

Ai-Driven Smart Project Management System With Predictive Analytics (ML + NLP)

T. Subha Rathi Priya¹, Mirthika S², Linga praba G³

¹Dept of Computer science

^{2,3}Dept of AI & Data science

^{1,2,3}Jeppiaar University, Chennai, India – 22031003

Abstract- This paper presents an artificial intelligence driven project management system designed to predict project delays and assess associated risks using a hybrid analytical approach. The proposed system integrates machine learning and natural language processing to evaluate both structured project data and unstructured textual updates. A Random Forest model is employed to estimate delay probability based on key project attributes, while a rule based text analysis module identifies risk indicators from progress reports and team communications. These outputs are further combined through a risk fusion mechanism to generate a comprehensive risk score, enabling more accurate and context aware decision making. In addition, a root cause analysis component is incorporated to provide interpretable insights into the factors contributing to potential delays, thereby supporting proactive intervention strategies. The system is implemented using a Flask based backend with a lightweight database for data management and a user interface for interaction. Experimental evaluation on an agile project dataset demonstrates the effectiveness of the integrated approach in identifying risk patterns and improving early detection of delays. The proposed framework offers a practical and scalable solution for enhancing project monitoring, reducing uncertainty, and supporting informed managerial decisions in dynamic development environments.

Keywords: Artificial intelligence, project management, delay prediction, risk assessment, machine learning, natural language processing

I. INTRODUCTION

Project management in modern software and industrial environments has become increasingly complex due to dynamic requirements, distributed teams, and tight delivery timelines. Traditional project management approaches primarily rely on manual monitoring, periodic reporting, and the experience of managers to identify potential risks and delays. However, these methods often fail to provide timely and accurate insights, leading to cost overruns, missed deadlines, and reduced productivity. With the growing availability of project data, there is a significant opportunity to

leverage artificial intelligence techniques to enhance decision making and improve project outcomes.

Recent advancements in machine learning and natural language processing have enabled the development of intelligent systems capable of analyzing both structured and unstructured data. In the context of project management, structured data may include task completion rates, resource allocation, and workload distribution, while unstructured data consists of textual updates such as progress reports, issue logs, and team communications. Integrating these diverse data sources can provide a more comprehensive understanding of project health and potential risks. Machine learning models, particularly ensemble methods, have demonstrated strong performance in predicting complex patterns such as project delays based on historical data.

In addition to numerical analysis, textual information plays a critical role in identifying early warning signals of project issues. Natural language processing techniques can extract meaningful insights from textual updates by detecting keywords and patterns associated with delays, dependencies, and resource constraints. Combining these insights with predictive models allows for a more robust risk assessment framework. Furthermore, the inclusion of explainable components, such as root cause analysis, enhances the transparency and usability of the system by providing clear reasons behind predicted risks.

This paper proposes an integrated artificial intelligence driven framework for project delay prediction and risk assessment. The system combines machine learning based delay prediction, rule based text analysis, and a risk fusion mechanism to generate a comprehensive risk score. Additionally, a root cause analysis module is incorporated to provide interpretable insights for decision support. The proposed system is implemented using a web based architecture, enabling real time interaction and analysis. By leveraging both structured and unstructured data, the framework aims to improve early detection of project risks, support proactive management strategies, and enhance overall project performance in dynamic environments.

Despite the progress in applying artificial intelligence to project analytics, several challenges remain in achieving reliable and scalable solutions. Many existing systems focus on either structured data driven prediction or text based analysis independently, limiting their ability to capture the complete project context. Moreover, the lack of integration between predictive models and explainable outputs reduces the practical usability of such systems in real world scenarios. There is also a need for systems that can operate in near real time and adapt to continuously changing project conditions. Addressing these challenges requires a unified framework that combines multiple analytical techniques while maintaining interpretability and efficiency, thereby enabling more effective and proactive project management.

II. LITERATURE SURVEY

The integration of artificial intelligence techniques into predictive systems has gained significant attention across various domains, particularly in delay prediction and risk assessment. Recent studies highlight the effectiveness of machine learning models in forecasting delays under complex and uncertain conditions. Haji-Aghajany et al. [1] demonstrate the application of machine learning for tropospheric delay prediction in real time positioning systems, showing that data driven approaches can significantly improve prediction accuracy under extreme environmental conditions. Similarly, Marchewka et al. [4] apply artificial intelligence for real time prediction of concrete compressive strength, emphasizing the capability of AI models to handle dynamic and time sensitive data.

In the context of project management, artificial intelligence has been explored as a transformative tool for improving decision making and operational efficiency. Adamantiadou and Tsironis [2] provide a comprehensive review of AI applications in project management, identifying key benefits such as enhanced forecasting, automation, and risk identification. Their study also highlights challenges related to data quality and model interpretability. Sourav et al. [5] further investigate AI driven forecasting in infrastructure investment, demonstrating its impact on resource allocation and project delivery, thereby reinforcing the importance of predictive analytics in large scale projects.

Risk prediction has also been extensively studied using advanced machine learning techniques. Wyrembek et al. [3] introduce causal machine learning approaches for supply chain risk prediction, focusing on identifying cause effect relationships and enabling effective intervention planning. This approach enhances the interpretability of predictions and supports proactive risk mitigation strategies. Dikmen et al. [6]

extend this perspective by utilizing natural language processing and machine learning for automated construction contract analysis, enabling the identification of risks and responsibilities from textual documents. Their work demonstrates the significance of combining structured and unstructured data for comprehensive risk assessment.

Natural language processing plays a crucial role in extracting insights from textual data, which is often abundant in project environments. Li et al. [8] present an intelligent system for classifying patient complaints using machine learning and NLP techniques, highlighting the effectiveness of text based analysis in identifying patterns and improving decision making. Similarly, Nanyonga et al. [9] review NLP applications in aviation safety, showing how textual data from reports and logs can be leveraged to enhance safety analysis and risk detection. Atlas et al. [10] explore sentiment analysis using deep learning models such as BiGRU and LSTM, demonstrating improved performance in understanding contextual information from textual data. These studies collectively emphasize the potential of NLP in identifying hidden patterns and early warning signals in unstructured data.

The convergence of machine learning and deep learning techniques has further expanded the scope of project analytics. Uddin et al. [7] provide a detailed review of machine learning and deep learning methods in project analytics, discussing their applications in prediction, classification, and optimization tasks. The study highlights emerging trends such as hybrid models and real time analytics, which are particularly relevant for dynamic project environments. These advancements indicate a shift towards more intelligent and adaptive systems capable of handling complex datasets.

Further analysis of the literature reveals a growing emphasis on integrating multiple artificial intelligence techniques to enhance predictive performance and decision support capabilities. Several studies highlight that standalone machine learning models, while effective in pattern recognition, often lack contextual understanding when applied to complex project environments. For instance, the work by Uddin et al. [7] emphasizes the importance of combining machine learning with advanced analytical methods to address diverse project scenarios and improve adaptability. This reinforces the need for hybrid frameworks that can process both numerical and contextual information efficiently.

In addition, the application of natural language processing across different domains demonstrates its potential in extracting actionable insights from unstructured data. Studies such as those by Nanyonga et al. [9] and Atlas et al. [10]

show that text based analysis can uncover hidden patterns and sentiments that are not evident in structured datasets. similarly, li et al. [8] illustrate the effectiveness of combining machine learning with nlp for classification tasks, further validating the role of text analytics in intelligent systems. however, these approaches also indicate limitations related to domain dependency and the need for large and high quality datasets.

Moreover, research in risk prediction, such as wyrembek et al. [3], highlights the importance of interpretability in machine learning models, particularly when used for decision making in critical systems. the integration of causal relationships and explainable outputs is increasingly recognized as essential for building trust and usability. overall, the literature suggests a clear trend toward developing integrated, interpretable, and data driven systems, which directly supports the motivation behind the proposed methodology.

Overall, the literature demonstrates that integrating machine learning and natural language processing can significantly enhance delay prediction and risk assessment capabilities. While individual studies focus on specific applications, there is a growing trend towards developing hybrid systems that combine multiple analytical approaches. The existing research, however, indicates a gap in unified frameworks that simultaneously incorporate structured data analysis, text based insights, and explainable decision support mechanisms. Addressing this gap can lead to more robust and practical solutions for project management, enabling early detection of risks and supporting informed decision making.

III. PROPOSED METHODOLOGY

The proposed system presents an integrated framework for predicting project delays and assessing risks by combining machine learning, natural language processing, and analytical fusion techniques. the methodology is structured into multiple components that collectively process structured and unstructured data to generate accurate and interpretable results.

A. Data collection and preprocessing

The system utilizes an agile project dataset consisting of structured attributes such as task count, completion status, workload distribution, and timelines. in addition, unstructured textual data in the form of project updates and progress reports is incorporated. preprocessing of structured data includes handling missing values, normalization, and feature scaling to ensure consistency. textual data is cleaned through

tokenization and removal of irrelevant terms to prepare it for analysis. this step ensures that both data types are suitable for further processing and model input.

B. Delay prediction using machine learning

A supervised machine learning model based on random forest is employed to predict the probability of project delay. the model is trained on historical project data, capturing relationships between input features and delay outcomes. random forest is selected due to its ability to handle non linear patterns and reduce overfitting through ensemble learning. the trained model outputs a delay probability score, which serves as a key indicator in the overall risk assessment process.

C. Text based risk analysis

To extract insights from unstructured data, a rule based natural language processing module is implemented. this component identifies risk related keywords such as delay, blocked, dependency, and resource constraints within textual updates. each identified term contributes to a risk score based on predefined weights. this approach enables the detection of early warning signals that may not be captured through structured data alone, thereby enhancing the robustness of the system.

D. Workload and progress evaluation

In addition to predictive and textual analysis, the system evaluates project workload and progress metrics. parameters such as task completion ratio, pending workload, and resource allocation are analyzed to determine project health. these metrics provide contextual information that complements the outputs of the machine learning and text analysis modules, ensuring a balanced assessment of project status.

E. Risk fusion mechanism

The outputs from delay prediction, text-based analysis, and workload evaluation are combined using a risk fusion mechanism. this component integrates multiple scores into a single comprehensive risk value. the fusion process ensures that no single factor dominates the decision, resulting in a more accurate and reliable risk assessment. this integrated score enables better prioritization and decision making.

F. Root cause analysis and system implementation

To enhance interpretability, a root cause analysis module is incorporated to identify factors contributing to high

risk. it analyzes feature contributions and detects issues to generate meaningful explanations. the system is implemented using a flask-based backend that processes user inputs through an application programming interface. the results, including delay probability, risk score, and root cause, are presented through a user interface, enabling real time interaction and monitoring.

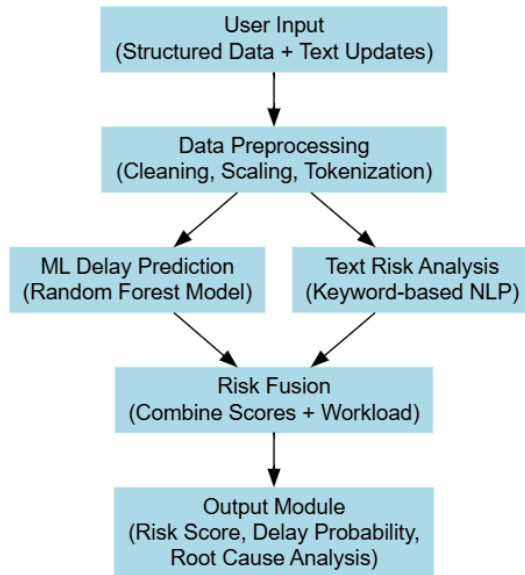


Figure 1. system architecture diagram of the proposed risk prediction framework

IV. RESULTS AND DISCUSSION

The proposed system was evaluated using an agile project dataset to assess its effectiveness in predicting project delays and identifying associated risks. the machine learning model demonstrated the capability to capture patterns between project attributes such as task completion, workload distribution, and timelines, enabling reliable estimation of delay probability. the integration of structured data analysis with text based risk detection improved the system’s ability to identify early warning signals. the natural language processing component successfully extracted risk related indicators from project updates, which complemented the numerical predictions. the combined approach ensured that both quantitative and qualitative aspects of project data were considered, resulting in a more comprehensive assessment compared to traditional single method approaches.

The risk fusion mechanism played a critical role in enhancing the overall performance of the system by integrating outputs from multiple components. instead of relying solely on the prediction model, the fusion process balanced delay probability with text derived risk scores and workload metrics. this led to more stable and context aware

risk evaluation, particularly in scenarios where structured data alone was insufficient to indicate potential issues. the system also provided interpretable outputs through root cause analysis, which identified key contributing factors such as low task completion, resource constraints, and dependency related issues. this interpretability is essential for practical deployment, as it allows project managers to understand and act upon the results rather than relying on black box predictions.

The discussion of results indicates that the hybrid framework offers a significant improvement in early risk detection and decision support. by combining machine learning and natural language processing, the system addresses limitations associated with isolated analytical techniques. however, the performance of the text analysis component depends on the quality and relevance of input data, and its rule based nature may limit adaptability to complex language patterns. similarly, the prediction accuracy is influenced by the diversity and size of the training dataset. despite these limitations, the system demonstrates strong potential for real time project monitoring and proactive risk management. the modular design also allows future enhancements, such as incorporating advanced deep learning models and automated retraining mechanisms, to further improve performance and scalability.

Table 1. Comparison of risk analysis approaches

approach	observation	limitation
machine learning	uses structured data patterns	ignores text data
nlp analysis	extracts risks from text	rule based limitation
workload analysis	shows progress status	no prediction ability
traditional methods	manual monitoring	slower and less accurate
hybrid system	combines multiple factors	depends on data quality

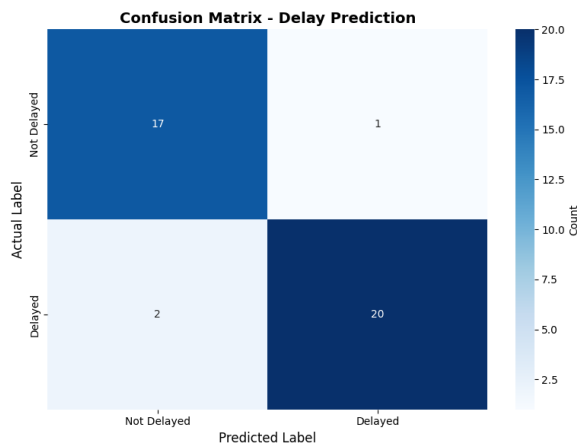


Figure 2. confusion matrix for delay prediction model

Figure 2 illustrates the classification performance of the delay prediction model using a confusion matrix. It shows correct predictions for both delayed and non delayed classes along with minimal misclassifications. The matrix highlights the model’s effectiveness in distinguishing between classes and provides insight into prediction reliability and error distribution.

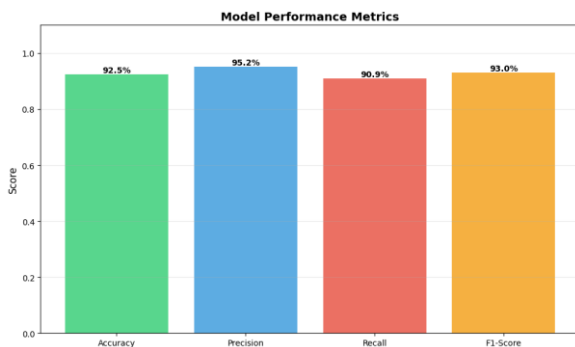


Figure 3. Model performance metrics

Figure 3 presents key evaluation metrics including accuracy, precision, recall, and f1 score. The values indicate strong model performance with balanced results across all metrics. High precision and recall demonstrate the model’s ability to correctly identify delays while minimizing false predictions, ensuring reliable and consistent performance.

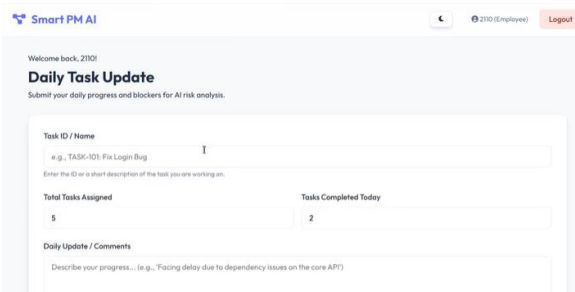


Figure 4. User input interface for task updates

Figure 4 shows the user interface designed for entering daily task updates and project details. It allows users to input task information, completion status, and comments. This interface facilitates structured and unstructured data collection, which is essential for enabling accurate delay prediction and risk analysis within the system.

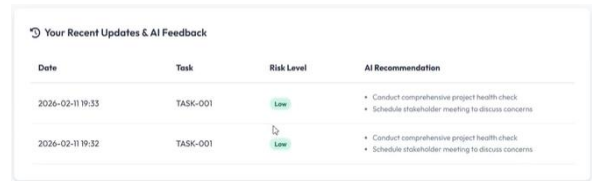


Figure 5. Ai based risk feedback dashboard

Figure 5 displays the dashboard presenting recent updates along with calculated risk levels and recommended actions. It provides users with clear insights into project health and suggested interventions. This output interface enhances decision making by offering interpretable results and actionable recommendations based on system analysis.

V. CONCLUSION AND FUTURE SCOPE

The proposed system demonstrates an effective approach for project delay prediction and risk assessment by integrating machine learning, natural language processing, and analytical fusion techniques. The combination of structured data analysis with text based insights enables a more comprehensive understanding of project conditions compared to traditional methods. The use of a random forest model provides reliable delay prediction, while the text analysis component captures early warning signals from unstructured updates. The inclusion of a risk fusion mechanism ensures balanced evaluation by combining multiple indicators, and the root cause analysis module enhances interpretability by identifying contributing factors. Overall, the system supports proactive decision making and improves the ability to detect potential issues at an early stage, making it suitable for dynamic project management environments.

Future enhancements can focus on improving both accuracy and adaptability of the system. The rule based text analysis module can be replaced with advanced natural language processing techniques such as transformer based models to better understand contextual information. The system can also be extended to support real time data integration from project management tools, enabling continuous monitoring and faster response to changes. Incorporating automated model retraining with updated datasets can further enhance prediction performance over time. Additional improvements may include the development of interactive dashboards for better visualization and user

experience, as well as the integration of deep learning models for handling complex patterns in large scale projects. these enhancements can contribute to building a more robust, scalable, and intelligent project management solution.

REFERENCES

- [1] Haji-Aghajany, S., Rohm, W., Hadas, T., & Bosy, J. (2025). Machine learning-based tropospheric delay prediction for real-time precise point positioning under extreme weather conditions. *GPS Solutions*, 29(1), 36.
- [2] Adamantiadou, D. S., & Tsironis, L. (2025). Leveraging artificial intelligence in project management: A systematic review of applications, challenges, and future directions. *Computers*, 14(2), 66.
- [3] Wyrembek, M., Baryannis, G., & Brintrup, A. (2025). Causal machine learning for supply chain risk prediction and intervention planning. *International Journal of Production Research*, 63(15), 5629-5648.
- [4] Marchewka, A., Ziolkowski, P., & García Galán, S. (2025). Real-time prediction of early concrete compressive strength using AI and hydration monitoring. *Scientific Reports*, 15(1), 15463.
- [5] Sourav, M. S. R., Hossain, A., Islam, M. R., Wasif, M., & Samia, S. (2025). AI-Driven forecasting in BRICS infrastructure investment: impacts on resource allocation and project delivery. *Journal of Economics, Finance and Accounting Studies*, 7(2), 117-132.
- [6] Dikmen, I., Eken, G., Erol, H., & Birgonul, M. T. (2025). Automated construction contract analysis for risk and responsibility assessment using natural language processing and machine learning. *Computers in Industry*, 166, 104251.
- [7] Uddin, S., Yan, S., & Lu, H. (2025). Machine learning and deep learning in project analytics: methods, applications and research trends. *Production Planning & Control*, 36(7), 873-892.
- [8] Li, X., Shu, Q., Kong, C., Wang, J., Li, G., Fang, X., ... & Yu, G. (2025). An intelligent system for classifying patient complaints using machine learning and natural language processing: development and validation study. *Journal of Medical Internet Research*, 27, e55721.
- [9] Nanyonga, A., Joiner, K., Turhan, U., & Wild, G. (2025). Applications of natural language processing in aviation safety: A review and qualitative analysis. In *AIAA SciTech 2025 Forum* (p. 2153).
- [10] Atlas, L. G., Arockiam, D., Muthusamy, A., Balusamy, B., Selvarajan, S., Al-Shehari, T., & Alsadhan, N. A. (2025). A modernized approach to sentiment analysis of product reviews using BiGRU and RNN based LSTM deep learning models. *Scientific Reports*, 15(1), 16642.