PRODUCTION PLANNING & CONTROL
THE MANAGEMENT OF OPERATIONS

Taylor & Francis
Taylor & Francis Group

Production Planning & Control
The Management of Operations

ISSN: 0953-7287 (Print) 1366-5871 (Online) Journal homepage: www.tandfonline.com/journals/tppc20

Machine learning and deep learning in project analytics: methods, applications and research trends

Shahadat Uddin, Sirui Yan & Haohui Lu

To cite this article: Shahadat Uddin, Sirui Yan & Haohui Lu (2025) Machine learning and deep learning in project analytics: methods, applications and research trends, Production Planning & Control, 36:7, 873-892, DOI: 10.1080/09537287.2024.2320790

To link to this article: https://doi.org/10.1080/09537287.2024.2320790

Published online: 11 Mar 2024.

PRODUCTION PLANNING & CONTROL
2025, VOL. 36, NO. 7, 873–892
https://doi.org/10.1080/09537287.2024.2320790

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Machine learning and deep learning in project analytics: methods, applications and research trends

Shahadat Uddin, Sirui Yan and Haohui Lu

School of Project Management, Faculty of Engineering, The University of Sydney, New South Wales, Australia

ABSTRACT

Project analytics refers to applying analytical techniques and methods to past and present data to gain insights into how the underlying project is performing. Machine learning (ML) and Deep learning (DL) have acquired extensive usage in various disciplines due to their analytical strength and the availability of high-speed computational devices. This article comprehensively surveys commonly used ML and DL algorithms for addressing project-related research problems. This study used author-selected keywords from article metadata to construct, analyse and visualise keyword co-occurrence networks to explore research trends. It has several notable observations: (a) Support vector machine and Random forest are the most used ML algorithms in project analytics; (b) although Artificial neural network remains a frequently used DL algorithm, its project-related applications have recently experienced a substantial decrease; (c) genetic algorithm and Fuzzy logic are the other advanced analytical methods frequently coined with ML and DL algorithms for addressing project-related problems; (d) there is a sharp increase of ML and DL applications in various project contexts; and (e) researchers used ML and DL algorithms for studying cost and time performance in construction and software project contexts. This article details these observations further and discusses their novelty and implications for research and practice.

ARTICLE HISTORY
Received 23 August 2023
Accepted 12 February 2024

KEYWORDS
Project analytics; machine learning; deep learning; research trends; research methods; research applications

1. Introduction

Due to their ability to identify trends and patterns, data-driven intelligence approaches have been prevalent and received applications in many areas (Zhang et al. 2011; Chih-Lin et al. 2017; Liu et al. 2019). Machine learning (ML) is the branch of artificial intelligence that applies known algorithms to the available data to simulate the way humans learn, which eventually helps unveil the hidden pattern and trend within that data (Janiesch et al. 2021). Deep learning (DL) is part of the broader family of ML, which teaches machines or computers to process data for learning purposes in a way inspired by the human brain (Janiesch et al. 2021). Although DL is an ML member, each group has distinct algorithms. For the brevity of analyses and research exploration, this study considers them as separate groups.

With the advent of the processing power of devices or computers, ML and DL have become the most commonly used data-driven approaches in recent years. The project analytics research domain went through a similar experience. Researchers applied a wide range of ML and DL algorithms to address various problems related to projects and their smooth management and execution (Bilal et al. 2019; Uddin et al. 2022, 2023). Therefore, exploring the project analytics research domain concerning its adaptation to data-driven intelligence approaches is crucial. This study aims to draw a broader picture regarding ML and DL applications for addressing project analytics-related problems and the associated research trends.

Project analytics, which works at five levels including descriptive, diagnostic, predictive, prescriptive and cognitive, uses analytical techniques on past and present project data to enable wise decisions on effective project delivery. Descriptive analytics is looking at current and historical data to account for what has happened. Diagnostic analytics is the exploration of underlying causes and effects. The third one, predictive analytics, predicts what could happen. Prescriptive analytics conjectures options from predictive analytics to determine the best course of action for the future. It also points out how to take advantage of opportunities or mitigate risks. Finally, cognitive analytics is the process of simulating human thoughts to learn from data and extract hidden cause-effect patterns among the attributes of a given project context. As evident in the current literature, researchers applied various qualitative and quantitative methods and approaches on these five project analytics levels. Kim et al. (2009) used a structural equation modelling approach to predict project performance for international construction projects. Using artificial intelligence and evolutionary computation, Tinoco et al. (2021) performed a prescriptive analysis for the equipment allocation optimisation problem in transportation projects. Different earned value measures are applicable at each project analytics level, from

CONTACT Shahadat Uddin shahadat.uddin@sydney.edu.au School of Project Management, Faculty of Engineering, The University of Sydney, New South Wales, Australia

descriptive to cognitive analytics (Chen et al. 2016). Kermanshachi et al. (2016) interviewed subject matter experts following the qualitative Delphi method to identify project complexity indicators.

ML and DL algorithms are appropriate for research investigation applications to the middle three project analytics levels (i.e. diagnostic, predictive and prescriptive). They are merely considered for descriptive and cognitive purposes. The advancement of information technologies facilitates project-related data collection throughout different implementation stages. High-speed computational facilities enable ML and DL to experience an unprecedented application of these data to unfold actionable and insightful information related to various performance-related project measures, including cost management, risk reduction and customer satisfaction (Spikol et al. 2018). Figure 1 illustrates how ML, DL and project data can drive project analytics at different levels to address various project-related problems.

ML and DL applications have recently been experiencing a considerable rise in other areas, including manufacturing, inventory analytics and optimisation. Abualsauod (2023) proposed a novel technique for fault and control management for the manufacturing industry using DL. Based on simulation experiments with two large-scale datasets, Lolli et al. (2019) achieved excellent classification accuracy for inventory systems using support vector machines and DL. Wang et al. (2023) applied various supervised ML techniques for parameter optimisation for the mechanical product design process. Overall, project management and other related areas see a massive rise in ML and DL applications for solving problems or improving the available solutions for better performance. However, there is a lack in the present literature of a comprehensive summary of these methods, their application areas and future trends in project analytics. The novelty of this study lies in fulfilling this gap.

The rest of the article is structured as follows. Section 2 describes the research approach. Section 3 briefly outlines the ML and DL approaches used in various project analytics contexts over time. This section also describes the differences between ML and DL algorithms and their usage statistics in the literature. Section 4 summarises the application type of ML and DL algorithms in diagnostic, predictive and prescriptive areas of project analytics. It also draws attention to the purpose and project context of their application. This section follows section 5, which provides a research trend analysis. After that, this study discusses the findings of this study in section 6, followed by the conclusion section. Such a structure of this article will ensure a nice flow of provided materials to readers and align with similar studies on methods, applications and research trends of ML and DL methods in other areas (Xin et al. 2018; Liu and Lang 2019).

2. Research approach

It is essential to focus on the current literature to explore the applications of certain methods or algorithms in a specific area (Snyder 2019). This research, therefore, first searches for published articles that apply ML and/or DL to address research problems in project analytics contexts. The main keywords for this search are ‘project’, ‘machine learning’, ‘deep learning’, ‘artificial intelligence’ and ‘AI’. This study considered ‘AI’ and ‘artificial intelligence’ as keywords since both are often used interchangeably in academic articles. The final search phrase is as follows:

> *“Project” AND (“AI” OR “Artificial Intelligence” OR “Machine learning” OR “Deep learning”).*

The abovementioned search phrase was searched in Scopus, PubMed, and IEEE Xplore databases for journals and conferences written in English. The search generated a total of 728 articles. Some articles appeared multiple times due to using the same search terms on different online search engines. This study followed the PRISMA guidelines for screening (Page et al. 2021). After removing duplicate articles, this study identified 322 unique articles. Next, after browsing through the titles of all articles, it is evident that many articles appeared in their titles that were irrelevant to the research topic, such as about the application of ML projects in other areas. Therefore, this study removed these irrelevant articles. Then, this study manually checked the remaining 94 articles by browsing through their abstracts, experiments, and results sections. It is noted that several pieces of literature were retrospective or presented frameworks that were not implemented. Considering the targeted objectives, this study evaluated only 59 of these articles. This

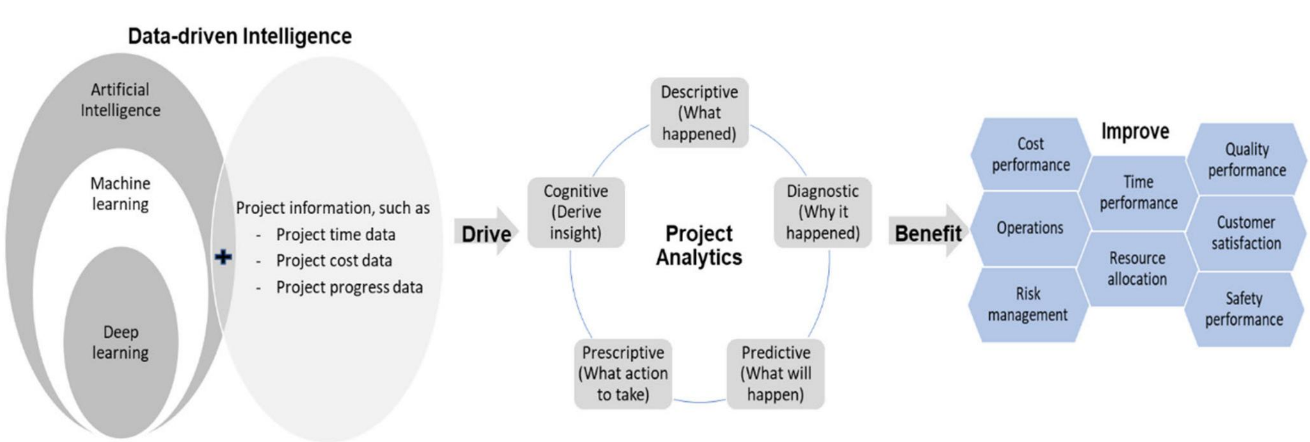

Figure 1. Illustration of how machine learning and deep learning drive project analytics to achieve competitive advantages.

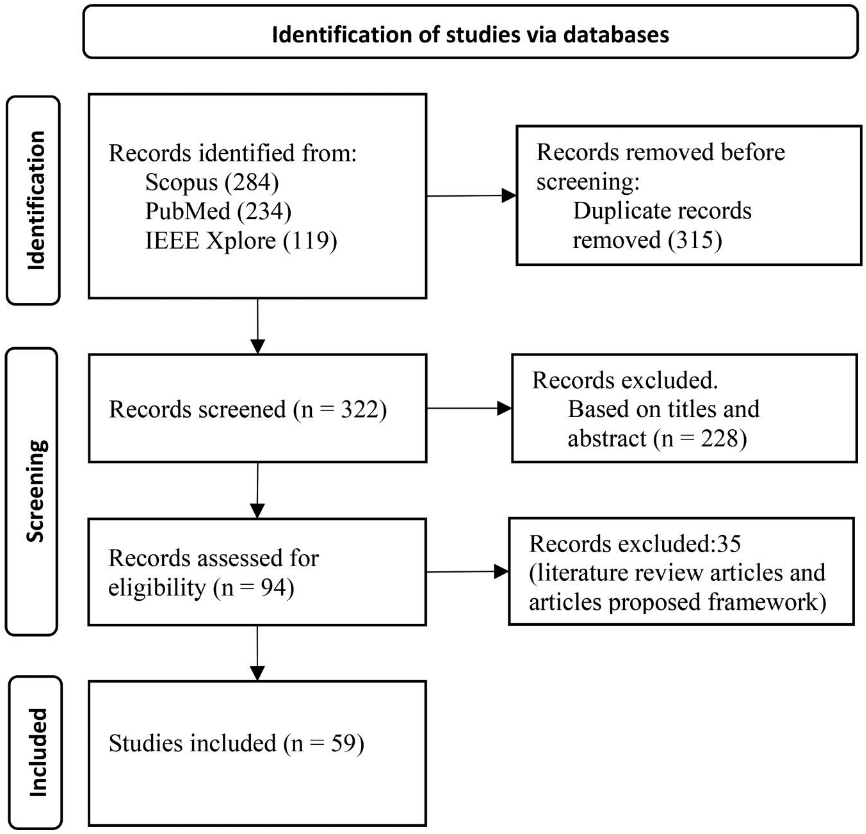

Figure 2. The article selection procedure followed in this study.

search was conducted on 25 April 2023. Figure 2 illustrates the entire search process.

This study then explored these 59 articles and their metadata to extract relevant information for the in-depth research analyses required to achieve the desired goals of this study. Apart from the main text, each article contains essential information about itself, including a list of authors and their affiliation details, author-defined and index keywords, volume, issue, pages, correspondence address, publisher, publication year and funding details. In addition to metadata analyses, this study investigates each article's abstract, methods and results sections. For example, it considers reading each article's methods and results sections to determine what ML and DL algorithms researchers applied in the underlying research. It also inspects the abstract of each article to figure out the aim and the context of the data collection of the underlying study. Table 1 presents a summary of these articles.

This study could follow other methods, including surveys, case studies, and analysis of other documents (e.g. annual reports), instead of a literature review for data collection. We chose a literature review as the research methodology, since this approach helps identify knowledge gaps quickly, avoid redundancy and conduct critical analysis (Snyder 2019).

3. Machine learning and deep learning for project analytics

The project analytics domain experienced the application of most of the classical ML algorithms. However, this number is only a few for the DL algorithms. Feature extraction is a crucial component of ML models, where the most important features must be picked and extracted manually for the model to learn appropriately (Zebari et al. 2020). On the other hand, DL can perform autonomous feature extraction from raw data (Kasongo and Sun 2020), which means it can learn to discover relevant characteristics straight from the input data without human interaction (Figure 3). DL can learn and build complicated representations, making it particularly useful for tasks involving high-dimensional raw input, such as image and language processing (Janiesch et al. 2021). This study considered only those algorithms for a brief description that are applied in addressing project-related problems.

3.1. Machine learning algorithms

3.1.1. Support vector machine

Support vector machine (SVM) applies to linear and non-linear data. Initially, it projects each data onto an *n*-dimensional feature space, where *n* denotes the number of features. Subsequently, the algorithm discerns a hyperplane for separating the data into two classes by concurrently optimising their margin distance and minimising classification discrepancies (Cortes and Vapnik 1995; Suthaharan and Suthaharan 2016). The marginal distance for a class is the distance between the decision hyperplane and its nearest instance. Figure 4(a) depicts an SVM classifier, further illustrating this concept. Chaudhary et al. (2016) employed SVM to categorise risk factors in software projects to reduce developers' workload and improve the accuracy of identifying harmful risk factors.

Table 1. A brief of the 59 articles considered by this study.

ID	Reference	Study aim	Algorithm used	Project context
1	Abbasianjahromi and Aghakarimi (2023)	Strategies to improve project security in the event of incorrect forecasts	DT and KNN	Construction
2	Al-Smadi and Al-Bdour (2023)	Forecast time and cost overruns	ANN	Construction
3	Uddin et al. (2023)	Apply machine learning for project performance modelling	Boosting, KNN, LR, RF and SVM	Engineering
4	Yu (2023)	Design and evaluation of a standard project engineering management system	ANN	Engineering
5	Golabchi and Hammad (2023)	Forecasting the utilisation of labour resources in construction projects	RNN	Construction
6	Zhang et al. (2023)	Predicting funding evaluation decisions based on personal characteristics	DT	R&D
7	Pang et al. (2022)	Predict project cost and duration	ANN, DT, LSTM, Multi regression, RF and SVM	IT
8	Mahmoodzadeh et al. (2022)	Predict project cost and duration	DT, Gaussian process regression and SVM	Construction
9	Taye and Feleke (2022)	Predict the failures in project management knowledge areas	DT, KNN, LR, NB and SVM	Software
10	Shoar et al. (2022)	Continuous evaluation of cost overruns for engineering services	Multi regression, RF and SVM	Building
11	Kusonkhum et al. (2022)	Cost projection for over-budget construction projects	KNN	Construction
12	Zhou et al. (2022)	Development of an accurate forecasting model for operation expense	ANN, KNN, RF and SVM	Transport
13	Greeshma and Edayadiyil (2022)	Implementation of a monitoring system for construction progress tracking	CNN	Construction
14	Venkata Ramana and Narsimha (2022)	Predicting success measures for software project outcome	Multi regression	Software
15	Rudra Kumar et al. (2022)	Contemporary resource-fitting predictions	SVM	Fintech
16	Rathod and Sonawane (2022)	Defining artificial intelligence role in time and cost management	ANN, Linear regression and SVM	Construction
17	Sampaio De Sousa and Villanueva (2022)	Estimating project revenue	ANN	Energy
18	Bharathi et al. (2022)	Selecting team members for projects with specific technical areas	KMC and KNN	Multiple
19	Sikimić and Radovanović (2022)	Predicting project efficiency in high-energy physics	ANN and Boosting	R&D
20	Hanci (2021)	Indicating risk groups for software projects	DT and NB	Software
21	Oliveira et al. (2021)	Minimise time spent and possible errors during issuing auto-assignment	KNN, LR, NB and SVM	Software
22	Ma et al. (2021)	Identify and assess potential risk factors in construction projects	CNN, KNN, NB and SVM	Construction
23	Gouthaman and Sankaranarayanan (2021)	Predict the percentage risk of software models	ANN, KNN, RF and SVM	Software
24	Malik et al. (2021)	Identify the customer needs from a customer-created information dataset	ANN	Multiple
25	Illahi et al. (2021)	Predict the success of crowdsourcing platform software projects	CNN, LR, LSTM, NB, RF and SVM	Software
26	Sousa et al. (2021)	Identify the risk level of project risk factors	ANN, KNN, NB, RF and SVM	Software
27	Karki and Hadikusumo (2021)	Identify the characteristics of qualified project managers	ANN, DT, LR, NB and RF	Construction
28	Zakaria et al. (2021)	Estimation of software project development effort	Classification and regression tree, RF and SVM	Software
29	Elmousalami (2021)	Produce accurate results for cost forecasting	ANN, Multi regression and RF	Multiple
30	Yue (2021)	Estimate the cost of construction projects	ANN	Construction
31	Egwim et al. (2021)	Predicting construction project delay	Bagging, Boosting, DT, NB and RF	Construction
32	Bogdan and Marginean (2020)	Predicting structure and clarity of software projects	ANN, LSTM and RF	Software
33	Herrero et al. (2020)	Predicting the final stage of a project in advance	ANN, DT, KNN, RF and SVM	Infrastructure
34	Sabahi and Parast (2020)	Predicting individual project performance	ANN, RF and SVM	Multiple
35	Chou and Lin (2020)	Estimate project duration	ANN and SVM	Engineering
36	Radliński (2020)	Predicting customer satisfaction	Ensemble approach, KNN, RF and SVM	Software
37	Yeh and Chen (2020)	Predicting the success of crowdfunding projects	ANN	Fintech
38	Sanni-Anibire et al. (2020)	Predicting the risk of delay in projects	ANN, KNN and SVM	Housing
39	Kanakaris et al. (2020)	Predict undesired situations	NB	Operation
40	Gondia et al. (2020)	Delay risk prediction of construction projects	DT and NB	Construction
41	Zhang et al. (2020)	Forecasting capital costs of projects	ANN	Mining

(*continued*)

Table 1. Continued.

ID	Reference	Study aim	Algorithm used	Project context
42	Yaseen et al. (2020)	Predict project delays	RF	Construction
43	Pena et al. (2019)	Propose a method for project evaluation	ANN	Software
44	Yaakobi et al. (2019)	Evaluate potentially valuable project ideas	Latent Dirichlet Allocation and Jaccard similarity	Organisation
45	Yurdakurban and Erdogan (2018)	Estimate software effort	DT, NB and Multi regression	Software
46	Pospieszny et al. (2018)	Effort and duration estimation for software projects	ANN and SVM	Software
47	Masoud et al. (2018)	Software project effort estimation	KMC	Software
48	Arage and Dharwadkar (2017)	Estimating the future costs of construction projects	Linear regression	Construction
49	Wauters and Vanhoucke (2017)	Predicting the actual duration of a project	KNN	Multiple
50	Iwata et al. (2016)	Development of the effort estimation model	ANN and SVM	Software
51	Chaudhary et al. (2016)	Classify and identify risk factors	SVM	Software
52	Wauters and Vanhoucke (2016)	Predicting the actual duration of a project	Bagging, Boosting, DT, RF and SVM	Multiple
53	Han et al. (2015)	Comparing machine learning algorithms for time prediction	ANN and DT	Software
54	Chou et al. (2015)	Estimate the cost and bid award amount of projects	ANN	Construction
55	Lopez-Martin et al. (2014)	Predict software practitioner productivity	ANN and Multi regression	Software
56	Chou et al. (2013)	Predicting project dispute resolution results	SVM	Construction
57	Cheng et al. (2012)	Assessing project success	ANN and KMC	Construction
58	Cheng et al. (2009)	Predicting project cash flow to enhance project cost management	ANN and KMC	Construction
59	Ko and Cheng (2007)	Dynamic prediction of project success	ANN	Construction

ANN: artificial neural network; CNN: convolutional neural network; DT: decision tree; KMC: K-means clustering; KNN: K-nearest neighbours; LR: logistic regression; LSTM: long short-term memory; NB: Naïve Bayes; RF: random forest; RNN: recurrent neural network; SVM: support vector machine; IT: information technology; R&D: research and development.

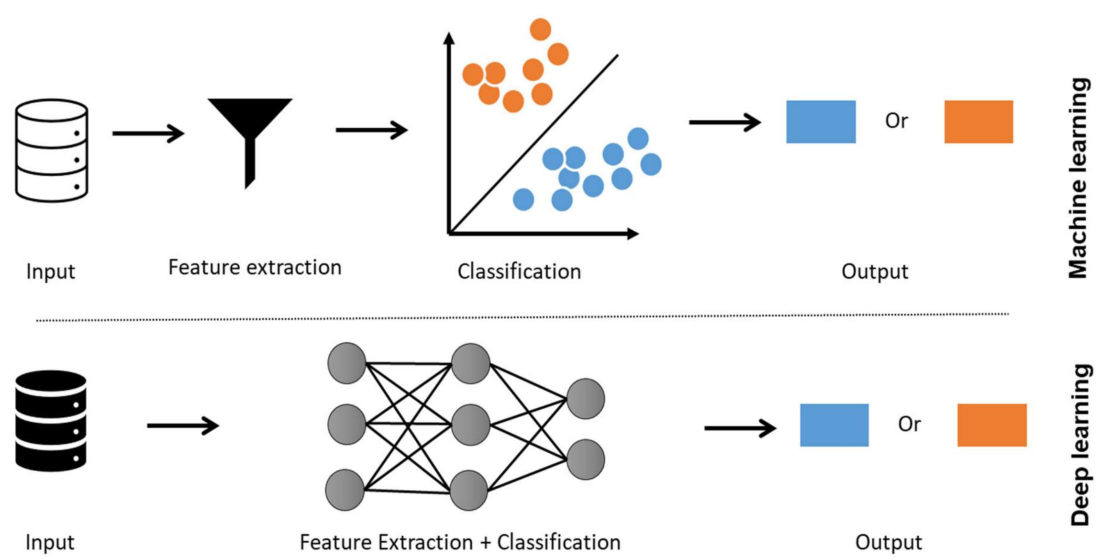

Figure 3. Difference between machine learning and deep learning method.

3.1.2. Logistic regression

Logistic regression (LR) is a sophisticated extension of traditional regression meant to model binary variables that often represent the presence or absence of an event (Hosmer et al. 2013; Sperandei 2014). LR determines the probability that a new instance belongs to a specific class. Because of its probabilistic character, the output range is between 0 and 1. Since LR is a binary classifier, a threshold value must be established to discriminate between the two classes. For example, if the probability value exceeds 0.50, it is predicted as 'class A'; otherwise, as 'class B'. Figure 4(b) illustrates an LR classifier. Using data from software companies, Taye and Feleke (2022) have applied LR to predict failures in project management knowledge domains.

3.1.3. Decision tree

The Decision tree (DT) algorithm is a tree-like data classification method. In a typical DT, nodes occur at several levels, with the first or uppermost node identified as the root node. All intermediary nodes (those with at least one child node) represent tests on input variables. The classification

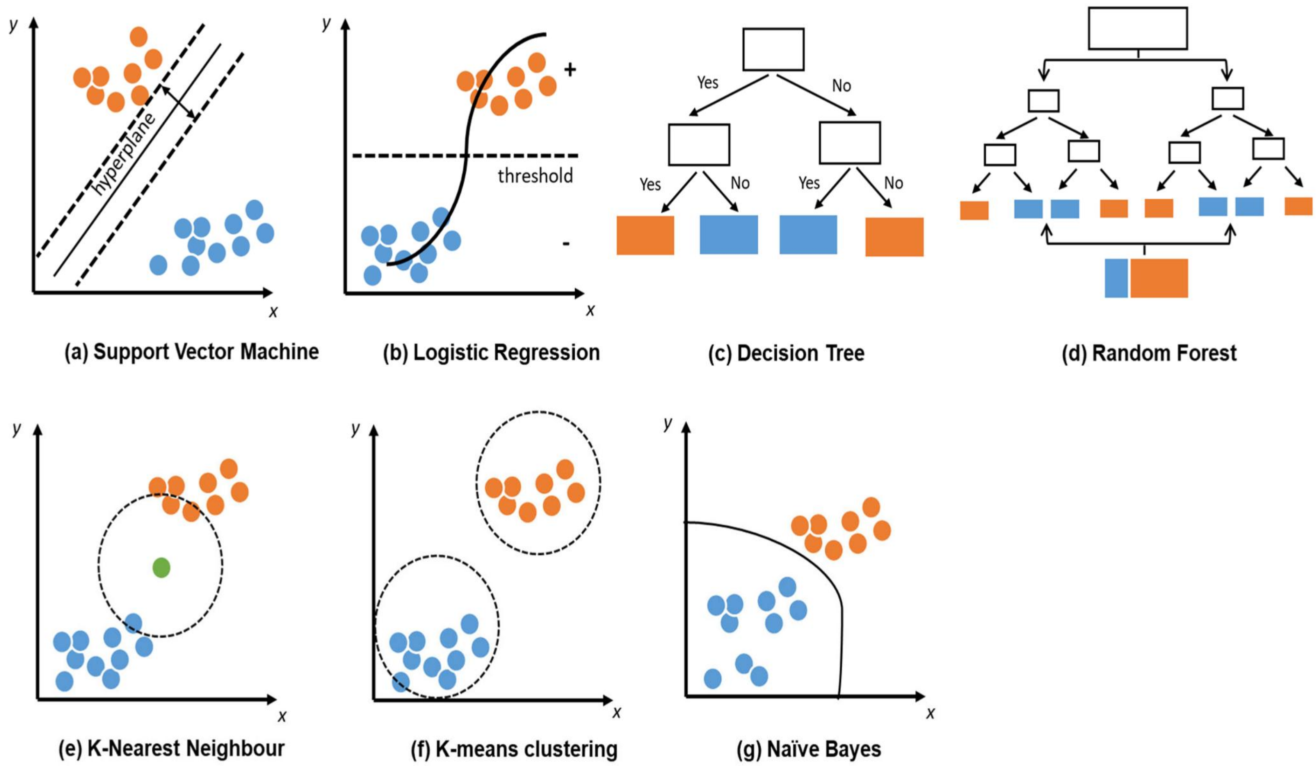

Figure 4. Illustration of different machine learning algorithms.

algorithm diverts towards the suitable child node based on the results of a particular test, repeating the testing and branching procedure until it reaches the leaf node (Quinlan 1986; Song and Ying 2015). The decision results are encapsulated in these terminal or leaf nodes. Figure 4(c) visualises a DT classifier. In their study, Zhang et al. (2023) applied DT to classify funding datasets and obtained satisfactory classification accuracy.

3.1.4. Random forest

Random forest (RF) is an ensemble classifier composed of several DTs, similar to the structure of a forest, which is a collection of many trees (Breiman 2001; Biau and Scornet 2016). RF is trained using different parts of the training dataset. The input vector of the data connected with a novel sample is sent through each DT in the forest for classification. Each DT investigates a separate part of the input vector, resulting in a classification result. The forest then aggregates these results, selecting the classification with the most 'votes' (for discrete classification outcomes) or computing the average across all trees (for numeric classification outcomes). Since the RF method includes results from several DTs, it can effectively mitigate the variation caused by a single DT's evaluation of the same information. Figure 4(d) shows the RF algorithm visually, clarifying its structure and operation. RF has longstanding applications in project management. For instance, Yaseen et al. (2020) applied a hybrid model combining RF and genetic algorithms to predict risky delays in construction projects to aid in monitoring and sustainability of construction project management.

3.1.5. K-nearest neighbours

The K-nearest neighbour (KNN) algorithm is one of the most straightforward and earliest classification algorithms. In KNN, *K* represents the nearest neighbours solicited for 'votes'. Different selections of *K* can yield varying classification outcomes for an identical object. The fundamental principle underlying KNN is that if most *K*-nearest samples in the feature space pertain to a specific category, the sample in question will likely belong to the same category and possess similar characteristics (Cunningham and Delany 2021; Uddin et al. 2022). A predetermined similarity metric, such as Euclidean distance, is used to identify the *K* most similar points from the training set to a given test point. Figure 4(e) shows the processes of the KNN algorithm. KNN demonstrates a broad spectrum of applications in project analytics. For example, Kusonkhum et al. (2022) applied KNN to predict over-budget projects in government construction projects.

3.1.6. K-means clustering

K-means clustering (KMC) is a well-known unsupervised ML approach that divides a dataset into separate groups. The *K* in K-means clustering refers to the number of clusters the user specifies (Hartigan and Wong 1979; Kodinariya and Makwana 2013). Iteratively, the method assigns each data point to the closest cluster centroid and then updates the centroid by computing the mean of all points inside the cluster. This method continues until the cluster assignment of data points no longer changes, signifying convergence. K-means clustering attempts to minimise intra-cluster distances, ensuring that data points within the same cluster are as comparable as feasible (Figure 4(f)). Bharathi et al. (2022) used KMC for industrial projects to recommend the most suitable team of potential employees for a project. Masoud et al. (2018) adopted KMC for resource forecasting and estimation in software project management.

3.1.7. Naïve Bayes

Naïve Bayes (NB) infers the probability that a new example belongs to a specific class based on the assumption that all attributes of a given class are independent (Duda and Hart 1973; Frank et al. 2000). Direct estimation of each relevant multivariate probability is unreliable. Combining prior and posterior probabilities, NB avoids subjective bias and the overfitting phenomenon of using sample information alone. The solid mathematical foundation allows for a low misspecification rate of NB, contributing significantly to its widespread popularity. Figure 4(g) illustrates the Naïve Bayes classifier. To give an example of NB's application, Hanci (2021) used NB to predict risk groups of software projects and obtained a certain level of accuracy. Gondia et al. (2020) applied NB to indicate the degree of delay in a construction project and showed good predictive performance.

3.2. Deep learning algorithms

3.2.1. Artificial neural network

An Artificial neural network (ANN) is a computing system inspired by the biological neural networks of the human brain. It consists of linked layers of nodes replicating the human brain's neurones. Each node receives input, processes it (usually with a non-linear function) and transmits the result to other nodes in the network (Figure 5(a)). Neural networks learn from input data by modifying the network's weights and biases, which are adjusted via a process known as back-propagation employing gradient descent optimisation (Rumelhart et al. 1986). These methods are frequently utilised in ML and AI-based applications ranging from image and audio recognition to natural language processing. Al-Smadi and Al-Bdour (2023) applied ANN to aid construction project time and cost overrun prediction.

3.2.2. Convolutional neural network

A Convolutional Neural Network (CNN) is a form of ANN that succeeds at image and text data processing by learning spatial feature hierarchies (Gu et al. 2018). CNNs are essential for image identification tasks, such as discriminating between photographs of cats and dogs, because they autonomously learn crucial visual features such as edges, textures, and shapes (Wang et al. 2020). CNNs are used in text data for sentiment analysis and document categorisation. For example, based on the content, they may classify movie reviews as favourable or bad by recognising important word patterns. CNNs automate the process of feature extraction from raw data in both picture and text applications, reducing the need for manual feature engineering. Figure 5(b) presents a CNN illustration. In practical project applications, CNNs prove highly beneficial. For instance, Greeshma and Edayadiyil (2022) constructed a supervised CNN classifier to monitor construction project progress.

3.2.3. Long Short-term Memory

Long Short-term Memory (LSTM) is another form of ANN that recognises long-term patterns in data sequences such as text, audio, or time series data. LSTM utilises the network loops that allow information to be transmitted from one stage in succession to the next, providing a type of memory (Medsker and Jain 2001). An LSTM, for example, may be utilised in natural language processing for tasks such as language translation. It would read a statement in one language, remember each word's context and where it was in the sentence, and then create a translation in another language. An LSTM with short memory power is known as a recurrent neural network. Figure 5(c) is an illustration of LSTM. Bogdan and Marginean (2020) constructed an LSTM model to predict the structure and clarity of software projects.

3.3. Usage statistics

Figure 6 illustrates the usage frequency of different ML and DL algorithms for project analytics. ANN has been applied the most time (28), followed by SVM (23) and RF (17). The other widely used ML algorithms are KNN (14) and DT (13). Although researchers developed several DL algorithms in the current literature, CNN and LSTM are the other DL algorithms (besides ANN) commonly used for project analytics applications. Interestingly, only one study used image data for research analysis. Greeshma and Edayadiyil (2022) applied

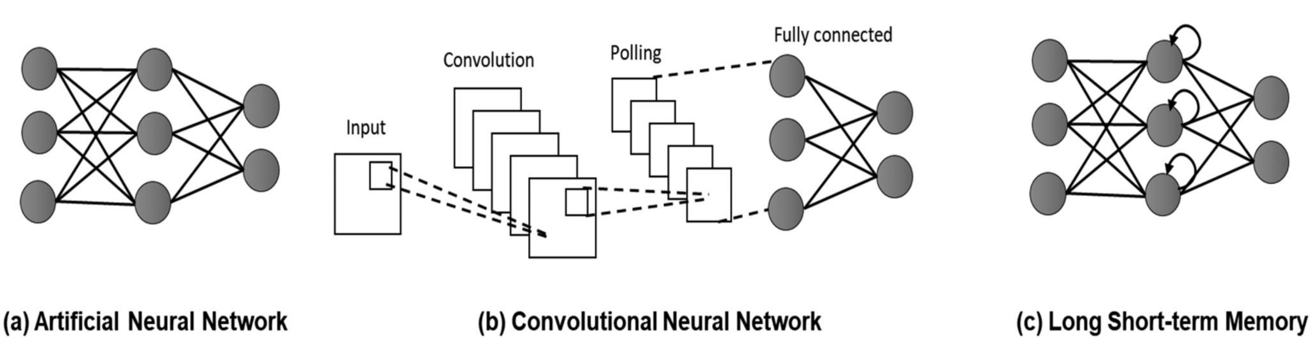

Figure 5. Illustration of different deep learning algorithms.

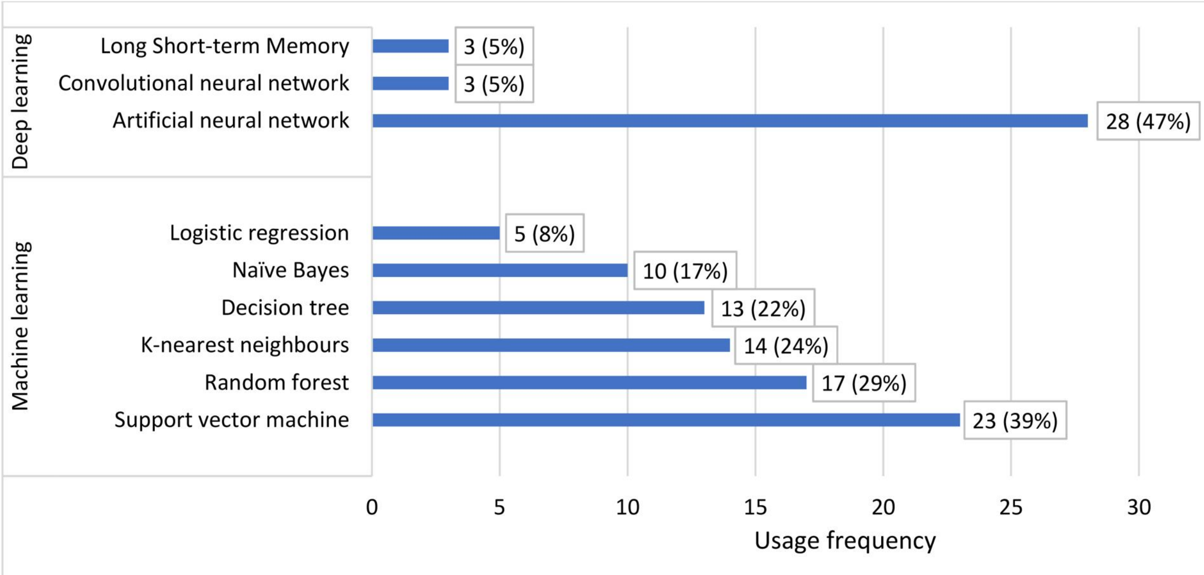

Figure 6. Usage frequency of different machine learning and deep learning algorithms.

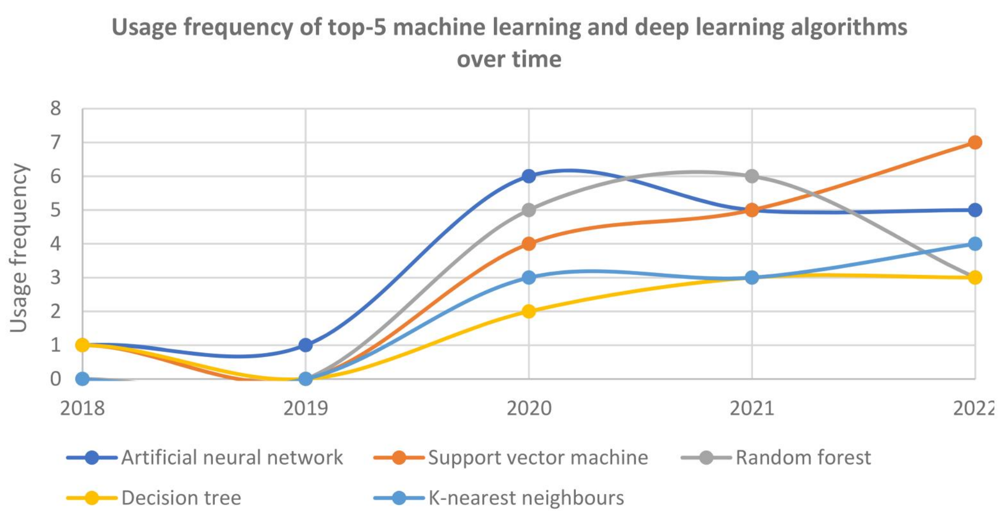

Figure 7. Usage frequency of top-5 machine learning and deep learning algorithms (as in Figure 6) over time (2018–2022). This study did not consider 2023, since it did not have the complete data for this year.

CNN to several images from worksites and websites to develop a construction progress tracking system.

Further exploration of the top five algorithms, as presented in Figure 7, reveals that there has been an increase in their applications in project analytics since 2020. SVM maintains the sharpest upward trajectory compared to the other four.

Fortunately, several learning packages, including open-source and commercial software, are available to the public, as summarised in Table 2. They facilitate the application of ML and DL algorithms in project-related investigations and applications.

4. Applications

4.1. Application type

According to Figure 1, each project analytics level addresses different questions. ML and DL methods are suitable for addressing sophisticated questions. For this reason, they are appropriate for diagnostics, predictive and prescriptive analytics.

4.1.1. Diagnostic analytics

Diagnostics analytics is employed after discovering what happened in a given context (e.g. project delay or cost overrun) using an earlier descriptive analysis step. It then uses advanced methodologies and techniques on the available data to answer the question –"Why did it happen?". Its primary purpose is to determine the root causes of an occurrence or trend. Rathod and Sonawane (2022) applied ANN and SVM to identify factors causing cost and time overrun within a project. They found that construction delay is the main reason for these two negative consequences. Ma et al. (2021) applied word association techniques of word2vec and skip-gram and CNNs on unstructured text data generated during project management to identify risk factors related to

Table 2. Summary of available machine learning and deep learning tools.

Tool/Library	Type	Description	Reference
Scikit-learn (sklearn)	ML	A free Python ML library with numerous classification, regression, and clustering techniques	Pedregosa et al. (2011)
TensorFlow	ML/DL	Google's end-to-end open-source ML platform includes a rich ecosystem for both ML and DL activities.	Abadi (2016)
PyTorch	ML/DL	An open-source ML library for Python, principally created by Facebook's AI Research team for applications such as NLP and AI.	Paszke et al. (2017)
Keras	DL	A Python-based neural network library designed for rapid experimentation and commonly used with TensorFlow.	Ketkar and Ketkar (2017)
Pandas	ML	A Python software package for data processing and analysis.	McKinney (2011)
NumPy	ML	A Python library that adds support for massive, multi-dimensional arrays and matrices, as well as high-level mathematical operations	Oliphant (2006)
Matplotlib	ML	A Python charting library and its numerical mathematics extension NumPy.	Hunter (2007)

ML: machine learning; DL: deep learning; NLP: natural language processing.

project safety. They then used SVM, KNN and NB to rank the importance of risk factors causing a task failure. Such a ranking of safety risk factors has great practical significance in investigating an undesired outcome for future projects, such as cost and budget overrun.

4.1.2. Predictive analytics

Predictive analytics, the third of the five levels on which project analytics works, predicts future outcomes by applying advanced analytical techniques, including statistical modelling, data mining and ML, on available historical data (Castro Miranda et al. 2022). Its primary goal is to go beyond knowing what has happened (descriptive analytics) and why it happened (diagnostic analytics) to offer the best assessment of what will happen. Most studies employing predictive analytics in project contexts mainly forecast attributes or variables related to one or more of these three factors - cost, duration and operations. Mahmoodzadeh et al. (2022) applied ML approaches of DT and SVM to 350 data points from 34 historical tunnelling projects to predict construction cost and time. In a survey-based study, Gouthaman and Sankaranarayanan (2021) applied ANN, KNN, RF and SVM to predict the risk percentage in software projects. Radliński (2020) also investigated software projects using several ML and ensemble approaches to predict user satisfaction. The author developed 40 models using 12 ML techniques and found RF the best-performing one.

4.1.3. Prescriptive analytics

Prescriptive analytics uses advanced processes and tools on the predictions from the earlier predictive analytics step to recommend the optimal course of action and strategies for moving forward (Poornima and Pushpalatha 2020). In any given context, it always seeks to answer this question—'"What should we do?'. Chaudhary et al. (2016) proposed an SVM-based framework first to identify risk factors for software projects and then prescribed their classification. They argued that their proposed framework could locate the most effective risks for any given software project context so that software developers can adopt mitigation actions as early as possible. Uddin et al. (2023) applied LR, KNN, RF and SVM to model project cost, time and quality performance. They also pointed out factors or attributes that facilitated improved project performance. Bharathi et al. (2022) proposed an AI-based recommendation system to prescribe employees from a pool with the required technical skills essential for the underlying project.

Table 3. Details of the study purpose. The sum of the right-hand column is higher than the number of articles considered in this study (59) since some explored factors related to more than one purpose.

Study purpose related to	Number of articles (%)
Cost performance	25 (42%)
Time performance	24 (41%)
Operations	10 (17%)
Quality performance	8 (14%)
Risk management	6 (10%)
Safety performance	4 (7%)
Resource allocation	3 (5%)
Customer satisfaction	2 (3%)

4.2. Application purpose

This study grouped all observed application purposes into eight areas. These areas relate to cost, time, quality, operations, safety, resource allocation, risk management and customer satisfaction. A few factors (e.g. progress monitoring, knowledge gaps and selecting members with appropriate qualities) are mapped into the operations area. The remaining seven areas are self-explanatory by their names. As detailed in Table 3, cost-related factors are the primary focus for research exploration for most articles (25) that used ML and DL algorithms, followed by time (24) and operations (10) related attributes.

Interestingly, quality, an essential component of the iron triangle (Pollack et al. 2018), did not get a place in the top three list. It came in fourth place. Only eight studies engaged in quality-related research using ML and DL approaches.

4.3. Application context

Figure 8 shows the frequency of the project context of the underlying research study for the 59 articles extracted for this study. ML and DL algorithms are applied chiefly to studies related to the construction project (19), followed by the software project (18). Few studies used research data collected from more than one project context. Malik et al. (2021) used online review data from multiple project contexts to develop a recommendation system for customer needs.

Figure 8 highlights the application of ML and DL algorithms across various project contexts, with the construction sector leading significantly with 19 studies. This dominance indicates the construction industry's keen interest in harnessing ML and DL algorithms to address its inherent complexities and optimise operations. Mahmoodzadeh et al. (2022) effectively applied ML techniques, particularly LR, to predict the cost and duration of tunnel construction projects, highlighting the importance of drilling systems and groundwater impact. Furthermore, Zhou et al. (2022) focused on improving operating expenditure predictions for the US Light Rail Transit Systems. They used ML to overcome the limitations of traditional cost estimation methods. Their work emphasised the need for accurate budgeting in public transit projects to prevent service cuts, demonstrating that ML could provide more dependable forecasting tools for early project planning stages. To give another example, Karki and Hadikusumo (2021) applied ML to identify the characteristics of competent project managers, such as tolerance and reliability in managing construction projects, identifying both constructive and destructive behaviours through interviews with industry professionals. Abbasianjahromi and Aghakarimi (2023) advanced an ML framework using DT and KNN algorithms to predict construction project safety performance, emphasising the need for safety training and commitment from the management. Meanwhile, Gondia et al. (2020) demonstrated the effectiveness of ML, particularly the NB model, in predicting delays in construction projects by analysing the complex interplay of various risk factors, aiding in proactive project management and risk mitigation strategies.

For the software project, Zakaria et al. (2021) explored the enhancement of software project estimation accuracy through ML algorithms, tackling the limitations of the traditional constructive cost model. Concurrently, Taye and Feleke (2022) investigated machine learning approaches to foresee project management knowledge area failures in software companies, underscoring the efficacy of SVM. Similarly, Hanci's research (2021) focused on using ML classifiers, such as LR and RF, to predict risk groups in global software development projects, focusing on identifying critical influencing factors. Finally, Oliveira et al. (2021) reported on an industrial initiative to automate issue assignments in software projects using ML. Their study offered a comparative analysis of algorithms and discussing the outcomes and insights from applying these techniques for a global electronics company.

5. Research trend

This study used the author-defined keywords from the extracted 59 articles to explore the research trend in applying ML and DL algorithms for project analytics. Keywords are a crucial component of publication metadata, which also play a pivotal role in creating the research impact of the underlying article (Uddin and Khan 2016). Author-defined keywords represent authors' perceptions of their research contribution within the thematic context of the underlying research domains. On the other side, index keywords are system-generated keywords extracted using a text mining tool from article texts. This study considered index keywords for trend analyses due to their ability to express authors' understandings of their work.

This study considered two time periods to analyse the research trend over time comprehensively. The first period (T_1) consists of 2019 and prior. The second period (T_2) spans 2020 to our data extraction date. Out of the 59 articles considered in this study (Table 1), Frank et al. (2000) were published during T_2. The remaining 17 were published during T_1. The splitting year 2020 is carefully considered since, from this year, the project analytics literature has been experiencing a sharp growth in applying ML and DL approaches for solving project-related research problems. This study explores KCNs for these two periods using a visual and a network analysis approach to analyse research trends.

5.1. Keyword co-occurrence network

This study constructs keyword co-occurrence networks (KCN) using the author-defined keywords to explore their

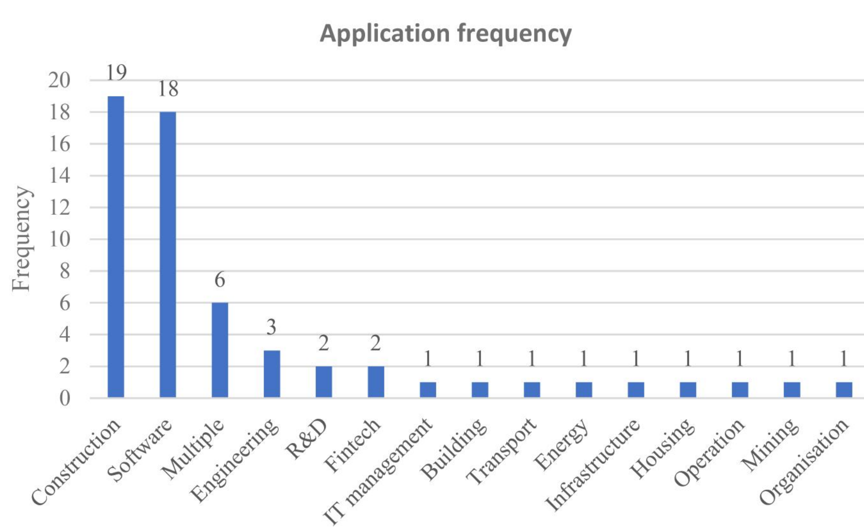

Figure 8. Frequency of the project context of the studies that used machine learning and deep learning.

associations and influence on each other. An analysis of such networks eventually helps extract the thematic contexts, research trends, and priority over time. A node in a KCN denotes an author-defined keyword, and an edge between two nodes indicates the co-occurrence of the underlying keywords represented by those nodes in an article. The number of simultaneous occurrences of a pair of keywords in multiple articles constitutes the weight of the link connecting them. Examining the keywords and their co-appearance in articles is essential to construct KCN. A network analysis of this KCN will enable us to understand how the research tendency changes over time. Figure 9 illustrates the KCN construction process from the abstract keyword data for two articles. In this figure, both articles have four keywords (K_1–K_4 for the first and K_3–K_6 for the second). Two common keywords (K_3 and K_4) exist between these two articles. There are two fully connected article-level keyword co-occurrence networks for the keyword list for each article (top-left is for the first article, and down-left is for the second article). The final KCN (right-hand side) is constructed by merging all links of these two networks. The appearance of one or more keywords in multiple articles is the key to this merging process. If there is no common keyword between two article-level keyword co-occurrence networks, these networks cannot be merged, resulting in two isolated networks. The weight between K_3 and K_4 is two since both articles have this keyword pair in common.

5.2. Visual approach

This study used VOSviewer to illustrate KCNs graphically. VOSviewer is a software tool to demonstrate a network as a two-dimensional map by considering keywords' relative position and density (Van Eck and Waltman 2010). The distance between a keyword pair in this map is roughly inversely proportional to the number of keyword co-occurrences. The number and strength of keyword connections to neighbouring keywords determine the font size of the keywords. The colour of each point on the map depends on the keyword density of that point. The higher the number of keywords in the point's neighbourhood and the higher the connectivity of adjacent keywords, the closer the point's colour is towards yellow (conversely, it appears blue). In addition, VOSviewer offers network diagrams, which involve three main elements: nodes, connecting lines and colours. The size of the nodes represents the number of keyword occurrences in the network data. The larger the node, the more times the keyword appears. The connecting lines represent the strength of the relationship between two nodes. VOSviewer can cluster keywords according to their strength of association and neighbourhood and use different colours to represent various clusters (Van Eck and Waltman 2017).

Figure 10 presents the KCN for the first period (T_1). The DL method of the *artificial neural network* and *artificial intelligence* are the most visible keywords on this visual map. Other easily pointable keywords on this map include *project management*, *predictive analytics*, *support vector machine*, *K-nearest neighbours* and *machine learning*. These keywords provide a quick overview of the thematic context of research efforts during T_1.

The *machine learning* keyword has become the most visible in KCN for T_2, as illustrated in Figure 11. Two other keywords (*artificial neural network* and *predictive analytics*) are almost equally visible in both networks. Some keywords, such as *big data*, *data analysis* and *deep learning*, are present only on this network. They are absent in T_1. Conversely, the *fuzzy logic* keyword appears only in the KCN for T_1. There are also differences in keywords' tendency towards forming clusters in these two networks. Overall, keywords related to data-driven intelligence have become the subject of a high volume of research in recent years.

5.3. Network analysis approach

This study considers several network analysis measures to compare the KCNs for T_1 and T_2 quantitatively. For network

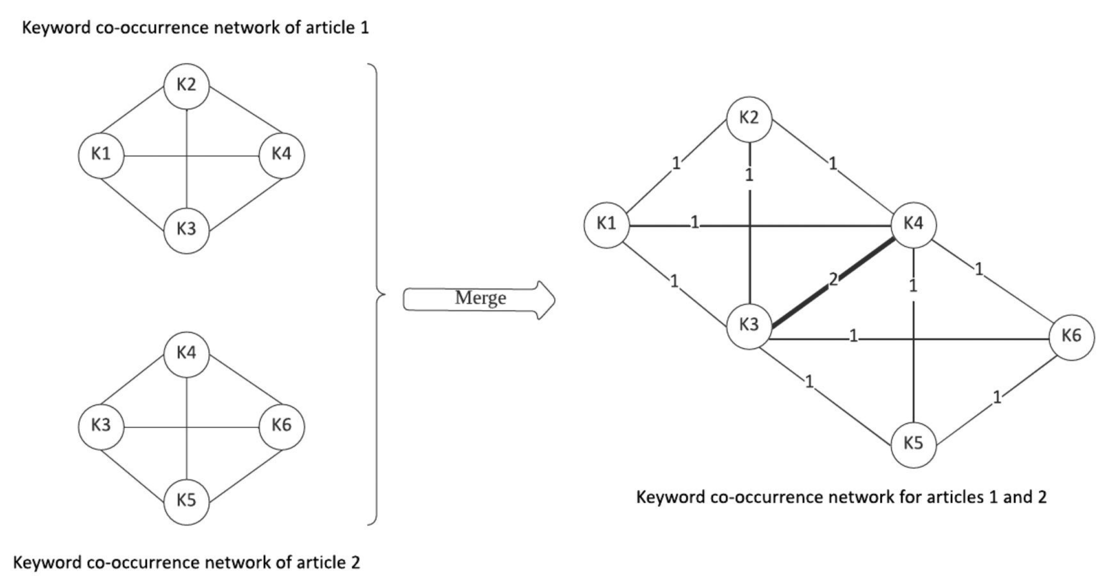

Figure 9. The construction process of keyword co-occurrence network from the keyword information of two articles.

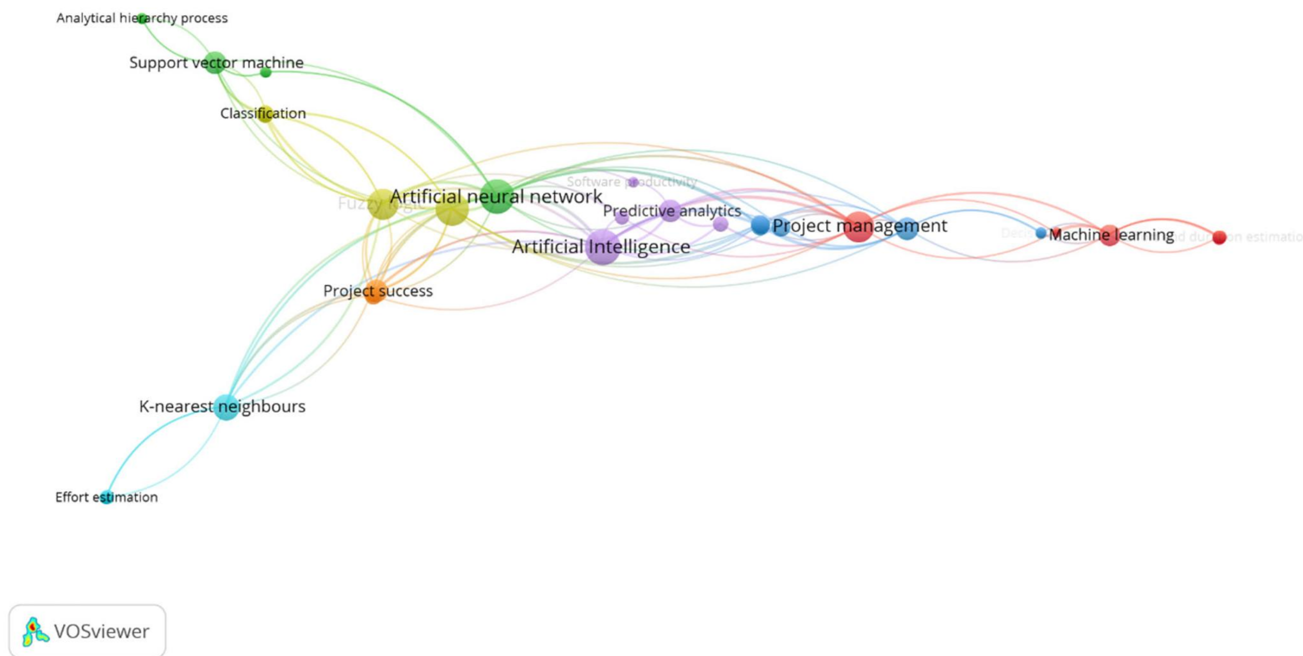

Figure 10. The keyword co-occurrence network for the first period (2019 and prior).

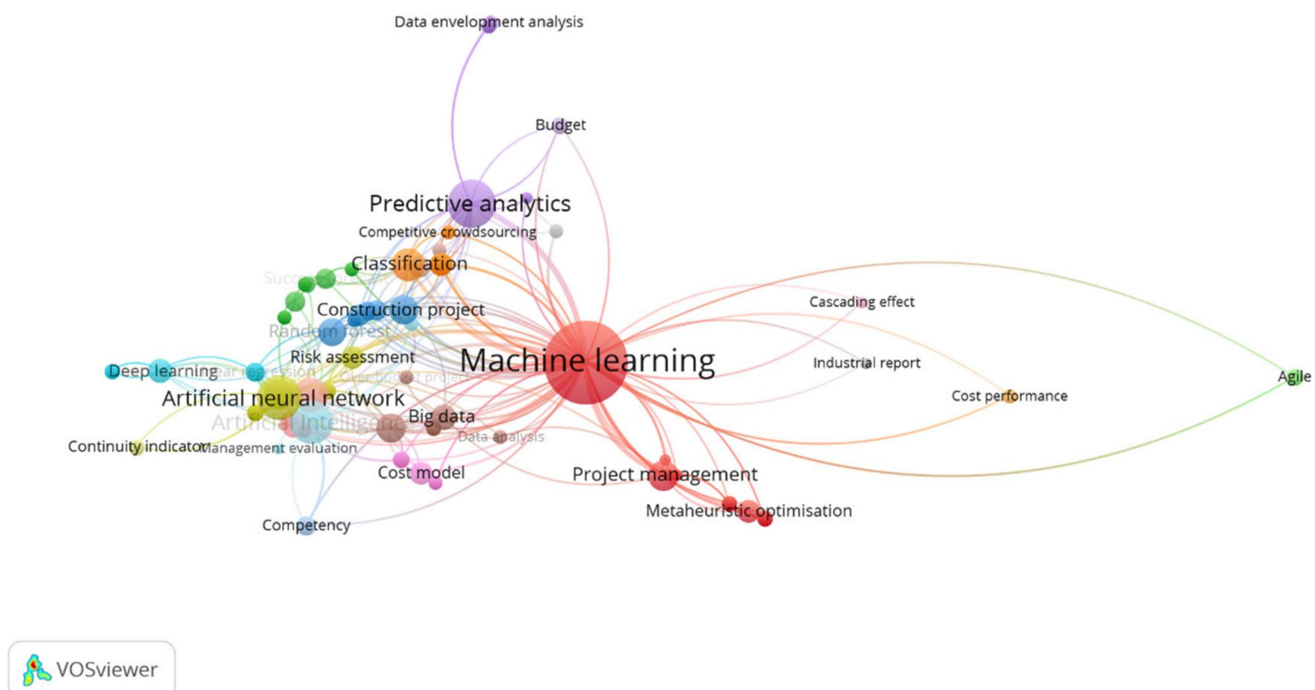

Figure 11. The keyword co-occurrence network for the second period (2020 and onwards).

analysis, this study used the Organisational Risk Analyser software tool (Altman et al. 2018). The following section briefly outlines those measures first. The subsequent section compares and contrasts KCNs for two periods using these measures.

5.3.1. Network measures

Network size. The size of a network is the number of nodes it has (Wasserman and Faust 2003). It is a network-level measure.

Network density. It is also a network-level measure. Network density is the ratio between the number of links a network has among its nodes and the maximum number of possible links (Wasserman and Faust 2003). The following formula can quantify the network density for a network with N nodes and e_t edges.

$$Density = \frac{2 \times e_t}{N(N-1)} \tag{1}$$

A KCN with high density indicates that its member keyword nodes frequently co-occur in articles and vice versa.

Degree centrality. For a node, degree centrality is the proportion of its direct connections with other network nodes compared to the highest number of connections it could have (Wasserman and Faust 2003). For a network with N nodes, any nodes could have a maximum of $(N-1)$ direction connections with other network nodes. Hence, if a node (n_i) has D_t direct links with its neighbours in a network with N nodes, then the following formula can calculate its degree centrality.

$$\text{Degree centrality } (n_i) = \frac{D_t}{N-1} \tag{2}$$

The degree centrality of a keyword in a KCN indicates its co-occurrence tendency and frequency with other keywords. Its value will be high for a frequently co-occurring keyword and vice versa.

Closeness centrality. Closeness centrality for a node in a network indicates how close it is to the remaining network nodes (Wasserman and Faust 2003). The following formula can calculate the closeness centrality on a scale between 0 and 1 for a node (n_i) within a network with size N.

$$\text{Closeness centrality}(n_i) = \frac{N-1}{\sum_{j=1}^{N} d(n_i, n_j)} \tag{3}$$

where $d(n_i, n_j)$ is the shortest distance between nodes n_i and n_j. A high value of this measure for a node represents its easy reachability by other network nodes. A keyword that facilitates the co-appearance of other keyword pairs in articles will have a high closeness centrality value and vice versa.

Betweenness centrality. This measure can quantify to what extent a node falls on the shortest paths between any other pair of network nodes (Wasserman and Faust 2003). The following formula quantifies the betweenness centrality for a node (n_i) within a network with size N.

$$\text{Betweenness centrality}(n_i) = \frac{2 \times \sum_{j<k} \frac{g_{jk}(n_i)}{g_{jk}}}{(N-1)(N-2)} \tag{4}$$

where g_{jk} is the number of shortest paths between nodes j and k, and $g_{jk}(n_i)$ is the number of shortest paths between nodes j and k having node i within the route. A high value for a keyword in a KCN represents its contribution to the knowledge evolution around the keywords of that KCN over time.

Figure 12 illustrates the calculation of these four measures using abstract network data.

5.3.2. Comparison using network measures

Table 4 presents the difference between the KCNs for two time periods through basic network-level measures. Network size and the number of edges have been increasing over time. A decrease in the network density measure is not unsurprising since scientific articles are limited to 4–6 keywords in most cases. Due to this limitation, a newly added node (keyword) to an existing KCN is unlikely to co-occur with most other available keywords.

Table 5 lists the top 10 keywords based on the three centrality measures in both KCNs. Four keywords (*Artificial neural network*, *Project management*, *Fuzzy logic* and *K-nearest neighbours*) are positioned in the top-10 list of each centrality for the first KCN (2019 and prior). This number has been reduced to three (*Predictive analytics*, *Artificial neural network* and *K-nearest neighbours*) for the second KCN. *Artificial neural network* and *K-nearest neighbours* are the keywords that appeared in the top-10 lists based on each centrality measure for both periods. Further exploration of the keywords in this table reveals a change in the keyword type within these top-10 lists between the two periods. For example, regarding the top-10 lists based on the closeness centrality, nine keywords are related to data-driven intelligence in the second network, whereas it is six for the first network. For the betweenness centrality, it is seven versus six in the same order. It becomes a tie with a value of eight for the degree centrality.

This study then compares the change in the node frequency values in two KCNs for the nine ML and DL algorithms commonly used in the current literature (Figure 6).

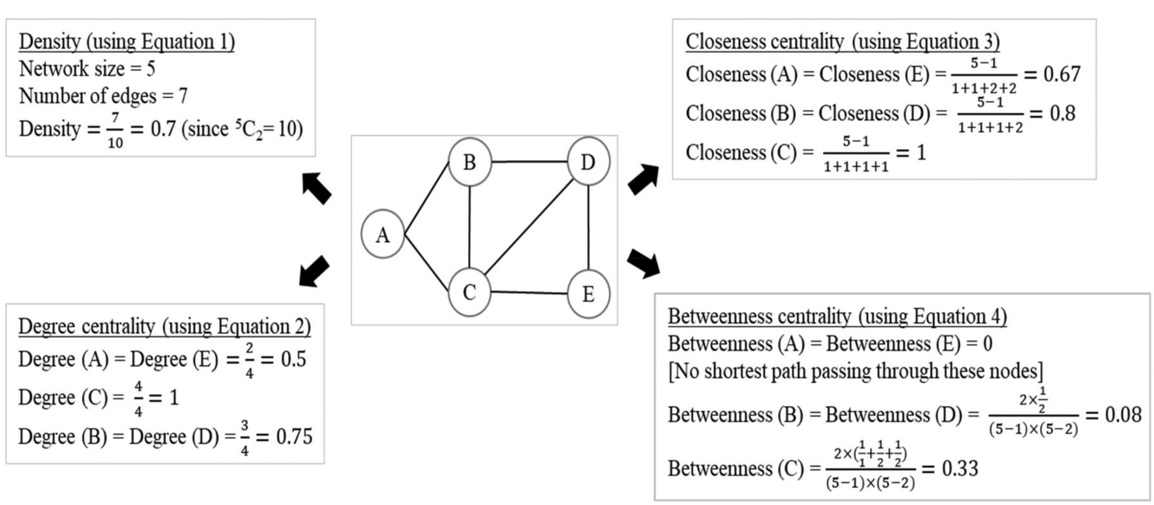

Figure 12. An illustration of four network measures calculation based on abstract network.

Table 6 details this comparison outcome. In this table, nodes are the frequently used ML and DL algorithms from Figure 6. In reporting the percentage values, this study used 17 and 42 for the first (2019 and prior) and second (2020 and onwards) periods, respectively, since they are the number of published articles during those times. All ML algorithms have experienced substantial growth over time, led by the *random forest*, which gained a 32% growth, followed by the *K-nearest neighbours* (23%). Interestingly, the *artificial neural network*, a DL approach, experienced a shrink of 8% over time. The other two DL approaches gained a 7% growth. Researchers tend to mention more specific DL approaches (e.g. *Convolutional neural network* and *Long Short-term Memory*) as keywords they adopted in their recent studies instead of the general *artificial neural network* keyword. For this reason, the formal group had no or little appearance during 2019 and prior, and the latter declined its appearance in recent years.

6. Discussion

This study provides an explorative review of ML and DL applications in the project analytics context. Such an explorative review of AI-based methods is standard in other areas, such as intrusion detection (Liu and Lang 2019) and cybersecurity (Xin et al. 2018). However, there is no study in the project analytics context. This article will potentially fill this gap. It aims to cater to readers from the beginner to the expert in project management. While some findings may appear evident to experts in computer science and other similar disciplines, consolidating these findings could be valuable to a broader project audience.

The novelty of this study lies in its notable findings concerning ML and DL applications in project analytics. First, the most commonly used ML algorithms are SVM and RF. It is the ANN among the DL algorithms. Second, there has been a sharp increase in the application of ML and DL algorithms in the project analytics context in recent years. Construction and software project contexts had been the primary application areas of these algorithms for researching attributes mostly related to cost and time performance. The third remarkable finding is that ANN application has recently decreased substantially due to a surge in applications of other sophisticated DL approaches (e.g. CNN) in project-related research problems. Despite this incorporation, this study did not notice any application of highly advanced graph-based DL algorithms, such as graph attention

Table 4. Comparison between two keyword co-occurrence networks (KCN).

Item	KCN-1(2019 and prior)	KCN-2 (2020 and onwards)
Network size	57	136
Number of edges	177	482
Network density	0.055	0.026

Table 6. Comparison of the frequency statistics of machine learning and deep learning algorithms in the two keyword co-occurrence networks over time.

Keywords	2019 and prior	2020 and onwards	Change (%)
Machine learning			
Support vector machine	5 (29%)	17 (43%)	14%
Random forest	1 (6%)	15 (38%)	32%
Decision tree	3 (18%)	10 (24%)	6%
K-nearest neighbours	1 (6%)	12 (29%)	23%
Naïve Bayes	1 (6%)	9 (24%)	18%
Logistic regression	0 (0%)	5 (12%)	12%
Deep learning			
Artificial neural network	9 (53%)	19 (45%)	−8%
Convolutional neural network	0 (0%)	3 (7%)	7%
Long Short-term Memory	0 (0%)	3 (7%)	7%

Table 5. Top 10 keywords for two periods based on the three centrality measures.

	Degree centrality		Closeness centrality		Betweenness centrality	
Rank	2019 and prior	2020 and onwards	2019 and prior	2020 and onwards	2019 and prior	2020 and onwards
1	Artificial Intelligence (0.069)	Machine learning (0.057)	Artificial Intelligence (0.008)	Artificial Intelligence (0.003)	Artificial neural network (0.054)	Machine learning (0.213)
2	Artificial neural network (0.065)	Predictive analytics (0.018)	Classification (0.008)	Machine learning (0.002)	Project management (0.039)	Artificial neural network (0.096)
3	Genetic algorithm (0.058)	Artificial Intelligence (0.017)	Artificial neural network (0.007)	Artificial neural network (0.002)	Genetic algorithm (0.026)	Predictive analytics (0.070)
4	Project management (0.051)	Artificial neural network (0.016)	Genetic algorithm (0.007)	Predictive analytics (0.002)	K-nearest neighbours (0.023)	Deep learning (0.015)
5	Fuzzy logic (0.051)	K-nearest neighbours (0.009)	Fuzzy logic (0.007)	Deep learning (0.002)	Machine learning (0.023)	Multi regression (0.014)
6	K-nearest neighbours (0.036)	Classification (0.009)	K-nearest neighbours (0.007)	Multi regression (0.002)	Fuzzy logic (0.014)	Construction project (0.012)
7	Regression (0.027)	Support vector machine (0.008)	Cost estimation (0.007)	Construction project (0.002)	Support vector machine (0.006)	K-nearest neighbours (0.010)
8	Support vector machine (0.027)	Project management (0.008)	Bid award amount (0.007)	K-nearest neighbours (0.002)	Cost estimation (0.003)	Cost overrun (0.007)
9	Predictive analytics (0.027)	Construction project (0.007)	Case-based reasoning (0.007)	Big data (0.002)	Bid award amount (0.003)	Success forecast (0.005)
10	Project success (0.027)	Ensemble model (0.007)	Cash flow (0.007)	Cost overrun (0.002)	Case-based reasoning (0.003)	Ensemble model (0.004)

networks (Zhang et al. 2020) and graph neural networks (Scarselli et al. 2009), for addressing a project-related problem. Other research domains (e.g. disease prediction, Zhang et al. 2020) widely adopted these graph-based DL algorithms for research investigations. Fourth, researchers employed ML and DL algorithms primarily to study cost and time performance factors. Although quality is an important performance factor, it became fourth after the operations-related performance factors (Table 3).

Like in other research domains, for example, disease prediction, SVM and RF are the two most used ML algorithms in project analytics (Figure 6). SVM is memory efficient and can choose the best line through its kernel function to classify the given data points (Noble 2006). RF is a tree-based classification approach and reveals the best classification accuracy on tabular data (Biau and Scornet 2016). ANN was followed as a research method in the first article, published in 2007, extracted for this study (Ko and Cheng 2007). Since then, it has maintained consistent applications over time. However, this study has yet to notice the application of advanced ANN-based DL algorithms, for example, graph attention networks (Velickovic et al. 2017), in the project analytics research domain. Although these algorithms are recently developed (Zhang et al. 2020), scholars have applied them in other research domains. It is expected to see their presence in the project analytics research field soon. This study also notices the application of graph ML in project analytics (Uddin et al. 2023). Graph ML is a new AI-driven approach based on ML and network analytics that usually generate improved classification outcome through feature engineering (Zhang et al. 2020). In addition to tabular data, this approach requires network data. Uddin et al. (2023) integrated ML and network analytics for extracting hand-crafted features to model project cost, time and quality performance.

Keyword co-occurrence networks enable extracting the other major keywords that co-occurred with ML and DL techniques for addressing project-related research problems quantitatively and visually. Genetic algorithm and Fuzzy logic are the further advanced techniques frequently coined with ML and DL techniques in project analytics. As in columns 2, 4 and 6 of Table 5 and Figure 10, these techniques have commonly been used with other ML and DL algorithms. However, they suffered a little decline in their applications for project analytics problems recently (2020 and onwards), as illustrated in columns 3, 5 and 7 of Table 5 and Figure 11. The dominance of other methods could be a reason, leaving scope for future research to address the underlying reasons for this declination specifically.

The two most crucial factors for any project are cost and time. It is highly desirable to complete a project within budget and time from the owner's perspective. Accordingly, most of the current project-related research emphasises research topics related to cost and time performance (Bititci et al. 2012). In the case of applying ML and DL algorithms to address project-related research issues, the same tendency was observed (Table 3). More than 80% of our extracted articles studied project cost and/or time-related research issues. A similar finding was then noticed concerning the project context. The project context of most of our reviewed articles is either construction or software development, which aligns with real-life frequency statistics of different types of projects considered for implementation.

The research trend analysis reveals that scholars have become increasingly inclined towards ML in addressing project-related research problems. ML algorithms are pertinent to tabular data. Due to the research context of the project analytics domain, most generated data are in tabular format. In the case of categorical data in text format, as in (Taye and Feleke 2022), researchers apply statistical approaches to converting them into numerical data. With the incorporation of new DL approaches (e.g. Convolutional neural network) for research data analysis, the artificial neural network has recently experienced a less frequent application (Table 6). However, it has maintained its position in the top-10 list. It has a lower place in the second period (2020 and onwards) than in the first period (2019 and prior) for centrality measures, except for the closeness centrality for which it has the same place (Table 4). KCN illustrations (Figures 10 and 11) further suggest that ML and DL have had applications in more thematic project analytics areas in recent years.

This study unfolds discerning insights regarding the practical relevance of its findings, especially for various project stakeholders. For example, project managers can think about appropriate ML and DL algorithms if they need to explore the underlying reasons for their underperforming ongoing projects. SVM and RF will be the default selection for them. Adopting the correct analytical approaches for project investigations will place them in a better position than their competitors. This study outlines a list of tools or software (Table 2) currently available to implement the required ML and DL approaches. Such readily available information would help them to make a quick analytical investigation plan if needed. This study argues that ML and DL algorithms are more applicable to the diagnostic, predictive and prescriptive levels. They are merely helpful for descriptive and cognitive analyses. This further guides different relevant project stakeholders on when to select ML and DL approaches considering the underlying explorative problem.

Like others, this study also has a few limitations. First, there could be selection bias in selecting and extracting 59 articles related to applying ML and DL algorithms in the project analytics context. Some studies in the current literature may not meet our search criteria but used ML, DL, or both to address project-related research problems. The inclusion of those articles may impact the numerical figures of our findings. However, the ignorance of those articles would not change the relative order of our results. Second, the considered classifications in the project application purpose (Table 3) and content (Figure 8) are subjective. Hence, employing other types of such categories for these two attributes would lead to numerically different results in those tables. Third, in constructing KCNs, this study considered author-defined keywords only. It did not explore KCNs based on index keywords.

7. Conclusion

The project analytics research domain has experienced tremendous growth in recent years concerning ML and DL applications to address project-related research problems. Classical ML algorithms primarily lead this consolidation. Analyses from the keyword co-occurrence networks, extracted from articles' meta-level data, further confirmed this growth of ML and DL applications for addressing data-driven project-related research problems. Although the current literature has yet to come across recently developed advanced DL algorithms (e.g. Deep neural networks), researchers have already applied others (e.g. Convolutional neural networks) to address research problems in the project analytics context.

The surge in utilising ML techniques underscores the evolving nature of project analytics. Depending on the nature of project-related problems, applying different methods offers fresh perspectives for academic discussions and further research. On the other hand, for industry professionals, understanding these trends means recognising the potential of specific ML and DL algorithms in enhancing the accuracy and efficiency of project analytics tasks. For instance, the discerning use of CNN hints at their potential in handling data with spatial hierarchies, a significant avenue in many project scenarios. Contrasting our findings with the existing literature, it becomes evident that, while some algorithmic applications are consistent with broader trends, there are unique challenges and opportunities for ML and DL applications within the project analytics domain, indicating the need for tailored solutions and a deeper understanding of the algorithms' strengths and limitations.

This study shows that AI-based ML and DL techniques have made fascinating promises in addressing data-driven solutions for project analytics-related research problems. Researchers and experts usually apply these techniques to data from multiple sources, which raises significant concerns from the perspective of development theory and practice (Bjola 2022). Development theory and practice is a collection of theories and practices that best help achieve desirable societal changes (Carmody 2019). AI techniques can extract more in-depth analytical insights from data, enabling them to quickly build socially impactful and sustainable solutions in specific development areas, including poverty, global health, and human rights (Bjola 2022). They can radically transform development theory and practice by integrating data and algorithms to generate insights into how development challenges are identified and researched. However, the most crucial concern AI-based solutions have been facing is to learn how to access and combine data from different sources in an ethically responsible fashion. They also lack in making a comprehensive interpretation of their suggested findings. These limitations of AI integration could add a new chapter to the current development theory and practice, making further research scopes for future researchers.

Author contributions

Shahadat Uddin: Study conception, Research plan, Formal analysis, Method, Supervision and Writing (original draft, review, and editing)
Sirui Yan: Data collection, Data analysis and Writing (original draft and review)
Haohui Lu: Method and Writing (original draft and review)
Emails: Sirui Yan: syan0201@hotmail.com
Haohui Lu: haohui.lu@sydney.edu.au

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The study was supported by the University of Sydney's Engineering Vacation Research Internship Program (2022, Winter).

Notes on contributors

Dr Shahadat Uddin is a Senior Lecturer at the School of Project Management in the Faculty of Engineering of the University of Sydney, Australia. He has research interests in Machine learning, Complex networks and Project analytics. Dr Uddin has published in several international and multi-disciplinary journals, including Neural Networks, Knowledge-based Systems, Expert Systems with Applications and Production Planning & Control. Dr Uddin has been awarded many academic awards for his outstanding research excellence, including the Top Researcher Award (Bangladesh University of Engineering & Technology Alumni Australia, 2020), Campus Director Leadership Award (Central Queensland University 2006), Certificate for Research Excellence (University of Sydney, 2010) and Dean's Research Award (University of Sydney, 2014). Dr Uddin has been placed in the Standard-Elsevier Top 2% Scientists list (Single year: 2021–2023 and Career-long: 2023) in the AI and Image Processing category.

Sirui Yan is an engineer in the IT industry. Previously, she received a Bachelor of Computer and Information Science from the Auckland University of Technology. Most recently, she was conferred a Master of Information Technology and a Master of Information Technology Management from the University of Sydney, specialising in Data Management and Analytics. She has consistently demonstrated a strong passion for researching Machine learning and Artificial intelligence.

Haohui Lu is pursuing a Doctor of Philosophy in Machine Learning at the University of Sydney. Previously, he received a Bachelor of Commerce degree in Operations Management and Decision Sciences, a graduate certificate in Data Science and a Master of Project Management from the University of Sydney. He has research interests in health informatics and graph machine learning. Recent work includes predicting the risk of chronic disease and modelling the spread of COVID-19 on road networks.

Data availability statement

The data supporting this study's findings are available from the corresponding author upon reasonable request.

References

Abadi, M. 2016. "TensorFlow: learning Functions at Scale." *Proceedings of the 21st ACM SIGPLAN* International Conference on Functional Programming. https://doi.org/10.1145/2951913.2976746

Abbasianjahromi, H., and M. Aghakarimi. 2023. "Safety Performance Prediction and Modification Strategies for Construction Projects via Machine Learning Techniques." *Engineering, Construction and Architectural Management* 30 (3): 1146–1164. https://doi.org/10.1108/ECAM-04-2021-0303

Abualsauod, E. H. 2023. "Machine Learning Based Fault Detection Approach to Enhance Quality Control in Smart Manufacturing." *Production Planning & Control* 1–9. https://doi.org/10.1080/09537287.2023.2175736

Al-Smadi, S., and H. Al-Bdour. 2023. "Machine Learning-Aided Time and Cost Overrun Prediction in Construction Projects: application of Artificial Neural Network." *Asian Journal of Civil Engineering* 24: 2583–2593.

Altman, N., K. M. Carley, and J. Reminga. 2018. *Ora User's Guide 2018. Center for the Computational Analysis of Social and Organizational Systems*. Pittsburgh: Carnegie Mellon University.

Arage, S. S., and N. V. Dharwadkar. 2017. "Cost Estimation of Civil Construction Projects Using Machine Learning Paradigm." *International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, 594–599. Palladam, India: IEEE.

Bharathi, S. S., A. Geetha, G. R. Kumar, M. D. K. Reddy, and A. Pandey. 2022. *Effective Way of Selecting the Industrial Project Team Based on Artificial Intelligence Methods*, 337–351. Singapore: Springer Nature.

Biau, G., and E. Scornet. 2016. "A Random Forest Guided Tour." *TEST* 25 (2): 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bilal, M., L. O. Oyedele, H. O. Kusimo, H. A. Owolabi, L. A. Akanbi, A. O. Ajayi, O. O. Akinade, and J. M. D. Delgado. 2019. "Investigating Profitability Performance of Construction Projects Using Big Data: A Project Analytics Approach." *Journal of Building Engineering* 26: 100850. https://doi.org/10.1016/j.jobe.2019.100850

Bititci, U., P. Garengo, V. Dörfler, and S. Nudurupati. 2012. "Performance Measurement: Challenges for Tomorrow." *International Journal of Management Reviews* 14 (3): 305–327. https://doi.org/10.1111/j.1468-2370.2011.00318.x

Bjola, C. 2022. "AI for Development: Implications for Theory and Practice." *Oxford Development Studies* 50 (1): 78–90. https://doi.org/10.1080/13600818.2021.1960960

Bogdan, D. M., and A. Marginean. 2020. "Predicting Structure & Clarity of Software Projects with Machine Learning." *IEEE 16th International Conference on Intelligent Computer Communication and Processing*. Cluj-Napoca, Romania: IEEE.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324

Carmody, P. 2019. *Development Theory and Practice in a Changing World*. UK: Routledge.

Castro Miranda, S. L., E. Del Rey Castillo, V. Gonzalez, and J. Adafin. 2022. "Predictive Analytics for Early-Stage Construction Costs Estimation." *Buildings* 12 (7): 1043. https://doi.org/10.3390/buildings12071043

Chaudhary, P., D. Singh, and A. Sharma. 2016. *Classification of Software Project Risk Factors Using Machine Learning Approach, in Advances in Intelligent Systems and Computing*, 297–309. New York: Springer International Publishing.

Chen, H. L., W. T. Chen, and Y. L. Lin. 2016. "Earned Value Project Management: Improving the Predictive Power of Planned Value." *International Journal of Project Management* 34 (1): 22–29. https://doi.org/10.1016/j.ijproman.2015.09.008

Cheng, M.-Y., H.-C. Tsai, and C.-L. Liu. 2009. "Artificial Intelligence Approachen b s to Achieve Strategic Control over Project Cash Flows." *Automation in Construction* 18 (4): 386–393. https://doi.org/10.1016/j.autcon.2008.10.005

Cheng, M.-Y., L.-C. Lien, H.-C. Tsai, and P. H. Chen. 2012. "Artificial Intelligence Approaches to Dynamic Project Success Assessment Taxonomic." *Life Science Journal* 9: 5156–5163.

Chih-Lin, I., Q. Sun, Z. Liu, S. Zhang, and S. Han. 2017. "The Big-Data-Driven Intelligent Wireless Network: architecture, Use Cases, Solutions, and Future Trends." *IEEE Vehicular Technology Magazine* 12 (4): 20–29. p. https://doi.org/10.1109/MVT.2017.2752758

Chou, J.-S., and J.-W. Lin. 2020. "Risk-Informed Prediction of Dredging Project Duration Using Stochastic Machine Learning." *Water* 12 (6): 1643. https://doi.org/10.3390/w12061643

Chou, J.-S., C.-W. Lin, A.-D. Pham, and J.-Y. Shao. 2015. "Optimized Artificial Intelligence Models for Predicting Project Award Price." *Automation in Construction* 54: 106–115. https://doi.org/10.1016/j.autcon.2015.02.006

Chou, J.-S., M.-Y. Cheng, and Y.-W. Wu. 2013. "Improving Classification Accuracy of Project Dispute Resolution Using Hybrid Artificial Intelligence and Support Vector Machine Models." *Expert Systems with Applications* 40 (6): 2263–2274. https://doi.org/10.1016/j.eswa.2012.10.036

Cortes, C., and V. Vapnik. 1995. "Support-Vector Networks." *Machine Learning* 20 (3): 273–297. https://doi.org/10.1007/BF00994018

Cunningham, P., and S. J. Delany. 2021. "k-Nearest Neighbour classifiers-A Tutorial." *ACM Computing Surveys* 54 (6): 1–25. https://doi.org/10.1145/3459665

Duda, R. O., and P. E. Hart. 1973. *Pattern Classification and Scene Analysis*. Vol 3. New York: Wiley.

Egwim, C. N., H. Alaka, L. O. Toriola-Coker, H. Balogun, and F. Sunmola. 2021. "Applied Artificial Intelligence for Predicting Construction Projects Delay." *Machine Learning with Applications* 6: 100166. https://doi.org/10.1016/j.mlwa.2021.100166

Elmousalami, H. H. 2021. "Comparison of Artificial Intelligence Techniques for Project Conceptual Cost Prediction: A Case Study and Comparative Analysis." *IEEE Transactions on Engineering Management* 68 (1): 183–196. https://doi.org/10.1109/TEM.2020.2972078

Frank, E., L. Trigg, G. Holmes, and I. H. Witten. 2000. "Naive Bayes for Regression." *Machine Learning* 41 (1): 5–25. https://doi.org/10.1023/A:1007670802811

Golabchi, H., and A. Hammad. 2023. "Estimating Labor Resource Requirements in Construction Projects Using Machine Learning." *Construction Innovation*. Advance online publication. https://doi.org/10.1108/CI-11-2021-0211

Gondia, A., A. Siam, W. El-Dakhakhni, and A. H. Nassar. 2020. "Machine Learning Algorithms for Construction Projects Delay Risk Prediction." *Journal of Construction Engineering and Management* 146 (1): 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Gouthaman, P., and S. Sankaranarayanan. 2021. "Prediction of Risk Percentage in Software Projects by Training Machine Learning Classifiers." *Computers & Electrical Engineering* 94: 107362. https://doi.org/10.1016/j.compeleceng.2021.107362

Greeshma, A., and J. Edayadiyil. 2022. "Automated Progress Monitoring of Construction Projects Using Machine Learning and Image Processing Approach." *Materials Today: Proceedings* 65 (2): 554–563. https://doi.org/10.1016/j.matpr.2022.03.137

Gu, Jiuxiang, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, et al. 2018. "Recent Advances in Convolutional Neural Networks." *Pattern Recognition* 77: 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Han, W., L. Jiang, T. Lu, and X. Zhang. 2015. "Comparison of Machine Learning Algorithms for Software Project Time Prediction." *International Journal of Multimedia and Ubiquitous Engineering* 10 (9): 1–8. 10https://doi.org/10.14257/ijmue.2015.10.9.01

Hanci, A. K. 2021. *Risk Group Prediction of Software Projects Using Machine Learning Algorithm. in 6th International Conference on Computer Science and Engineering*. Ankara, Turkey: IEEE.

Hartigan, J. A., and M. A. Wong. 1979. "Algorithm as 136: A k-Means Clustering Algorithm." *Applied Statistics*28 (1): 100–108. https://doi.org/10.2307/2346830

Herrero, Á., S. Bayraktar, and A. Jiménez. 2020. "Machine Learning to Forecast the Success of Infrastructure Projects Worldwide." *Cybernetics and Systems* 51 (7): 714–731. https://doi.org/10.1080/01969722.2020.1798645

Hosmer, D. W. Jr,., S. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. Vol. 398. NJ: John Wiley & Sons.

Hunter, J. D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. https://doi.org/10.1109/MCSE.2007.55

Illahi, I., H. Liu, Q. Umer, and N. Niu. 2021. "Machine Learning Based Success Prediction for Crowdsourcing Software Projects." *Journal of Systems and Software* 178: 110965. https://doi.org/10.1016/j.jss.2021.110965

Iwata, K., T. Nakashima, Y. Anan, and N. Ishii. 2016. "Effort Estimation for Embedded Software Development Projects by Combining Machine Learning with Classification." In *4th Intl Conf on Applied Computing and Information Technology*. Las Vegas, NV: IEEE. https://doi.org/10.1109/acit-csii-bcd.2016.058

Janiesch, C., P. Zschech, and K. Heinrich. 2021. "Machine Learning and Deep Learning." *Electronic Markets* 31 (3): 685–695. https://doi.org/10.1007/s12525-021-00475-2

Kanakaris, N., N. Karacapilidis, G. Kournetas, and A. Lazanas. 2020. *Combining Machine Learning and Operations Research Methods to Advance the Project Management Practice*, 135–155. United States: Springer International Publishing.

Karki, S., and B. Hadikusumo. 2021. "Machine Learning for the Identification of Competent Project Managers for Construction Projects in Nepal." *Construction Innovation* 23 (1): 1–18. https://doi.org/10.1108/CI-08-2020-0139

Kasongo, S. M., and Y. Sun. 2020. "A Deep Learning Method with Wrapper Based Feature Extraction for Wireless Intrusion Detection System." *Computers & Security* 92: 101752. https://doi.org/10.1016/j.cose.2020.101752

Kermanshachi, S., B. Dao, J. Shane, and S. Anderson. 2016. "Project Complexity Indicators and Management Strategies–A Delphi Study." *Procedia Engineering* 145: 587–594. https://doi.org/10.1016/j.proeng.2016.04.048

Ketkar, N., and N. Ketkar. 2017. "Introduction to Keras." *Deep Learning with Python: a Hands-on Introduction* 97–111.

Kim, D. Y., S. H. Han, H. Kim, and H. Park. 2009. "Structuring the Prediction Model of Project Performance for International Construction Projects: A Comparative Analysis." *Expert Systems with Applications* 36 (2): 1961–1971. https://doi.org/10.1016/j.eswa.2007.12.048

Ko, C.-H., and M.-Y. Cheng. 2007. "Dynamic Prediction of Project Success Using Artificial Intelligence." *Journal of Construction Engineering and Management* 133 (4): 316–324. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:4(316)

Kodinariya, T. M., and P. R. Makwana. 2013. "Review on Determining Number of Cluster in K-Means Clustering." *International Journal* 1 (6): 90–95.

Kusonkhum, W., K. Srinavin, N. Leungbootnak, P. Aksorn, and T. Chaitongrat. 2022. "Government Construction Project Budget Prediction Using Machine Learning." *Journal of Advances in Information Technology* 13 (1): 29–35. https://doi.org/10.12720/jait.13.1.29-35

Liu, H., and B. Lang. 2019. "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey." *Applied Sciences* 9 (20): 4396. https://doi.org/10.3390/app9204396

Liu, Q., J. Liu, W. Le, Z. Guo, and Z. He. 2019. "Data-Driven Intelligent Location of Public Charging Stations for Electric Vehicles." *Journal of Cleaner Production* 232: 531–541. https://doi.org/10.1016/j.jclepro.2019.05.388

Lolli, F., E. Balugani, A. Ishizaka, R. Gamberini, B. Rimini, and A. Regattieri. 2019. "Machine Learning for Multi-Criteria Inventory Classification Applied to Intermittent Demand." *Production Planning & Control* 30 (1): 76–89. https://doi.org/10.1080/09537287.2018.1525506

Lopez-Martin, C., A. Chavoya, and M. E. Meda-Campana. 2020. "A Machine Learning Technique for Predicting the Productivity of Practitioners from Individually Developed Software Projects." *15th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. Las Vegas, NV: IEEE.

Ma, G., Z. Wu, J. Jia, and S. Shang. 2021. "Safety Risk Factors Comprehensive Analysis for Construction Project: Combined Cascading Effect and Machine Learning Approach." *Safety Science* 143: 105410. https://doi.org/10.1016/j.ssci.2021.105410

Mahmoodzadeh, A., H. R. Nejati, and M. Mohammadi. 2022. "Optimized Machine Learning Modelling for Predicting the Construction Cost and Duration of Tunnelling Projects." *Automation in Construction* 139: 104305. https://doi.org/10.1016/j.autcon.2022.104305

Malik, H., A. Afthanorhan, N. A. Amirah, and N. Fatema. 2021. "Machine Learning Approach for Targeting and Recommending a Product for Project Management." *Mathematics* 9 (16): 1958. https://doi.org/10.3390/math9161958

Masoud, M., W. Abu-Elhaija, Y. Jaradat, I. Jannoud, and L. Dabbour. 2018. *Software Project Management: Resources Prediction and Estimation Utilizing Unsupervised Machine Learning Algorithm, in Lecture Notes in Mechanical Engineering*, 151–159. New York: Springer International Publishing.

McKinney, W. 2011. "Pandas: A Foundational Python Library for Data Analysis and Statistics." *Python for High Performance and Scientific Computing* 14 (9): 1–9.

Medsker, L. R., and L. Jain. 2001. "Recurrent Neural Networks." *Design and Applications* 5: 64–67.

Noble, W. S. 2006. "What is a Support Vector Machine?" *Nature Biotechnology* 24 (12): 1565–1567. https://doi.org/10.1038/nbt1206-1565

Oliphant, T. E. 2006. *A Guide to NumPy*. Vol. 1. United States: Trelgol Publishing.

Oliveira, P., R. M. C. Andrade, I. Barreto, T. P. Nogueira, and L. Morais Bueno. 2021. "Issue Auto-Assignment in Software Projects with Machine Learning Techniques." *IEEE/ACM 8th International Workshop on Software Engineering Research and Industrial Practice.* Madrid, Spain: IEEE.

Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." *International Journal of Surgery (London, England)*88: 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Pang, D.-J., K. Shavarebi, and S. Ng. 2022. *Development of Machine Learning Models for Prediction of IT project Cost and Duration*. IEEE 12th Symposium on Computer Applications & Industrial Electronics Penang, Malaysia: IEEE.

Paszke, A., S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. 2017. *Automatic differentiation in pytorch.*

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg. 2011. "Scikit-Learn: Machine Learning in Python." *The Journal of Machine Learning Research* 12: 2825–2830.

Pena, A. B., G. F. Castro, D. M. L. Alvarez, I. A. M. Alcivar, G. L. Nunez, D. S. Cevallos, and J. L. Z. Santa. 2019. "Method for Project Execution Control based on Soft Computing and Machine Learning." *XLV Latin American Computing Conference*. Panama: IEEE. https://doi.org/10.1109/CLEI47609.2019.235097

Pollack, J., J. Helm, and D. Adler. 2018. "What is the Iron Triangle, and How Has It Changed?" *International Journal of Managing Projects in Business* 11 (2): 527–547. https://doi.org/10.1108/IJMPB-09-2017-0107

Poornima, S., and M. Pushpalatha. 2020. "A Survey on Various Applications of Prescriptive Analytics." *International Journal of Intelligent Networks* 1: 76–84. https://doi.org/10.1016/j.ijin.2020.07.001

Pospieszny, P., B. Czarnacka-Chrobot, and A. Kobylinski. 2018. "An Effective Approach for Software Project Effort and Duration Estimation with Machine Learning Algorithms." *Journal of Systems and Software* 137: 184–196. https://doi.org/10.1016/j.jss.2017.11.066

Quinlan, J. R. 1986. "Induction of Decision Trees." *Machine Learning* 1 (1): 81–106. https://doi.org/10.1007/BF00116251

Radliński, Ł. 2020. "Stability of User Satisfaction Prediction in Software Projects." *Procedia Computer Science* 176: 2394–2403. https://doi.org/10.1016/j.procs.2020.09.308
Rathod, K., and A. Sonawane. 2022. "Application of Artificial Intelligence in Project Planning to Solve Late and over-Budgeted Construction Projects." *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. IEEE. https://doi.org/10.1109/ICSCDS53736.2022.9761027
Rudra Kumar, M., R. Pathak, and V. K. Gunjan. 2022. *Machine Learning-Based Project Resource Allocation Fitment Analysis System (ML-PRAFS)*, 1–14. Singapore: Springer Nature.
Rumelhart, D. E., G. E. Hinton, and R. J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323 (6088): 533–536. https://doi.org/10.1038/323533a0
Sabahi, S., and M. M. Parast. 2020. "The Impact of Entrepreneurship Orientation on Project Performance: A Machine Learning Approach." *International Journal of Production Economics* 226: 107621. https://doi.org/10.1016/j.ijpe.2020.107621
Sampaio De Sousa, B. J, and Villanueva, J. M. M. 2022. "Methodology for Evaluating Projects Aimed at Service Quality Using Artificial Intelligence Techniques." *Energies* 15 (13): 4564. https://doi.org/10.3390/en15134564
Sanni-Anibire, M. O., R. M. Zin, and S. O. Olatunji. 2020. "Machine Learning Model for Delay Risk Assessment in Tall Building Projects." *International Journal of Construction Management* 22 (11): 2134–2143. https://doi.org/10.1080/15623599.2020.1768326
Scarselli, F., M. Gori, T. Ah Chung, M. Hagenbuchner, and G. Monfardini. 2009. "The Graph Neural Network Model." *IEEE Transactions on Neural Networks* 20 (1): 61–80. https://doi.org/10.1109/TNN.2008.2005605
Shoar, S., N. Chileshe, and J. D. Edwards. 2022. "Machine Learning-Aided Engineering Services' Cost Overruns Prediction in High-Rise Residential Building Projects: Application of Random Forest Regression." *Journal of Building Engineering* 50: 104102. https://doi.org/10.1016/j.jobe.2022.104102
Sikimić, V., and S. Radovanović. 2022. "Machine Learning in Scientific Grant Review: Algorithmically Predicting Project Efficiency in High Energy Physics." *European Journal for Philosophy of Science* 12 (3): 1–21. https://doi.org/10.1007/s13194-022-00478-6.
Snyder, H. 2019. "Literature Review as a Research Methodology: An Overview and Guidelines." *Journal of Business Research* 104: 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039
Song, Y.-Y., and L. Ying. 2015. "Decision Tree Methods: Applications for Classification and Prediction." *Shanghai Archives of Psychiatry* 27 (2): 130.
Sousa, A., J. P. Faria, J. Mendes-Moreira, D. Gomes, P. C. Henriques, and R. Graça. 2021. *Applying Machine Learning to Risk Assessment in Software Projects*. United States: Springer International Publishing. https://doi.org/10.1007/978-3-030-93733-1_7
Sperandei, S. 2014. "Understanding Logistic Regression Analysis." *Biochemia Medica* 24 (1): 12–18. https://doi.org/10.11613/BM.2014.003
Spikol, D., E. Ruffaldi, G. Dabisias, and M. Cukurova. 2018. "Supervised Machine Learning in Multimodal Learning Analytics for Estimating Success in Project-Based Learning." *Journal of Computer Assisted Learning* 34 (4): 366–377. https://doi.org/10.1111/jcal.12263
Suthaharan, S., and S. Suthaharan. 2016. *Support vector machine.* Machine learning models and algorithms for big data classification: thinking with examples for effective learning, 207–235.
Taye, G. D., and Y. A. Feleke. 2022. "Prediction of Failures in the Project Management Knowledge Areas Using a Machine Learning Approach for Software Companies." *SN Applied Sciences* 4 (6): 165. https://doi.org/10.1007/s42452-022-05051-7
Tinoco, J., M. Parente, A. G. Correia, P. Cortez, and D. Toll. 2021. "Predictive and Prescriptive Analytics in Transportation Geotechnics: Three Case Studies." *Transportation Engineering* 5: 100074. https://doi.org/10.1016/j.treng.2021.100074
Uddin, S., and A. Khan. 2016. "The Impact of Author-Selected Keywords on Citation Counts." *Journal of Informetrics* 10 (4): 1166–1177. https://doi.org/10.1016/j.joi.2016.10.004
Uddin, S., I. Haque, H. Lu, M. A. Moni, and E. Gide. 2022. "Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction." *Scientific Reports* 12 (1): 6256. https://doi.org/10.1038/s41598-022-10358-x
Uddin, S., S. Ong, and H. Lu. 2022. "Machine Learning in Project Analytics: A Data-Driven Framework and Case Study." *Scientific Reports* 12 (1): 15252. p. https://doi.org/10.1038/s41598-022-19728-x
Uddin, S., S. Ong, H. Lu, and P. Matous. 2023. "Integrating Machine Learning and Network Analytics to Model Project Cost, Time and Quality Performance." *Production Planning & Control* 1–15. https://doi.org/10.1080/09537287.2023.2196256
Van Eck, N. J., and L. Waltman. 2010. "Software Survey: VOSviewer, a Computer Program for Bibliometric Mapping." *Scientometrics* 84 (2): 523–538. https://doi.org/10.1007/s11192-009-0146-3
Van Eck, N. J., and L. Waltman. 2017. "Citation-Based Clustering of Publications Using CitNetExplorer and VOSviewer." *Scientometrics* 111 (2): 1053–1070. https://doi.org/10.1007/s11192-017-2300-7
Velickovic, P., G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio. 2017. "Graph Attention Networks." *Stat* 1050 (20): 10–48550.
Venkata Ramana, B., and G. Narsimha. 2022. *Identification of the Ideal Team Capabilities and Predictive Success Measure for Software Projects Using Machine Learning*. Singapore: Springer, 593–608.
Wang, C., C. Ye, Y. Bi, J. Wang, and Y. Han. 2023. "Application of Mechanical Product Design Parameter Optimization Based on Machine Learning in Identification." *Production Planning & Control* 1–15. https://doi.org/10.1080/09537287.2022.2160388
Wang, L., Z.-H. You, Y.-M. Li, K. Zheng, and Y.-A. Huang. 2020. "GCNCDA: A New Method for Predicting circRNA-Disease Associations Based on Graph Convolutional Network Algorithm." *PLoS Computational Biology* 16 (5): e1007568-e1007568. https://doi.org/10.1371/journal.pcbi.1007568
Wasserman, S., and K. Faust. 2003. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
Wauters, M., and M. Vanhoucke. 2016. "A Comparative Study of Artificial Intelligence Methods for Project Duration Forecasting." *Expert Systems with Applications* 46: 249–261. https://doi.org/10.1016/j.eswa.2015.10.008
Wauters, M., and M. Vanhoucke. 2017. "A Nearest Neighbour Extension to Project Duration Forecasting with Artificial Intelligence." *European Journal of Operational Research* 259 (3): 1097–1111. https://doi.org/10.1016/j.ejor.2016.11.018
Xin, Y., L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang. 2018. "Machine Learning and Deep Learning Methods for Cybersecurity." *IEEE Access*. 6: 35365–35381. https://doi.org/10.1109/ACCESS.2018.2836950
Yaakobi, A., M. Goresh, I. Reychav, R. McHaney, L. Zhu, H. Sapoznikov, and Y. Lib. 2019. "Organisational Project Evaluation via Machine Learning Techniques: An Exploration." *Journal of Business Analytics* 2 (2): 147–159. https://doi.org/10.1080/2573234X.2019.1675478
Yaseen, Z. M., Z. H. Ali, S. Q. Salih, and N. Al-Ansari. 2020. "Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model." *Sustainability* 12 (4): 1514. https://doi.org/10.3390/su12041514
Yeh, J.-Y., and C.-H. Chen. 2020. "A Machine Learning Approach to Predict the Success of Crowdfunding Fintech Project." *Journal of Enterprise Information Management* 35 (6): 1678–1696. https://doi.org/10.1108/JEIM-01-2019-0017
Yu, L. 2023. "Project Engineering Management Evaluation Based on GABP Neural Network and Artificial Intelligence." *Soft Computing* 27 (10): 6877–6889. https://doi.org/10.1007/s00500-023-08133-9
Yue, J. 2021. *Construction Project Cost Management Mode Based on Artificial Intelligence Technology*, 580–587. United States: Springer International Publishing.
Yurdakurban, V., and N. Erdogan. 2018. "Comparison of Machine Learning Methods for Software Project Effort Estimation." *26th Signal Processing and Communications Applications Conference.* Izmir, Turkey: IEEE.
Zakaria, N. A., A. R. Ismail, A. Y. Ali, N. H. M. Khalid, and N. Z. Abidin. 2021. "Software Project Estimation with Machine Learning." *International Journal of Advanced Computer Science and Applications* 12 (6): 726–734. https://doi.org/10.14569/IJACSA.2021.0120685

Zebari, R., A. Abdulazeez, D. Zeebaree, D. Zebari, and J. Saeed. 2020. “A Comprehensive Review of Dimensionality Reduction Techniques for Feature Selection and Feature Extraction.” *Journal of Applied Science and Technology Trends* 1 (2): 56–70. https://doi.org/10.38094/jastt1224

Zhang, C., J. Yao, G. Hu, and X. Cao. 2023. “A Machine Learning Based Funding Project Evaluation Decision Prediction.” *Computer Systems Science and Engineering* 45 (2): 2111–2124. https://doi.org/10.32604/csse.2023.030516

Zhang, H., H. Nguyen, X.-N. Bui, T. Nguyen-Thoi, T.-T. Bui, N. Nguyen, D.-A. Vu, V. Mahesh, and H. Moayedi. 2020. “Developing a Novel Artificial Intelligence Model to Estimate the Capital Cost of Mining Projects Using Deep Neural Network-Based Ant Colony Optimization Algorithm.” *Resources Policy* 66: 101604. https://doi.org/10.1016/j.resourpol.2020.101604

Zhang, J., F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen. 2011. “Data-Driven Intelligent Transportation Systems: A Survey.” *IEEE Transactions on Intelligent Transportation Systems* 12 (4): 1624–1639. https://doi.org/10.1109/TITS.2011.2158001

Zhang, J., Z. Jiang, X. Hu, and B. Song. 2020. “A Novel Graph Attention Adversarial Network for Predicting Disease-Related Associations.” *Methods (San Diego, Calif.)*179: 81–88. https://doi.org/10.1016/j.ymeth.2020.05.010

Zhou, G., A. Etemadi, and A. Mardon. 2022. “Machine Learning-Based Cost Predictive Model for Better Operating Expenditure Estimations of U.S. light Rail Transit Projects.” *Journal of Public Transportation* 24: 100031. https://doi.org/10.1016/j.jpubtr.2022.100031