purchase	62130	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	54.13	CALTEX ST.	7.49E+22	NULL	NULL	NULL	NULL	61907		########	554582	9.23E+08	1	NULL	NULL
62130	Purchase	62129	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	268.73	CALTEX W	7.49E+22	NULL	NULL	NULL	NULL	61906		########	816452	9.23E+08	1	NULL	NULL
62003	Purchase	62002	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	16.3	MCDONAL	7.46E+22	NULL	NULL	NULL	NULL	61779		########	242195	9.23E+08	1	NULL	NULL
62000	Purchase	61999	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	139	GIANT CHI	7.46E+22	NULL	NULL	NULL	NULL	61776		########	178783	9.23E+08	1	NULL	NULL
61999	Purchase	61998	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	188.98	BAIADA PT	7.46E+22	NULL	NULL	NULL	NULL	61775		########	975739	9.23E+08	1	NULL	NULL
61997	Purchase	61996	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	77.17	SUPER CHI	7.46E+22	NULL	NULL	NULL	NULL	61773		########	537684	9.23E+08	1	NULL	NULL
61994	Purchase	61993	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	18	WOOLWOI	7.46E+22	NULL	NULL	NULL	NULL	61770		########	633587	9.23E+08	1	NULL	NULL
61993	Purchase	61992	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	2.5	WOOLWOI	7.46E+22	NULL	NULL	NULL	NULL	61769		########	116989	9.23E+08	1	NULL	NULL
61990	Purchase	61989	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	80.05	UNIQLO AI	7.46E+22	NULL	NULL	NULL	NULL	61766		########	168223	9.23E+08	1	NULL	NULL
61988	Purchase	61987	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	6	7-ELEVEN	7.46E+22	NULL	NULL	NULL	NULL	61764		########	347669	9.23E+08	1	NULL	NULL
61955	Purchase	61954	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	4	STAND IN	7.46E+22	NULL	NULL	NULL	NULL	61731		########	462449	9.23E+08	1	NULL	NULL
61954	Purchase	61953	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	5	BARONS H	7.46E+22	NULL	NULL	NULL	NULL	61730		########	919552	9.23E+08	1	NULL	NULL
61952	Purchase	61951	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	4.5	ELAINES G	7.46E+22	NULL	NULL	NULL	NULL	61728		########	248871	9.23E+08	1	NULL	NULL
61951	Purchase	61950	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	42.51	BP ROSSM	7.46E+22	NULL	NULL	NULL	NULL	61727		########	223746	9.23E+08	1	NULL	NULL
61941	Purchase	61940	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	168.96	BCF AUSTF	7.46E+22	NULL	NULL	NULL	NULL	61717		########	547859	9.23E+08	1	NULL	NULL
61926	Purchase	61925	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	70	CALTEX ST.	7.49E+22	NULL	NULL	NULL	NULL	61702		########	727283	9.23E+08	1	NULL	NULL
61925	Purchase	61924	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	20	CALTEX AC	7.49E+22	NULL	NULL	NULL	NULL	61701		########	299432	9.23E+08	1	NULL	NULL
61923	Purchase	61922	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	20.34	TEMORA F	7.49E+22	NULL	NULL	NULL	NULL	61699		########	633128	9.23E+08	1	NULL	NULL
61877	Purchase	61876	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	57.71	BUNNINGS	7.49E+22	NULL	NULL	NULL	NULL	61653		########	278173	9.23E+08	1	NULL	NULL
61828	Purchase	61827	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	25.62	SECURE PA	7.49E+22	NULL	NULL	NULL	NULL	61604		########	143148	9.23E+08	1	NULL	NULL
61827	Purchase	61826	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	3.5	WESTFIELD	7.49E+22	NULL	NULL	NULL	NULL	61603		########	752988	9.23E+08	1	NULL	NULL
61727	Purchase	61726	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	90	R & J BATT	7.46E+22	NULL	NULL	NULL	NULL	61503		########	491691	9.23E+08	1	NULL	NULL
61726	Purchase	61725	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	64.3	TASTE OF S	7.46E+22	NULL	NULL	NULL	NULL	61502		########	378645	9.23E+08	1	NULL	NULL
61724	Purchase	61723	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	20.98	PLINE PH N	7.46E+22	NULL	NULL	NULL	NULL	61500		########	544216	9.23E+08	1	NULL	NULL
61723	Purchase	61722	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	30.9	PIZZA HUT	7.46E+22	NULL	NULL	NULL	NULL	61499		########	527825	9.23E+08	1	NULL	NULL
61720	Purchase	61719	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	14.25	AAA WATE	7.46E+22	NULL	NULL	NULL	NULL	61496		########	376124	9.23E+08	1	NULL	NULL
61718	Purchase	61717	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	20	WOOLWOI	7.46E+22	NULL	NULL	NULL	NULL	61494		########	426642	9.23E+08	1	NULL	NULL
61717	Purchase	61716	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	28	THE REJEC	7.46E+22	NULL	NULL	NULL	NULL	61493		########	826166	9.23E+08	1	NULL	NULL
61716	Purchase	61715	SUCCESS	482003	########	NULL	NULL	NULL	205	AUD	69.65	SHELL KEL	7.46E+22	NULL	NULL	NULL	NULL	61492		########	484733	9.23E+08	1	NULL	NULL

Fig. 17 Real-world dataset dictionaries

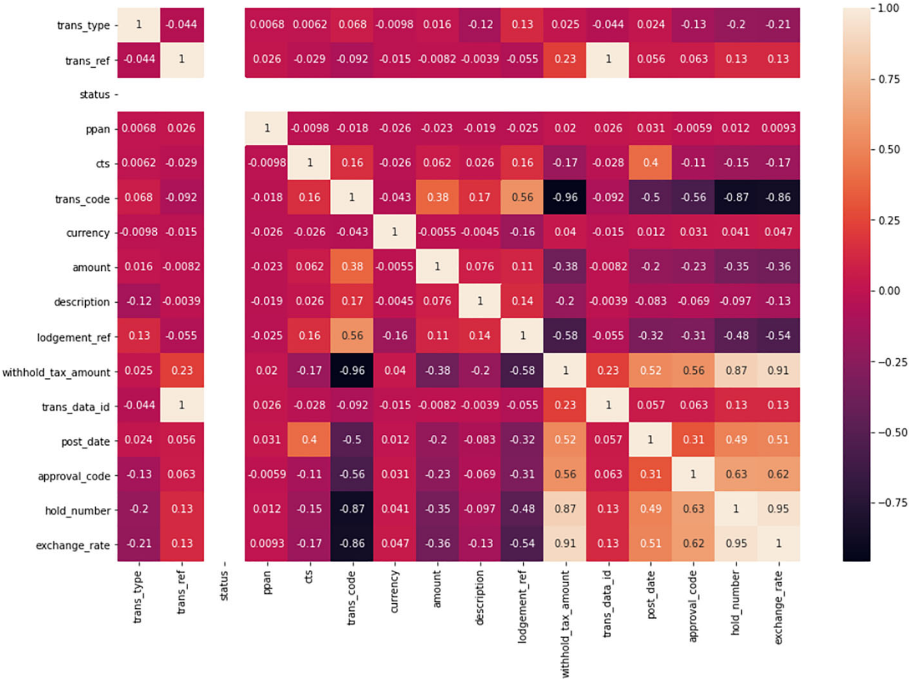


Fig. 18 The heatmap on real-world dataset detecting the untold fact

the y-axis feature. Examining interaction effects between variables inside an ML model provides significant insight into the collaborative influence of features on prediction outcomes. The analysed K-Means model reveals a notable interaction effect between the variables “trans_ref” and “description”. The significance and characteristics of this interaction effect

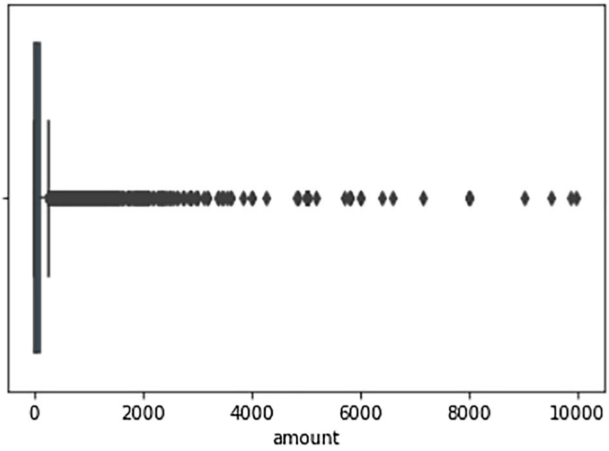


Fig. 19 The outliers in amounts of the transactions

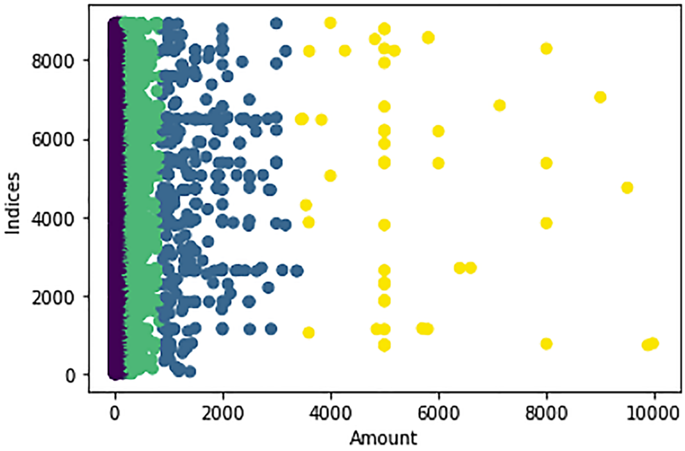


Fig. 20 Detecting anomalies in the transaction data

indicate that the association between these two variables greatly influences the model's predictive performance, exceeding what could be explained by examining each component separately.

4 Discussion

AI can equip financial managers with analytical tools to facilitate data-driven decision-making by leveraging their intuitive, social, and strategic intelligence regarding their interaction with socio-economic systems and tacit experience (Brock & Von Wangenheim, 2019; Morse, 2020). According to Choo (1991) uncertainty, complexity and ambiguity are the three obstacles that hamper organizational decision-making. However, data-driven decision-making

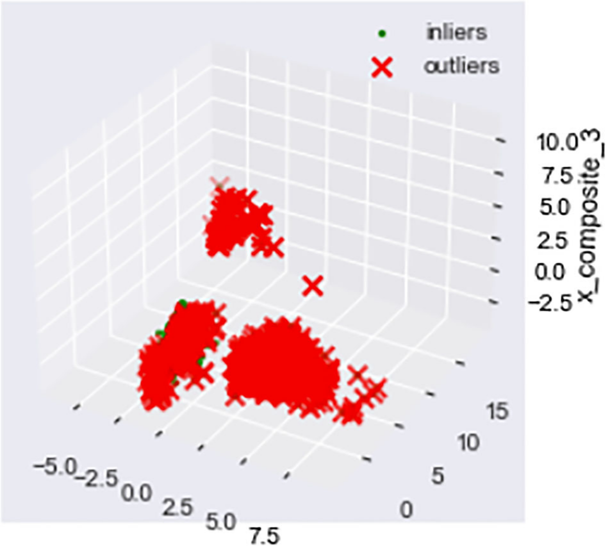


Fig. 21 Detecting anomalies in the amount of data

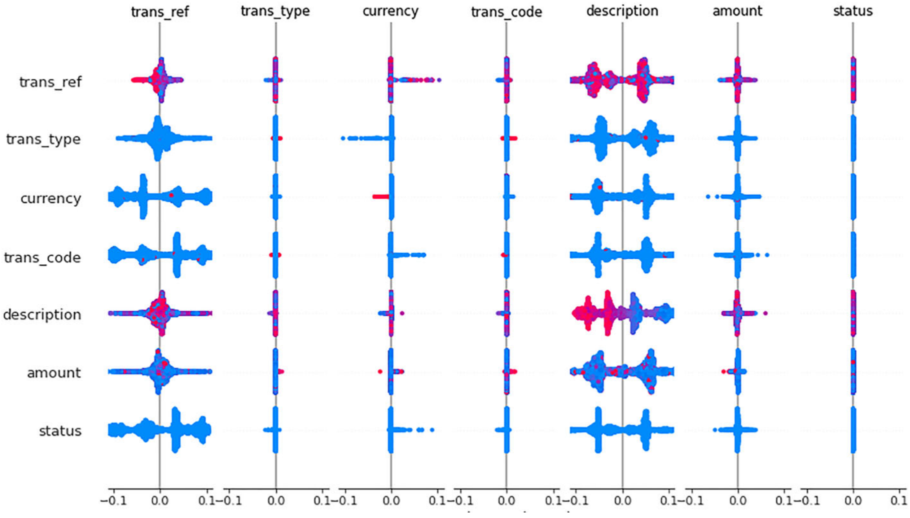


Fig. 22 Shap interaction value

also demands imaginative foresight, an awareness of an organisation's complex social and political processes, and social methods such as persuasion and bargaining to manage ambiguity and uncertainty. It seems improbable that AI, at its current level, can replicate human problem-solving in all these domains. Additionally, sense-making, i.e., interpreting information (Weick, 1995), and sense-giving, i.e., communicating the outcomes of sense-making, are of the utmost importance for managers aiming to make decisions based on data and analytics. However, opaque AI systems and a lack of AI literacy among managers make understanding the logic underlying the automatic recommendations AI systems give, challenging (Choudhury et al., 2020).

The preceding section's findings provide evidence of explainable AI and give significant insight into how the AI provides distinct complementing traits required for effective data-driven decision-making to detect anomalies for future fraud prevention. Analytical and intuitive methods are optimal for formulating data-based decisions (Hung, 2003; Martin & Martin, 2009). In this context, representing AI output responses derived from incorporated transparency will assist managers in comprehending these outputs' significance. Such understanding that results from AI transparency, will be crucial for strengthening trust in AI systems, allowing people to make more efficient and accurate judgments. Given the rise of AI-based solutions in financial business processes and practices (Vrontis et al., 2022), our research adds a paradigm for incorporating transparency into AI-based systems, which can boost managers' trust.

4.1 Contribution to theory

Our research expands the theoretical framework by applying the ADR lens to investigate explainable AI-based ML models for anomaly detection in finance. This perspective is intended to solve a fundamental difficulty posed by AI technologies that are difficult to comprehend and that prevents them from becoming a vital strategic resource for businesses. Adopting ADR highlights the significance of technology as a strategic resource for sustaining a competitive advantage in the research onion framework (Alturki, 2021; Bilandzic & Venable, 2011). Regardless of how effective the decision support technology or model may be, it will fail without managerial adoption and trust. Consequently, explainable AI-based ML models can enable the confident realization of their value, overcoming trust concerns connected to AI output in the managerial decision-making arena, which are caused by the opaqueness of AI algorithms (Andriopoulos & Lewis, 2009; Raisch & Krakowski, 2021). In this study, the ADR viewpoint moves the focus away from adopting new-age technologies and tools (AI systems) to the trustworthiness of AI technology (via transparency and explainability) as a strategic intangible resource for achieving sustained corporate competitiveness.

Explainable AI will enhance the capability of finance managers within organizations to comprehend, interpret, and explain automated outputs, enabling them to understand why anomalies occur and to devise effective prevention measures. It will also boost managers' confidence in the AI's output which delivers answers, thus producing new information for process and resource efficiency strategizing (Tambe et al., 2019). This will increase the organization's potential to develop capabilities that favourably impact employee and business productivity (Makarius et al., 2020). In addition, AI model defects can be easily recognized, evaluated, and systematically corrected, thereby enhancing the accuracy of output answers (Satell & Sutton, 2019). Lastly, explainable AI will enhance justice and accountability, as managers can explain the plans, practises, and policies developed based on AI output to the workforce.

4.2 Contribution to practice and policy

Our study provides practical insights for developing AI systems by emphasizing the imperative of incorporating cross-disciplinary teams. Teams consisting of data scientists, AI professionals and domain specialists play a crucial role in understanding technical and business aspects of AI applications. (In addition to domain experience, the implementation team needs to grasp the relevance of the data utilized in the analysis and to assess the correctness of the suggestions to evaluate the algorithms training process's efficacy (Keding, 2021). While the demand for data scientists, ML specialists, and robot engineers is rising, creativity, leadership, emotional intelligence, domain expertise, and tacit experience are the essential skills required to push AI transparency and ultimately to enhance ML algorithms (Correani et al., 2020; Jarrahi, 2018). This study refers to areas where AI can augment rather than replace people in decision-making, showing how collective complementary intelligence might evolve AI models through explainability resulting form AI transparency. Our findings lead to suggestions for organizations contemplating AI implementations, as follows:

Organizations should first evaluate the necessary decision-making activities, determine the essential skills and competencies required to execute these jobs, and finally make strategic decisions that divide the tasks between people and AI.

Investments in AI-literacy and AI-skills training for managers will help businesses gain the benefits of a human-AI symbiosis from explainable AI. This will help to consolidate and maximize the suitable use of human talents such as creativity, communication, empathy, negotiation, intuition, persuasion and negotiation, which are necessary for the organization to grow and which currently show up the limitations of AI (Davenport & Bean, 2017).

To strike a good balance between investing in intelligent technology and preserving established business procedures, organizations will need to understand which skills managers use in decision-making, as well as evaluate the AI limitations within a given environment (Davenport, 2018). Additionally, new AI governance mechanisms are required to ensure that automated decisions adhere to legislative criteria and are ethical. This will result in a redefinition and reconsideration of the decision-making process regarding accountability, rewards, risks, investments, and long-term viability (Kiron & Schrage, 2019). In the organizational policy context, we present novel governance methods to ensure that regulatory norms of AI outputs are ethical and adhered to. This will also lead to organizational decision-making procedures being reconfigured.

4.3 Limitations of the study and future research directions

Our research is not without limitations. Primarily, emphasizing the financial sector raises concerns about the applicability of our findings to other businesses, which indicates a need for further investigation. Further, managers' different levels of AI literacy could be a notable obstacle to the efficient implementation and use of AI technologies in organizational decision-making.

These constraints point to numerous possible directions for further research. A compelling research direction is to expand the scope of our explainable AI models to encompass a wide range of industries. Thereby our AI models' effectiveness in various organizational settings could be assessed. A further area of potential exploration is one examining the impact of managers' AI literacy levels on adopting and proficiently utilizing explainable AI systems.

Additionally, we recommend that future research endeavours investigate these ML models' potential efficacy in analysing diverse datasets to reveal latent information or anomalies. This could be relevant in cases where outcomes are predetermined (observed in supervised learning models) or entirely uncertain (observed in unsupervised learning models).

The recent progress in deep learning technology presents compelling opportunities, particularly in its potential to uncover previously concealed information in financial transactions by identifying intricate hidden patterns. This promising investigation opportunity particularly concerns the dynamic field of financial technology. In summary, our study adds significantly to the current body of literature and also stimulates further investigation, presenting various avenues for both theoretical and empirical enquiry.

5 Conclusion

The globalization of multinational corporations and their activities have resulted in strategic management and retention of human resources becoming crucial factors in organizations' overall performance and productivity. Despite the interest in and claims surrounding the benefits of AI systems in financial processes, creating regulatory compliance, detecting fraud, improving investment appraisal, and reducing operational costs and risks, we have limited research on how to achieve explainable AI (Budhwar & Malik, 2020; Budhwar et al., 2022). To adopt AI in management decision-making, developing the foundations of trust in these systems is required. This has become a focal point in finance and general finance research literature. This article contributes to the finance literature by presenting an implementation framework that demonstrates the use of SHAP in explaining the illustrative instance of AI-based ML models' anomaly detection. Thereby we demonstrate a method to enhance AI transparency and explainability. These explanations will increase the dependability and credibility of AI-based ML models among finance managers and executives, thus improving the effectiveness and efficiency of data-driven strategic financial decision-making to produce long-term value for organizations.

AI technologies are used extensively in data analysis, producing different classification and regression techniques. However, having reviewed the related works, we structured our experiment to explain how AI technologies are used transparently to prepare ML algorithms in detecting anomalies for decision-making. We conclude our work with a transparent assessment and brief explanation of the proposed ML-based analytical artefacts model, the proposed model's results, and the implications this study has for further research.

We developed a new explainable AI technology, an ML-based artefact to detect anomalies, as shown through the ML framework, to improve our knowledge of how anomalies can be detected using synthetic and real-world datasets. With the transparent AI technology, utilizing supervised and unsupervised ML techniques, we explained both synthetic and real-world datasets and developed ML-based artefacts. The experimental case studies offered methodological steps that any design researcher could follow in designing data analytics. We recognized that the SVC and decision tree methods presented the most accurate results in the training and testing the synthetic dataset's data. This ML framework can be put to industrial use and used in academic research because all the ML-based algorithms are transparent on GitHub for its reusability and further applications (https://github.com/MachineLearning-UON/ML-Financial_transaction.git).

Gregor and Hevner (2013) provided four quadrants to categorize research design, variously named improvement, invention, routine design, and exaptation. The first research type develops a new solution for known problems (providing research opportunities and knowledge contribution); the second invents new solutions for new issues (again providing research opportunities and knowledge contribution); the third applies known solutions to known problems (providing no major knowledge contribution); the fourth extends known solutions to new situations (also providing research opportunities and knowledge contribution). According to these four study contribution quadrants, this research falls in the first, the improvement quadrant. Our study, therefore, proposes a new solution for known problems, which is to detect anomalies in customer transactions in financial organizations.

The study's contribution is three-fold: First, limited research has attempted to capture financial transaction behaviour. We investigated a synthetic data set and a real-world dataset of a financial organization in capturing financial transactions to detect anomalies using ADR methodology in the research onion framework. Second, we proposed a five-step ML framework that should enable new researchers in this field to use a similar analytical design using supervised and unsupervised ML. Third, we developed an explainable AI technology-based ML model with algorithms in which seven supervised ML routines are carried out, i.e., we used the KNeighbors classifier, GaussianNB, Logistic Regression, SVC, Decision Tree classifier, Random Forest classifier, Gradient Boosting classifier, and two unsupervised ML systems – K-Means and Isolation Forest. This is an ML problem-solving solution strategy for developing further analytical solutions. We, therefore, argue that this paper makes a unique contribution.

Acknowledgements The author(s) received financial support, access to an operational payment platform, and access to data for the research from the Novatti Group Ltd, Melbourne, Australia. The listed third author contributed to the study by supplying anonymized transactional big data without disclosing human identities or personal information, so that none would be compromised. All researchers ensured the ethical integrity of the research.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflict of interests The author(s) declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. *Journal of Network and Computer Applications, 68*, 90–113.

Ahmed, M., Choudhury, N., & Uddin, S. (2017). Anomaly detection on big data in financial markets. Paper presented at the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).

Ahmed, M., Mahmood, A. N., & Islam, M. R. (2016). A survey of anomaly detection techniques in financial domain. *Future Generation Computer Systems, 55*, 278–288.

Alalie, H. M., Harada, Y., & Noor, I. M. (2018). A resource-based view: How information technology creates sustainable competitive advantage to improve organizations. *Journal of Advance Management Research, 6*(12), 1–5.

Al-Anqoudi, Y., Al-Hamdani, A., Al-Badawi, M., & Hedjam, R. (2021). Using machine learning in business process re-engineering. *Big Data and Cognitive Computing, 5*(4), 61.

Alturki, R. (2021). Research onion for smart IoT-enabled mobile applications. *Scientific Programming, 2021*, 1–9.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989.

Andriopoulos, C., & Lewis, M. W. (2009). Exploitation–exploration tensions and organizational ambidexterity: Managing paradoxes of innovation. *Organization Science, 20*(4), 696–717.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82–115.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE, 12*(7), e0180944.

BBC. (2020). Apple's 'sexist' credit card investigated by US regulator. Retrieved September 20, 2020, from https://www.bbc.com/news/business-50365609

Belizón, M. J., & Kieran, S. (2022). Human resources analytics: A legitimacy process. *Human Resource Management Journal, 32*(3), 603–630.

Bieda, L. (2020). How organizations can build analytics agility. *MIT Sloan Management Review*, Issue October, 2020, Available: https://sloanreview.mit.edu/article/how-organizations-can-build-analytics-agility/

Bilandzic, M., & Venable, J. (2011). Towards participatory action design research: Adapting action research and design science research methods for urban informatics. *Journal of Community Informatics, 7*(3), 1–15.

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Bishop, C. M., & Nasrabadi, N. M. (2007). Pattern recognition and machine learning. *Journal of Electronic Imaging*, *16*(4), 049901.

Boxall, P. (1996). The strategic HRM debate and the resource-based view of the firm. *Human Resource Management Journal, 6*(3), 59–75.

Brock, J.K.-U., & Von Wangenheim, F. (2019). Demystifying AI: What digital transformation leaders can teach you about realistic artificial intelligence. *California Management Review, 61*(4), 110–134.

Bromiley, P., & Rau, D. (2016). Operations management and the resource based view: Another view. *Journal of Operations Management, 41*, 95–106.

Budhwar, P., & Malik, A. (2020). Call for papers for the special issue on Leveraging artificial and human intelligence through Human Resource Management. In *Human Resource Management Review*. Retrieved June, 24, 2020.

Budhwar, P., Malik, A., De Silva, M. T., & Thevisuthan, P. (2022). Artificial intelligence–challenges and opportunities for international HRM: A review and research agenda. *The International Journal of Human Resource Management, 33*(6), 1065–1097.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics, 57*, 203–216.

Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020). Supporting human autonomy in AI systems: A framework for ethical enquiry. In C. Burr & L. Floridi (Eds.), *Ethics of digital well-being* (pp. 31–54). Springer.

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News, 538*(7623), 20.

Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

Chang, W.-H., & Chang, J.-S. (2010). Using clustering techniques to analyze fraudulent behavior changes in online auctions. Paper presented at the 2010 International Conference on Networking and Information Technology.

Chang, J.-S., & Chang, W.-H. (2012). A cost-effective method for early fraud detection in online auctions. Paper presented at the 2012 Tenth International Conference on ICT and Knowledge Engineering.

Chen, K., Zhai, X., Sun, K., Wang, H., Yang, C., & Li, M. (2021). A narrative review of machine learning as promising revolution in clinical practice of scoliosis. *Annals of Translational Medicine, 9*(1), 67.

Cheng, M. M., & Hackett, R. D. (2019). A critical review of algorithms in HRM: Definition, theory, and practice. In *Academy of management proceedings* (Vol. 2019, No. 1, p. 18018). Academy of Management.

Choo, C. W. (1991). Towards an information model of organizations. *The Canadian Journal of Information Science, 16*(3), 32–62.

Chowdhury, S., Budhwar, P., Dey, P. K., Joel-Edgar, S., & Abadie, A. (2022b). AI-employee collaboration and business performance: Integrating knowledge-based view, socio-technical systems and organisational socialisation framework. *Journal of Business Research, 144*, 31–49.

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2022a). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review, 33*, 100899.

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2022c). Embedding transparency in artificial intelligence machine learning models: Managerial implications on predicting and explaining employee turnover. *The International Journal of Human Resource Management, 34*, 1–32.

Ciechanowski, L., Jemielniak, D., & Gloor, P. A. (2020). TUTORIAL: AI research without coding: The art of fighting without fighting: Data science for qualitative researchers. *Journal of Business Research, 117*, 322–330.

Correani, A., De Massis, A., Frattini, F., Petruzzelli, A. M., & Natalicchio, A. (2020). Implementing a digital strategy: Learning from the experience of three digital transformation projects. *California Management Review, 62*(4), 37–56.

Davenport, T. H. (2018). From analytics to artificial intelligence. *Journal of Business Analytics*, *1*(2), 73–80.

Davenport, T. H., & Bean, R. (2017). How P&G and American express are approaching AI. *Harvard Business Review*, 1–6. Available at https://hbr.org/2017/03/how-pg-and-american-express-are-approaching-ai

Elshawi, R., Al-Mallah, M. H., & Sakr, S. (2019). On the interpretability of machine learning-based model for predicting hypertension. *BMC Medical Informatics and Decision Making, 19*(1), 1–32.

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and Engineering Ethics, 26*(6), 3333–3361.

Glancy, F. H., & Yadav, S. B. (2011). A computational model for financial reporting fraud detection. *Decision Support Systems, 50*(3), 595–601.

Grandinetti, J. (2021). Examining embedded apparatuses of AI in Facebook and TikTok. *Ai & Society, 38*, 1–14.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly, 37*, 337–355.

Gunasekaran, A., Papadopoulos, T., Dubey, R., Wamba, S. F., Childe, S. J., Hazen, B., & Akter, S. (2017). Big data and predictive analytics for supply chain and organizational performance. *Journal of Business Research, 70*, 308–317.

Gunning, D., & Aha, D. (2019). DARPA's explainable artificial intelligence (XAI) program. *AI Magazine, 40*, 44–58. https://doi.org/10.1609/aimag.v40i2.2850

Gupta, S., Modgil, S., Bhattacharyya, S., & Bose, I. (2022). Artificial intelligence for decision support systems in the field of operations research: Review and future scope of research. *Annals of Operations Research, 308*, 1–60.

Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review, 61*(4), 5–14.

Huang, D., Mu, D., Yang, L., & Cai, X. (2018). CoDetect: Financial fraud detection with anomaly feature detection. *IEEE Access, 6*, 19161–19174.

Hung, S.-Y. (2003). Expert versus novice use of the executive support systems: An empirical study. *Information & Management, 40*(3), 177–189.

Hwang, S., Kim, J., Park, E., & Kwon, S. J. (2020). Who will be your next customer: A machine learning approach to customer return visits in airline services. *Journal of Business Research, 121*, 121–126.

Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*. https://doi.org/10.1007/s10479-021-04187-w

Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons, 61*(4), 577–586.

Jarrin, E. P., Cordeiro, F. B., Medranda, W. C., Barrett, M., Zambrano, M., & Regato, M. (2019). A machine learning-based algorithm for the assessment of clinical metabolomic fingerprints in Zika virus disease. Paper presented at the 2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI).

Kafai, Y. (1996). Learning through artifacts: Communities of practice in classrooms. *AI & SOCIETY, 10*(1), 89–100.

Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons, 63*(1), 37–50.

Keding, C. (2021). Understanding the interplay of artificial intelligence and strategic management: Four decades of research in review. *Management Review Quarterly, 71*(1), 91–134.

Kersting, K., & Meyer, U. (2018). From big data to big artificial intelligence? Algorithmic challenges and opportunities of big data. *KI-Künstliche Intelligenz, 32*, 3–8.

Khan, W., Walker, S., & Zeiler, W. (2022). Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy, 240*, 122812.

Kiron, D., & Schrage, M. (2019). Strategy for and with AI. *MIT Sloan Management Review*, *60*(4), 30–35.

Lasaga, D., & Santhana, P. (2018). Deep learning to detect medical treatment fraud. Paper presented at the KDD 2017 Workshop on Anomaly Detection in Finance.

Le Khac, N. A., & Kechadi, M.-T. (2010). Application of data mining for anti-money laundering detection: A case study. Paper presented at the 2010 IEEE International Conference on Data Mining Workshops.

Lee, T. H., & Boynton, L. A. (2017). Conceptualizing transparency: Propositions for the integration of situational factors and stakeholders' perspectives. *Public Relations Inquiry, 6*(3), 233–251.

Magomedov, S., Pavelyev, S., Ivanova, I., Dobrotvorsky, A., Khrestina, M., & Yusubaliev, T. (2018). Anomaly detection with machine learning and graph databases in fraud management. *International Journal of Advanced Computer Science and Applications, 9*(11), 33–38.

Makarius, E. E., Mukherjee, D., Fox, J. D., & Fox, A. K. (2020). Rising with the machines: A sociotechnical framework for bringing artificial intelligence into the organization. *Journal of Business Research, 120*, 262–273.

Mardiana, S. (2020). Modifying research onion for information systems research. *Solid State Technology, 63*(4), 5304–5313.

Martin, R., & Martin, R. L. (2009). *The design of business: Why design thinking is the next competitive advantage*. Harvard Business Press.

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management, 58*(3), 103434.

Morse, G. (2020). Harnessing artificial intelligence. *Harvard Business Review*. October, 23, 2021.

Motiwalla, L. F., Albashrawi, M., & Kartal, H. B. (2019). Uncovering unobserved heterogeneity bias: Measuring mobile banking system success. *International Journal of Information Management, 49*, 439–451.

Müller, V. C. (2020). Ethics of artificial intelligence and robotics.

Munk, A. K., Olesen, A. G., & Jacomy, M. (2022). The thick machine: Anthropological AI between explanation and explication. *Big Data & Society, 9*(1), 20539517211069892.

Nelson, G. S. (2019). Bias in artificial intelligence. *North Carolina Medical Journal, 80*(4), 220–222.

Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569.

Pourhabibi, T., Ong, K.-L., Kam, B. H., & Boo, Y. L. (2020). Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems, 133*, 113303.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*(1), 137–141.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210.

Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence, 76*, 130–157.

Šabić, E., Keeley, D., Henderson, B., & Nannemann, S. (2021). Healthcare and anomaly detection: Using machine learning to predict untold fact in heart rate data. *AI & SOCIETY, 36*(1), 149–158.

Sample, I. (2017). Computer says no: Why making AIs fair, accountable and transparent is crucial. *The Guardian, 5*, 1–15.

Sariannidis, N., Papadakis, S., Garefalakis, A., Lemonakis, C., & Kyriaki-Argyro, T. (2020). Default avoidance on credit card portfolios using accounting, demographical and exploratory factors: Decision making based on machine learning (ML) techniques. *Annals of Operations Research, 294*(1), 715–739.

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-driven cybersecurity: An overview, security intelligence modeling and research directions. *SN Computer Science, 2*(3), 1–18.

Satell, G., & Sutton, J. (2019). We need AI that is explainable, auditable, and transparent. *Harvard Business Review*.

Saunders, M., Lewis, P., & Thornhill, A. (2007). Research methods. *Business Students 4th Edition Pearson Education Limited, England, 6*(3), 1–268.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly, 35*, 37–56.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277–284.

Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review, 61*(4), 66–83.

Shrestha, Y. R., Krishna, V., & von Krogh, G. (2021). Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. *Journal of Business Research, 123*, 588–603.

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review, 61*(4), 15–42.

Vrontis, D., Christofi, M., Pereira, V., Tarba, S., Makrides, A., & Trichina, E. (2022). Artificial intelligence, robotics, advanced technologies and human resource management: A systematic review. *The International Journal of Human Resource Management, 33*(6), 1237–1266.

Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. *Business Process Management Journal, 26*(7), 1893–1924.

Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Sage.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, *5*(2), 171–180.

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security, 57*, 47–66.

Yang, G., Ye, Q., & Xia, J. (2022). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Information Fusion, 77*, 29–52.

Yaram, S. (2016). Machine learning algorithms for document clustering and fraud detection. Paper presented at the 2016 International Conference on Data Science and Engineering (ICDSE).

Zhou, F., Ayoub, J., Xu, Q., & Jessie Yang, X. (2020). A machine learning approach to customer needs analysis for product ecosystems. *Journal of Mechanical Design, 142*(1), 1–13.