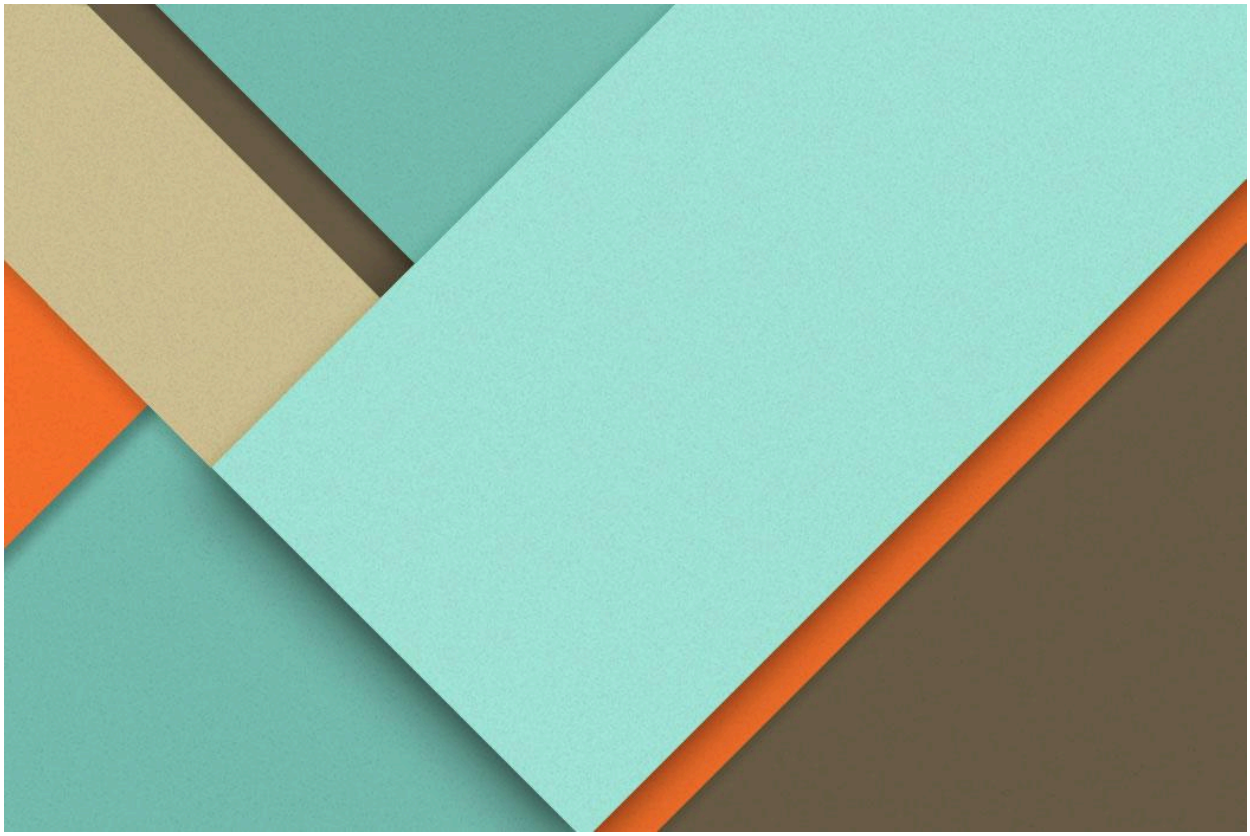


STARTUP FUNDING PREDICTOR

05-01-2026

ARCHANA.S

Project Type: Mini Project – Lasso Regression & Random Forest Regression

Title

Startup Funding Prediction using Machine Learning

Description:
This project predicts the **expected funding amount (in Indian Rupees Crore)** for a startup based on its key characteristics using **Machine Learning regression models**

Objective

- To predict the **funding amount (₹ Crore)** for a startup.
- To identify **important factors affecting startup funding** using **Lasso Regression**.
- To improve prediction accuracy using **Random Forest Regression**.
- To provide an **interactive Streamlit-based web application** for real-time funding predictions.

Dataset

Number of records: 65 startups

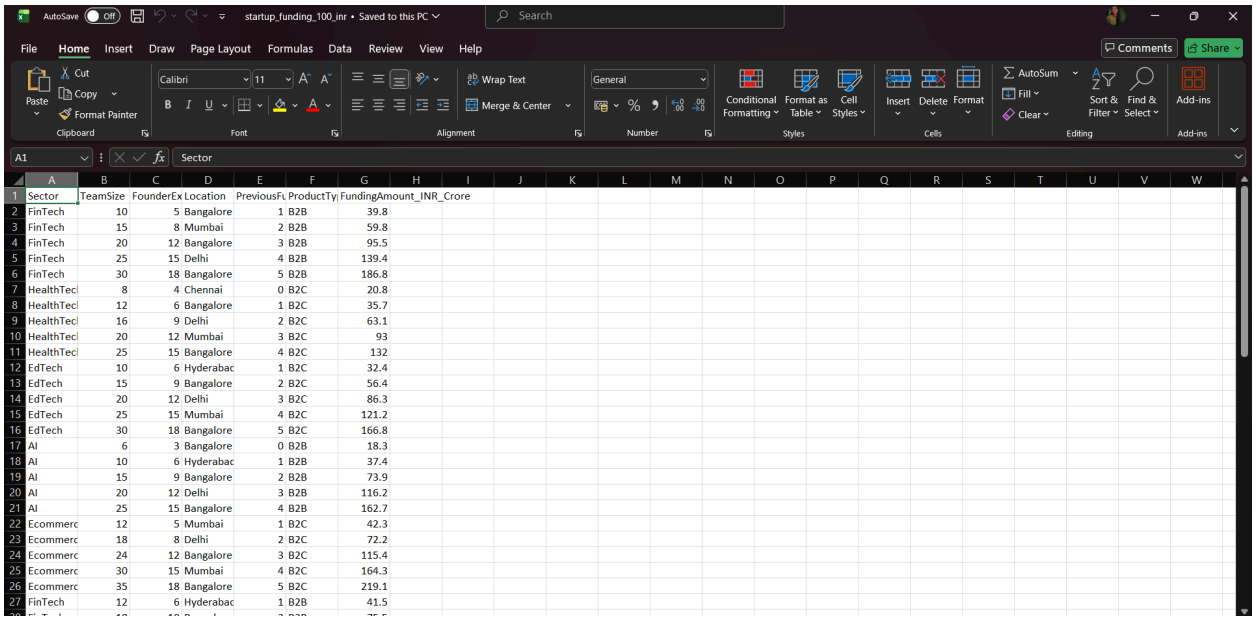
Features (Independent Variables):

FEATURE	DESCRIPTION
Sector	Industry of the startup(AI,FinTech,etc.)
TeamSize	Number of employees
FounderExperience	Experience of founders (in years)
Location	City of operation
PreviousFundingRounds	Number of previous funding rounds
ProductType	Type of product (B2B or B2C)

Target Variable:

FundingAmount_INR_Crore

Dataset Preview:



Sector	TeamSize	FounderEx	Location	PreviousFu	ProductTy	FundingAmount_INR_Crore
FinTech	10	5	Bangalore	1	B2B	39.8
FinTech	15	8	Mumbai	2	B2B	59.8
FinTech	20	12	Bangalore	3	B2B	95.5
FinTech	25	15	Delhi	4	B2B	139.4
FinTech	30	18	Bangalore	5	B2B	186.8
HealthTecl	8	4	Chennai	0	B2C	20.8
HealthTecl	12	6	Bangalore	1	B2C	35.7
HealthTecl	16	9	Delhi	2	B2C	63.1
HealthTecl	20	12	Mumbai	3	B2C	93
HealthTecl	25	15	Bangalore	4	B2C	132
EdTech	10	6	Hyderabac	1	B2C	32.4
EdTech	15	9	Bangalore	2	B2C	56.4
EdTech	20	12	Delhi	3	B2C	86.3
EdTech	25	15	Mumbai	4	B2C	121.2
EdTech	30	18	Bangalore	5	B2C	166.8
AI	6	3	Bangalore	0	B2B	18.3
AI	10	6	Hyderabac	1	B2B	37.4
AI	15	9	Bangalore	2	B2B	73.9
AI	20	12	Delhi	3	B2B	116.2
AI	25	15	Bangalore	4	B2B	162.7
Ecommerc	12	5	Mumbai	1	B2C	42.3
Ecommerc	18	8	Delhi	2	B2C	72.2
Ecommerc	24	12	Bangalore	3	B2C	115.4
Ecommerc	30	15	Mumbai	4	B2C	164.3
Ecommerc	35	18	Bangalore	5	B2C	219.1
FinTech	12	6	Hyderabac	1	B2B	41.5

Prediction

- **Problem Type:** Regression
- **Prediction Type:** Continuous numeric value (Funding Amount)

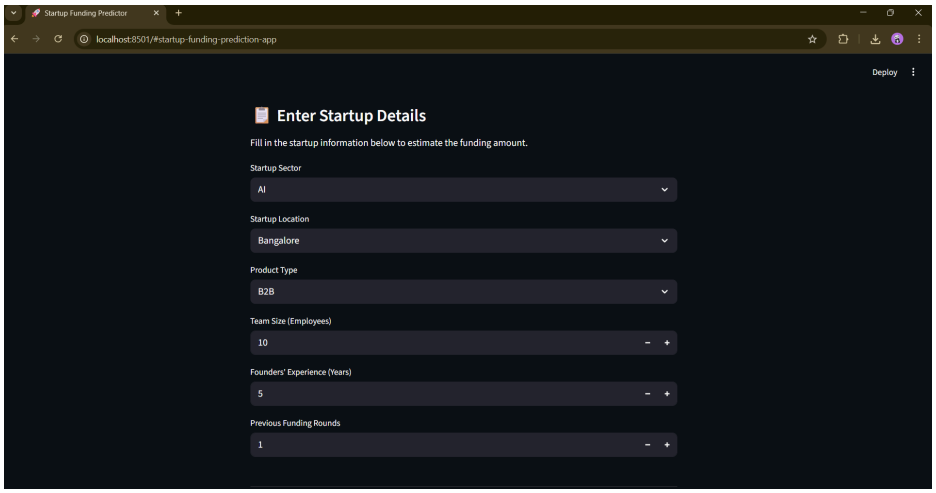
Models Used:

- **Lasso Regression:**
  Used to identify the most important factors influencing startup funding.
- **Random Forest Regression:**
  Used to provide accurate and reliable funding predictions.

Workflow:

1. Data preprocessing and encoding of categorical features
2. Training Lasso and Random Forest regression models
3. Building a Streamlit web application for user input
4. Predicting funding amount and displaying important features

App Screenshot:



Code

app.py:

Streamlit-based web application for funding prediction.

Trains regression models and evaluates performance.

Example Code Snippet:

```
input_data = pd.DataFrame([{
    "Sector": le_sector.transform([sector])[0],
    "TeamSize": team_size,
    "FounderExperience": founder_experience,
```

```
    "Location": le_location.transform([location])[0],
    "PreviousFundingRounds": previous_rounds,
    "ProductType": le_product.transform([product_type])[0]
}], columns=feature_columns)

prediction = rf.predict(input_data)[0]
st.success(f"₹ {prediction:,.2f} Crore")
```

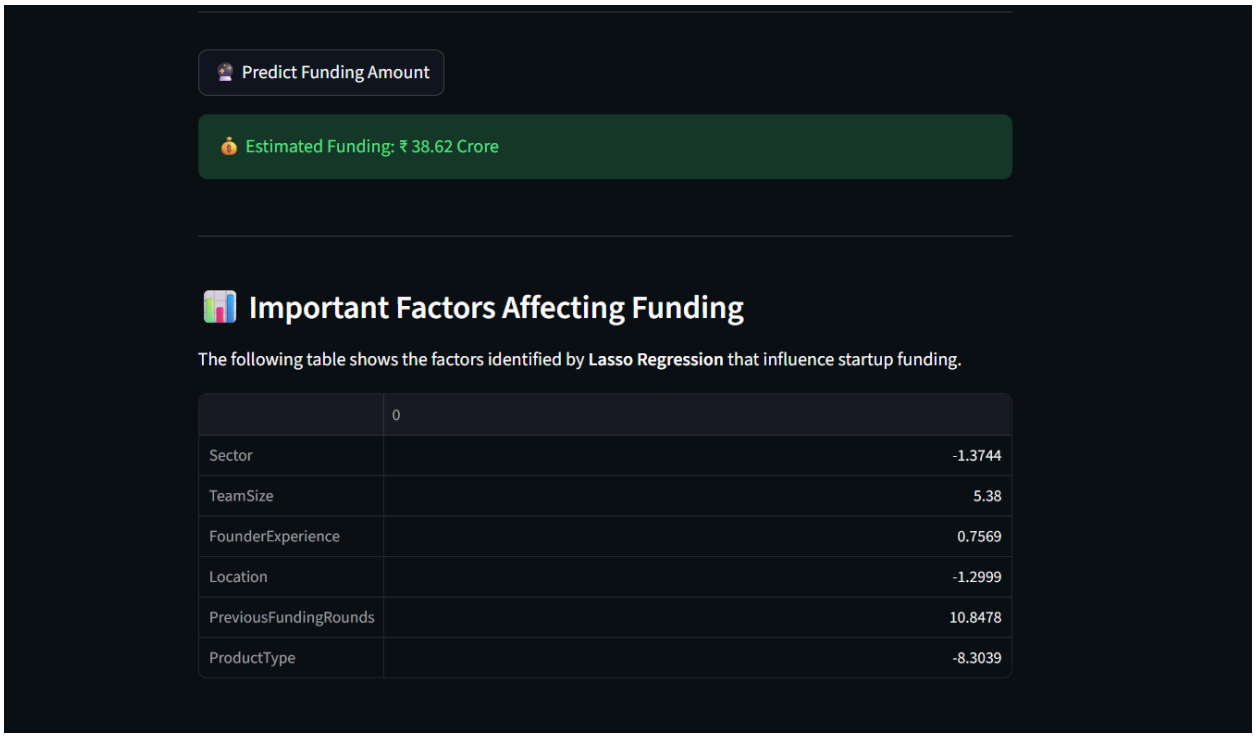
Expected Output

User Input Example:

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

Predicted Funding Amount:

```
₹ 38.62 Crore
```



Conclusion

- Random Forest Regression provides **accurate startup funding predictions**.
- Lasso Regression successfully identifies **key factors influencing funding**.
- The Streamlit application enables **interactive and real-time prediction**.
- This project is useful for **entrepreneurs and investors** to estimate startup funding potential.