International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Canalyser: An Intelligent Project Risk Analysis and Visualization System Using Machine Learning

Mr. Mehul Patil, Ms. Riddhi Mehta

Department of computer science and engineering, Parul Institute of Technology, Parul university, Gujarat, india

Abstract—Project risk management plays a crucial role in ensuring the successful execution of projects across domains such as software development, construction, and enterprise systems. Traditional risk assessment techniques rely heavily on manual evaluation, expert judgment, and qualitative models, which are often time-consuming, inconsistent, and prone to human bias. These limitations reduce the effectiveness of decision-making, particularly in large-scale and data-intensive environments. This paper presents Canalyser, an intelligent web-based system designed to automate project risk analysis using machine learning and interactive visualization techniques. The system accepts multiple project parameters, including budget, duration, team size, complexity, and resource availability, and predicts the associated risk level categorized as low, medium, or high. Furthermore, it provides real-time graphical insights through dashboards, enabling users to analyze trends and make informed decisions. Experimental observations indicate that the proposed system improves efficiency, reduces manual effort, and enhances predictive accuracy compared to traditional approaches.

Index Terms—Project Risk Analysis, Machine Learning, Predictive Analytics, Data Visualization, Risk Management, Streamlit

I. INTRODUCTION

The rapid growth of digital technologies and increasing project complexity have significantly transformed modern project management practices. Projects today operate in dynamic environments characterized by uncertainty, resource constraints, and evolving requirements. These factors introduce various forms of risk that can adversely affect project outcomes, including delays, cost overruns, and performance failures. Effective risk management is therefore essential to ensure successful project execution.

Traditional project risk analysis methods rely on techniques such as expert judgment, historical comparisons, and qualitative risk matrices. Although these approaches have been widely adopted, they suffer from several limitations. The reliance on human expertise introduces subjectivity and inconsistency, while the manual nature of analysis makes it inefficient for handling large datasets. Additionally, these methods lack real-time adaptability, making them unsuitable for rapidly changing project environments.

With advancements in machine learning and data analytics, it is now possible to automate risk prediction and enhance decision-making processes. Machine learning algorithms can analyze patterns within historical project data and generate predictive insights with higher accuracy and consistency. However, many existing systems focus primarily on prediction and lack effective visualization mechanisms for interpreting results.

To address these challenges, this paper proposes Canalyser, an intelligent project risk analysis system that integrates machine learning with interactive dashboards. The system provides automated risk prediction along with real-time visualization, thereby improving efficiency, reducing manual effort, and enhancing decision-making capabilities.

II. LITERATURE REVIEW

Project risk analysis has been extensively studied using various methodologies ranging from traditional qualitative techniques to advanced machine learning approaches. Early methods such as SWOT analysis, expert judgment, and risk matrices were primarily qualitative and relied on domain expertise. While these methods are simple and easy to implement, they lack objectivity and scalability, making them less suitable for complex projects.

Quantitative approaches introduced statistical models and probability-based techniques such as Monte Carlo simulation to improve accuracy. These methods provide more reliable results but require significant computational effort and domain knowledge, limiting their practical applicability in real-time scenarios.

Recent research has focused on the application of machine learning techniques for risk prediction. Algorithms such as Logistic Regression, Decision Trees, and Random Forest have been used to classify project risks based on historical data. These models demonstrated improved prediction accuracy; however, most existing implementations lack integration with user-friendly visualization tools, and many systems fail to provide real-time insights.

The primary research gap identified is the lack of a unified system that combines predictive modeling with interactive visualization in a user-friendly interface. This paper addresses this gap by proposing an integrated solution that enhances both prediction and interpretability.

III. SYSTEM ARCHITECTURE

The Canalyser system is designed using a modular architecture that integrates multiple components to ensure efficient data processing, prediction, and visualization. The architecture consists of four primary layers: user interface, data processing, machine learning, and visualization.

A. User Interface Layer

The user interface layer is developed using Streamlit, which provides an interactive platform for users to input project-related parameters. This layer ensures ease of use and accessibility, allowing both technical and non-technical users to interact with the system effectively.

B. Data Processing Layer

The data processing layer handles tasks such as data cleaning, normalization, and transformation. Libraries such as Pandas and NumPy are used to manage structured data efficiently. This layer ensures that the input data is properly formatted and suitable for machine learning models.

C. Machine Learning Layer

The machine learning layer is responsible for training and applying predictive models. Algorithms such as Logistic Regression, Decision Trees, and Random Forest are used to classify project risks. The model analyzes input features and generates predictions based on learned patterns.

D. Visualization Layer

The visualization layer presents results using graphical tools such as Matplotlib and Plotly. This layer provides dashboards displaying risk levels, trends, and comparisons, enabling users to interpret results quickly.

IV. METHODOLOGY

The methodology of the Canalyser system involves a structured pipeline consisting of data collection, preprocessing, model training, and prediction. Each stage is designed to ensure accuracy and efficiency in risk analysis.

A. Data Collection

Data collection involves gathering project-related inputs such as budget, duration, team size, complexity, and resource availability. These parameters serve as features for the predictive model.

B. Data Preprocessing

Data preprocessing is performed to improve data quality and consistency. This includes handling missing values, normalizing numerical features, and encoding categorical variables. Proper preprocessing is essential to ensure reliable model performance.

C. Model Training

Model training is conducted using supervised learning techniques. The input features are used to train models that classify projects into different risk categories. Multiple algorithms are evaluated to determine the most effective model.

D. Prediction

Prediction involves applying the trained model to new input data to determine the risk level. The output is categorized into low, medium, or high risk, providing actionable insights for decision-making.

V. IMPLEMENTATION

The system is implemented using Python and follows a modular design approach. The input module collects user data through the Streamlit interface, while the processing module prepares the data for analysis. The prediction module applies machine learning algorithms to generate risk classifications. The visualization module displays results using charts and graphs, providing an intuitive understanding of project risks. The integration of these modules ensures smooth data flow and real-time analysis. The system is designed to be scalable and can be extended to support additional features and datasets in the future.

VI. RESULTS AND ANALYSIS

The system was evaluated using sample datasets to assess its performance and effectiveness. The results indicate that the proposed system provides faster and more consistent predictions compared to traditional manual methods. Visualization further enhances the interpretability of results, allowing users to identify patterns and trends easily. Performance evaluation of different models shows that Logistic Regression achieves approximately 82% accuracy, Decision Tree achieves around 85% accuracy, and Random Forest achieves up to 90% accuracy. Among these models, Random Forest demonstrates the highest accuracy and stability, making it the preferred choice for risk prediction in the Canalyser system.

VII. ADVANTAGES AND LIMITATIONS

The proposed system offers several advantages, including automated risk prediction, reduced human error, faster processing, and an intuitive user interface. These features collectively improve the overall efficiency and effectiveness of project risk management.

However, the system also has certain limitations. Its performance depends on the quality and availability of training data, and the accuracy of predictions can be further improved with larger and more diverse datasets.

VIII. FUTURE SCOPE

Future enhancements of the system may include the integration of real-time data sources, implementation of advanced deep learning models, deployment on cloud platforms, and support for multi-project analysis. These improvements will further enhance the scalability and effectiveness of the system in handling complex project environments.

IX. CONCLUSION

This paper presented Canalyser, an intelligent project risk analysis system that integrates machine learning with visualization techniques. The system automates risk prediction, improves efficiency, and enhances decision-making capabilities. By combining predictive analytics with interactive dashboards, the proposed solution provides a scalable and effective approach to modern project risk management.

REFERENCES

[1] Project Management Institute, A Guide to the Project Management Body of Knowledge (PMBOK® Guide), 6th ed. Newtown Square, PA: PMI, 2017.

[2] T. Mitchell, Machine Learning. New York: McGraw-Hill, 1997.

[3] Scikit-learn Documentation, "Machine Learning in Python," Scikit-learn.org. [Online]. Available: https://scikit-learn.org

[4] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 3rd ed. Waltham, MA: Morgan Kaufmann, 2011.

[5] Various research studies on project risk prediction using machine learning methodologies and data analytics frameworks.