iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54673

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Prediction System for Startup Acquisition Status Using Random Forest

Rutvika S. Mane[1], Ankit R. Waikar[2], Janhavi J. Mishra[3], Disha A. Lahoti[4], Prof. Shobha S. Raskar[5]
Computer Engineering Department, Savitribai Phule University, Pune, Maharashtra

Abstract: ***Startup acquisition is the process where companies interested in investing in startups buy and own some or all of the startup's shares. However, investing in startups requires extensive research as many startups fail within a short period of time, and other factors such as IPO and acquisition by other companies also need to be considered. The recent advances in machine learning have made it possible to predict a startup's acquisition status based on financial statistics, historic data and market trends. The model uses supervised learning to find patterns and correlations between the input features (startup characteristics) and the related acquisition status by learning from labelled data. The prediction model is constructed using machine learning algorithms like KNN and random forests which can save time and costs for investors and speed up their investment process. Also, it aims to classify a startup's acquisition status as Operating, Acquired, IPO, or Closed based on input data. When using the Random Forest approach, the suggested system for predicting acquisition status achieves an accuracy score of 93%, whereas using the KNN (K-Nearest Neighbours) strategy, it earns a score of 91%.***

Keywords: Startup, Acquisition, IPO, Operating, Acquired, Machine Learning.

I. INTRODUCTION

The number of startups worldwide has significantly increased in recent years, leading to an upsurge in startup acquisition. This process involves companies interested in investing in startups purchasing some or all of their shares. However, identifying which startups to invest in requires extensive research, as many startups fail within a short period of time. Other factors, such as whether the startup has been acquired by another company or has an IPO, are also taken into account.

Acquisitions can be driven by various reasons, such as a desire for expansion or access to innovative technologies. For entrepreneurs looking to sell their businesses, navigating the acquisition process requires careful strategy and consideration. Acquisitions not only indicate a startup's growth trajectory but also help bridge financial gaps. However, many startups face challenges and fail due to factors like poor product-market fit and insufficient funding. In the Indian startup ecosystem, where the number of startups is rapidly increasing, choosing the right investment opportunities is crucial to avoid capital losses. The venture capital industry offers investment opportunities in early-stage companies, but current tools for risk reduction and uncertainty management are limited.

Machine learning and data-driven approaches have the potential to address these challenges by leveraging available data from platforms like Crunchbase. Previous efforts focused on predicting specific outcomes, such as acquisitions or public offerings, but broader predictions considering various outcomes can provide venture investors with valuable information to construct portfolios with lower risk and potentially higher returns. Machine learning can aid in decision-making processes, helping investors identify opportunities and assess investment risks more effectively. The decision-making process for investing in startups can be time-consuming, especially for companies with limited funds. To make the most of their investments, these companies are seeking more efficient ways to filter potential startups. With the recent advancements in machine learning, prediction tasks can be solved using various machine learning approaches. This project aims to predict a startup's acquisition status based on its financial statistics. The machine learning model will take input data about the startup in which the investing company is interested and classify its acquisition status as Operating, Acquired, IPO, or Closed. The objective is to assist investors in their decision-making process, saving them time and costs while speeding up their investment process.

II. LITERATURE REVIEW

Jaiesh Singhal et. al. [1] has developed a comprehensive solution that simplifies startup analysis for the investors. As in today's world, investing in startups is becoming increasingly popular. However, many investors lack the expertise and resources required to analyze startups and make data-driven investment decisions. The unique 3-in-1 analysis, which includes founder analysis, investment trend analysis, and success prediction, is crucial for investing in new-age startups, particularly during these uncertain times where investing in such avenues is riskier than ever.

© IJRASET: All Rights are Reserved |

Miao Chen et. al. [2] has proposed a new investment target recommendation system, which utilizes both semantic connectivity information and sequential previous investment information. The system involves a preprocessing step that involves analyzing the sequential historic investment records and the heterogeneous graph. The aim is to create a more accurate and effective investment target recommendation system.

M.A. Cusumano [3] has explored the significance of key factors in the success of startups, including the founders' background, technological expertise, and relevance to startup funding. As there is limited research on this specific problem, insights can be gained from previous methodologies to guide the approach. However, no practical implementation was conducted. This highlights the need for further research to develop effective methods for predicting startup success.

S. Lyu et. al. [4] has proposed a solution for the challenging task of predicting the success of early-stage startups (IPO or acquisition) using incremental graph representation learning. This approach involves utilizing self-attention and fine-tuning through supervised link prediction and node classification to learn improved node-level representations and account for sequential interdependence across periods. The model incorporates network structure information as well as investor, startup, and management member attributes, resulting in state-of-the-art performance compared to various baselines. The study also highlights the importance of factors such as gender, education, and networking in predicting startup success, despite only incorporating degree and institution data for the founder/team. Additionally, the study demonstrates that the proposed approach is particularly effective for nodes in the Largest Connected Component.

Ivy Wang [5] has developed a model for predicting a startup's acquisition status based on its financial statistics. To address the challenge of biased data, a novel ensemble model is utilized instead of under/oversampling the data. This approach combines a high precision model with a high accuracy model that is trained on a dataset transformed by the first model. Preliminary experiments indicate that this new model has the potential to achieve higher precision predictions while maintaining performance with respect to accuracy and weighted recall.

S. Ragothaman et. al. [6] highlights the potential of machine learning as a tool for early-stage investors to screen potential investments based on specific features that may signal a company's potential to outperform its peers. While not a substitute for other evaluation methods, machine learning models can help reduce risk and uncertainty for investors. Previous approaches include rule-based expert systems, but machine learning models have the advantage of being able to learn from data and identify patterns that may not be apparent to human experts, although their success depends on the quality and relevance of the data used and the choice of algorithms and techniques employed.

Xiang et. al. [7] attempted to predict acquisitions using machine learning, focusing on companies founded between 1970 and 2007. They used various firm descriptors, including information on management teams, finance sources, and news from TechCrunch, but had to drop approximately 20,000 companies due to incomplete descriptions, highlighting the issue of data sparsity. With a dataset of 60,000 companies described by 22 features and segmented according to business sector, they enriched the data with news distribution in the five most representative topics for each sector, using a corpus of over 38,000 news. However, only around 5,000 companies had a presence in the corpus. The study found that considering news information improved the prediction results, achieving a success rate of 70%.

F. R. da Silva Ribeiro Bento [8] states that the use of Crunchbase data in predicting the success of early-stage startups has been explored in several works. One study analyzed a dataset of over 80,000 startups from five US states founded between 1985 and 2014 to predict either an MA or an IPO. The author compared the performance of different machine learning models such as logistic regression, SVM, and random forests in predicting startup success. To avoid the sparsity problem found in other works, the proposed approach included a data acquisition and selection stage, followed by a pre-processing stage. The results suggest that using machine learning to analyze Crunchbase data has potential in creating a decision support system for early-stage startup prediction.

P. Sathya Worawong et al.'s [9] study, factor analysis was used to categorize parameters and identify relationships. The 14-parameter matrix was reduced to 7 factors through PCA rotation. The study emphasized the significance of managerial expertise but did not make predictions based on fundraising rounds.

K. Regmi et al.[10] conucted, the focus was on examining the influence of accelerators on startups. The study explored different factors that impact a company's performance as perceived by startup accelerators. However, the analysis did not consider revenue, exit value, or IPO return analysis of the startups.

III. EXISTING WORK

[1] This project aims to simplify startup analysis for investors who lack the knowledge and resources needed to make data-driven investment decisions. It offers a unique 3-in-1 analysis approach, including founder analysis, investment trend analysis, and a success predictor, which are crucial for investing in new-age startups, especially in times of heightened risk.

© IJRASET: All Rights are Reserved |

Through comprehensive research, analysis, and a user-friendly system, it provides hassle-free and relatively low-risk investment guidance by considering a wide range of factors and offering data-backed recommendations. The results are presented on a choropleth map, visually indicating the success/failure rates of different organizations based on regional trends, global macroeconomic conditions, and geopolitical factors. This project focuses on three key pillars for evaluating startups: assessing founders, analysing previous investments in the industry, and predicting future growth prospects. It should be noted that the study is limited to venture capitalists interested in investing in a single startup rather than a portfolio. Additionally, the data acquisition process only includes startups that have achieved unicorn status, ensuring a more reliable reference point. Future advancements could involve considering "soonicorns" (soon-to-be unicorns) to cater to a broader customer base and increase the data points for more accurate predictions. While our project provides machine learning-generated analysis, the user retains the discretion to interpret the results and take appropriate actions.

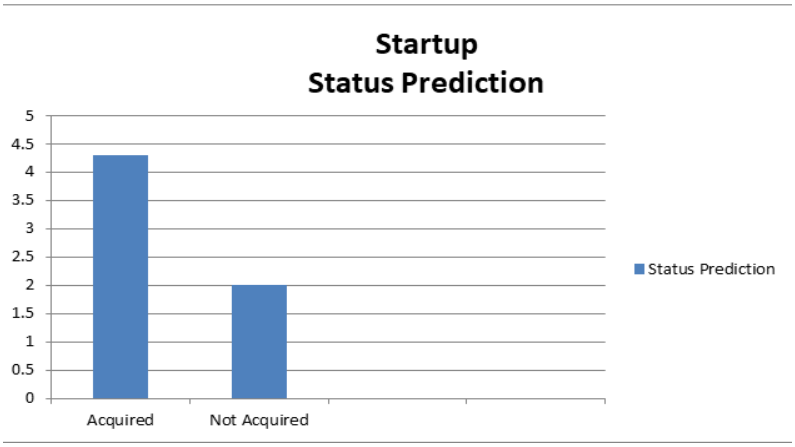


Fig 3.1: Existing System

Implementing the aforementioned approach poses potential challenges for investors regarding the status of startup acquisitions. One of the primary concerns is the possibility of encountering problems when it comes to determining whether a startup has been acquired or not. This issue becomes particularly significant if investors only become aware of a startup's closure after a considerable amount of time has passed. Such a scenario can result in wasted time and financial resources for the investors involved. Additionally, relying solely on the binary classification of a startup's acquisition status as "acquired" or "not acquired" may not provide the comprehensive information necessary for making effective investment decisions. By solely focusing on the binary acquisition status, investors may overlook crucial factors that can impact their investment outcomes. Therefore, it is essential to supplement the acquisition status with additional indicators and analysis to enhance the decision-making process and mitigate potential risks associated with investing in startups.

IV. PROPOSED MODEL

The proposed model for "startup acquisition status prediction" aims to predict the acquisition status of startups, including whether they are acquired, have gone public through an IPO, are closed, or still operating. This model utilizes machine learning algorithms and techniques to analyze various factors that influence a startup's acquisition status and provide accurate predictions.

1) *Acquired:* Refers to the situation where a larger, more established company purchases a smaller startup. When a startup is acquired, it means that the ownership of the startup is transferred to the acquiring company.
2) *IPO:* "IPO" stands for Initial Public Offering. An IPO refers to the process through which a private company offers its shares to the public for the first time, making them available for trading on a stock exchange.
3) *Operating:* Refers to the status of a startup that is actively conducting its business operations. When a startup is classified as "operating," it means that it is functioning, generating revenue, and carrying out its core activities as intended.
4) *Closed:* Refers to the status of a startup that has ceased its operations and is no longer conducting business. When a startup is classified as "closed," it means that it has permanently shut down its operations, often due to various reasons such as financial difficulties, lack of market viability, inability to sustain growth, or strategic decision-making.

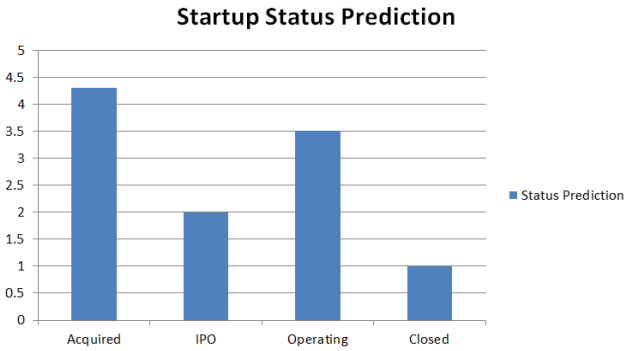


Fig 4.1: Proposed System

To develop the model, a diverse dataset of startups is collected from crunchbase, it comprises information about their financial statistics, market performance, industry trends, funding history, and other relevant factors. This dataset is then used to train the machine learning model. The model employs a supervised learning approach, where it learns from labeled data to identify patterns and relationships between the input features (startup characteristics) and the corresponding acquisition status. Machine learning algorithms, such as KNN, random forests are employed to build the predictive model.

A. System Model

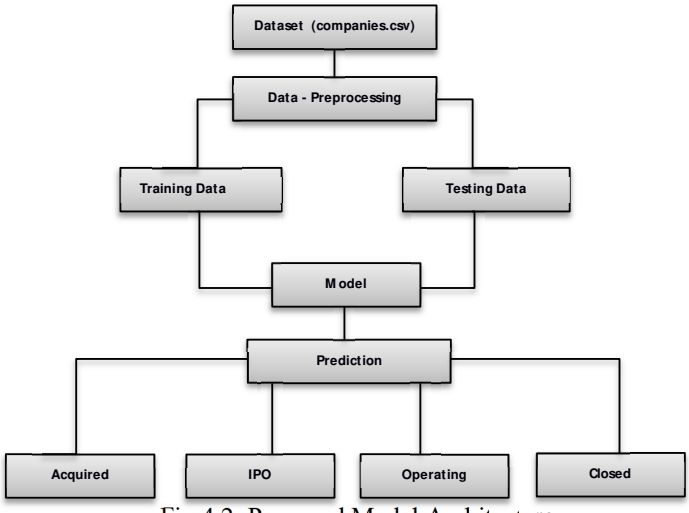


Fig 4.2: Proposed Model Architecture

The process of developing a startup acquisition model includes:

1) *Dataset:* The dataset used in this project is sourced from Crunchbase, a comprehensive database containing information about companies, investors, and various other details related to companies. The dataset exclusively comprises information derived from the respective company's website database, providing a reliable and accurate representation of the company's attributes. The dataset encompasses a wide range of features, including the company's name, permalink, category, funding dates, funding rounds, funding amount, city, state, founding dates, last milestone dates, and numerous other relevant parameters that offer valuable insights for analysis and prediction purposes.
2) *Data – Preprocessing:* Before training the model, the data undergoes a comprehensive cleaning and transformation process. This involves addressing issues such as inconsistent data, missing values, and duplicate values. Various techniques, including normalization, Principal Component Analysis (PCA), and outlier removal, are employed to ensure the data is of high quality. Inconsistent data is rectified to ensure uniformity, missing values are handled through imputation or removal, and duplicate values are eliminated. Additionally, the data is transformed to follow a normal distribution, dimensionality is reduced using PCA, and outliers are addressed. These steps collectively prepare the data for effective model training.

3) *Model Building:* To ensure an effective evaluation of our machine learning model's performance, we divided the dataset into training and testing data. The training data was utilized to train the model by exposing it to various patterns and relationships within the data. Subsequently, the model's performance was evaluated using the testing data, which had been kept separate from the training process. In order to maintain an appropriate balance between the training and testing datasets, we allocated approximately 70% of the data for training and the remaining 30% for testing. This division allowed us to assess the model's ability to generalize and make accurate predictions on unseen data, ensuring the reliability and robustness of our analysis. In order to train our acquisition model, we employed two popular machine learning algorithms, namely Random Forest and K-Nearest Neighbors (KNN). These algorithms were selected based on their demonstrated high prediction accuracy. By using both Random Forest and KNN algorithms, we aimed to leverage the strengths of each model and increase the overall accuracy of our acquisition prediction model. The combined utilization of these algorithms allowed us to benefit from their complementary approaches to pattern recognition and prediction, enabling more reliable and precise outcomes in identifying the acquisition status of startups.
4) *Prediction:* The "startup acquisition status prediction" model is to forecast the status of startups in terms of acquisition, including whether they have been acquired, undergone an IPO, closed down, or are still operating.

V. RESULTS

The proposed system for predicting acquisition status achieves an impressive accuracy score of 93% when utilizing the Random Forest algorithm, and a score of 91% when employing the KNN (K-Nearest Neighbors) algorithm. These high accuracy scores highlight the effectiveness and reliability of the system in accurately predicting the acquisition status of startups. By leveraging advanced machine learning techniques and algorithms, the system provides valuable insights to investors and stakeholders, aiding them in making informed decisions regarding their investment strategies and potential acquisitions.

VI. CONCLUSION

In conclusion, the use of machine learning algorithms in predicting startup acquisition status can provide valuable insights and advantages for investors, entrepreneurs, and other stakeholders. By analyzing a wide range of data sources and identifying patterns and correlations that may not be immediately apparent to human analysts, ML models can help predict which startups are most likely to be acquired.

The project outlined above involves developing an ML model to predict the likelihood of a startup being acquired based on various input features. Historical data on startup acquisitions will be used to train and validate the model, and it will be applied to new data to generate predictions on which startups are most likely to be acquired in the future.

Overall, the project has the potential to assist in making informed investment decisions, driving innovation and growth in the startup ecosystem, and potentially contributing to the success of startups and larger companies alike. The use of machine learning in predicting startup acquisition status represents an exciting and promising area of research that has the potential to transform the business landscape.

VII. FUTURE WORK

In future work, there are several potential directions to enhance the startup acquisition status prediction using machine learning project.

1) *Incorporating Additional Features:* Currently, the model relies on a set of predefined features to predict the acquisition status. In future work, more relevant features can be explored and incorporated into the model which can include financial indicators, industry-specific metrics, and social media sentiment analysis.
2) *Incorporating time-series Analysis:* Considering the temporal nature of startup data, time-series analysis techniques can be integrated into the model. This would enable capturing trends, seasonality, and other time-dependent patterns that may influence the acquisition status.
3) *Expand the Dataset:* While the current model is trained on a specific dataset, expanding the dataset to include a larger and more diverse set of startups can enhance the model's generalizability. Including startups from various industries, geographical locations, and stages of development can help capture a broader range of acquisition scenarios and improve the predictive capability of the model.

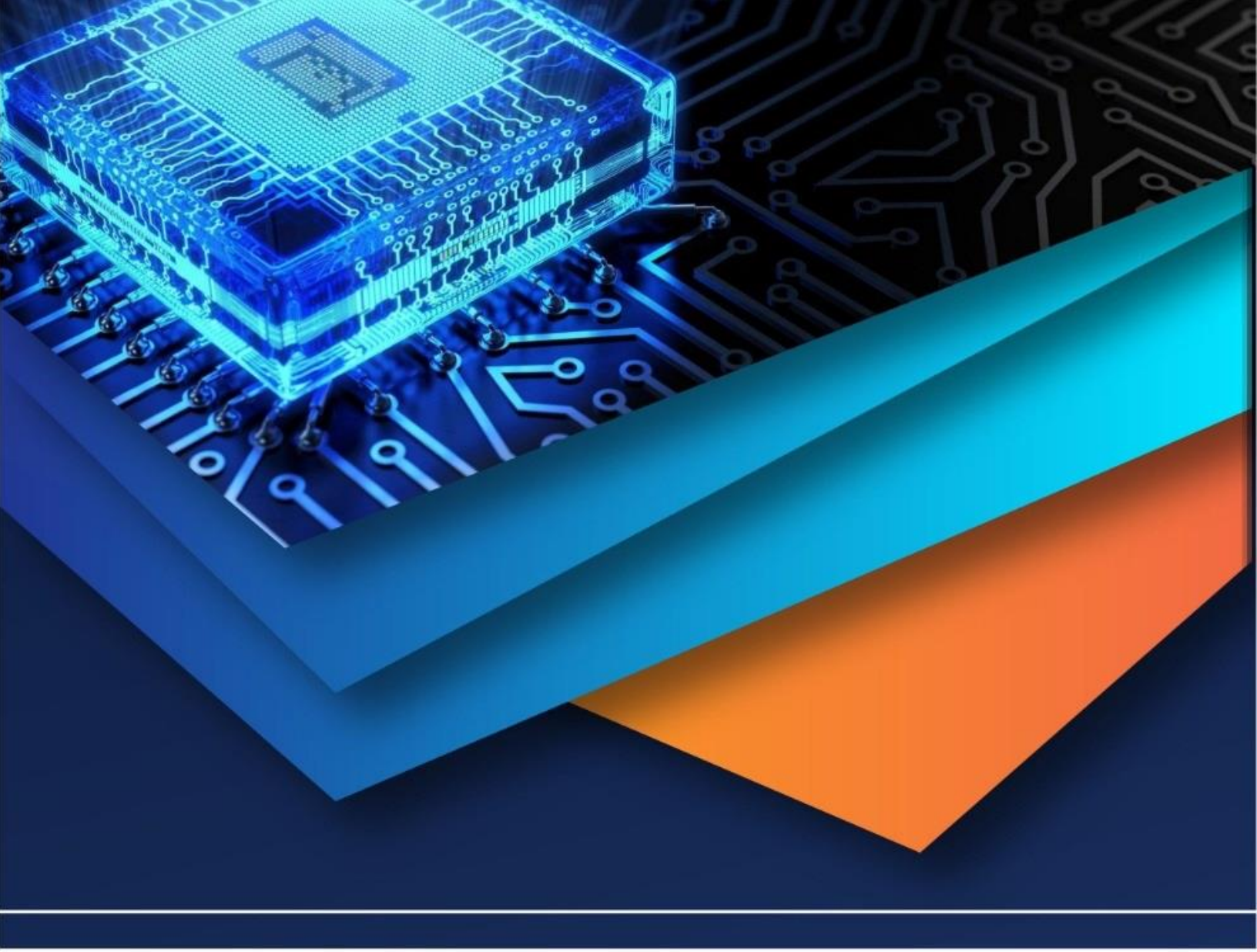
By exploring these avenues for future work, the startup acquisition status prediction project can be further enhanced, leading to more accurate predictions and valuable insights for investors and stakeholders in the startup ecosystem.

© IJRASET: All Rights are Reserved |

REFERENCES

[1] J. Singhal, C. Rane, Y. Wadalkar, M. Joshi and A. Deshpande, "Data Driven Analysis for Startup Investments for Venture Capitalists," 2022 International Conference for Advancement in Technology (ICONAT), 2022, pp. 1-6, doi: 10.1109/ICONAT53423.2022.9725892.
[2] M. Chen et al., "A Trend-aware Investment Target Recommendation System with Heterogeneous Graph," 2021 International Joint Conference on Neural Networks (IJCNN), 2021, pp. 1-8.
[3] M.A. Cusumano, "Evaluating a Startup Venture Considering the key elements of successful startups," Communications of the ACM, Vol. 56, No.10, pp. 26-29, October 2013.
[4] S. Lyu, S. Ling, K. Guo, H. Zhang, K. Zhang, S. Hong, Q. Ke and J. Gu, "Graph Neural Network Based VC Investment Success Prediction" arXiv e-prints, 2021.
[5] Ivy Wang "Predicting a Startup's Acquisition Status" ,Stanford University ,2019.
[6] S. Ragothaman, B. Naik, and K. Ramakrishnan, "Predicting corporate acquisitions: An application of uncertain reasoning using rule induction," Inf. Syst. Frontiers, vol. 5, no. 4, pp. 401–412, Dec. 2003.
[7] G. Xiang, Z. Zheng, M. Wen, J. Hong, C. Rose, and C. Liu, "A supervised approach to predict company acquisition with factual and topic features using profiles and news articles on techcrunch," in Proc. Proc. 6th Int. AAAI Conf. Weblogs Social Media, May 2012, pp. 2690–2696.
[8] F. R. da Silva Ribeiro Bento, "Predicting start-up success with machine learning," M.S. thesis, Dept. Inf. Manage., Universidade Nova do Lisboa, Lisbon, Portugal, 2018.
[9] P. Sathya Worawong, N. Thawesaengskulthai and K. Saengchote, "Determinant of Startups' Fund-raising Value: Entrepreneur and Firm Characteristic," 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 2018, pp.1309-1314.
[10] K. Regmi, S. A. Ahmed and M. Quinn, "Data Driven Analysis of Startup Accelerators," Universal Journal of Industrial and Business Management 3, pp. 54-57, 2015.
[11] G. Ross, S. Das, D. Sciro and H. Raza, " CapitalVX: A machine learning model for startup selection and exit prediction,The Journal of Finance and Data Science," The Journal of Finance and Data Science, vol. 7, pp. 94-114, 2021.
[12] A. Krishna, A. Agrawal and A. Choudhary, "Predicting the Outcome of Startups: Less Failure, More Success." IEEE 16th International Conference on Data Mining Workshops, 2016, pp. 798-805.
[13] J. Arroyo, F. Corea, G. Jimenez-Diaz and J. A. Recio-Garcia, "Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments," in IEEE Access, vol. 7, pp. 124233-124243, 2019.

© IJRASET: All Rights are Reserved |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)