Prosperity Prognosticator

Machine Learning for Startup Success Prediction

1. Project Overview

Objective: Develop a machine learning-based system to predict startup success probability using historical and financial data. Assist investors and entrepreneurs in making data-driven decisions.

2. Problem Statement

Startup failure rates are high. The system aims to analyze patterns in historical data and provide probability-based success predictions.

3. Scope of Project

In Scope: Data preprocessing, model training, evaluation, prediction module. Out of Scope: Real-time streaming data, automated investments, legal advisory.

4. Functional Requirements

• Data upload (CSV) • Data cleaning and preprocessing • Feature engineering • Train ML models (Logistic Regression, Decision Tree, Random Forest, SVM, Gradient Boosting, Neural Networks) • Model evaluation (Accuracy, Precision, Recall, F1, ROC-AUC) • Prediction output with probability and risk level

5. Non-Functional Requirements

• Prediction time < 2 seconds • Accuracy target $\geq$ 75% • Scalable up to 100,000+ records • Modular and maintainable code

6. Data Requirements

Features: Funding amount, funding rounds, sector, founder experience, team size, market size, revenue, location, years in operation, burn rate. Target: Startup Success (1 = Success, 0 = Failure)

7. Technology Stack

Python, NumPy, Pandas, Scikit-learn, Matplotlib, TensorFlow (optional), Flask/Streamlit. Tools: Git, GitHub, VS Code, Jupyter Notebook.

8. Success Criteria

• Achieve ≥ 75% accuracy • Provide probability-based risk assessment • Proper documentation and reproducibility

9. Constraints

Limited dataset availability, class imbalance, overfitting risk, market unpredictability.

10. Future Enhancements

Explainable AI (SHAP/LIME), real-time API integration, cloud deployment, automated retraining.

11. Project Timeline

Phase	Duration
Requirement Analysis	1 Week
Data Collection	1 Week
Data Preprocessing	1 Week
Model Building	2 Weeks
Evaluation & Optimization	1 Week
Deployment	1 Week