PAPER • **OPEN ACCESS**

Data-Driven Process Mining Framework for Risk Management in Construction Projects

To cite this article: A Khodabakhshian and F Re Cecconi 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1101** 032023

View the article online for updates and enhancements.

IOP Conf. Series: Earth and Environmental Science 1101 (2022) 032023 doi:10.1088/1755-1315/1101/3/032023

Data-Driven Process Mining Framework for Risk Management in Construction Projects

A Khodabakhshian[1] and F Re Cecconi[1]

[1]Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Via Ponzio, 31, 20133, Milan Italy

Ania.khodabakhshian@polimi.it

Abstract. Construction Projects are exposed to numerous risks due to their complex and uncertain nature, threatening the realization of the project objectives. However, Risk Management (RM) is a less efficient realm in the industry than other knowledge areas given the manual and time-consuming nature of its processes and reliance on experience-based subjective judgments. This research proposes a Process Mining-based framework for detecting, monitoring, and analysing risks, improving the RM processes using evidence-based event logs, such as Risk Registers and Change-Logs within previous projects' documents. Process Mining (PM) is a data-driven methodology, well established in other industries, that benefits from Artificial Intelligence(AI) to identify trends and complex patterns among event logs. It performs well while intaking large amounts of data and predicting future outputs based on historical data. Therefore, this research proposes a Bayesian Network (BN)-based Process Mining framework for graphical representation of the RM processes, intaking the conditional dependence structure between Risk variables, and continuous and automated risk identification and management. A systematic literature review on RM, PM, and AI forms the framework theoretical basis and delineates the integration areas for practical implementation. The proposed framework is applied to a small database of 20 projects as the case study, the scope of which can be tailored to the enterprise requirements. It contributes to creating a holistic theoretical foundation and practical workflow applicable to construction projects and filling the knowledge gap in inefficient and discrete conventional RM methods, which ignore the interdependencies between risk variables and assess each risk isolated.

1. Introduction

Construction projects are complicated, uncertain, unique, and dynamic in nature, and they require substantial investment and construction time, which increases the likelihood of risks [1]. Prince2 Standard defines risk as an uncertain event or set of events that, should it occur, will affect the achievement of the project/portfolio objectives. Risk is measured by the combination of the probability of a perceived threat or opportunity occurring and the magnitude of its impact on project objectives [2], like time and cost overruns and scope creep. Therefore, efficient Risk Management practices to exploit or enhance positive risks (opportunities) while avoiding or mitigating negative risks (threats) are vital to meet project objectives.

However, risk analysis is in poor condition in construction compared to other industries such as finance or insurance [3]. Many practitioners are still not aware of the importance of integrating RM in their projects. Moreover, most RM practices in the industry are based on intuition, judgments, and experience of project managers and are conducted in a manual, time-consuming, and ineffective fashion.

Risks and their severity vary from one project to another based on the project's characteristics. Therefore, identifying risk drivers and their inter-relationship with a knowledge-driven approach can serve as a Decision Support System for improving projects' RM productivity and performance [4].

Various novel RM models and techniques are being implemented in practice on project and enterprise levels: e.g., data mining, business scorecard analysis, system dynamics, or simulation approaches [5]. However, these methods are being used in a scattered manner and for specific issues. Hence, a holistic, coherent, and systematic framework to integrate the benefits of these methods is lacking. On the other hand, given the great current tendency to use Industry 4.0 technologies in Construction Engineering and Management for digitalization, automation, and increasing productivity of the projects throughout their entire life cycle [6], the RM realm can significantly benefit from these advancements as well.

This research proposes the Process Mining approach for improving RM processes in the construction industry. It seeks common areas and techniques between RM and Process Mining realms. From this viewpoint and as a process, RM faces various problems and issues throughout the project's life cycle that can be addressed and solved by AI techniques like Machine Learning and Deep learning algorithms. Consequently, a Bayesian Network-based graphical presentation model is presented, embodying the research's systematic and process-driven viewpoint, which depicts the risk tree of a small set of 15 construction projects in Italy as the case study of the Risk Identification stage. Given the small number of the input samples, the BN is simplified and is focused merely on financial risks. For this purpose, two BN models, namely Naive Bayes and Bayesian Search, are proposed, which are validated by five new projects.

2. Literature Review

A comprehensive literature review has been conducted to find interrelations between Process Mining, RM, and AI. Though no similar publications on Process Mining for RM in the Construction Industry are found, some remarkable adjunct research works are mentioned in this section.

2.1. Process Mining for Risk Management

Process Mining is an analytical, evidence-based, and data-driven method to detect, monitor, and improve business or project processes [7]. As a theme of Business Project Management, Process Mining is a combination of data-mining and traditional model-driven Business Project Management that analyses processes simply and systematically by exploiting current and historical process data stored as event logs in the information systems. For this purpose, Machine Learning and Data Mining algorithms are used to understand the current performance of the processes and their deviations concerning a normative model [8]. In contrast to traditional methods like process mapping, Process Mining is fact-based, objective, continuously enhancing, self-operating, efficient, less time-consuming, holistic, and detailed [9].

PM has an outstanding performance in ever-changing and uncertain environments like construction projects and can intake Industry 4.0 novel technologies for the Digital Transformation of the industry. Digital transformation is taking place at a high pace thanks to the advancement of AI and Information Systems to support the firms' business processes, causing digital and the physical process flows to be wholly intertwined. Digital connectivity among designers, managers, workers, consumers, and physical industrial assets can unlock enormous value and provide significant benefits. PM can bring integrity into the application of data-driven approaches, most of which suffer from unstructured processes and fail to provide an adequate procedural model [10].

Like Project Management, PM has standard implementation steps based on the application objectives, e.g., 1) process discovery, 2) process conformance checking, 3) process enhancement [8]. Therefore, it can be implemented in enterprises and projects with various project management maturity levels. Moreover, Project Management processes are of structured type since they are well-defined, predictable, repeatable, and with predetermined inputs and outputs, making the PM techniques implementation more viable and manageable.

Based on Kulakli and Birgun (2021) [7] comprehensive literature review, there has been a significant increase in publication numbers of Process Mining starting from 2015. It is studied for various business applications with multidimensional techniques, like transaction verification level, accident emergency rescue, fraud detection, information system audits and reengineering the export process, information navigation, production planning in ERP environment, assessing the life cycle phase of the business process redesign, and accounting control flow [11].

Integration of PM and RM has been limitedly studied in previous research works. Taroun (2014) [12] studied different risk models and measures in construction projects by assessing the various definitions, risk elements, and allied concepts of risk models. Caron, Vanthienen, and Baesens (2013) [13] provided a full exploration of the applicability of PM in the context of the eight components of the COSO Enterprise RM Framework, which was illustrated based on the risks involved in insurance claim handling processes. Lamine et al. (2020) [15] researched to establish the Business Process-Risk Integrated Method (BPRIM) framework to address risks considering enterprise engineering. In PM integration with enterprise and project processes, Liu et al. (2012) [16] proposed a generic approach of business process simulation for operational decision support by simulating credit card applications.

2.2. Artificial Intelligence for Risk Management

Data-driven approaches like AI algorithms can improve analytical capabilities across RM due to the ability to analyze enormous amounts of data while offering a high granularity and depth of predictive analysis [17]. The prediction is made by identifying relationships or patterns in a data sample and the effect of each parameter on the outcome, learning from previous data and generalizing the rules to new data. Though specific algorithms perform better for each problem (e.g., supervised learning, unsupervised learning, reinforcement learning), the application of AI approaches is highly context dependent. Moreover, their application for construction project RM is still relatively new and unknown [18]. Therefore, it is required to identify all problems throughout the process using PM and find the proper algorithm. For this purpose, Probabilistic Graphical Models are suggested.

Probabilistic Graphical Models are statistical techniques based on probability and graph theory that enables modeling of stochastic systems to perform risk and probability analysis. Among these models, Bayesian Networks are the most implemented one for analyzing causal influences. BNs, also known as Directed Acyclic Graphical model (DAG), present the causal relationship between random variables by a directed graph based on prior knowledge [19]. They are composed of a BN structure, including nodes as random variables and directed arcs as the indicator of the causal relationships among the variables, and a Conditional Probability Distribution table, which presents the influence of a parent node on a child node. The subjective causal relationships between variables are determined by learning algorithms, expert opinion, or both [20], the former of which is the focus of this research. For this purpose, AI algorithms are required to learn the relationship between the variables and their effect on the final outcome.

There are remarkable studies on AI applications in Risk Management. Liu et al. (2021) [21] developed a risk evaluation method based on the Bayesian network for Urban Rail Transit PPP project construction. Mittnik and Starobinskaya (2010) [22] presented a hybrid BN-based operational-risk taxonomy for modeling common shocks and mapping causal dependencies between frequencies and severity of risk events. Chattapadhyay, Putta, and Mohan Rao (2021) [23] developed a risk prediction system based on a cross analytical-Machine Learning model for construction megaprojects, using K-means clustering based on high-risk factors and allied sub-risk component identification. Harl et al. (2020) [24] proposed a Predictive Business Process Monitoring technique using Gated Graph Neural Networks (GGNNs) to forecast the future behavior of a running process instance or the value of process-related metrics and to make a prediction more explainable by visualizing how much the different activities included in a process impacted the prediction.

About the integration of AI and Process Mining, Ou-Yang and Winarjo (2011) [25] proposed a Petri Net integration approach to support multi-agent PM by using the alpha algorithm. Stefanini et al. (2020) [10] proposed a PM-based methodology to achieve a procedural, comprehensive, and computable process model to deal with unstructured processes by assessing and combining the outcomes of different PM algorithms. Butt (2020) [26] offered an integrative business process management framework used as a reference model by academics and manufacturing organizations in their journey towards a successful transition from traditional manufacturing to Industry 4.0.

3. Research Methodology

To depict the framework stages and workflow, it is essential to analyze RM implementation steps from the practical point of view as our target process. Therefore, the latest versions of RM practices in project management standards (e.g., PMBOK 6, Prince2, and ICB 4.0) have been reviewed. Based on them, RM has seven sequential stages; 1)Plan Risk Management, 2)Identify Risks, 3)Perform Qualitative Risk Analysis, 4)Perform Quantitative Risk Analysis, 5)Plan Risk Responses, 6)Implement Risk Responses, 7)Monitor Risks [27]. Based on the enterprise's status, Project Management and RM maturity level, these seven stages can form the basis for process discovery, process conformance checking, or process enhancement procedure.

This research aims to define a PM implementation framework to improve the RM practices' performance. However, the main challenge is the common perspective towards projects as a one-time, temporary, and unique phenomenon. The project-driven perspective towards RM hinders effective data and lessons learned register and causes data scarcity for systematic analysis and decision making. In most cases, risk analysis is merely done during the tender and contract awarding phase. The risk tables are not updated throughout the construction phase by real-time data. PM tries to emphasize that though projects are unique and different, similar risk patterns and factors occur in them. Therefore, data science and AI algorithms can highlight these patterns learned from previous projects. This approach can automate and improve current manual RM processes and risk control technique to ascertain the optimum identification, assessment, and mitigation of the risks.

The proposed framework architecture, presented in figure 1, consists of data sources, event logs, Process Mining levels and purposes, RM steps, and techniques. Data sources consist of data at project, strategic, and enterprise levels. This raw data needs to be gathered and stored systematically in Project Management Information Systems for analysis and decision making. Event logs are the data required for each RM stage, vary based on the requirements of the specific problem and the technique used. However, the main ones are Risk register, Risk report, change logs, Issue logs, Project performance information, etc. It is noteworthy that the type of project documents might differ from one enterprise or project to another, but the contents are quite the same. Therefore, the disciplines and practices of each PM level (Process Discovery, Process Conformance Checking, and Process Enhancement) should be tailored to all or some RM steps, based on the need and maturity level of the project or enterprise.

Some problems will arise throughout the tailoring process that needs to be addressed and solved by proper AI techniques. These techniques can include a wide range of Machine Learning and Deep Learning algorithms. The requirement of the model input for each RM step is different based on the problem type, e.g., classification, regressions, etc., precision requirement, and available data. The input data from event logs should be pre-processed, normalized, and train and test datasets should be determined to train the model for learning from previous data and make a prediction for future data. figure 2. presents the data flow between Event logs and the proposed model in training and application phases. It can be noticed that Event logs fid the input data to the model in the training phase and vice versa happens in the application phase, resulting in automatic identification and analysis of project risks by the model.

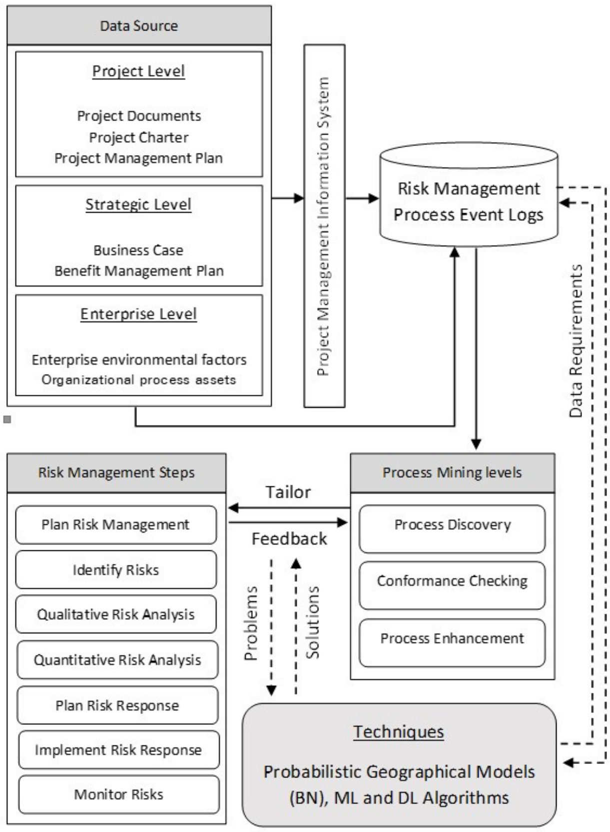

Figure 1. Process Mining Architecture for the proposed Risk Management framework

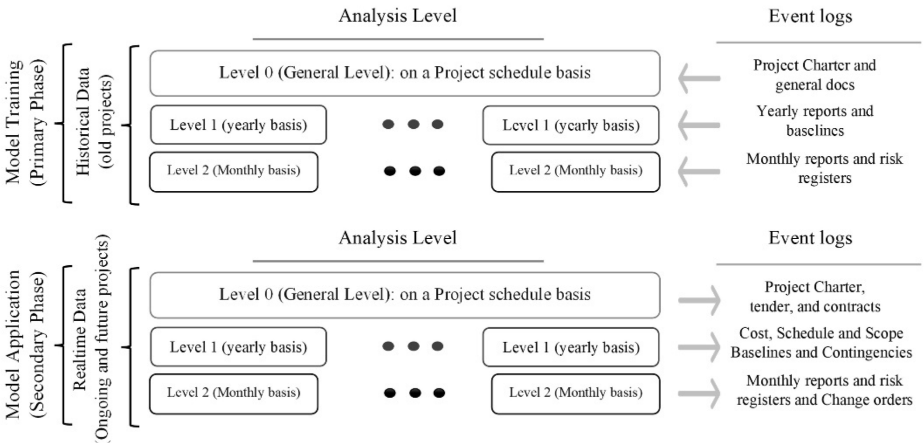

Figure 2. Process Mining Model implementation procedure with respect to the Event Logs

3.1. Framework for Process Discovery

Figure 3 presents the workflow for PM-based RM approach for Process Discovery, Conformance Checking, and Process Enhancement phases sequentially from left to right. For Process Discovery, it is assumed that the enterprise has no established RM processes, techniques, or templates (within Organizational Process Assets [27]). Therefore, the main goal is to propose an actual process model and establish real-time links to the key data sources and event logs [9]. The "Plan Risk Management" step is highly important at this level. The required event logs are the Project Charter, Project Management Plan, Project documents like stakeholder register, Enterprise Environmental Factors, and Organizational Process Assets.

3.2. Framework for Conformance Checking

At this level, it is assumed that the enterprise has established RM processes. However, the procedures, techniques, formats, inputs, and outputs should be confirmed and evaluated compared to business requirements or policies. Therefore, the main goal is to use data-driven and statistical methods to understand the root cause (and enhancement opportunity) of every process variation, determine its operational significance, and prioritize responses according to its impact on meaningful KPIs tied to key business outcomes [9].

3.3. Framework for Process Enhancement

The main goal is to improve an existing process model at this level by approaches like process KPI reporting and control and bottlenecks discovery and resolving. This research uses AI algorithms to automate and optimize the steps throughout the RM process that are conducted manually and inefficiently. Process automation drives excellence throughout a digital workforce, cutting process costs and freeing up teams to focus on value-generating work. AI modeling uncovers patterns (and their root causes and impacts) across data from previous cases and systems and adds an accurate analytical insight to the process. The performance enhancement is the level where the application of AI can be most beneficial and tangible. They shift the process towards being objective and evidence-based, which solves the subjectivity of traditional RM practices. It is noteworthy that each algorithm is fed by input data from event logs of previous projects, like the project’s features, associated risks, and probability. Once trained properly, it can automatically perform the target process.

3.4. Bayesian Network for the Case Study

Bayesian Networks can present a holistic, interrelated, and graphical overview of the entire risks associated with a project or enterprise. Fifteen construction projects in Italy were thoroughly analyzed to create a practical risk tree consisting of all possible risk categories associated with a database of projects. As the result of the data gathering from event logs, the general risk tree (figure 4) was created in GENIE Academic software. BNs were chosen for risk modeling because of their excellent performance in small datasets and their process-based analysis, aligned with the PM nature.

The Risk Identification realm for Financial Risks is chosen for practical implementation of the network. Based on historical data, there are 14 main project features affecting the financial risk, and the financial risks fall into 16 main types, all of which are presented in the financial risk BN (figure 6). Moreover, the Financial Risk types and their distribution is presented in figure 5. The structure and parameters of the risk network, presenting the relationships between parent and child nodes, are learned using two different algorithms, Naive Bayes and Bayesian Search For this purpose, the network is populated with the financial risk information of the previous 15 projects. Naive Bayes is conducted separately for each risk in a binary fashion (yes/no), while the Bayesian Search is conducted for the entire risks altogether. As a result, each input and output variable's probability is automatically learned and presented in figure 6.

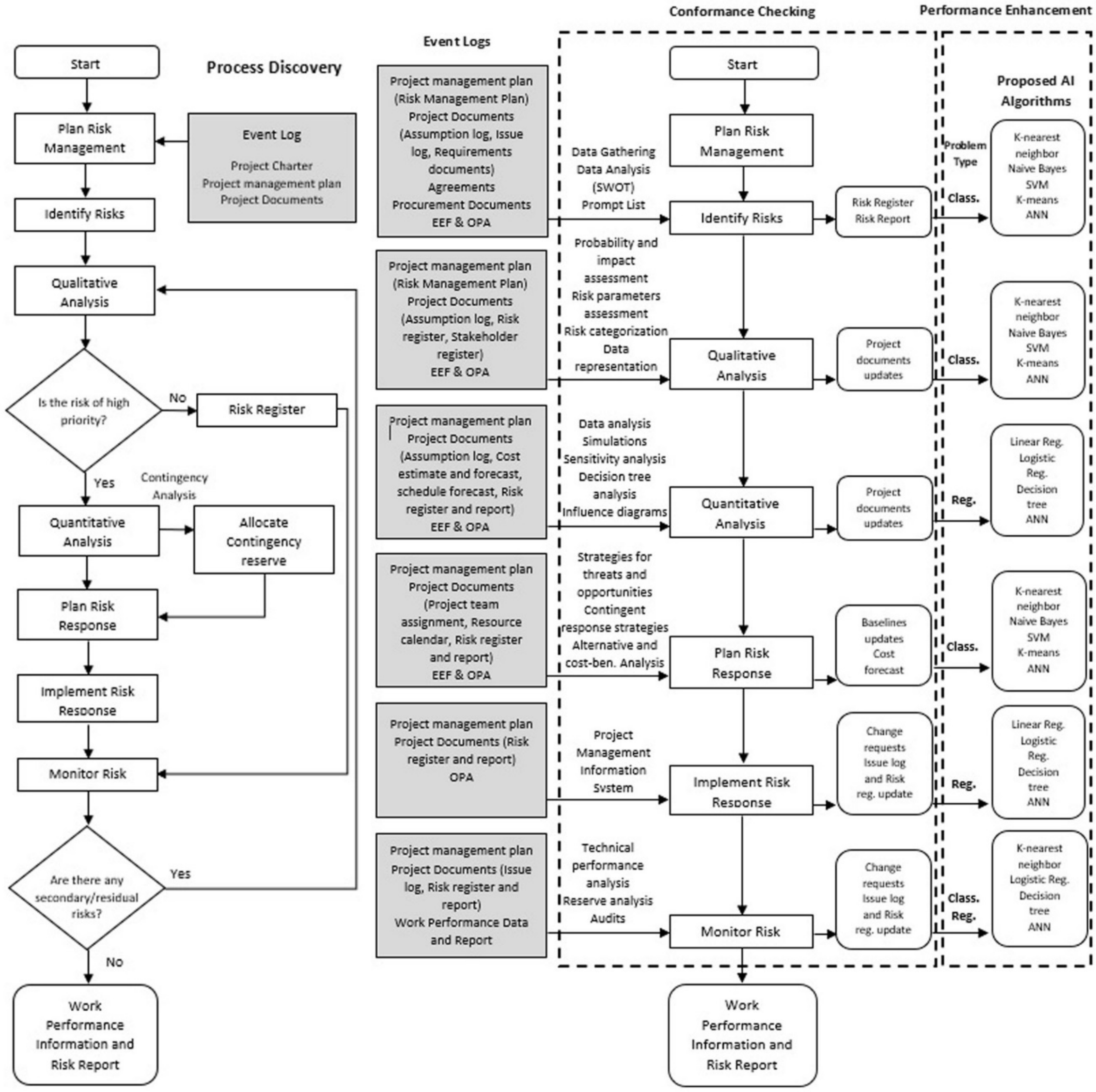

Figure 3. Framework workflow for in Process Discovery, Conformance Checking, and Performance Enhancement Phases

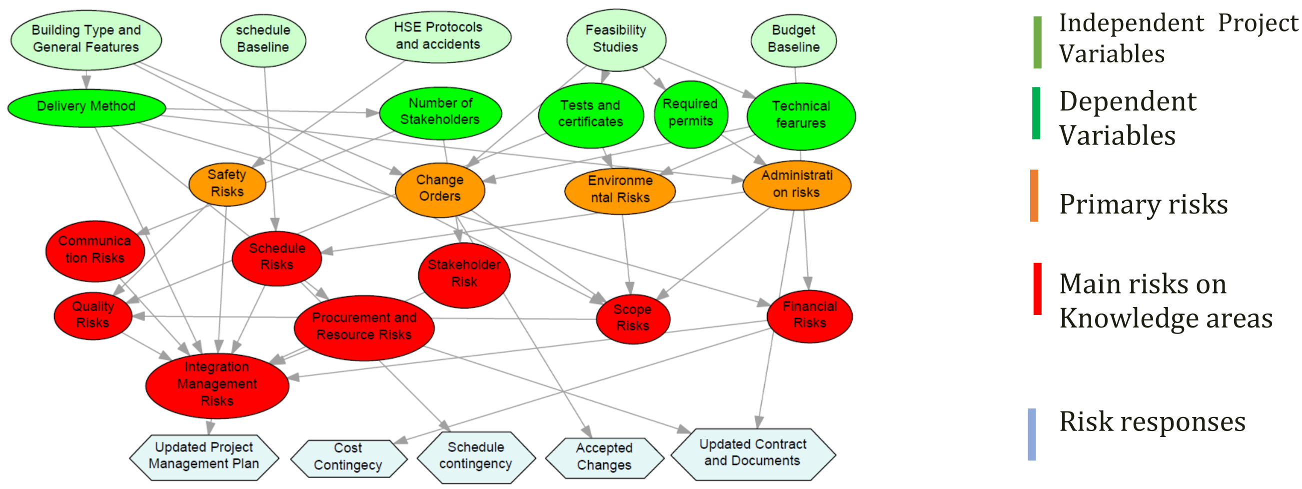

Figure 4. Risk Tree of the Case Study created by Bayesian Network

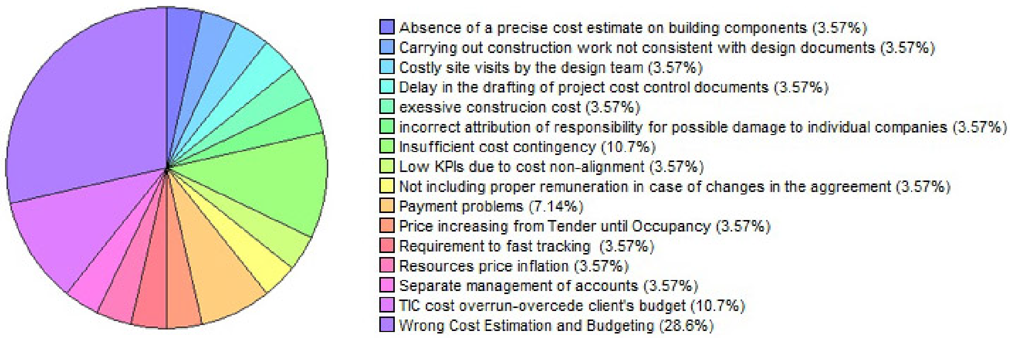

Figure 5. Financial Risk groups identified in the 15 case study projects

4. Findings and Discussion

The results of BN implementation for the case study are presented in figure 6. In the case of the Naive Bayes model, we have only one outcome with two states of yes and no (the "Wrong Budget Estimation" risk), and the probability of the "yes" state shows its possibility of occurrence and necessity for consideration in the project's risk register. In the Bayesian Search model, the Risk Description node has 16 different states (each representing one of the identified risks). The probability of each state after inserting the evidence or characteristics (event log data) of a new project shows whether that specific risk should be identified in the risk list or not. The network was generated by learning from the input data and was validated by five new projects. The validation results 80% accuracy (4 out of 5 correct prediction) for the Naive Bayes model and 40 % accuracy for Bayesian Search model (4 out of 5 correct prediction), which was a predictable result based on the limited number of projects. Figure 6a presents the probability of each parent node given the evidence that the "Wrong Budget Estimation" risk has happened and figure 6b presents the risk exposure of the network using the Bayesian Search algorithm.

As evident in the results and flowcharts, the proposed framework can easily be applied in practice for different purposes with minimum conflicts and disorders due to PM and RM disciplines' high conformity, implementation steps, and data requirements. Due to RM steps' structured and well-defined nature, it is easy to determine event logs, inputs, techniques, and outputs for each step. Moreover, it enables benefiting from AI methods like Bayesian Network for automatically and efficiently solving PM-related problems.

An advantage of the Bayesian Network-based risk tree is its ability to update immediately based on new priors, beliefs, and knowledge. That is, if the RM performance got better during the PM application and the probability of the risks occurring diminished, the network could immediately update its priors and probabilities to fit the current status. This is precisely the research objective to constantly monitor and control the risks' status and continually improve the RM practices performance. Hence, the proposed framework can improve the PM maturity level for the RM practices in a project or enterprise. Moreover, for a given number of projects, specific structure learning algorithms perform properly. For example, for our case of a small dataset, the automatic interdependency learning might not be so accurate. Therefore, we can either choose a simple Naive Bayes model ignoring the interdependencies or integrate the expert's subjective opinion with the objective data for more complex algorithms implementation.

5. Conclusions and Further Research

Despite the extensive adoption of Artificial Intelligence in different fields of engineering, science, and construction management, its practical application in Risk Management is scattered, problem-based, unsystematic, and without established frameworks. Moreover, as a subjective knowledge area, Risk Management is not being conducted effectively in construction projects and requires novel technologies to solve the shortcoming of traditional, time-consuming, and experience-based methods. Therefore, this paper proposes a Process Mining-based framework for construction RM, which benefits from Bayesian Networks and Probabilistic Graphical Models in a process-oriented and systematic mode. This framework learns the interdependencies and causal impacts of risk variables from historical data registered in event logs. It is applicable to all seven steps of the RM process, e.g., identification, etc., and to projects with different PM maturity levels, e.g., Conformance Checking, etc. Integrating the disciplines of RM, PM, and AI, this framework can serve as a data-driven decision support system and serve the construction industry's digital transformation goal, which significantly improves the project's controllability, efficiency, agility, and success rate.

The main limitation of the research was the small number of projects and the existence of unretrievable missing values, which obliged us to shrink down the scope to merely financial risks and simplify the network. Future steps of this ongoing research will be including a broader database of projects, which will enable implementing more advanced BN models and other AI algorithms like Artificial Neural Networks for higher precision rates. Moreover, as mentioned in figure 4, this framework should integrate all risk categories to achieve the best-desired results and improve the RM processes' performance. Therefore, the research will expand its implementation scope to include other risk categories and study their interdependencies.

a)

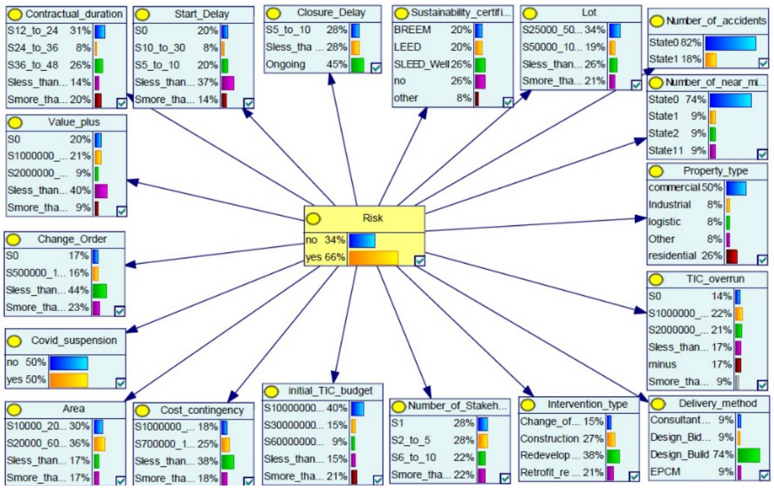

Figure 6a) Case study Naïve Bayes Network for “Wrong Budget Estimation” financial risk identification

b)

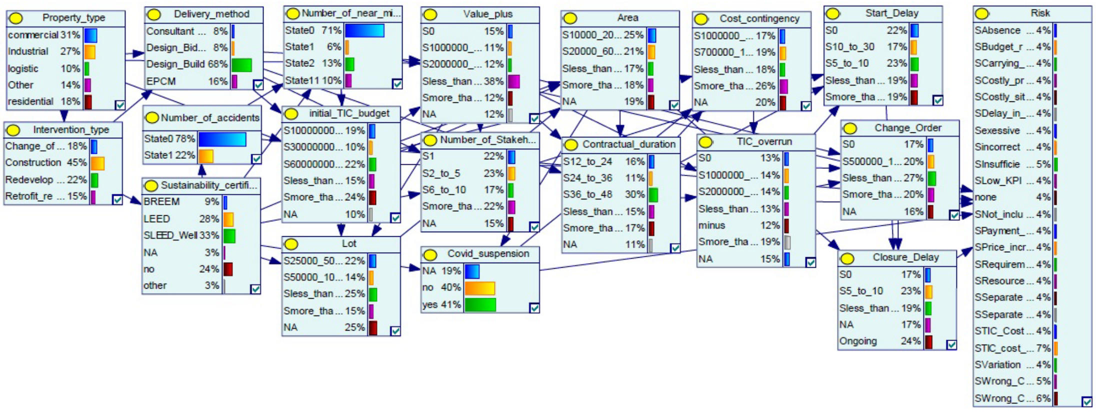

Figure 6b) Case study Bayesian Search Network for all financial risks’ identification

6. References

[1] Kardes I, Ozturk A, Cavusgil ST and Cavusgil E 2013 Managing global megaprojects:Complexity and risk management. *Int. Bus. Rev*. **22** 905–17

[2] Court R, Andrews S and Park B 2017 *Managing Successful Projects with Prince2* (Norwich: TSO (The Stationery Office))

[3] Laryea S. 2008 Risk pricing practices in finance, insurance, and construction. *COBRA 2008. Constr. Build. Res. Conf.* R. Inst. Chart. Surv., Dublin: University of Reading.

[4] De Palma A, Picard N and Andrieu L 2012 Risk in Transport Investments. *Netw. Spat. Econ.* **12** 187–204

[5] Wu DD, Kefan X, Hua L, Shi Z and Olson DL 2010 Modeling technological innovation risks of an entrepreneurial team using system dynamics: An agent-based perspective. *Technol. Forecast Soc. Change* **77** 857–69
[6] Pan Y and Zhang L 2021 Roles of artificial intelligence in construction engineering and management: A critical review and future trends. *Autom. Constr.* **122** vol 103517
[7] Kulakli A and Birgun S 2021 Process Mining Research in Management Science and Engineering Fields: The Period of 2010–2019 In: Durakbasa NM and Gençyilmaz MG (eds) *Digital Conversion on the Way to Industry 4.0. ISPR 2020. Lecture Notes in Mechanical Engineering* (Cham.: Springer)
[8] van der Aalst WMP 2016 *Process mining: data science in action*. 2nd ed. (Berlin: Springer)
[9] Miller M 2014 *The Ultimate Guide to Process Mining* p. 240 (Celonis)
[10] Stefanini A, Aloini D, Dulmin R and Mininno V 2020 A process mining methodology for modeling unstructured processes. *Knowl. Process. Manag.* **27** vol 4 294-310
[11] He Z, Wu Q, Wen L and Fu G. A 2019 process mining approach to improve emergency rescue processes of fatal gas explosion accidents in Chinese coal mines. *Saf. Sci.* **111** 154–66
[12] Taroun A 2014 Towards a better modelling and assessment of construction risk: Insights from a literature review. *Int. J. Proj. Manag.* **32** 101–15
[13] Caron F, Vanthienen J and Baesens B 2013 A comprehensive investigation of the applicability of process mining techniques for enterprise risk management. *Comput. Ind.* **64** 464–75
[14] Lamine E, Thabet R, Sienou A, Bork D, Fontanili F and Pingaud H 2020 BPRIM: An integrated framework for business process management and risk management. *Comput. Ind.* **117** vol 103199
[15] Kraus MA and Drass M 2020 Artificial intelligence for structural glass engineering applications overview, case studies and future potentials. *Glas. Struct. Eng.* **5** 247–85
[16] Liu Y, Zhang H, Li C and Jiao RJ 2012 Workflow simulation for operational decision support using event graph through process mining. *Decis. Support. Syst.* **52** 685–97
[17] Guzman-Urbina A, Aoyama A and Choi E 2018 A polynomial neural network approach for improving risk assessment and industrial safety. *ICIC Express Lett.* **12** 97–107
[18] Basaif AA, Alashwal AM, Mohd-Rahim FA, Karim SBA and Loo SC 2020 Technology awareness of artificial intelligence (AI) application for risk analysis in construction projects. *Malaysian Constr. Res. J.* **9** 182–95
[19] Borujeni SE, Nannapaneni S, Nguyen NH, Behrman EC and Steck JE 2021 Quantum circuit representation of Bayesian networks. *Expert Syst. Appl*. **176** vol 114768
[20] Garvey MD, Carnovale S and Yeniyurt S 2015 An analytical framework for supply network risk propagation: A Bayesian network approach. *Eur. J. Oper. Res*. 243 618–27
[21] Liu Z, Jiao Y, Li A and Lui X 2021 Risk assessment of urban rail transit PPP project construction based on bayesian network. *Sustain*. **13** 11507
[22] Mittnik S and Starobinskaya I 2010 Modeling dependencies in operational risk with hybrid Bayesian networks. *Methodol. Comput. Appl. Probab.* **12** 379–90
[23] Chattapadhyay DB, Putta J and Rama Mohan Rao P 2021 Risk identification, assessments, and prediction for mega construction projects: A risk prediction paradigm based on cross analytical-machine learning model. *Buildings* **11** 172
[24] Harl M et al. 2020 Explainable predictive business process monitoring using gated graph neural networks gated graph neural networks. *J. Decis. Syst.* **29** sup 1 312-327
[25] Ou-Yang C and Winarjo H 2011 Petri-net integration - An approach to support multi-agent process mining. *Expert Syst. Appl.* **38** 4039–4051
[26] Butt J 2020 A conceptual framework to support digital transformation in manufacturing using an integrated business process management approach. *Designs* **4** 1–39
[27] Project Management Institute(PMI) 2017 *A guide to the project management body of knowledge (PMBOK guide).* 6th ed. (Pennsylvania, USA: Project Management Institute)