International Journal of Computing Sciences Research (ISSN print: 2546-0552; ISSN online: 2546-115X)
Vol. 9, pp. 3364-3391
doi: 10.25147/ijcsr.2017.001.1.223
https://stepacademic.net

Long Paper

Building Trust in Predictive Analytics: A Review of ML Explainability and Interpretability

Shruthi Sajid
EEMCS, University of Twente, Netherlands
shruthisajid97@gmail.com

Jeewanie Jayasinghe Arachchige
Department of Computer Science, Vrije University, Netherlands
j.jayasinghearachchige@vu.nl
(corresponding author)

Faiza Allah Bukhsh
EEMCS, University of Twente, Netherlands
f.a.bukhsh@utwente.nl

Abhishta Abhishta
BMS, University of Twente, Netherlands
s.abhishta@utwente.nl

Faizan Ahmed
EEMCS, University of Twente, Netherlands
faizan.ahmed@utwente.nl

Date received: June 19, 2024
Date received in revised form: September 30, 2024
Date accepted: October 12, 2024

Recommended citation:

Sajid, S., Arachchige, J. J., Bukhsh, F. A., Abhishta, A., & Ahmed, F. (2025). Building trust in predictive analytics: A review of ML explainability and interpretability. *International Journal of Computing Sciences Research, 9*, 3364-3391. https://doi.org/10.25147/ijcsr.2017.001.1.223

Abstract

Purpose – The purpose of the manuscript is to explore the previous literature to reveal the trust and interpretability of predictive analytical models that use ML /AI techniques.

Method – The methodology applied for the study is the guidelines of Kitchenham et al. (2007).

Results – The results reveal that past research explicitly discussed the usage of predictive analytics. However, ML models are considered black boxes and suffer from transparency. The study proposed a typical process to ensure that predictions made by AI/ML models can be interpreted and trusted.

Conclusion – The literature review conducted predictive analytics and AI/ML techniques in business decision-making, highlighting their usage in industries. The study reveals a significant gap exists in research on the explainability and interpretability of these ML models within a business context.

Recommendations – Recommended the need for more research on transparency and interpretability of ML models by developing sector-specific explainability frameworks to bridge technical insights and business decisions. Further, it is recommended to integrate ethical and regulatory considerations into explainability frameworks and study collaboration methods between AI/ML experts and business leaders to align ML models with business goals.

Research Implications – The research highlights the significant gap in the literature explainability and interpretability of ML and AI models in the business context. Therefore the research stresses the need for future investigations into improving model transparency and creating industry-specific and ethical frameworks that help organizations derive more meaningful, trusted, and interpretable insights from data-driven models.

Practical Implications – It should focus on improving transparency, trust, and collaboration in using predictive analytics. By addressing explainability issues and incorporating ethical, regulatory, and industry-specific considerations, businesses can more effectively use the power of AI and ML to drive data-informed decisions.

Social Implications – This study highlights the importance of ethical and regulatory concerns related to AI and ML, such as data privacy, and fairness.

Keywords – explainability, interpretability, machine learning, predictive analytics, trust

INTRODUCTION

Data in an organization is table stakes. However, the key to data management is when organizations can make the most optimal use of the data amassed to generate insights for product improvements, track business performance, and drive efficiency overall. It also implies that the whole concept of big data isn't novel anymore. Data has the potential to change how businesses operate and seamlessly implement data-driven solutions. However, with a growing demand in the advent of trends in technology and IT systems, the speed and efficiency at which data is generated is increasing by the second. Owing to the massive proliferation of data, there is an intense need to establish a data governance standard that allows the automation of data to be managed, accessed, and interpreted conveniently.

Slowly, organizations across the globe are moving towards leveraging Artificial Intelligence/Machine Learning (AI/ML) techniques and tools at every step of their operations to optimize business value, improve performance, and drive efficiency. With such technologies in place, organizations can pull all the stops to extract value, deliver business insights, automate tasks, and make data-informed business decisions. While the umbrella term Artificial Intelligence can include Expert Systems, Machine Vision, Natural Language Processing, and Speech Recognition, among others, Machine Learning is indeed the core of it. ML algorithms have proven to be advantageous in areas such as medical diagnosis, financial risk prediction, safety evaluation of robots, aircraft manufacturing, and so on (Fu et al., 2021). These algorithms are designed to acquire knowledge from real historical data and generate actionable insights to facilitate decision-making. Data-driven insights not only help automate business operations but also identify key drivers for growth and ways to increase business revenue.

Predictive Analytics is one of the fields of advanced analytics that makes use of AI and ML models for data-informed decision-making. The core of predictive analytics is to make predictions, and forecast activity, behavior, and trends about future outcomes using real and historical data coupled with data mining and machine learning techniques. The approach not only helps find useful patterns in data but also identifies risks and opportunities for the business. The research hypothesis (Wach & Chomiak-Orsa, 2021) claims that predictive analytics can support strategic decision-making by detecting projects endangered by misappropriate budget execution. This helps the board of management in making decisions that focus on minimizing negative consequences and liabilities. Likewise, predictive analytics is also widely used in forecasting sales predictions for a business by analyzing historical data, trends, behavior, etc. Organizations rely on forecasts to make informed decisions, foresee market performances, and manage resources like cash flow, project funds, plans, inventory, workforce, inventory, etc. (Ayyagari, 2018). Since organizations typically rely on software tools, expert systems, and algorithms, the quality of a business decision and what it entails depends on the quality of the forecast (Lackes et al., 2020). Moreover, forecasting also involves analyzing critical and sensitive business data. It is of paramount importance to bring in a component of trust so that business leaders can make the right decisions to drive overall business performance.

With rapid advances in analytics and cognitive capabilities in IT, algorithmic decision-making is critical in influencing business decisions and deciding the granularity of information people are exposed to. While it is indeed a revolution in how data is processed and used in organizations, it also dramatically increases the complexity of breaking down and understanding the insights of model predictions. The paper (Lampathaki et al., 2021) mentioned that despite the many benefits that AI and ML algorithms can bring about, humans typically have zero to little insight concerning the knowledge of how these systems make any decisions or predictions due to their "black-box" effect. The paper of Burkart and Huber (2021) states the same idea, highlighting that predictions obtained by AI algorithms have high accuracy. However, humans often perceive the models as black boxes. Often, the training data that the model learns from could be skewed and riddled with errors, introducing components of bias and unfairness in results. Since these black boxes suffer from a certain degree of opacity while explaining themselves, it often raises the question of how to 'trust' the predictions that they make (Adadi & Berrada, 2018). This adds to the growing concern about handing over the task of critical decision-making to algorithms that suffer from interpretability and explainability. An example can be seen in the field of healthcare- care where the adoption of an AI-based medical diagnosis system hasn't gained traction among healthcare professionals due to the lack of interpretability and explainability in model predictions (Fan et al., 2020; Cheng et al., 2021). It could also be attributed to a lack of trust in how the model makes certain predictions. In Kaur et al. (2022), the authors claim that AI/ML models can be deemed trustworthy provided they have a certain degree of fairness, explainability, accountability, reliability, and acceptance.

With a rise in speculations across different sectors on the nuances of entrusting expert systems and algorithms to making key decisions, there has also been an increase in governance frameworks and processes to address these concerns. To this end, the Defense Advanced Research Projects Agency (DARPA), responsible for the development of emerging technologies for use by the military of the United States, curated an Explainable Artificial Intelligence (XAI) program (Gunning et al, 2021). The program aims to create a suite of new or modified machine-learning techniques that produce explainable models. When these models are combined with effective explanation techniques, they help end users understand, trust, and effectively manage IT systems. Likewise, Europe has adopted the GDPR, an ambitious set of comprehensive regulations for collecting, storing, and using personal information, with the hopes of making the European Union fit for the digital age. The law spans many provisions, some of which are related to automated decision-making.

Under the GDPR, Article 22, the Right to Explanation says that 'The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.' (GDPR, 2018). This means that subjects impacted by algorithmic decisions have the right to request to be informed by organizations on how algorithms have made automated decisions (Goodman & Flaxman, 2017). It gives every individual the choice to engage or disengage from something if they find the decisions are biased or unfair. While the law aims to clear the disparity in decision-making, it also puts organizations and their business at stake in a plethora of ways. To expect the company and its IT team to explain why their algorithm has made a certain decision seems close to impossible. This is because an AI or ML algorithm trains itself on massive amounts of data while performing multiple bouts of micro-decisions and complex mathematical equations. A human expert, even a programmer, would find it extremely difficult to explain the accurate reasoning behind an algorithm's exact functioning. This warrants the need to establish a strong and well-structured governance framework that organizations can adopt to interpret and explain the decisions made by the algorithms. While there exists a wealth of literature on the quantification of interpretability and explainability for models, the quantification of trust as a component is fairly new and unexplored.

Therefore, one of the aims of this literature review is to examine the usage of predictive analytics in a business context and how an organization can use state-of-the-art techniques like AI and ML algorithms to derive insights from data. The caveat to be highlighted is the fact that it doesn't end with just making predictions and having an exhaustive set of insights. The key to driving business efficiency is understanding the insights, trusting the predictions, and making informed decisions as the next steps. To this end, this research's objective is three-fold. First, it will focus on understanding predictive analytics as a technique across various industries. It then progresses to identify the AI and ML techniques that are applied to predict future outcomes based on historical business data and the most used frameworks followed by industries to do the same. This is followed by a focus on understanding the watchwords 'explainability', 'interpretability', and 'trust' in predicting outcomes.

The remainder of the paper is structured as follows: The next section describes the research methodology followed for the literature study along with the research questions, the strategy followed to select articles from the different databases, etc. This is followed by the Section "Results", which discusses relevant literature and results obtained as part of the literature study. Following this, "Discussion" attempts to answer the research questions by summarizing the results obtained from the study. The section "Practical and Theoretical Implications" discusses the research gap identified and the steps to address it as part of the forthcoming research. Finally, the conclusion and recommendation are provided.

METHODOLOGY

The research methodology applied for the study is a Systematic Literature Review (SLR) based on the guidelines of Kitchenham et al. (2007). This research methodology aims to summarize existing evidence, identify gaps in existing research, and provide a framework or approach to include new research possibilities. It also focuses on a search strategy that is less likely to be biased with the search results. This section discusses the research method and search strategy conducted to identify and analyze relevant studies. It also explains the research questions identified along with the search queries used in the databases, including- inclusion and exclusion criteria and quality assessment of included studies.

Research Questions

The research questions aim to understand the existing work in the area of predictive analytics, forecasting, explainable artificial intelligence, and interpretable machine learning. In addition to revealing existing works related to the interpretability and explainability of AI/ML approaches, the research questions aim to identify the research gaps, which can be directed to future research as well. The research questions formulated for the study are:

- RQ1: What is the definition of predictive analytics in existing literature?
- RQ2: How do available AI/ML techniques support predictive analytics?
- RQ3: What methodologies exist in the literature to make data-informed business decisions using predictive analytics?
- RQ4: How are the explanations provided by AI and ML in model predictions validated and trusted?

To answer these questions, relevant literature articles are identified from different databases and digital libraries, collated, and analyzed to understand the scope of current work in the areas mentioned above.

Search Strategy

The search strategy employed in this review focuses on gathering literature from the databases. Among the 106 databases in the repository, IEEE Xplore, Scopus, and Web of Science are chosen as the most relevant to the study.

Once the relevant databases and digital libraries are selected for the study, the next step is to create a search string with relevant keywords. To ensure the results are inclusive and accurate, the search string has been formulated by experimenting with multiple combinations of keywords and synonyms relevant to the scope of the study. The query is as follows:

("Predictive analytics" OR "Forecasting" OR "Predictive business process management" OR "Predictive process monitoring") AND ("Business decisions" OR "Actionable insights" OR "Data-informed decisions")

Using the string mentioned above, the search has been performed on the title, abstract, and keywords in the digital libraries. The search string returned a total of 755 articles from the three databases. Zotero (Zotero, n.d), an open-source easy-to-use tool to help with the collection, organization, annotation, and citation of research was used for this purpose. The collected articles have been exported to Zotero to remove duplicates and screening. It is implied that the search results have to be carefully filtered again based on specific search boundaries to get the right results and take the study forward. Selection criteria have been carefully formulated to narrow down the search results.

Inclusion and Exclusion Criteria

The studies selected thus far are based on a simple keyword search on abstract, title, and keywords which returned many results. Next, an attempt has been made to concise the search results further with the following boundaries:

- Range of Publication Years: 2010 to 2022
- Subject Area: Business and Computer Science

The range of publication years for the search strategy has been chosen by keeping in mind the inception of the Explainable Artificial Intelligence (XAI) program in 2015 by DARPA (Gunning et al, 2021). The search string returned approximately 450 articles from the three selected databases. To further scope it down, a comprehensive set of Inclusion Criteria (ICs) and Exclusion Criteria (ECs) has been formulated. The Inclusion Criteria (ICs) are:

- The paper is in English.
- The source type of material should be conference papers and journal articles.
- The paper is available for download.
- The paper should complement the scope of the study.
- The paper should include modifications to an existing approach or introduce new means to approach the scope of the study.
- The paper should address one or more research questions.

The Exclusion Criteria (ECs) are:

- Titles with little or no similarity to the scope of the study.
- Papers with limited access or only abstracts available
- “Articles in Press” status of Publication Stage.
- Papers with abstracts that have little or no similarity to the scope of the study.

The criteria mentioned above have been carefully applied to screen the remaining papers, while also reading their abstracts. This resulted in approximately 83 articles to review and analyze further.

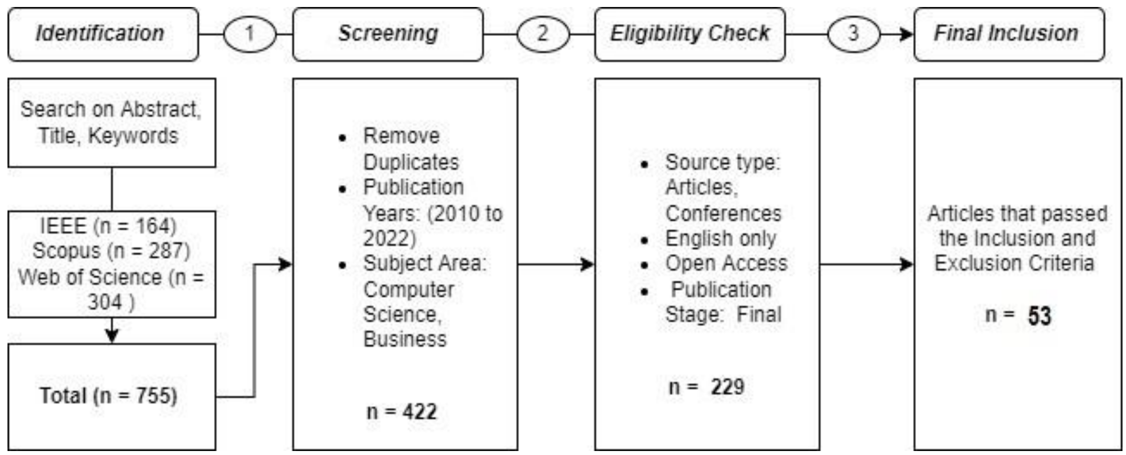

Figure 1. Literature selection process

Data Extraction Strategy

The search strategy used for the literature review can be seen in Fig.1. In order to answer the four search questions mentioned in the "Research Questions "section, it has been decided to formulate an inclusive data extraction strategy that helps with extracting relevant information from the literature.

1. For RQ1: it is important to understand if there exists a general-purpose or industry-standard definition of predictive analytics.
2. For RQ2: literature materials have been screened to understand the different techniques employed by AI/ML-based models to make predictions and generate actionable insights.
3. For RQ3: which is also an extension of RQ2, the methodologies applied, their limitations and the validation of results have been analyzed.
4. For RQ4: papers that explicitly discuss explainability, interpretability, and trust have been analyzed. It is also critical to categorize the various strategies that have been used to validate the degree of explainability in the models.

RESULTS

This section presents an overview of the findings from the literature consumed over the course of the systematic literature review. This will help answer the research questions mentioned in the "Research Questions" section.

Understanding Predictive Analytics

Simply put, 'Analytics' in general is a process of discovering, analyzing, interpreting, and explaining significant data trends and patterns. This makes it easy for key stakeholders and decision-makers to consume the data in a digestible way and make better business decisions. Four different types of analytical models co-exist and supplement each other. The researchers (Attri et al., 2021) and (Mello & Martins, 2019) classify them as Descriptive analytics, Diagnostic analytics, Predictive Analytics, and Prescriptive analytics. Summarizing the findings from the three materials, it can be stated that,

- Descriptive analytics, commonly used in management reports for sales, marketing, and finance focuses on past facts to identify previous successes and failures.
- Diagnostic analytics, as the name suggests, not only attempts to explain events that happened but also goes a step further to explain the reason behind them.
- Predictive analytics couples historical data with rules and advanced algorithms to determine the likelihood of an event occurring in the future.
- Prescriptive analytics follows closely behind and attempts to determine a course of action based on the findings.

The core of forecasting outcomes and trends is analyzing data from the past and present. In (Haran & Moore, 2014), the authors explain that forecasting can be done using point prediction where an attempt is made to guess what the future will hold. This seems highly unlikely and often leads to inaccurate results. Another possible method is to identify a range of plausible estimates. This leaves room for some margin of error by guessing the range within which the possible outcome resides. However, predictive analytics goes a step further than merely guessing. It often applies machine learning to predict the future states of a running business process (Hsieh et al., 2021). For example, a simple sales margin forecast for specific products can be done by analyzing data from the previous year. By observing the trend, the number of products to be supplied to the market can be determined. However, with predictive analytics and advanced algorithm-based decision-making, relevant key drivers, patterns, trends, and other insights can be identified to the extent of generating a potential customer base for specific products.

Predictive analytics and machine learning are often confused as the same in the industry. Predictive analytics, at its core, employs statistics on both historical and current data to estimate or forecast future events. These statistical approaches include machine learning, predictive modeling, and data mining. The paper (Yefimenko, 2018) defines predictive analytics as a process that includes multiple steps. The initial steps revolve around understanding the scope of the project, setting expectations with relevant stakeholders, getting familiar with the data, and performing exploratory data analysis. This is followed by employing statistical techniques, creating machine learning models, and making predictions. Predictive analytics has a suite of models and algorithms it can make use of for predictions. The most used ones as specified in the literature (Yefimenko, 2018; Zytek et al., 2021) are mentioned in the next section.

Use of ML in Predictive Analytics

Classification Models

Classification machine learning uses a set of rules according to which a new object can be assigned to the relevant class (Wach & Chomiak-Orsa, 2021). In the context of predictive analytics, the classification model predicts a class label for a given example of input data.

Decision Trees: A decision tree is one of the supervised learning methods that makes use of a tree-like approach for decisions and demonstrates the chance of event outcomes (Mahbooba, 2021). Decision trees are also recognized as one of the most popular interpretable methods for supervised classification, expressing models in terms of if-then rules (Vélez et al., 2020). This algorithm builds predictive models from a given set of observations by following a top-down approach with a decision node at the top and subsequent decision and leaf nodes. The root nodes are significant predictors, and the leaf nodes help with making the final decision. The internal nodes have both outgoing and incoming branches in contrast to the leaves (terminal nodes), which are assigned to probable target variables (Wach & Chomiak-Orsa, 2021).

A decision tree generates rules to predict the value of a target variable based on some splitting criteria. Some of the most used algorithms are ID3 and CART. They make use of statistical metrics such as information gain, entropy, and GINI to compute how well attributes separate training examples based on the target data.

Support Vector Machine: Support Vector Machine (SVM) is a binary classifier that can separate a set of n-dimensional (where n is the number of features) data points into two classes (Wach & Chomiak-Orsa, 2021). This can be useful in the context of predictive analytics as it takes groups of observations and constructs boundaries to predict which class future observations belong to. Each data item is plotted as a point in the n-dimensional space, following which optimal hyper-plane attempts to segregate the classes. It makes use of non-linear kernel methods or similarity functions to transform input data to a high-dimensional feature space. This is called the Kernel trick. The goal of SVM is to maximize the margin or distance between the two classes. SVM is also quite versatile as it can be used for both Classification and Regression problems. They're most advantageous when working with high dimensional data i.e. when the number of potential predictors is large and flexible with any shape including linear, radial, and polynomial, among others.

Time Series Model

A time series model is often referred to as a collection of data points that has time as the input parameter to make scientific predictions. Most likely, the input parameter is historical time-stamped data which makes it easier to predict future outcomes and observations by identifying patterns. It involves building models through this historical analysis to drive future strategic decision-making (Lin & Ying, 2011), (Dael et al., 2022).

The papers (Shaikh, 2021; Kumar et al., 2021a; Poleneni et al., 2021) explicitly discuss the use of time series analysis for the prediction of COVID-19 outbreak in the future. Owing to the exponential and rapid speed of the disease, it has been increasingly difficult to control and analyze the situation with merely human resources and observations. The authors claim that it has become obligatory to work on an automated decision-making model to make informed decisions and ease the spread of the virus. In the three papers, future forecasts are predicted by analyzing the COVID-19 epidemic occurrence and employing relevant algorithms to derive conclusions. As per the literature, the different types of time series models include:

Autoregressive (AR): AR models are regression models where the dependent or response variable is a linear function of past values of the dependent or response variable. AR typically restores the values from a variable belonging to the previous periods as input for the current regression equation (Poleneni et al., 2023). This equation, in turn, predicts the output for the upcoming period.

Moving Average (MA): The MA model is one of the types of time series models that accounts for the possibility of a relationship between a variable and the residuals from the preceding periods (Poleneni et al., 2021). Unlike AR, here the dependent or response variable is a linear function of past values of the error term. MA computes the average of a subset of numbers. This is done multiple times for several subsets of data. However, this may be disadvantageous as it always assumes that the trend is linear. As a result, it cannot be used to forecast long-term time series data (Kumar et al., 2021b).

Autoregressive Moving Average (ARMA): ARMA is a combination of AR and MA. In this model, the dependent or response variable is a linear function of past values of both the dependent or response variable and the error term.

Autoregressive Integrated Moving Average (ARIMA): ARIMA is a generalization of the ARMA model as it enhances the features of AR and MA models. In comparison with the other models, ARIMA performs better in terms of accuracy (Kumar et al., 2021b). In an ARIMA model, the future value of a variable is assumed to be a linear function of several past observations and random errors. ARIMA includes three iterative steps (Zheng & Zhong, 2011). They are as follows:

1. Model Identification
2. Parameter Estimation
3. Diagnostic Checking

This three-step process is repeated until a satisfactory model is generated for predictions.

Explainability, Interpretability, and Trust

Predictive analytics, underpinned by advanced AI and ML techniques, provides business process intelligence in organizations. While the most commonly used metric to measure the predictive capability of a model is accuracy, the resulting model is still considered a ’black box’ when presented to business stakeholders (Wickramanayake et al., 2022), (Sindhgatta et al., 2020b). One can think of an ML model as a black box when the model lacks transparency and is uninterpretable. On the contrary, a model can also be called a glass box when the predictions are fully interpretable and true to its name. Being unable to interpret results from a black box model can be a roadblock for businesses as the models are unable to provide further insights into why a certain business process prediction was made (Zytek et al., 2021).

Within the ML community, despite the many attempts by researchers to distinguish ‘interpretability’ and ‘explainability’, the two terms are still used interchangeably. The authors of Interpretability (Rosenfeld & Richardson, 2019) presented the relationship between Explainability and Interpretability as follows.

1. Explainability - The ability to discover meaning between input data and model outputs i.e. to take an ML model and explain the behavior in human understandable terms.
2. Interpretability - To be able to understand the inner workings of a model i.e. to understand exactly why and how the model is making predictions.

Recent studies in (Alhomsi & Vivacqua, 2021), and (Sokol & Flach, 2020) have identified two types of explainability namely local and global explainability. In local explainability, an explanation behind an individual decision or prediction is provided, while in global explainability only a single explanation is given for the whole dataset. Added to this, explanation techniques can also be classified as ante-hoc and post-hoc. Ante-hoc methods such as linear regression, decision trees, and random forest have a degree of explainability incorporated in the model itself while post-hoc techniques rely on other models to be trained and provide explanations.

In addition to the above observations, there are several dimensions mentioned in (El-Khawaga et al., 2022) to explainability that can be used to validate the explanations that the model predicts. These are as follows:

1. How to explain?
   This relates to how a predictive model derives predictions based on the inputs. It can be done in the form of a proxy model, feature importance, or visualization.
2. How much to explain?
   Relates to the level of granularity to which models can be explained i.e. local explanations and global explanations.

3. How to present?
This relates to choosing ways to present an explanation to the stakeholders based on the characteristics of the end user, their level of expertise, the scope of the explanation, and the purpose it brings forth. It can be done by visual explanations, verbal explanations, or analytical explanations.
4. When to explain?
Relates to the point in time an explanation should be provided to stakeholders. This can be intrinsic, where explainability is introduced during the process of building the model, or post-hoc where explainability is established on the basis of model outcomes.

The next sub-sections list some techniques in the existing literature to introduce explainability, interpretability, and trust which are commonly used.

Feature Engineering

In an ML life cycle, the steps leading up to predicting a desired outcome can include a variety of iterations such as gathering and preparing data cleaning and removing irrelevant features, and finally presenting the improved results. In ML, a 'feature' is typically considered as a measurable input that is used in predictive models to predict future outcomes. These features, when dealt with the right way, can be very beneficial leading to improvements such as boosting predictive results, decreasing computational times, reducing excessive noise, and increasing decision-making transparency (Chatzimparmpas et al., 2022), (Krause et al., 2014). This is commonly known as Feature Engineering. Simply put, it is the process of converting raw observations and insights into desired features using statistical and machine-learning approaches. Likewise (Chatzimparmpas et al., 2020), StackGenVis is a system that helps users dynamically adapt performance metrics, managing data instances and selecting the most important features for a given dataset. To get the most accurate results, the importance of choosing the right features is emphasized largely. This is done in 3 ways:

1. Univariate: Identical for all models but different for each feature.
2. Permutation: Observe how random re-shuffling of each predictor influences model performance.
3. Accuracy: Similar to Permutation, it removes features one by one but retrains the model based on the accuracy as feedback.

The authors (Chatzimparmpas et al., 2022) divided the processes of feature engineering into four phases:

1. Feature Ideation: This is ideally the first step in feature engineering where new features are created from the raw data.
2. Feature Generation: This follows the ideation phase and supports the creation of more relevant features from the combination of already existing ones. Typical interactions during this process are addition, subtraction, multiplication, and division.

3. Feature Transformation: This phase includes modifications over the features such as binning, scaling, logarithmic transformations, etc.
4. Feature Selection: This phase selects a subset of features from the pool of features that are available. Some of the methods for this are filter, wrapper, embedded, and hybrid.

Visual Analytics

With the many advances in state-of-the-art IT systems backed by AI and ML technologies, humans are more dependent on expert systems for decision-making. However, even with these advancements, the lack of transparency and explainability regarding decisions made by these models is an increasing concern. It also induces a sense of fear because of the inability to fully comprehend the internal workings of the models and their decisions. Recent research claims that visual metaphors seem to address the problem of interpreting ML algorithms and go beyond judging a model's performance simply based on its accuracy score (Krause et al., 2016). Similar to how humans have excelled in understanding data using visualization techniques, it is natural to apply the same principle to interpreting ML models. Visual analytics can help with understanding the inner workings of a model, extract information from a model (post-hoc interpretability), and enable performance diagnosis for building accurate models. According to (Kaouni et al., 2021), a visual analytics platform can be created as a web application with the help of Python programming language and libraries such as Pandas, PM4py, or Plotly.

One such approach can be seen in (Xie et al., 2029), where the authors claim that visualizations combined with causal relations can help with exploring and validating decisions made by the models. Added to this, there is a causal graph explanation that supports a set of intuitive user controls to perform what-if analyses and make action plans. This system consists of three components, the data processing component for processing high-dimensional data, the causal detection component for computing the causal graph, and the visualization component for supporting the causal graph exploration and what-if analysis. This type of system is introduced to improve a user's confidence about the results from an ML model and thereby take actionable decisions.

Likewise, a visual analytics tool developed by (Legg et al., 2019), demonstrates the benefit of human-benefit collaboration that promotes transparency, inspection, understanding, and trust in the learning process. This approach is typically called active learning, where at first the machine is trained on a small sample of labels given as input by the humans. As and when it crosses paths with new unlabeled data, the machine has the choice to classify the new data or query the human for a class label based on its confidence in classification.

Local Interpretable Model-agnostic Explanation (LIME)

Black-box models are often recognized as complex and not straightforward to interpret while a certain degree of trust is placed in machines to predict future outcomes that can drive the efficiency of an organization, it is also important to understand the underlying mechanics of the model and the predictions that come with it. Research and current literature explore ways in which explainability methods can interpret the inner workings of a black-box model without having direct access to it (Lin et al., 2019). To this end, post-hoc explanations provide insights after a model is trained. One way to do this is by using the Local Interpretable Model-Agnostic Explanation model (LIME). Here, model-agnostic refers to the flexibility that comes with the application of LIME. Staying true to the name, LIME is model-agnostic, such that it can be applied to any machine learning model. The technique attempts to understand the model by perturbing the input of data samples and understanding how the predictions change. LIME attempts to offer a solution for the trade-off between interpretability and performance i.e. complex models that can handle large and versatile data sets and less complex models that are much easier to interpret but fall short on performance. The paper (Chowdhury et al., 2023) elucidates the benefits of using LIME in model interpretability and explainability. They are as follows:

1. Interpretability: LIME provides an intuitive understanding of the relation between input variables and model response in a way that's convenient even without ML expertise. This is most beneficial for business stakeholders who would like to dive deep into the results to make actionable decisions.
2. Local Fidelity: LIME focuses specifically on how the model behaves in the vicinity of the individual observation being predicted. This may lead to only a handful of variables that relate to a local/individual prediction, even if a model has hundreds of variables globally.
3. Model-Agnostic: As explained earlier, LIME can explain any model and treats all of them as black boxes.

The authors of (Ribeiro et al., 2016) explain the core concepts of LIME with the help of an example. Predictions made by expert systems need to be validated and trusted before making decisions based on the outcomes.

Existing Frameworks

Machine Learning algorithms aim to bring a component of discipline to decision-making because they uncover relevant factors that humans might tend to overlook. With recent breakthroughs in various disciplines, organizations, and industries are leaning towards automated decision-making more than ever. However, to ensure the decisions made by these systems are tangible and can be applied to a real-world context, frameworks and methodologies need to be put in place. This section describes the most commonly used methodologies to build interpretability and trust according to the results of the literature review.

Fairness, Accountability, and Transparency (FAT)

True to its name, FAT is responsible for ensuring the algorithm backing a decision stay Fair, Accountable, and Transparent. This framework primarily came into existence to tackle ethical challenges and biases in automated decisions. The research of Shin and Park (2019) presents a conceptual model of FAT as an antecedent variable affecting satisfaction while trust is a moderator influencing this relationship. For example, in the case of expert and recommendation systems, transparency and fairness are key indicators in algorithms to build trust amongst users.

According to (Zhdanov et al., 2022), the three components of FAT are defined as follows:

1. Fairness: To ensure a lack of bias in model prediction. Bias can occur in two ways data bias and model bias. Data bias refers to the dataset serving as input and model bias is a result of the unfair influence of model fitting.
2. Accountability: To ensure prediction accuracy and coverage of data points with reliable predictions. Accountability bridges the gap between an algorithm generating a high number of unreliable predictions and a lower number of high-quality predictions.
3. Transparency: Relates to the function of the modeling method. This focuses specifically on the black box models that can generate accurate predictions but cannot explain the logic underlying the predictions.

While prediction accuracy is important for decision-making, it also requires an increase in computational effort and complexity. It boils down to the business leaders having to arrive at a trade-off and decide on the level of accuracy that is satisfactory for business requirements. The paper (Zhdanov et al., 2022) also suggests an iterative approach to creating fair, accountable, and transparent models for business.

1. To tackle transparency, select a general modeling method such as linear models and decision trees that are more transparent.
2. Formulate a set of predictors with a workable solution.
3. Set of checks based on accountability conditions such as prediction accuracy, coverage, model stability, replicability, etc.
4. Next, include fairness criteria to check for bias. The suggested approach is to remove outliers and partition the dataset into relevant subsets for calculation. Re-processing the dataset and repeating the computation of predictors with different approaches could also be another way to deal with model bias.

Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM is a methodology that focuses on aligning data mining processes with business goals. It is an iterative approach with opportunities to evaluate the progress of the project and ensures business goals remain the core of the project rather than an afterthought. In addition to this, it serves as a roadmap by offering best practices for data engineers, analysts, and business stakeholders to adhere to and carry out a project (Bohanec et al., 2017). Authors Wirth and Hipp (2000) mentioned the CRISP-DM model can be seen as a hierarchical process model with four levels of abstraction: phases, generic tasks, specialized tasks, and process instances. The specialized task goes one level further and explains how the actions in the generic tasks should be carried out in specific scenarios. The fourth level is more of a record of actions, decisions, and results of the whole project. The lifecycle of a typical data mining project in CRISP-DM has 6 phases as listed below.

1. Business Understanding: The first phase focuses on understanding the core problem, objectives, and requirements and then aligning them in accordance with the strategy and goals of the business. A detailed plan can also be during this phase.
2. Data Understanding: The second phase revolves around getting familiar with the data, discovering insights, and checking the quality.
3. Data Preparation: In this phase, the raw data will be manipulated into a form that can be analyzed and used as input to the ML model.
4. Modeling: This phase is to select various modeling techniques to be applied to the prepared and pre-processed data.
5. Evaluation: Before deploying the model, it's important to evaluate the model and check if business objectives have been met. This phase determines if there are gaps that the model fails to bridge while aligning with business goals.
6. Deployment: In the last phase, the models are pushed into production by running them on a live environment. This makes the model's predictions available to users, developers, or systems, so they can make business decisions based on data, interact with their application, and so on.

Likewise, Studer et al. (2021) propose a process model for the development of machine learning applications by making use of the CRISP-DM methodology as the baseline. It is specifically designed for ML models that are developed and maintained as part of a product or service. The proposed process model, CRISP-ML(Q) with the addition of quality assurance, covers the six phases of CRISP-DM. CRISP-ML(Q) extends the scope of process models with an additional phase, monitoring, and maintenance, to address risks of model degradation in a changing environment. Further, the literature also has various quality measures such as robustness, scalability, explainability, model complexity, and resource demand to understand the success of the ML models.

The two approaches mentioned above, have two different focuses however they act as complementary. FAT focuses on ethical principles and ensuring that ML models are fair, transparent, and accountable, which is critical in building trust and addressing ethical concerns. Othe other hand, CRISP-DM is a framework for aligning the data mining or ML process with business goals, ensuring that technical projects deliver meaningful outcomes while maintaining structure and repeatability. In practice, they can use both together.

DISCUSSION

This section discusses the most important observations and is structured according to the research questions mentioned in the section “Research Questions”. This serves as the preliminary identifier for the research gap identified.

RQ1: What is the definition of predictive analytics according to the literature?

Modern technological advances have given way to the automation of data-driven decision-making backed by predictive analytics and machine learning techniques. In the context of business process management, predictive analytics is observed as the process of predicting future observations of a business by learning from historical data and events of the past (Sindhgatta et al., 2020a). This literature review observed several definitions of predictive analytics. Authors (Wach & Chomiak-Orsa, 2021), defined predictive analytics as a process that uses advanced mathematical formulas, statistical algorithms, as well as IT tools and services to identify dependencies, relationships, and patterns in data sets and reduce their complexity. In simpler terms, predictive analytics, with the help of historical data, statistical algorithms, and machine learning techniques, aim to identify the likelihood of future outcomes.

RQ2: How do AI/ML techniques support predictive analytics according to the literature?

Predictive analytics combined with AI/ML techniques and algorithms facilitate organizations' progress up the Business Intelligence (BI) maturity curve. These analytics solutions operate in real time to improve performance and drive efficiency by predicting future outcomes based on trends and patterns observed in the data. Models are typically trained to learn from previous iterations using specialized algorithms to produce reliable and accurate insights and results.

The available techniques in the literature support predictive analytics, ML algorithms such as decision trees, and support vector machines trained on historical data to provide outcomes for business decisions. The decision tree is a qualitative way of performing analysis and forecasting as it does not require a lot of statistical information but relies more on the intuition of decision factors. Support vector machines, on the other hand, can be used for data classification to obtain the most classified surface to make easier predictions. In addition to these algorithms, outside the scope of the identified literature, there are various algorithms to choose from, such as XGBoost, Random Forest, KMeans, etc. Another interesting approach to predictive analytics is understanding how data changes over time and the direction in which it changes. This can be done using Time series analysis where clean and time-stamped data is used to identify trends and patterns in historical data. This helps analysts and other relevant stakeholders identify fluctuations, outliers, seasonal variations, residuals, and more from the data.

RQ3: What methodologies exist in the literature to make data-informed business decisions using predictive analytics?

It's implied that the driving force behind good business decision-making lies in the experience and instincts of business leaders. Now, with organizations transitioning to the cloud and making progress with digital transformation, research confirms that businesses that base decisions on data are likely to be more trustworthy. This is powered by automated algorithmic decisions that involve the use of statistical models to output results and influence decisions.

In the literature, there were many references to methodologies that could help with making data-informed business decisions. However, the scope of the literature aims to address the basis on which these predictions are translated to actionable decisions by businesses. This is made possible by ensuring stakeholders can trust the predictions and are easily interpretable. For example, Carta et al. (2021) attempt to predict stock market forecasting by exploiting news and domain-specific lexicon. They propose an approach as follows:

1. Feature Engineering: Create an extended set of features by extracting data from the news.
2. Make use of ML-based predictive algorithms to create forecasts.
3. The decision tree algorithm provides explanations of the outcomes and predictions to improve explainability.
4. The approach is validated through an experimental study.

Similar to the above-mentioned approach, the summary of the findings for research question 3 has been visualized in Figure 2. It shows the typical process followed to ensure predictions made by AI/ML models can be interpreted and trusted to make informed decisions.

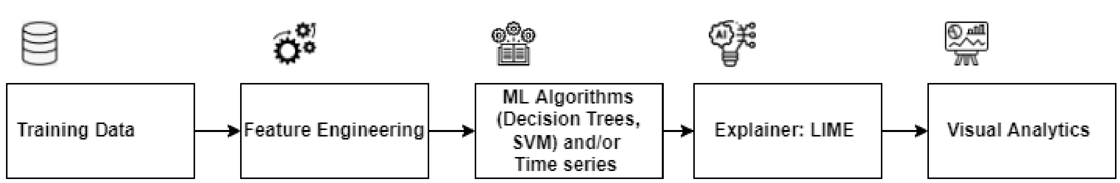

Figure 2. The typical lifecycle for Interpretability and Trust

RQ4: How are the explanations provided by AI and ML in model predictions validated and trusted?

According to the literature, most of the results are validated by relevant stakeholder and expert interviews along with case studies. In addition to this, two frameworks have been identified as part of the literature review that help conform the explanations provided by AI and ML in a standardized way. They are FAT and CRISP-DM. While the former is more focused on ensuring the algorithms stay accountable to the results they provide, the latter is a systematic and streamlined process that's followed in a typical project lifecycle that involves ML and AL algorithms. The validation of model explanations can also be facilitated via techniques such as feature engineering, LIME, and relevant visual analytics of the data. In addition to this, through the literature review, table 1 summarizes how the results of model predictions can be presented to business stakeholders (Mueller et al., 2019; Alhomsi & Vivacqua, 2021).

Table 1. Explanations for Business Stakeholders

Format (How it is expressed)	**Reference(*What it is about*)**
Visualization e.g. Heatmaps	Examples include misclassifications, counter-examples, outliers, clear cases, and close competitors.
Text (Statements, Narratives or Stories, Answers to queries, Human-machine dialogs)	Patterns, Classes, Ontologies
Formal Expressions (Logical expressions, Matrices)	Features, Weights, Probabilities, Ranks, Parameters
Conceptual Process Models (Diagrams)	Decisions, Strategies, Goals
Graphs, Networks	Algorithms, Computational Processes, Proofs
Tables	Incidents, Events (includes self-explanations or stories)
Abstractions, generalizations	Cause-effect relations
Timelines	
Hierarchies (Trees)	

PRACTICAL AND THEORETICAL IMPLICATIONS

This section discusses the research gap identified as part of the literature review, along with the future research focus. The advancement of AI and ML in model predictions and high-stakes decision-making has seen great success as organizations across the globe are slowly adopting them as the norm. The adoption of these advanced analytics and algorithms has also faced skepticism from stakeholders for its impediments. In a typical AI/ML context, when investors, consumers, business stakeholders, and end-users come together to make informed decisions based on a model, the first question that comes up is 'How much can I trust the predictions that a model makes?'

While the scope of the literature review spans various topics like predictive analytics, explainability, interpretability, and trust in model predictions, limitations still need to be bridged. Most of the literature materials identified as part of this review offer significant opportunities in predictive analytics by providing means to analyze, diagnose, finetune, and improve models and the predictions that come with them. However, one of the key weaknesses identified in the existing literature is that the interpretations are not extracted in a way that will be meaningful in a business context i.e. there is little effort on understanding model explanations in a way that will drive efficiency for businesses and their revenue. Although most of the materials focus on forecasting and predictions, understanding the nuances of black-box algorithms, they don't necessarily discuss this in light of how the business can derive insights. The dream is to tap into key drivers as a result of insights generated so that business stakeholders can easily make sense of them, look into financing options, and demand planning for operations.

In addition, the literature review lacks the discovery of formal applications of the FAT framework and CRISP-DM in business or information systems research. Most of the existing literature highly focuses on the need for a governance framework without going into the specifics of how these algorithms can actually conform to them. Therefore, this research gap creates a necessity that assists organizations in interpreting and trusting the AI/ML approaches that are used in decision-making.

Social Implications

Beyond business impacts, AI/ML adoption has important social implications:

1. **Bias and Fairness:** AI models, if not transparent, can perpetuate bias, particularly in areas like finance and healthcare. Future research should explore how explainability can reduce bias and ensure fairness across different demographic groups.
2. **Ethical Accountability:** With AI systems making critical decisions, ethical accountability becomes essential. Research is needed to investigate how businesses can implement frameworks to ensure transparency and ethical compliance in AI-driven decisions.
3. **Workforce Impact:** AI adoption may lead to job displacement. Future studies should explore strategies for balancing automation with employment, such as reskilling programs that support workforce adaptation.
4. **Trust and Public Perception:** Trust in AI is crucial for its acceptance. Research should focus on how explainable AI can improve public understanding and trust in AI systems, especially as they become more integrated into daily life.

CONCLUSION AND RECOMMENDATIONS

The literature review has been conducted to understand predictive analytics and how AI/ML techniques and algorithms are being leveraged to make data-informed business decisions. The research questions defined in the previous section have been answered with the results obtained as part of the literature review. It is observed that plenty of literature exists that explicitly discusses the usage of predictive analytics in industries such as automotive and healthcare. In addition, ML models considered as black boxes for the lack of transparency are addressed along with the complexity of interpreting them. Several approaches were observed to address this issue such as feature engineering, visual analytics, and relevant explainer algorithms. However, a significant gap was identified: the lack of research focusing on the explainability and interpretability of ML models specifically within a business context. Given the importance of transparent and interpretable models for decision-making in areas such as investment, resource allocation, and operational efficiency, this is a critical area for further exploration. To address this, future research should prioritize the development of a robust framework or set of guidelines that business leaders can rely on to interpret automated predictions confidently.

Recommendations for future research include:

1. **Development of Business-Specific Explainability Frameworks:** Research should focus on creating frameworks tailored to different business sectors that ensure ML models are not only accurate but also interpretable by non-technical stakeholders. This would help bridge the gap between technical insights and actionable business decisions.
2. **Evaluation of Frameworks Across Diverse Business Scenarios:** Studies should evaluate the effectiveness of these frameworks in various business contexts, such as finance, retail, and logistics, to ensure that they can be generalized across industries.
3. **Integration of Ethical and Regulatory Considerations:** As transparency becomes increasingly important, future research should also investigate how explainability frameworks can integrate ethical and regulatory guidelines, ensuring compliance with emerging laws such as AI transparency requirements.
4. **Collaboration Between AI/ML Experts and Business Leaders: Further** studies should examine how collaboration between technical and business teams can improve the alignment of ML models with business goals, ultimately making predictions more actionable

ACKNOWLEDGEMENT

Not applicable

FUNDING

The study did not receive funding from any institution.

DECLARATIONS

Conflict of Interest

The researcher declares no conflict of interest in this study.

Informed Consent

Not applicable. No personal data is used in this research.

Ethics Approval

No ethical approval is required as the research doesn't involve human data or any company data.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138-52160.

Alhomsi, R., & Vivacqua, A. S. (2021). The Explainable Business Process (XBP)-An Exploratory Research. *Revista de Informática Teórica e Aplicada, 28*(1), 78-96. https://doi.org/10.22456/2175-2745.107964

Attri, V., Batra, I., & Malik, A. (2021, April). A Relative Study on Analytical Models. In 2021 *2nd International Conference on Intelligent Engineering and Management* (ICIEM) (pp. 52-57). IEEE. https://doi.org/10.1109/ICIEM51511.2021.9445372

Ayyagari, R. (2018, October). Comparative Analysis of Different Forecasting Techniques for Ford Mustang Sales Data. In 1st *African International Conference on Industrial Engineering and Operations Management*. https://doi.org/10.46254/AF01.20180139.

Bohanec, M., Robnik-Šikonja, M., & Borštnar, M. K. (2017). Organizational learning supported by machine learning models coupled with general explanation methods: A Case of B2B sales forecasting. *Organizacija*, 50(3), 217-233. https://doi.org/10.1515/orga-2017-0020

Burkart, N., & Huber, M.F.(2021). A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, *70*, 245-317, https://doi.org/10.48550/arXiv.2011.07876

Carta, S. M., Consoli, S., Piras, L., Podda, A. S., & Recupero, D. R. (2021). Explainable machine learning exploiting news and domain-specific lexicon for stock market forecasting. *IEEE Access*, *9*, 30193-30205. https://doi.org/10.1109/ACCESS.2021.3059960

Chatzimparmpas, A., Martins, R. M., Kucher, K., & Kerren, A. (2020). StackGenVis: Alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Transactions on Visualization and Computer Graphics, 27*(2), 1547-1557. https://doi.org/10.1109/TVCG.2020.3030352

Chatzimparmpas, A., Martins, R. M., Kucher, K., & Kerren, A. (2022). Featureenvi: Visual analytics for feature engineering using stepwise selection and semi-automatic extraction approaches. *IEEE Transactions on Visualization and Computer Graphics, 28*(4), 1773-1791. https://doi.org/10.1109/TVCG.2022.3141040

Cheng, F., Liu, D., Du, F., Lin, Y., Zytek, A., Li, H., ... & Veeramachaneni, K. (2021). Vbridge: Connecting the dots between features and data to explain healthcare models. *IEEE Transactions on Visualization and Computer Graphics, 28*(1), 378-388. https://doi.org/10.1109/TVCG.2021.3114836

Chowdhury, S., Joel-Edgar, S., Dey, P. K., Bhattacharya, S., & Kharlamov, A. (2023). Embedding transparency in artificial intelligence machine learning models: managerial implications on predicting and explaining employee turnover. *The International Journal of Human Resource Management, 34*(14), 2732-2764., https://doi.org/10.1080/09585192.2022.2066981

Dael, F. A., Yavuz, U., & Almohammedi, A. A. (2022, March). Performance evaluation of time series forecasting methods in the stock market: A comparative study. In *2022 International Conference on Decision Aid Sciences and Applications* (DASA) (pp. 1510-1514). IEEE. https://doi.org/10.1109/DASA54658.2022.9765177

El-Khawaga, G., Abu-Elkheir, M., & Reichert, M. (2022). Xai in the context of predictive process monitoring: An empirical analysis framework. Algorithms, 15(6), 199.

Fan, W., Liu, J., Zhu, S., & Pardalos, P. M. (2020). Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). *Annals of Operations Research, 294*(1), 567-592.

General Data Protection Regulation (GDPR). (2018). *Article 22 GDPR: Automated individual decision-making, including profiling*. GDPR-info.eu. https://gdpr-info.eu/art-22-gdpr/

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine, 38*(3), 50-57. https://doi.org/10.48550/arXiv.1606.08813

Gunning, D., Vorm, E. S., Wang, Y., & Turek, M. (2021). *Darpa's explainable AI (xai) program: a retrospective*. https://doi.org/10.22541/au.163699841.19031727/v1

Haran, U., & Moore, D. A. (2014). A better way to forecast. *California Management Review, 57*(1), 5-15. https://doi.org/10.1525/cmr.2014.57.1.5

Hsieh, C., Moreira, C., & Ouyang, C. (2021, October). Dice4el: interpreting process predictions using a milestone-aware counterfactual approach. In *2021 3rd International Conference on Process Mining* (ICPM) (pp. 88-95). IEEE. https://doi.org/10.48550/arXiv.2107.08697

Kaouni, A., Theodoropoulou, G., Bousdekis, A., Voulodimos, A., & Miaoulis, G. (2021). Visual analytics in process mining for supporting business process improvement. In *Novelties in Intelligent Digital Systems* (pp. 166-175). IOS Press.

Kaur, D., Uslu, S., Rittichier, K. J., & Durresi, A. (2022). Trustworthy artificial intelligence: a review. *ACM computing surveys (CSUR), 55*(2), 1-38. https://doi.org/10.1145/3491209

Kitchenham, B. & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering*. https://doi.org/10.1016/j.infsof.2008.09.009/

Krause, J., Perer, A., & Bertini, E. (2014). INFUSE: Interactive feature selection for predictive modeling of high dimensional data. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 1614-1623. https://doi.org/10.1109/TVCG.2014.2346482

Krause, J., Perer, A., & Ng, K. (2016, May). Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5686-5697). https://doi.org/10.1145/2858036.2858529

Kumar, R., Jain, A., Tripathi, A. K., & Tyagi, S. (2021a, January). Covid-19 outbreak: An epidemic analysis using time series prediction model. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence) (pp. 1090-1094). IEEE. https://doi.org/ 10.1109/Confluence51648.2021.9377075

Kumar, R., Kumar, P., & Kumar, Y. (2021b, January). Analysis of financial time series forecasting using deep learning model. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence) (pp. 877-881). IEEE. https://doi.org/10.1109/Confluence51648.2021.9377158

Lackes, R., Siepermann, M., & Vetter, G. (2020). What drives decision makers to follow or ignore forecasting tools: A game-based analysis. *Journal of Business Research, 106*, 315-322. https://doi.org/10.1016/j.jbusres.2019.02.036

Lampathaki , F., Agostinho, C., Glikman, Y., & Sesana, M. (2021, June). Moving from 'black box 'glass box' artificial intelligence in manufacturing with XMANAI. In *2021 IEEE International Conference on Engineering, Technology, and Innovation* (ICE/ITMC) (pp. 1-6). IEEE.

Legg, P., Smith, J., & Downing, A. (2019). Visual analytics for collaborative human-machine confidence in human-centric active learning tasks. *Human-centric Computing and Information Sciences,* 9(1), 5. https://doi.org/10.1186/s13673-019-0167-8

Lin, Z. Q., Shafiee, M. J., Bochkarev, S., Jules, M. S., Wang, X. Y., & Wong, A. (2019). Do explanations reflect decisions? A machine-centric strategy to quantify the performance of explainability algorithms. *arXiv preprint* arXiv:1910.07387. https://doi.org/10.48550/arXiv.1910.07387

Mahbooba, B., Timilsina, M., Sahal, R., & Serrano, M. (2021). Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. *Complexity, 2021*(1), 6634811, https://doi.org/10.1155/2021/6634811

Mello, R., & Martins, R. A. (2019). Can big data analytics enhance performance measurement systems?. *IEEE Engineering Management Review, 47*(1), 52-57. https://doi.org/10.1109/EMR.2019.2900645

Mueller, S. T., Hoffman, R. R., Clancey, W., Emrey, A., & Klein, G. (2019). Explanation in human-AI systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable AI. *arXiv preprint* arXiv:1902.01876. https://doi.org/10.48550/arXiv.1902.01876

Poleneni, V., Rao, J. K., & Hidayathulla, S. A. (2021, January). COVID-19 prediction using ARIMA model. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence) (pp. 860-865). IEEE. https://doi.org/ 10.1109/Confluence51648.2021.9377038

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

Rosenfeld, A., & Richardson, A. (2019). Explainability in human-agent systems. *Autonomous Agents and Multi-agent Systems, 33*, 673-705. https://doi.org/10.1007/s10458-019-09408-y .

Shaikh, S., Gala, J., Jain, A., Advani, S., Jaidhara, S., & Edinburgh, M. R. (2021, January). Analysis and prediction of COVID-19 using regression models and time series forecasting. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence) (pp. 989-995). IEEE. https://doi.org/10.1109/Confluence51648.2021.9377137

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior, 98*, 277-284. https://doi.org/10.1016/j.chb.2019.04.019

Sindhgatta, R., Moreira, C., Ouyang, C., & Barros, A. (2020a). Exploring interpretable predictive models for business processes. In *Business Process Management: 18th International Conference, BPM 2020, Seville, Spain, September 13–18, 2020, Proceedings 18* (pp. 257-272). Springer International Publishing.

Sindhgatta, R., Ouyang, C., & Moreira, C. (2020b, December). Exploring interpretability for predictive process analytics. In *International Conference on Service-Oriented Computing* (pp. 439-447). Cham: Springer International Publishing.

Sokol, K., & Flach, P. (2020). One explanation does not fit all: The promise of interactive explanations for machine learning transparency. *KI-Künstliche Intelligenz, 34*(2), 235-250. https://doi.org/10.1007/s13218-020-00637-y

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K. R. (2021). Towards CRISP-ML (Q): a machine learning process model with quality assurance methodology. *Machine Learning and Knowledge Extraction, 3*(2), 392-413. https://doi.org/10.3390/make3020020

Vélez, D., Ayuso, A., Perales-González, C., & Rodríguez, J. T. (2020). Churn and Net Promoter Score forecasting for business decision-making through a new stepwise regression methodology. *Knowledge-Based Systems, 196*, 105762. https://doi.org/10.1016/j.knosys.2020.10576

Wach, M., & Chomiak-Orsa, I. (2021). The application of predictive analysis in decision-making processes on the example of mining company's investment projects. *Procedia Computer Science, 192*, 5058-5066. https://doi.org/10.1016/j.procs.2021.09.284

Wickramanayake, B., He, Z., Ouyang, C., Moreira, C., Xu, Y., & Sindhgatta, R. (2022). Building interpretable models for business process prediction using shared and specialised attention mechanisms. *Knowledge-Based Systems, 248,* 108773. https://doi.org/10.1016/j.knosys.2022.108773

Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (Vol. 1, pp. 29-39).

Xie, X., Du, F., & Wu, Y. (2020). A visual analytics approach for exploratory causal analysis: Exploration, validation, and applications. *IEEE Transactions on Visualization and Computer Graphics, 27*(2), 1448-1458., https://doi.org/ 10.1109/TVCG.2020.3028957

Yefimenko, S. (2018). Advances in GMDH-based predictive analytics tools for business intelligence systems.

Zheng, F., & Zhong, S. (2011, August). Time series forecasting using an ensemble model incorporating ARIMA and ANN based on combined objectives. In *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce* (AIMSEC) (pp. 2671-2674). IEEE. https://doi.org/ 10.1109/AIMSEC.2011.6011011

Zotero. (n.d.). Zotero. Retrieved from https://www.zotero.org/.

Zytek, A., Liu, D., Vaithianathan, R., & Veeramachaneni, K. (2021). Sibyl: Understanding and addressing the usability challenges of machine learning in high-stakes decision-making. *IEEE Transactions on Visualization and Computer Graphics, 28*(1), 1161-1171. https://doi.org/ 10.1109/TVCG.2021.3114864

Zhdanov, D., Bhattacharjee, S., & Bragin, M. A. (2022). Incorporating FAT and privacy-aware AI modeling approaches into business decision-making frameworks. *Decision Support Systems, 155,* 113715., https://doi.org/10.1016/j.dss.2021.113715

Author's Biography

Shruthi Sajid is currently working as a Business Consultant in an IT Consulting company in the Netherlands. She helps clients across various industries leverage ServiceNow capabilities to identify and combat business challenges and processes. She has a Master's degree from the University of Twente with a focus in Business and IT. Her academic background keeps her on her toes and passionate about converting data into insights using data-driven tools and languages.

Jeewanie Jayasinghe Arachchige holds a PhD in Information Systems and Management from Tilburg University in 2013. She obtained her master's degree in Information Technology in 2002 from Keele University, United Kingdom. Her main research interests are Conceptual Modeling, Ontologies, Service-oriented Designing, and Process Mining. She is a lecturer at the Department of Computer Science at Vrije University, Netherlands. Further, she was a researcher and a lecturer at the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Netherlands. She has served as a PC member of several international conferences.

Dr. Faiza Bukhsh is an Associate Professor in the Data Management and Biometrics group at the University of Twente, Netherlands. Her research covers machine learning, event-driven intelligence, and data privacy. She leads national women-in-computing initiatives and has contributed to research in privacy, machine learning, and process mining. Dr. Bukhsh has worked on various industry projects with partners like KPN, Vodafone, Philips, and ING. She completed her postdoc in cyber security at the University of Twente and earned her PhD from Tilburg University, focusing on ontological norms and process mining for compliance checking.

Abhishta Abhishta is an assistant professor at the University of Twente, specializing in the economic and financial impacts of cyber-attacks. He develops data-driven techniques to assess these impacts, aiming to guide organizations in their security investments. His research has been supported by NWO grants for projects on cloud security (MASCOT) and the development of a Responsible Internet (CATRIN), as well as a project with the Dutch police on cyber-criminal collaboration. Abhishta serves on the program committees for ACM/IEEE/IFIP conferences and reviews articles for communication journals. He teaches courses on Enterprise Security and Information Systems and is part of the behavioral data science incubator at his university, applying data science to social science problems. Through his work, Abhishta contributes to enhancing digital security and reliability.

Dr. Faizan Ahmed earned his PhD in Applied Mathematics from the University of Twente and has since dedicated his research to advancing the fields of applied mathematics and machine learning. His current research is centered on applied machine learning and explainable AI, where he strives to make AI models more interpretable, transparent, and trustworthy. By developing techniques that enhance the interpretability of complex models, Dr. Ahmed aims to bridge the gap between AI's technical capabilities and its practical applications across industries.