Analysis and Forecasting of Startup Investments

Kavishka Fernando-20020384

February 23, 2025

GitHub Repository: `https://github.com/kavishkafer/BIS_Assignment`

Dataset: `https://www.kaggle.com/datasets/adilshamim8/startup-growth-and-investment-data?resource=download`

Introduction

In the modern business domain, investment trends play a crucial role in determining the success of startups across various industries. There are many types of investments. This analysis explores investment patterns, compares funding distributions across industries, and forecasts future investment trends using historical data.

Business Domain and Dataset

The financial investment sector, particularly startup funding, is the focus of this analysis. The dataset, *Startup Growth and Investment Data*, was obtained from Kaggle. It includes details such as funding rounds, investment amounts, valuation, number of investors, country, industry, and year founded.This dataset contains 5,000 records of startup growth and investment data across various industries. It provides insights into funding trends, valuation, and investor activity for startups globally.

Analytical Process

The primary question addressed is: *Which industries receive the highest investment, and how can we predict future investment trends?* The analysis follows these steps:

1. **Data Collection:** Extracted startup funding data from Kaggle.

2. **Data Preprocessing:** Cleaned the dataset by handling missing values and duplicates.

3. **Feature Engineering:** Encoded categorical variables and normalized numerical features.

4. **Exploratory Data Analysis (EDA):** Visualized industry-wise funding and yearly investment trends.

5. **Time Series Forecasting:** Used ARIMA and SARIMA models to predict future investment trends.

Exploratory Data Analysis (EDA)

Investment Distribution by Industry

The dataset was grouped by industry to calculate total investment. Figure 1 shows the top 10 industries with the highest funding. Industries like Blockchain, SaaS, and EdTech received the most investment.
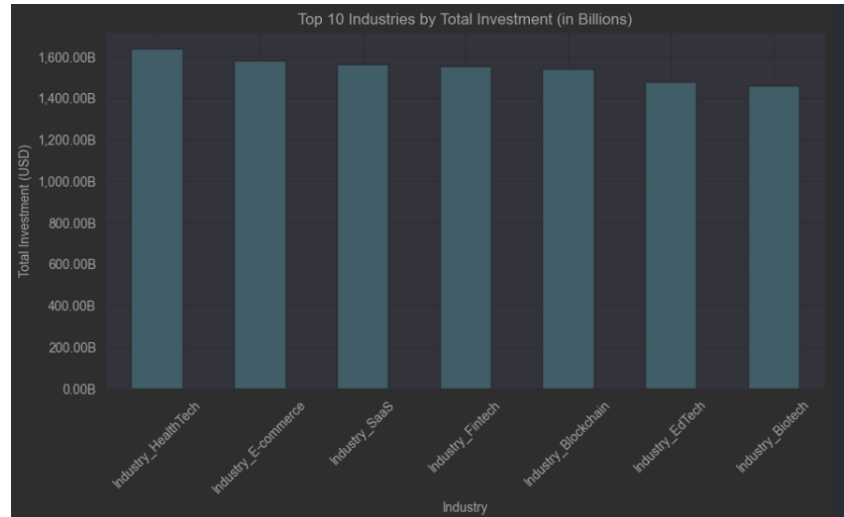

Figure 1: Top 10 Industries by Total Investment

Investment Trends Over Time

Yearly investment trends were analyzed to determine growth. Figure 2 shows a general upward trend, with peaks in certain years indicating high investor confidence.

Figure 2: Total Startup Investment Over Years

Investment Distribution

A histogram was plotted to analyze the distribution of investment amounts (Figure 3). Log transformation was applied to handle extreme values.
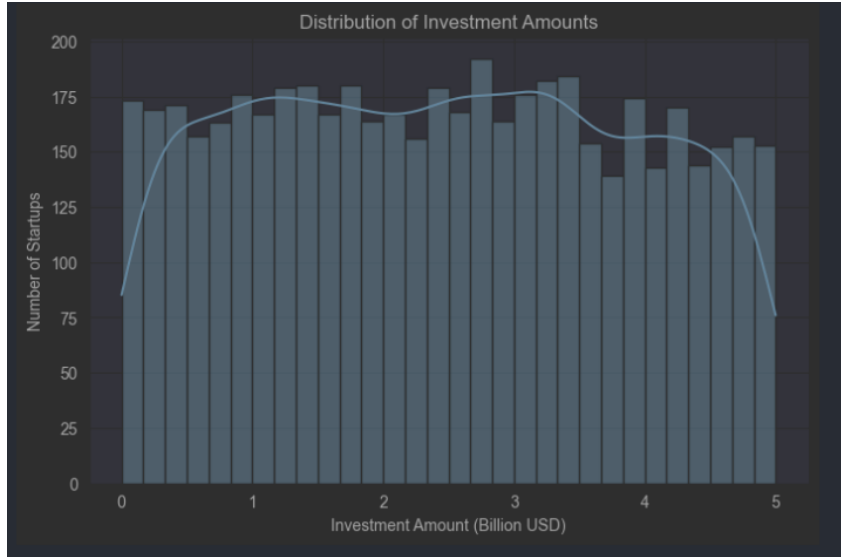
Figure 3: Distribution of Investment Amounts

Feature Engineering

Categorical data, such as industry type, was handled using One-Hot Encoding. This transformation ensures compatibility by representing categories as binary vectors. After that we Decompose time series into trend, seasonality, and residuals. The image is in the notebook; it is published in the end of report.

Time Series Forecasting

ARIMA Model

The ARIMA model was fitted to historical investment data to capture trends and seasonality. Figure 4 shows the forecasted results.
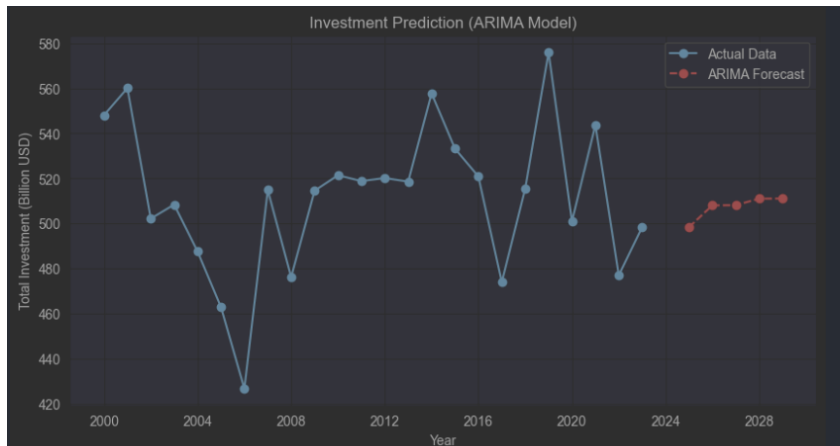

Figure 4: Investment Forecasting Using ARIMA

SARIMA Model

The SARIMA model accounted for seasonality in investment patterns. Figure 5 shows the results, which demonstrated better predictive accuracy than ARIMA. As I mentioned in the

earlier paragraph there was an visual seasonal trend as we decompose time series into trend, seasonality, and residuals. SARIMA more suitable for datasets with seasonal trends.
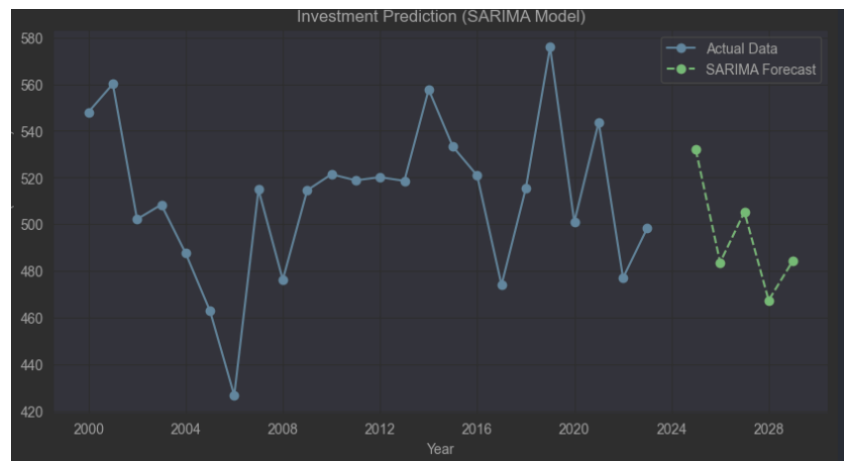

Figure 5: Investment Forecasting Using SARIMA

Conclusion

This study analyzed startup funding patterns across industries, identified trends over time, and implemented forecasting models to predict future investments. The findings highlight industries with the highest funding and provide insights for businesses. The SARIMA model outperformed ARIMA due to its ability to account for seasonal variations hence proved our graph regarding seasonal decomposition.
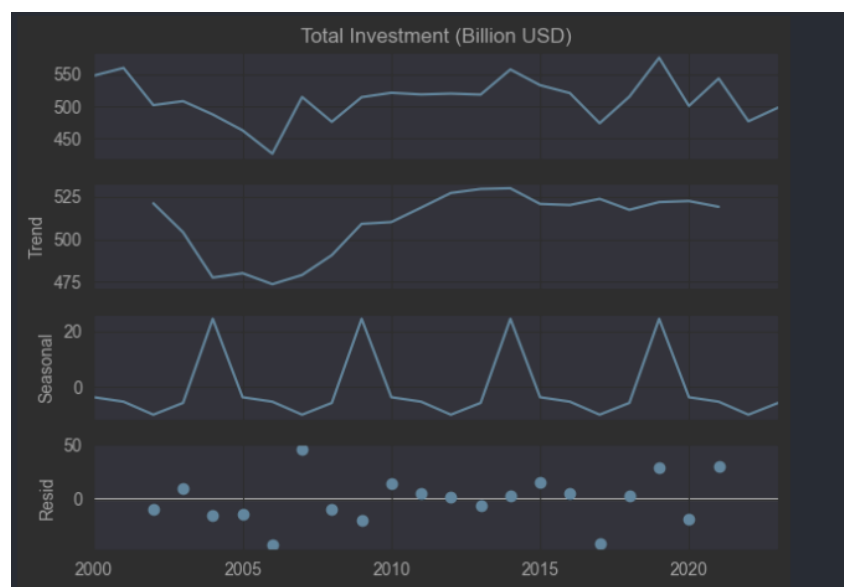

Figure 6: Seasonal trends