

Startup Funding Prediction

A Machine Learning-Based Web Application using Regression Techniques

Abstract

With the rapid growth of startups across various domains such as FinTech, HealthTech, Artificial Intelligence, and E-commerce, estimating startup funding has become a challenging task. Funding decisions depend on multiple factors including team size, founder experience, location, previous funding rounds, and product type. Entrepreneurs and investors often struggle to assess funding potential accurately.

This project, **Startup Funding Prediction**, uses **Machine Learning regression techniques** to predict the expected funding amount (in Indian Rupees Crore) for a startup. **Lasso Regression** is applied to identify the most influential factors affecting funding, while **Random Forest Regression** is used to improve prediction accuracy. The system is implemented using Python and Streamlit, providing an interactive and user-friendly web application for real-time funding prediction.

Introduction

Startup funding plays a critical role in determining the growth and success of a business. Due to the complexity of funding criteria and varying market conditions, predicting funding amounts manually is difficult and often inaccurate.

Machine learning techniques provide a data-driven approach to analyze startup characteristics and predict funding more reliably. This project applies regression-based machine learning models to estimate startup funding and support decision-making for entrepreneurs and investors.

Objectives of the Project

The main objectives of this project are:

- To predict startup funding amount (in ₹ Crore)
- To analyze factors influencing startup funding
- To identify important features using Lasso Regression
- To improve prediction accuracy using Random Forest Regression
- To demonstrate practical application of regression models
- To build an interactive Streamlit-based web application

Scope of the Project

This project focuses on:

- Startup-level funding prediction
- Educational and real-world application of machine learning
- Supporting entrepreneurs and investors in decision-making
- A scalable framework that can be extended using real startup data

Technologies Used

Programming Language

- Python

Libraries & Tools

- Pandas – Data handling and manipulation
- NumPy – Numerical computation
- Scikit-learn – Machine learning modeling
- Streamlit – Web application framework

Machine Learning Algorithms

- Lasso Regression (L1 Regularization)
- Random Forest Regression

Dataset Description

The dataset contains structured information related to startup characteristics and funding details.

Dataset Size

- Number of records: **65 startups**

Input Features

Feature	Description
Sector	Startup domain (AI, FinTech, HealthTech, etc.)
TeamSize	Number of employees
FounderExperience	Founders' experience (years)
Location	City of operation
PreviousFundingRounds	Number of previous funding rounds
ProductType	Business type (B2B or B2C)

Methodology

1. Data Loading

The startup dataset is loaded and analyzed using Python libraries.

2. Data Preprocessing

Categorical variables such as sector, location, and product type are encoded for model compatibility.

3. Feature Selection

Lasso Regression is used to identify important factors affecting startup funding.

4. Model Training

Random Forest Regression is trained to predict funding amounts accurately.

5. Prediction

The trained model predicts startup funding based on user input.

6. Deployment

A Streamlit web application provides real-time predictions.

Why Regression is Used

Regression techniques are used because:

- The output variable is continuous
- Funding amount varies numerically

- Regression models capture relationships between features
- Suitable for financial prediction problems

Why Lasso Regression is Used

Lasso Regression is used because:

- It performs automatic feature selection
- Reduces overfitting using L1 regularization
- Produces interpretable models
- Identifies key influencing factors

System Architecture

Startup Details (User Input)



Data Preprocessing & Encoding



Lasso Regression (Feature Importance)



Random Forest Regression



Funding Amount Prediction (₹ Crore)

Variable	Coefficient
Sector	0.3744
TeamSize	5.38
FounderExperience	0.7399
Location	-1.2989
PreviousFundingRounds	12.0478
IndustryType	-4.3000

Application Features

- Interactive user interface
- Real-time funding prediction
- Displays funding in Indian Rupees (Crore)
- Highlights important influencing factors
- Simple and user-friendly design

Results and Output

- Predicts startup funding accurately
- Provides real-time results
- Identifies key funding factors
- Displays output clearly using Streamlit

Sample Output

Input:

Sector: AI
Location: Bangalore
Product Type: B2B
Team Size: 10
Founder Experience: 5 years
Previous Funding Rounds: 1

Predicted Output:

₹ 38.62 Crore (Approximate Value)

Advantages

- Accurate funding estimation
- Data-driven decision support
- Easy to use
- Practical regression application
- Suitable for academic projects

Limitations

- Dataset is simulated
- Accuracy depends on data quality
- Real-world funding may vary

Conclusion

The **Startup Funding Prediction** project demonstrates the effective use of machine learning regression techniques in solving real-world financial prediction problems. **Lasso Regression** identifies key factors influencing funding, while **Random Forest Regression** improves prediction accuracy. The Streamlit web application provides a simple and interactive interface for real-time predictions.

This project is suitable for academic submission and provides valuable insights into regression-based machine learning applications