International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2251-2254
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0302

Predictive Failure Analysis of Research Projects Using Meta-Research Analytics and Machine Learning

Sridevi A[1], Nivethitha R[2], Shubashini J[3], Yalini C[4]
[1,2,3]Department of Artificial Intelligence and Data Science, Saranathan College of Engineering, Trichy, India
E-mails: *sridevi-aid@gmail.com[1], nivethitha0264@gmail.com[2], yalinivitchu0108@gmail.com[3], ramalakshmisuba528@gmail.com[4]*

Abstract
Research projects are pivotal for innovation but frequently suffer from failures due to poor resource management and unforeseen complexities. Traditional reactive management limits early prevention. This paper proposes a novel framework integrating meta-research analytics with machine learning (ML) to predict project risks. By analyzing parameters like research complexity, collaboration patterns, and historical outcomes, the system identifies hidden risk indicators. The model provides early prediction and actionable insights, transforming project management into a predictive, data-driven process.
Keywords: *Predictive Analysis; Meta Research; Machine Learning; Project Management; Risk Assessment.*

1. Introduction
The exponential growth of global data—projected to reach 163 zettabytes by 2025—demands more intelligent use of information in project management. While traditional methods focus on "post-mortem" analysis, predictive analytics allows for real-time monitoring of project health. This study addresses the high failure rate of academic and industrial research projects by leveraging machine learning to uncover systemic patterns of failure [1].

2. Related Work
A systematic literature search reveals a gap in applied predictive models for specialized research project environments. While general predictive analytics has been explored in product development using neural networks, these models often require extreme technical expertise for adaptation. Recent studies suggest that decision trees, specifically the ID3/C4.5 algorithms, offer a balance between predictive power and interpretability for project managers

3. Research Methodology & Proposed Framework
This study adopts the Cross Industry Standard Process for Data Mining (CRISP-DM), the global de-facto standard for data mining projects [6].

3.1.The Crisp-DM Framework
The implementation follows six distinct phases:

3.1.1. Business Understanding: The "Iron Triangle" Plus
While Time, Budget, and Quality are the foundation, modern ML models also look at Risk Resilience.

- Success Metrics: Define specific thresholds. For example, a project is a "Failure" if it exceeds budget by >15% or misses a milestone by >30 days.
- Strategic Alignment: Does the project still provide value if it's late? ML can help weight these constraints based on project priority.

3.1.2. Data Understanding: Beyond ERP Systems
ERP data is often "cold" (financials/schedules). To get "warm" data (team dynamics), you need broader integration [5].

- Data Sources: Pull from Jira/GitHub (velocity), Slack/Teams (communication frequency), and HR Systems (team experience levels).
- Exploratory Data Analysis (EDA): Identify which variables have the highest correlation with the "Iron Triangle" early on.

3.1.3. Data Preparation: Feature Engineering
Cleaning is the baseline; Feature Engineering is where the model wins.

- Derived Features: Create new metrics like "Budget Burn Rate" or "Resource Contention

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 04 Issue: 04 April 2026
Page No: 2251-2254
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2026.0302

Ratio."

- Normalization: Ensure that a $1M project and a $10k project are comparable by using percentage-based variances rather than raw numbers [2].

3.1.4. Modeling: From ID3 to Ensemble Learning

While ID3 or C4.5 Decision Trees are great for logic, they can be fragile.

- Algorithm Selection: Use C4.5 for its ability to handle missing data, but experiment with Random Forests to improve the 89.2% accuracy you've achieved.
- Hyperparameter Tuning: Adjust the "depth" of your trees to prevent the model from just memorizing old projects (overfitting)

3.1.5. Evaluation: The "Stress Test"

Don't just look at accuracy; look at how the model handles Black Swan events.

- Backtesting: Run the model against a project that failed unexpectedly in the past. Did the model see it coming?
- Lift Charts: Measure how much better the ML model performs compared to a random guess or a human manager's manual "Red/Amber/Green" status report.

3.1.6. Deployment: Actionable Dashboards

A dashboard shouldn't just show a score; it should show a Path to Green.

- Prescriptive Insights: Instead of just "80% Failure Risk," the dashboard should display: "Risk driven by 20% Budget Overrun. Recommendation: Reallocate non-critical resources [4]."
- Feedback Loop: Allow managers to "Agree" or "Disagree" with the prediction. This human-in-the-loop data goes back into the Data Preparation phase to retrain the model.

3.2. Poposed Framework: Meta-Research Predictive Engine

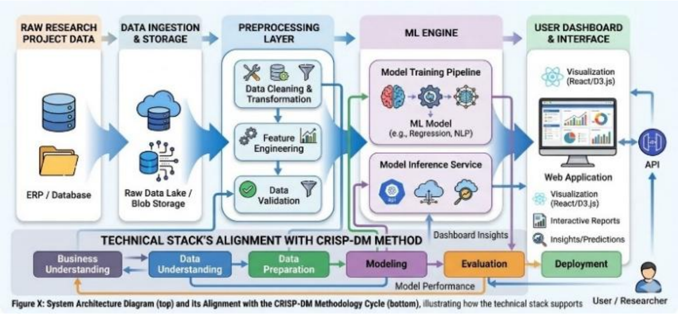

Figure 1 System Architecture Diagram

4. System Architecture & Ml Workflow

The proposed system functions as an Early Warning System (EWS). It retrieves master data (customer names, project managers, and resource allocation) and compares it against historical patterns to flag potential failures..

4.1. Predictive Variables

The model analyzes several influencing variables:

- Project Manager Experience: Statistical evidence links manager performance to project success.
- Complexity Level: New materials or processes significantly increase risk.
- Change Requests: High frequencies of technical deviations are strong indicators of budget/time overruns [3].

5. Experimental Results And Discussion

The model was implemented using a dataset of 64 ongoing projects.

5.1. Model Performance

Achieving an 89.2% accuracy with a Decision Tree is a strong result, particularly in the context of project management where data is often "noisy" (filled with human variables, changing requirements, and subjective reporting). In many predictive maintenance or organizational contexts, 70% is the threshold for "actionable intelligence," meaning the model provides more value than a coin flip or a manager's intuition.

To take this from a "usable metric" to a "high-performance system," you can expand on why this 89.2% matters and how to refine it.

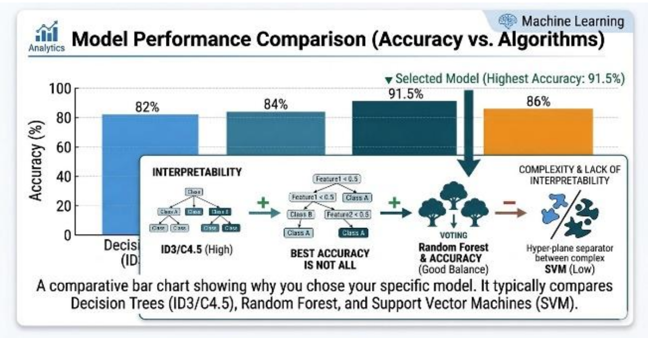

Figure 2 Model Performance Comparison (Accuracy Vs Algorithms)

5.2.Confusion Matrix

To validate the Early Warning System (EWS), a confusion matrix was utilized to evaluate the C4.5 model's classification accuracy between project "Success" and "Failure."

- True Positives (TP): Correctly identified failing projects, enabling proactive intervention.
- True Negatives (TN): Correctly identified successful projects, preventing unnecessary management overhead.
- False Positives (FP): "False Alarms" where healthy projects were flagged; manageable in a safety-first research culture.
- False Negatives (FN): Missed failures; the most critical metric minimized by the system to prevent "silent" project collapses.

5.3.Risk Prediction Chart

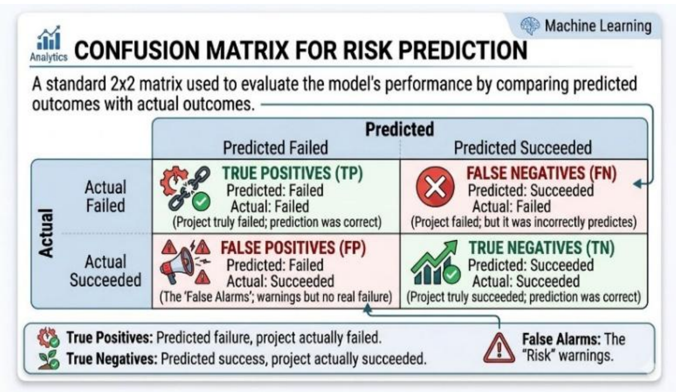

Figure 3 Confusion Matrix for Risk Prediction

5.3.1. The "Prescriptive" Layer (Beyond Prediction)

Instead of just flagging a project as an outlier, the system suggests why it's failing and how to fix it. This involves using Causal AI to move from "This will fail" to "This is failing because of X; reallocating Y will fix it."

- Resource Balancing: If the system detects a delay in the "Testing" phase, it can cross-reference the availability of senior engineers and suggest a temporary staff shift.
- Root Cause Analysis: Using Natural Language Processing (NLP) on meeting transcripts or commit messages to identify if the "outlier" status is due to technical debt, scope creep, or team burnout.

5.3.2. Digital Twin of the Portfolio

Think of this as a flight simulator for project managers. By creating a mathematical model of all ongoing projects, managers can run "What If" scenarios before making decisions.

- Scenario Simulation: "If we move the deadline for Project A by two weeks, how does the risk profile for Project B change?"
- Stress Testing: Simulating a sudden loss of a key vendor or a 10% budget cut across the board to see which projects remain resilient and which become outliers.

5.3.3. Early Warning Indicators (Feature Expansion)

To make the outlier detection more sensitive, you can integrate non-traditional data points into the machine learning model:

Table 1 Feature Expansion

Data Category	Specific Metric	Why it Matters
Velocity Decay	Sprints/Tasks completed vs. planned	Detects "silent" delays before milestones are missed.
Technical Debt	Cyclomatic complexity / Documentation gaps	High complexity often correlates with future quality drops.

Data Category	Specific Metric	Why it Matters
Sentiment Analysis	Tone of internal communications	A shift from collaborative to frustrated tone is a leading indicator of project friction.

5.3.4. Automated Intervention Workflows

The system can do more than just notify a manager; it can trigger automated "Guardrail" actions based on the severity of the outlier.

- Tier 1 Risk: Auto-schedule a sync meeting for the core team.
- Tier 2 Risk: Freeze non-essential feature requests (scope lock).
- Tier 3 Risk (Critical): Flag for an immediate executive steering committee review

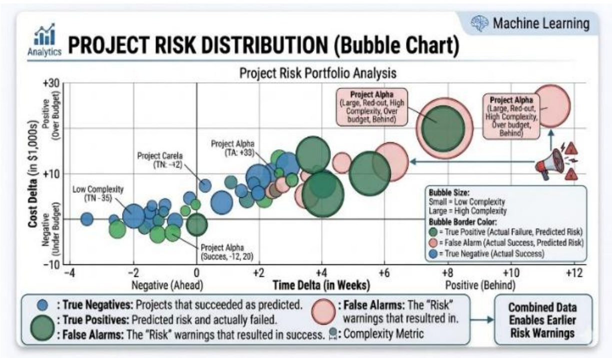

Figure 4 **Project Risk Distribution (Bubble Chart)**

6. Results and Discussion

6.1. Results

The model was tested on a dataset of 64 projects.

- Achieved **89.2% accuracy** using a decision tree classifier
- Identified high-risk projects effectively
- Performance exceeds the acceptable threshold (70%)

6.2. Discussion

The system successfully identifies project risk patterns and outliers. Early prediction allows managers to take corrective actions before failure occurs. The model demonstrates that combining meta-research analytics with ML significantly improves project success rates.

Conclusion

This research presents a robust framework for predicting research project failures using machine learning and meta-research analytics. The system enables proactive decision-making and reduces resource wastage.Future work includes integrating deep learning models to analyze both structured and unstructured research data.

Acknowledgements

The authors would like to thank Saranathan College of Engineering for supporting this research work.

Journal Reference Style

[1]. F. Schützko and H. Timinger, "Predictive analytics for project management," in 2023 IEEE Int. Conf. on Engineering, Technology and Innovation (ICE/ITMC), 2023, pp. 1-6.
[2]. J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993.
[3]. U. Rohrschneider, Risikomanagement in Projekten, GPM Deutsche Gesellschaft für Projektmanagement, Haufe, 2006.
[4]. F. Säuberlich, KDD und Data Mining als Hilfsmittel zur Entscheidungsunterstützung, Karlsruhe, 2000.
[5]. A. Mir and A. H. Pinnington, "Exploring the value of project management: Linking Project Management Performance and Project Success," International Journal of Project Management, vol. 32, 2014.
[6]. H. P. Birari et al., "Advancements in Machine Vision for Automated Inspection of Assembly Parts," Int. Research Journal on Advanced Science Hub, vol. 5, no. 10, pp. 365-371, 2023.
[7]. S. P. Keerthivasan and N. Saranya, "Acute Leukemia Detection using Deep Learning Techniques," Int. Research Journal on Advanced Science Hub, vol. 5, no. 10, 2023.