

AI STARTUP RISK RADAR: FINAL REPORT

Predicting Entrepreneurial Risks Before They Happen

Prepared by: **Gervásio Magul**

Université de Lorraine – IAE Nancy

Master's in Entrepreneurship and Project Management

Specialization: AI for Software Engineering

Date: November 2025

Abstract

Startups face high failure rates, often due to unmanaged risks in finance, operations, market positioning, and team capabilities. This project introduces the AI Startup Risk Radar, a machine learning tool designed to predict inherent risk levels (Low, Medium, High) based on key startup attributes. Using a synthetic dataset of 150 records, models including Decision Tree, Random Forest, and Logistic Regression were systematically tested. The Random Forest model was selected for its robust performance and interpretability, with feature importance analysis highlighting budget, runway, and competition as the critical risk drivers. The results are visualized using Radar charts to display risk distribution across Financial, Operational, Market, and Team categories, and practical case studies offer actionable mitigation strategies. This project successfully blends AI, entrepreneurship, and project management, serving as both a pedagogical tool for students and a pragmatic decision-support framework for entrepreneurs.

I. Introduction

1.1 Background and Problem Statement

Empirical research consistently shows that up to 90% of startups fail within five years. This high rate of failure is largely attributed to unmanaged and unforeseen risks across multiple domains. Early-stage entrepreneurs frequently lack accessible, objective, and timely tools to anticipate these risks and understand their specific project vulnerabilities at the outset of the venture.

1.2 Project Objectives

The core mission of this project is the development of the AI Startup Risk Radar, an interpretable Artificial Intelligence tool to address this critical knowledge gap. The specific

objectives achieved are:

- To predict startup risk levels (Low, Medium, High) using supervised machine learning models.
- To visualize risk across four critical dimensions: Financial, Operational, Market, and Team.
- To develop practical case studies offering evidence-based risk mitigation recommendations.
- To deliver a professional project that effectively integrates technical rigor (AI) with practical business application (Entrepreneurship and Project Management).

II. Methodology

2.1 Dataset Generation

To ensure controlled experimentation and interpretability, a synthetic dataset of 150 records was generated.

- Features: The dataset comprises eight quantifiable startup attributes: budget, teamsize, deadlinemonths, marketscore, productmaturity, foundereexperience, competitionintensity, and runway_months.
- Target Variable: The dependent variable, risk_level (Low/Medium/High), was determined using a predefined, rule-based scoring system where critical combinations (e.g., low budget and short runway) automatically led to higher risk classification.

2.2 Machine Learning Models

Three algorithms were evaluated for predictive accuracy and explanatory power:

- Decision Tree: Employed as an initial baseline due to its simple, flowchart-like interpretability.
- Logistic Regression: Used as a linear baseline for comparison of non-linear performance gains.
- Random Forest: Selected as the final primary model for its ensemble robustness and superior ability to handle non-linear relationships while providing clear feature importance metrics.

2.3 Workflow and Tools

The project followed a standard data science pipeline: Data Ingestion (CSV) → Model Training/Testing → Prediction → Visualization. All computational tasks were executed in Python using core libraries (pandas, scikit-learn) within the Google Colab environment, with version control managed via GitHub.

III. Results and Analysis

3.1 Model Performance Comparison

The Random Forest classifier yielded the highest overall performance, demonstrating superior accuracy and balanced classification metrics (precision and recall) across the Low, Medium, and High risk classes compared to the baseline models.

3.2 Key Feature Importance

Analysis of the Random Forest model revealed the quantitative impact of each input feature on the risk prediction. The three factors identified as the most critical drivers of a startup's predicted risk were:

- Budget (Highest impact)
- Runway Months
- Competition Intensity

3.3 Risk Radar Visualizations

The model output was translated into an intuitive Risk Radar chart format, displaying the magnitude of vulnerability across the four established categories.

- **Case A (High Risk):** Visualization shows significant weakness, primarily concentrated in Financial and Market sectors.
- **Case B (Medium Risk):** Visualization indicates a pressurized but balanced risk profile, with operational stress highlighted by tight deadlines.
- **Case C (Low Risk):** Visualization confirms a strong and robust profile, underpinned by substantial financial and team capacity.

IV. Case Studies and Mitigation

4.1 Case A: High Risk

- **Prediction:** High Risk
- **Vulnerabilities:** Low budget, short runway, and intense market competition.
- **Actionable Mitigation:** Immediate priority is securing follow-on funding to extend the operational runway. Strategically, the venture must identify a niche or pivot to reduce direct competitive exposure.

4.2 Case B: Medium Risk

- **Prediction:** Medium Risk
- **Vulnerabilities:** Moderate resources, but high Operational pressure from tight deadlines.
- **Actionable Mitigation:** Management should urgently seek to extend the project timeline, focus on strengthening team capacity to meet deliverables, and conduct rapid validation of market assumptions.

4.3 Case C: Low Risk

- **Prediction:** Low Risk
- **Vulnerabilities:** Minimal risk, supported by a strong budget, experienced founder, and long runway.
- **Actionable Mitigation:** The focus shifts to disciplined execution: maintaining financial governance and investing in continuous, iterative customer and market validation to solidify traction.

V. Discussion and Future Work

5.1 Strengths and Practical Relevance

The AI Startup Risk Radar's core strength lies in its high interpretability, provided by both the feature importance scores and the visually accessible radar charts. This tool holds immense practical relevance for entrepreneurship education and functions as a valuable, objective decision-support mechanism for founders and investors.

5.2 Limitations

The principal limitation is the reliance on a synthetic dataset. The underlying simplified, rule-based risk scoring may not fully replicate the complexity, non-linearity, and exogenous factors of real-world startup failure dynamics.

5.3 Future Directions

To significantly enhance the model's predictive power and utility, future research should pursue:

- Integration and validation using real-world datasets (e.g., from databases like Crunchbase).
- Expansion of the feature set to include dynamic variables such as funding stage, customer churn rates, and specific regulatory environments.
- Development of APIs for real-time integration with active project management tools (e.g., Jira, Trello) for continuous risk monitoring.

VII. Conclusion

The AI Startup Risk Radar successfully demonstrates the application of machine learning to critical entrepreneurship and project management challenges by providing data-driven risk prediction and insightful vulnerability visualizations. By combining technical rigor with strong visual communication and practical mitigation strategies, this project sets a new standard for student work—delivering tangible impact in the field of entrepreneurial decision-making.