Software Design Document (SDD)

Project Title: AI Project Failure Predictor
Version: 1.0
Authors: Tirupati Naidu Kolli
Date: 19-FEB-2026

1. Introduction

1.1 Purpose:

The AI Project Failure Predictor is a professional-level tool designed to **analyze project plans** and **predict risks at the task and project level**. It leverages AI/ML to generate risk scores, classify tasks as Low/Medium/High risk, highlight critical tasks, and provide actionable suggestions for mitigation.

1.2 Scope:

- Input: Project plan data via CSV, Excel, or manual entry
- Output: Task-level and project-level risk scores, critical tasks, mitigation suggestions, and reports
- Prototype: Interactive Streamlit app with professional analytics, visualizations, and export capabilities
- Audience: Project managers, stakeholders, academic evaluators

1.3 Objectives:

- Predict task and project-level risks
- Provide actionable suggestions to reduce project failure likelihood
- Offer rich analytics and visualizations for decision support
- Ensure scalability to enterprise-sized project plans

2. System Overview / Architecture

High-Level Architecture:

[User Input / CSV Upload]
↓
[Data Preprocessing & Validation] → [Feature Engineering]
↓
[ML Risk Prediction Module] → [Risk Scores, Feature Importance]

↓

[Analytics & Visualization Module] → [Critical Tasks, Charts, Reports]

↓

[Streamlit Dashboard / UI] → [Interactive Visualization, Mitigation Suggestions, Export]

Modules:

1. **Input Module:** CSV/Excel upload, manual task entry, input validation
2. **Data Processing & Feature Engineering:** Clean, encode categorical data, scale numeric features, compute derived features (e.g., task complexity, dependency weights)
3. **ML Risk Prediction Module:**
    a. Random Forest, Gradient Boosting, or ensemble models
    b. Hyperparameter tuning
    c. Cross-validation and model evaluation (accuracy, precision, recall, F1-score)
4. **Analytics & Visualization Module:**
    a. Risk heatmaps, bar charts of top risky tasks, Gantt-style charts, overall project risk gauges
    b. Feature importance visualization
    c. Summary statistics and dependency analysis
5. **UI Module (Streamlit):**
    a. Sidebar inputs, filters, CSV upload, manual task input
    b. Main panel displays predictions, visualizations, critical tasks, and mitigation suggestions
    c. Export options (CSV, PDF)

3. Functional Requirements

- Load project plan data (CSV/Excel or manual input)
- Preprocess and clean data
- Generate task-level and project-level risk scores
- Highlight critical/high-risk tasks
- Provide actionable mitigation suggestions
- Generate rich analytics and visualizations
- Interactive Streamlit dashboard for exploring results
- Exportable risk report (CSV and PDF)
- Allow filtering and sorting by risk, task attributes, and dependencies

4. Non-Functional Requirements

- **Performance:** Handle large datasets (hundreds of tasks) efficiently
- **Usability:** Professional, intuitive Streamlit interface
- **Reliability:** Robust handling of missing, inconsistent, or erroneous data
- **Scalability:** Easily extendable to larger project plans, additional features, or integration with other project management tools
- **Maintainability:** Modular code structure for ease of updates

5. Data Design

Input Features:

Feature	Type	Description
Task ID / Name	String	Unique task identifier
Start Date / End Date	Date	Planned schedule
Planned Duration	Numeric	Days between start and end
Dependencies	List/String	IDs of dependent tasks
Team Member Allocation	Numeric	Number of people assigned
Budget Allocation	Numeric	Task budget
Historical Task Success Rate	Numeric	Past performance (0–1)
Complexity Score	Numeric	Task complexity (1–10)
Criticality	Boolean/Numeric	Whether task is on critical path
Progress %	Numeric	Completion % (optional for ongoing projects)

Outputs:

Output	Type	Description
Task Risk Score	0–1	Probability of task failure
Task Risk Category	Low/Medium/High	Risk classification
Overall Project Risk Score	0–1	Aggregated project risk
Critical Tasks Highlighted	List	Tasks with highest risk scores
Actionable Suggestions	Text	Recommendations for risk mitigation

6. Analytics & Visualizations

- **Summary statistics:** total budget, task distribution, timelines, team allocations
- **Risk heatmap:** tasks vs. risk scores
- **Bar chart:** top 5 risky tasks
- **Feature importance chart:** factors contributing to risk
- **Gantt-style charts:** visualize dependencies and critical path
- **Overall project risk gauge:** quick visual summary
- **Trend analytics:** if historical project data is available

7. UI / Prototype Design (Streamlit)

- **Sidebar:** CSV/Excel upload, manual input, filters for attributes
- **Main Panel:**
  - Task-level risk table (sortable/filterable)
  - Visualizations: heatmaps, bar charts, timelines
  - Critical tasks highlighted
  - Actionable suggestions displayed dynamically
- **Export Options:** CSV, PDF
- **Professional Styling:** Color-coded risks (green/yellow/red), clean headers, structured layout

8. Process Flow

1. User uploads project plan (CSV/Excel) or inputs tasks manually
2. Data preprocessing & feature engineering
3. ML models predict task & project-level risks
4. Analytics module calculates summaries, generates charts, highlights critical tasks
5. Streamlit dashboard displays results and suggestions interactively
6. Exportable report is generated for stakeholders
7. Optionally, iterate with additional models or updated data

9. Implementation Plan

- **Tech Stack:** Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly, Streamlit, ReportLab/PDF generation

- **Full Project Deliverables:**
    - Production-ready Streamlit app
    - Multiple ML models with evaluation metrics
    - Full analytics dashboard
    - Exportable reports
    - Extensible and modular architecture for future updates

10. Testing Plan

- Validate risk predictions on real and synthetic datasets
- Verify visualizations are correct and update dynamically
- Test Streamlit interface for usability and responsiveness
- Ensure handling of missing, inconsistent, or invalid input
- Evaluate ML models using standard metrics (accuracy, F1-score, precision, recall)

11. Future Enhancements / Scalability

- Support integration with project management tools (e.g., Jira, MS Project)
- Incorporate time-series analysis for ongoing projects
- Advanced visualization dashboards (interactive Gantt charts)
- Automated recommendation system for resource reallocation

12. Appendices / References

- Sample dataset: project_plan_full.csv
- References: scikit-learn, Streamlit, Matplotlib, Seaborn, Plotly, project risk management literature