International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

# FORECASTING STARTUP SUCCESS OR FAILURE USING CRUNCHBASE DATA

J.VASAVI, M.PRIYANKA, C.SATHWIKA, J.VAMSHI, M.PREM KUMAR, MR.K.SUNIL KUMAR

## ABSTRACT

Startups play a crucial role in driving innovation and economic growth, but they also encounter significant challenges, with a considerable proportion failing within their initial years. The traditional method of forecasting startup success or failure involves decision-makers relying on historical financial data, market analysis, and subjective evaluations of the founding team and business concept. While these approaches do provide some insights, they are limited in scalability, subject to human biases, and struggle to handle large volumes of diverse data. Consequently, the accuracy and reliability of traditional methods are compromised, leading to less effective identification of potential successful startups or missed opportunities to prevent investment in failing ventures. The necessity for accurate and efficient startup success prediction is of utmost importance for various stakeholders, including investors, entrepreneurs, and policymakers. Investing in early-stage startups involves high risks, and Artificial Intelligence (AI) can play a vital role in identifying high-potential startups, reducing the likelihood of financial losses. Entrepreneurs can greatly benefit from AI-based forecasting as it helps them assess their venture's viability, refine their strategies, and improve their chances of achieving long-term success. Policymakers can also leverage AI-driven insights to create targeted and effective support programs that foster a thriving startup ecosystem. AI has emerged as a promising solution to enhance the accuracy and efficiency of startup forecasting. AI is capable of processing vast amounts of structured and unstructured data, including social media trends, industry reports, and company news, with remarkable speed and precision. By uncovering hidden patterns and adapting to dynamic market conditions, AI enables better-informed decision-making in the ever-competitive startup landscape. This project focuses on using AI to predict startup success or failure by analyzing data from Crunchbase, a comprehensive database of startup information. The proposed system highlights the significance of AI in this context, underscoring its potential benefits for investors, entrepreneurs, and policymakers alike. Through AI, this system aims to revolutionize startup forecasting, providing valuable insights that empower all stakeholders involved in the startup ecosystem.

## INTRODUCTION

### Overview

The impact of artificial intelligence on forecasting startup success or failure using Crunchbase data represents a convergence of cutting-edge technology and entrepreneurial innovation in the modern business landscape. Startups are integral to economic growth and innovation, yet they face substantial challenges and high failure rates. Consequently, accurately predicting which startups are likely to succeed or fail is of paramount importance to investors, entrepreneurs, and the broader business community. This endeavour involves harnessing the power of artificial intelligence (AI) and machine learning to analyze vast datasets from Crunchbase, a comprehensive platform that tracks information about startups, including funding rounds, team details, industry focus, and market trends. By applying AI algorithms and predictive modeling to this rich dataset, it becomes possible to develop robust predictive models that can assess the viability and potential trajectory of startups [1]. The motivation behind this topic is multifaceted. Firstly, startup investments are inherently risky, and investors seek ways to mitigate those risks by making informed decisions. AI-driven predictive models can provide valuable insights into which startups have a

higher likelihood of success based on historical data and a wide array of factors, including funding history, market conditions, and team composition .

Secondly, for entrepreneurs and founders, gaining insights into the factors that influence startup success can inform strategic decisions, resource allocation, and business pivots. Understanding the key drivers of success or failure can be a valuable guide in navigating the challenging startup landscape. Additionally, AI-powered forecasting has the potential to shape the startup ecosystem as a whole . By identifying trends, patterns, and common characteristics among successful startups, it can inform startup incubators, accelerators, and policymakers in their efforts to foster entrepreneurship and innovation. Furthermore, this topic contributes to the broader conversation about the impact of AI on decision-making processes across industries. AI is increasingly being leveraged to make predictions, optimize resource allocation, and automate complex tasks, and its application in the startup realm exemplifies its transformative potential in the world of business and finance.

In summary, the exploration of how artificial intelligence can forecast startup success or failure using Crunchbase data is at the intersection of technology, entrepreneurship, and data-driven decision-making. It addresses pressing challenges in the startup ecosystem, empowers investors and founders with valuable insights, and showcases the transformative potential of AI in shaping the future of business and innovation .

**Motivation**

The research on "The Impact of Artificial Intelligence on Forecasting Startup Success or Failure Using Crunchbase Data" is motivated by several compelling factors . Firstly, the startup ecosystem plays a pivotal role in driving innovation, job creation, and economic growth. However, the high failure rate among startups poses a significant challenge. Predicting which startups are likely to succeed or fail is of great interest to investors, entrepreneurs, policymakers, and the broader business community .

Traditional methods of startup evaluation often rely on subjective assessments and historical data, which may not capture the dynamic nature of the startup landscape. Artificial Intelligence (AI) offers the potential to revolutionize this process by leveraging advanced algorithms, machine learning techniques, and vast datasets such as Crunchbase . By harnessing AI, we can develop more accurate and data-driven predictive models, enabling stakeholders to make more informed decisions regarding investments, resource allocation, and support for startups. Crunchbase, as a comprehensive repository of startup information, provides a rich source of data for this research. AI-driven analysis of this data can uncover hidden patterns, correlations, and early indicators of startup success or failure . Moreover, AI can adapt and evolve its predictions as new data becomes available, making it a valuable tool for ongoing monitoring and decision-making in the dynamic startup landscape. Furthermore, the research addresses broader implications, including the ethical considerations and potential biases associated with AI-driven predictions in the startup domain. It is essential to understand how AI impacts decision-making and whether it can provide a fair and unbiased assessment of startups, ensuring that the benefits of AI are accessible to a wide range of stakeholders . In summary, this research is driven by the need to enhance our understanding of the startup ecosystem, improve prediction accuracy, and leverage the power of AI to make more informed decisions about startup investments and support. It also aims to shed light on the ethical considerations surrounding AI in startup forecasting, contributing to the responsible and equitable development of AI technologies in the business world .

**Applications**

Investment Decision-Making: Investors, including venture capitalists, angel investors, and private equity firms, can use AI-powered predictive models to make more informed decisions about where to allocate their capital. This can lead to better investment portfolios with startups that have a higher likelihood of success, potentially increasing their returns on investment.

Startup Support and Mentorship: Incubators, accelerators, and startup support organizations can use AI-based forecasts to identify startups that may benefit the most from their programs. This ensures that limited resources are directed toward startups with a higher probability of success, maximizing the impact of support initiatives.

Policy and Economic Development: Governments and policymakers can utilize AI-driven insights to shape policies and incentives that promote entrepreneurship and economic growth. This could lead to more targeted programs and regulatory frameworks to support startups in areas where they are most likely to thrive.

Risk Mitigation for Lenders: Financial institutions that provide loans or credit to startups can use AI to assess the creditworthiness of these businesses more accurately. This reduces the risk associated with lending to startups and can potentially lead to more favourable lending terms.

Market Research and Competitive Analysis: Companies and market research firms can employ AI to gain deeper insights into emerging trends and competitive landscapes within specific industries. Understanding which startups are likely to disrupt or succeed in a market can inform strategic decisions.

Incubator and Accelerator Program Design: Incubators and accelerators can optimize the design of their programs based on AI insights, tailoring mentorship, resources, and networking opportunities to address the specific needs and challenges faced by startups in their portfolios.

Entrepreneurial Ecosystem Development: Economic development agencies and organizations focused on building entrepreneurial ecosystems can use AI forecasts to strategically nurture startups that align with their regional or sectoral priorities.

Corporate Innovation: Large corporations seeking innovation through partnerships or investments in startups can benefit from AI-based assessments to identify startups with technologies or solutions that align with their strategic goals.

## SYSTEM REQUIREMENTS

The functional requirements or the overall deworkion documents include the product perspective and features, operating system and operating environment, graphics requirements, design constraints and user documentation.

The appropriation of requirements and implementation constraints gives the general overview of the project in regard to what the areas of strength and deficit are and how to tackle them.

- Python IDLE 3.7 version (or)
- Anaconda 3.7 (or)
- Jupiter (or)
- Google colab

**Hardware Requirements**

Minimum hardware requirements are very dependent on the particular software being developed by a given Enthought Python / Canopy / VS Code user. Applications that need to store large arrays/objects in memory will require more RAM, whereas applications that need to perform numerous calculations or tasks more quickly will require a faster processor.

- Operating system : Windows, Linux
- Processor : minimum intel i3
- Ram : minimum 4 GB
- Hard disk : minimum 250GB

## Conclusion

In conclusion, the data preprocessing steps undertaken in this work play a pivotal role in preparing the startup dataset for the subsequent machine learning model training and prediction of startup success or failure. These steps are indispensable in ensuring the quality and reliability of the data, enhancing its consistency, and facilitating meaningful insights. First and foremost, the handling of missing values through visualizations, such as heatmaps, aids in identifying and addressing gaps in the dataset, contributing to more robust model training. Eliminating rows with missing values and startups with uncertain funding forms fosters data integrity. Furthermore, standardizing date representations by converting 'founded_at,' 'first_funding_at,' and 'last_funding_at' columns to datetime format not only ensures uniformity but also allows for accurate calculations, such as determining the funding period in years.

The categorization of startup statuses into 'success' and 'failure' classes simplifies the classification task, providing a clearer distinction between operational outcomes. This categorization helps in framing the machine learning problem effectively. The calculation of the funding period adds a temporal dimension to the dataset, enabling the models to consider the duration of funding when predicting startup success or failure. Subsequently, the model training and evaluation stages, employing Decision Tree and Random Forest models, allow for the assessment of predictive performance. Metrics such as accuracy, confusion matrix, and classification report provide a comprehensive evaluation of model effectiveness. The incorporation of a grid search for hyperparameter tuning optimizes the Random Forest model, enhancing its predictive power by identifying the best combination of hyperparameters. Lastly, the use of the trained model to predict the status of operational startups based on their funding period offers valuable insights into the potential success or failure of these enterprises.

## REFERENCES

[1] Han, JungYun, and Charles Eesley. "Impacts of Accelerator Cohort Composition on Startup Performance." 2023 Portland International Conference on Management of Engineering and Technology (PICMET). IEEE, 2023.

[2] Xu, Ruiyun, Hailiang Chen, and J. Leon Zhao. "SocioLink: Leveraging Relational Information in Knowledge Graphs for Startup Recommendations." Journal of Management Information Systems 40.2 (2023): 655-682.

[3] Pandya, Darshanaben Dipakkumar, et al. "Forecasting Number of Indian Startups using Supervised Learning Regression Models." 2023 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2023.

[4] Singhal, Jaiesh, et al. "Data driven analysis for startup investments for venture capitalists." 2022 International Conference for Advancement in Technology (ICONAT). IEEE, 2022.

[5] Mishra, Ajai, Dharm Singh Jat, and Durgesh Kumar Mishra. "An Experimental Study of Machine Learning Algorithms for Predicting Start-Up Success." Intelligent Sustainable Systems: Selected Papers of WorldS4 2022, Volume 1. Singapore: Springer Nature Singapore, 2023. 813-825.

[6] Di Giannantonio, Rocco, Matthias Murawski, and Markus Bick. "The Impact of Machine Learning-Based Techniques on the Scouting and Screening Processes of Early-Stage Venture Capital Firms." Conference on e-Business, e-Services and e-Society. Cham: Springer International Publishing, 2022.