Online ISSN: 3107 - 7676

IJMR 2026; 2(1): 62-64

2026 January - February

www.allmultiresearchjournal.com

Received: 13-11-2025

Accepted: 20-12-2025

Published: 18-01-2026

DOI: https://doi.org/10.54660/IJMR.2026.2.1.62-64

AI Powered Startup Investment Analysis

Amann Bhatia [1]*, Suhani Jain [2], Rahul Gurbani [3], Shruti Patil [4]
School of Computing, MIT ADT University, Pune, Maharashtra, India

Corresponding Author; **Amann Bhatia**

Abstract
Startup investments have always been characterized by high uncertainty and the lack of quantitative decision-making tools. In this project, we present an AI-driven approach for startup investment analysis using a Random Forest classification model trained on verified Indian startup data from Kaggle (Startup Funding Dataset 2023). Our web-based application, built using Python and Streamlit, predicts the probability of a startup's success or failure based on key financial, industrial, and funding metrics. The system also integrates a Generative AI module that identifies top-performing industries in specific sectors. The project blends business analytics and machine learning, providing investors and entrepreneurs with actionable insights.

Keyword: Startup Analysis, Investment Prediction, Random Forest, Machine Learning, Generative AI, Indian Startup Ecosystem, Data Pipeline.?

1. Introduction

Startups play a vital role in economic development by creating innovation and employment. However, predicting a startup's long-term success remains a complex challenge due to dynamic market conditions, variable funding patterns, and the absence of historical data for new ventures.

Traditional investors rely heavily on intuition and expert judgment, which often leads to bias or inconsistency. To address this, we designed an AI-powered analytical platform that learns from past investment data and provides data-driven predictions. The Indian startup ecosystem, with more than 90,000 registered startups as of 2023, offers an ideal case study due to its diversity in industry, funding, and regional focus.

2. Literature Review

Existing research on startup prediction includes both financial and machine learning perspectives. Early studies focused on ratio-based analysis or regression models, which provided limited predictive power. Recent work incorporates deep learning and ensemble methods for higher accuracy. Random Forest and Gradient Boosting have shown promise in modeling complex relationships in high-dimensional datasets. Data from sources such as Crunchbase and Kaggleâ€™s Indian Startup Funding Dataset offer valuable information about in-dustry trends, investment stages, and funding rounds. However, most existing models fail to combine technical analysis with business relevance. Our project aims to bridge this gap by incorporating both

predictive modeling and real-world industry interpretation through Generative AI.

3. System Architecture

The proposed system architecture is designed in modular layers, shown in Fig. 1. It integrates data preprocessing, model training, and a visualization layer through Streamlit.

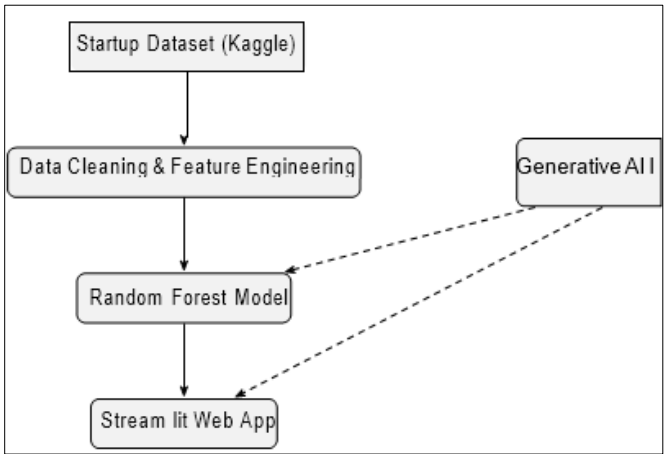

Fig 1: System Architecture of the Proposed Startup Investment Prediction Platform

4. Data Pipeline

The data pipeline ensures that raw data is transformed into high-quality, model-ready input. It involves cleaning, handling missing values, encoding categorical data, and feature scaling.

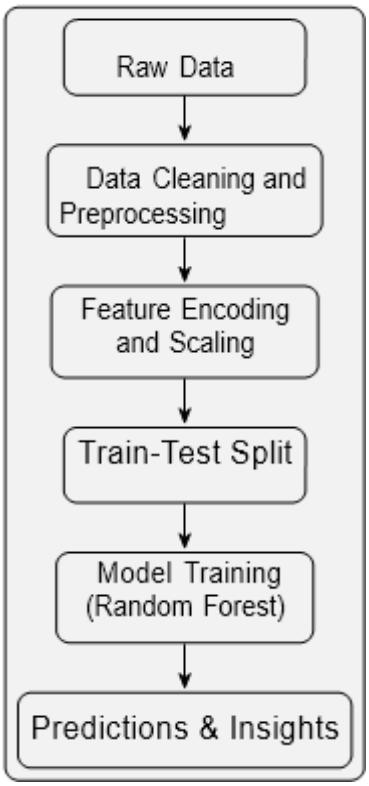

Fig 2: Compact data processing and model training pipeline

5. Machine Learning Methodology

The model uses a Random Forest Classifier, an ensemble technique that builds multiple decision trees and merges their outputs to improve predictive accuracy and control overfitting. Mathematically, the final class prediction is given by:

$$Y = \text{mode}(h_1(x), h_2(x), ..., h_n(x))$$

where each hi(x) represents an individual decision tree prediction.

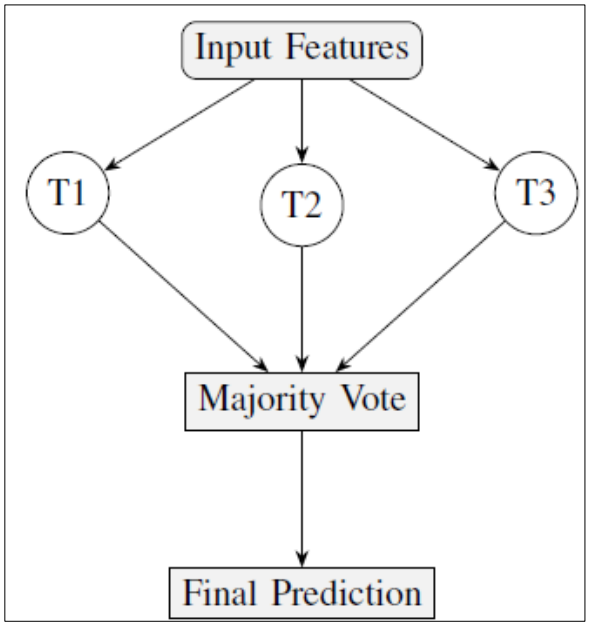

Fig 3: Random Forest Ensemble Prediction Flow

6. Implementation and Results

The system was developed using Python libraries such as scikit-learn, pandas, and matplotlib. Streamlit was used to create an interactive front-end that allows users to input startup parameters (e.g., funding amount, sector, number of investors) and receive real-time predictions.

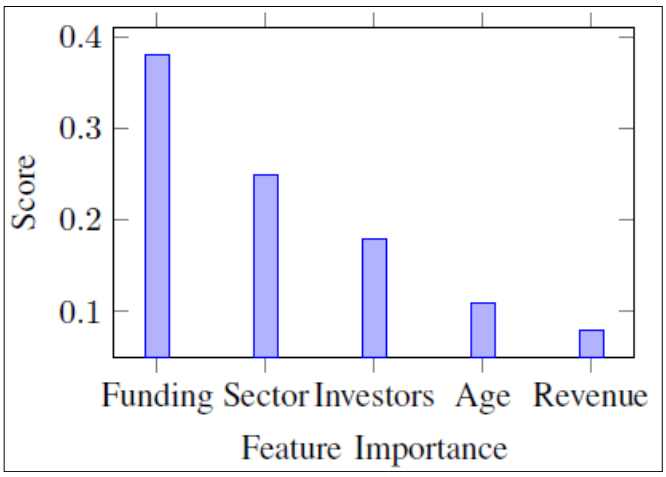

Fig 4: Feature Importance from Random Forest Model

The Random Forest model achieved an accuracy of approximately 89.3% on test data, outperforming baseline Logistic Regression and Decision Tree models. Precision and recall values suggest balanced predictive performance.

7. Generative AI Integration

The Generative AI (GenAI) tool enhances the application by identifying top-performing industries based on user-defined

filters. It uses pre-trained language models to summarize recent market data and suggest promising sectors.

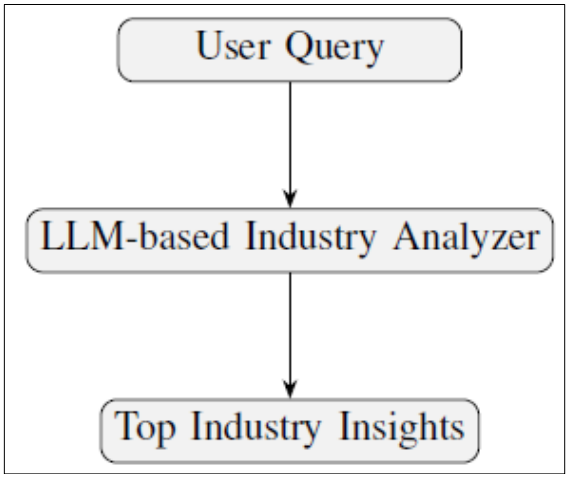

Fig 5: Generative AI Module Workflow

This integration distinguishes the system from conventional predictive tools, adding interpretability and future trend prediction for investors.

8. Business Relevance

From a business standpoint, the proposed platform reduces investment risk by combining quantitative analysis with explainable AI-driven insights. Venture capitalists can simulate scenarios before funding a startup, while entrepreneurs can benchmark their business metrics against successful peers. The platform's Streamlit dashboard supports real-time visualization of predicted outcomes, sector-wise funding trends, and growth forecasts. This bridges the gap between technical predictions and actionable business intelligence.

9. Conclusion

This project demonstrates how machine learning and Generative AI can be combined to improve startup investment decisions. The Random Forest model achieved robust accuracy, and the Generative AI feature provided contextual industry insights.

Future improvements include integrating real-time financial APIs, adding regional language support, and incorporating time-series forecasting models to track evolving market trends.

10. Acknowledgment

The authors thank the faculty and mentors of the School of Computing, MIT ADT University, Pune, for their continuous guidance and support throughout this project.

11. References

1. Kaggle. Indian startup funding dataset; c2023 [cited 2026 Feb 6]. Available from: https://www.kaggle.com/datasets/sudalairajkumar/indian-startup-funding
2. Breiman L. Random forests. Machine Learning. 2001;45(1):5–32.
3. Crunchbase. Startup database and analytics platform. [cited 2026 Feb 6]. Available from: https://www.crunchbase.com
4. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997;9(8):1735–1780.
5. Raschka S, Mirjalili V. Python machine learning. Birmingham (UK): Packt Publishing; c2019.
6. Ng A. Machine learning yearning. deeplearning.ai; 2018 [cited 2026 Feb 6]. Available from: https://www.deeplearning.ai

How to Cite This Article
Bhatia A, Jain S, Gurbani R, Patil S. AI Powered Startup Investment Analysis. International Journal of Multi Research. 2026;2(1):62-64.