

Prosperity Prognosticator: Machine Learning for Startup Success Prediction

1. INTRODUCTION

1.1 Project Overview

- Prosperity Prognosticator predicts startup success using supervised machine learning.
- The system analyzes structured startup data such as funding, market category, and business metrics.
- A Random Forest classification model is trained to predict startup success or failure.

1.2 Purpose

- To reduce uncertainty in startup investment decisions.
- To assist investors with data-driven predictions.
- To automate startup evaluation using machine learning models.

2. IDEATION PHASE

2.1 Problem Statement

- Startup investments involve high financial risk.
- Traditional decision-making relies on intuition.
- A predictive system is required to evaluate startup success probability.

2.2 Empathy Map

- Says: Is this startup worth investing in?
- Thinks: How can I reduce financial loss?
- Does: Reviews reports and funding data.
- Feels: Uncertain about outcomes.

2.3 Brainstorming

- Develop a supervised ML classification model.
- Compare Logistic Regression and Random Forest.
- Select Random Forest for strong predictive performance.

3. REQUIREMENT ANALYSIS

3.1 Functional Requirements

- Accept CSV startup dataset input.
- Perform preprocessing and cleaning.
- Train and evaluate ML model.
- Generate success probability output.

3.2 Non-Functional Requirements

- High prediction accuracy.
- Fast processing time.
- Scalability for large datasets.
- User-friendly system.

3.3 Technology Stack

- Programming Language: Python
- Libraries: Pandas, NumPy, Scikit-learn
- Algorithm: Random Forest Classifier
- Environment: Jupyter Notebook / VS Code

4. PROJECT DESIGN

- Perform Exploratory Data Analysis (EDA).
- Apply preprocessing techniques such as encoding and scaling.
- Train Random Forest model and save as .pkl file.
- Generate and visualize predictions.

5. PROJECT PLANNING & SCHEDULING

- Phase 1: Data Collection & Cleaning
- Phase 2: Exploratory Data Analysis
- Phase 3: Model Training
- Phase 4: Model Evaluation
- Phase 5: Result Visualization

6. FUNCTIONAL AND PERFORMANCE TESTING

- Evaluate accuracy score.
- Generate confusion matrix.
- Calculate precision and recall.
- Test model on unseen dataset.

7. RESULTS

- Random Forest achieved strong predictive performance.

- Model successfully classified startups as success or failure.
- Results visualized for interpretation.

8. ADVANTAGES & DISADVANTAGES

Advantages

- Reduces investment risk.
- Automated evaluation system.
- Data-driven decision support.

Disadvantages

- Dependent on dataset quality.
- Cannot predict unexpected market shifts.

9. CONCLUSION

Prosperity Prognosticator demonstrates how machine learning models such as Random Forest can effectively predict startup success and assist investors in making informed decisions.

10. FUTURE SCOPE

- Integrate real-time financial APIs.
- Develop interactive web dashboard.
- Implement advanced ensemble and explainable AI models.