

# Global Startup Ecosystem Analysis

## Funding & Growth Analysis with Machine Learning Insights

---



### Overview

This project explores the global startup ecosystem by analyzing funding trends, geographic distribution, industry-wise investments, and debt financing. It also includes a **predictive modeling component** to determine a startup's likelihood of receiving **Series A funding**, using seed-stage characteristics.

Built using **Python** for data preprocessing and modeling, and **Power BI** for dashboarding, this project combines **exploratory data analysis** and **machine learning insights** in one cohesive solution.

---



### Key Objectives

- Clean and prepare a real-world startup funding dataset
  - Visualize global startup funding patterns via a Power BI dashboard
  - Train machine learning models to predict Series A funding success
  - Evaluate and compare model performance
  - Provide business insights for investors and entrepreneurs
- 



### Power BI Dashboard Features

An interactive and visually engaging dashboard presenting:

- **Total Funding Overview**
- **Funding Over Time** (1900–2020)
- **Top Markets by Funding**
- **Funding Distribution by Country** (Map view)
- **Debt Financing by Industry** (Treemap)
- **Dynamic filters** for Year, Market, and Country

✓ Technologies: Power BI, DAX, Power Query

---



### Data Cleaning & Preparation

- Removed duplicate entries and standardized inconsistent country/market labels
- Handled missing values using median imputation (for numerical) and mode (for categorical)

- Converted funding columns to numeric types and corrected date formats
- Created binary target column: **Raised\_Series\_A** (1 = Yes, 0 = No)

✓ Tools: Python (Pandas, NumPy)

---

## Exploratory Data Analysis (EDA)

- Identified **Biotechnology, Mobile, and Software** as top funded markets
- **North America** and **Europe** dominate in terms of total funding
- A notable rise in funding occurred during **1999–2008** and **2010–2015**
- Debt financing is **concentrated in Mobile and Clean Tech** sectors

✓ Libraries: Seaborn, Matplotlib, Plotly

---

## Machine Learning: Predicting Series A Success

**Goal:** Predict whether a startup will secure Series A funding using seed-stage data.

### Features Used:

- Seed funding amount
- Number of funding rounds
- Number of investors
- Market sector
- Country
- Time to next funding round

### Models & Performance:

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	95.8%	0.96	0.96	0.96
Random Forest Classifier	96.5%	0.97	0.96	0.97

- Confusion matrix shows **strong performance across both classes**
- **Random Forest outperformed Logistic Regression** by a small margin
- Cross-validation was used to validate stability

✓ Libraries: Scikit-learn, XGBoost, Seaborn (for confusion matrix)

---

## Tech Stack

- **Languages:** Python
  - **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Plotly
  - **BI Tool:** Power BI
  - **ML Techniques:** Supervised Learning (Classification), Model Evaluation
- 

## Future Enhancements

- **Time-Series Forecasting:** Predict future funding trends per country
  - **Clustering:** Segment startups by funding, stage, and geography
  - **Investor Profiling:** Identify top venture capitalists by portfolio success
- 

## Business Benefits & Insights

This project delivers actionable insights for investors, entrepreneurs, and startup ecosystem analysts. Here's what we gained from the analysis:

### For Investors:

- **Improved Decision-Making:** Predictive models help identify promising startups early, increasing the success rate of Series A investments.
- **Market Opportunity Mapping:** Geographic and industry-based funding visualizations highlight emerging hotspots and underserved sectors.
- **Risk Mitigation:** Historical patterns of failed and successful funding rounds guide smarter portfolio allocation.

### For Startups:

- **Self-Assessment Tool:** Startups can benchmark their seed-stage metrics against successful Series A cases to improve their positioning.
- **Market Targeting Strategy:** Insights on where and in which markets startups are more likely to get funded help focus go-to-market strategies.
- **Timing Optimization:** Patterns around funding rounds offer clues about optimal time windows to approach investors.

### For Ecosystem Stakeholders:

- **Macro-Level Trends:** Dashboard provides a high-level view of how startup funding evolved globally over decades.
- **Policy Support:** Policymakers can identify funding gaps in specific regions or industries and take targeted actions to support innovation.

✓ Overall, this analysis empowers data-driven funding strategies, reduces guesswork, and enhances visibility into the complex dynamics of global startup success.