

HIDDEN PATTERN MINING IN FAILED BUSINESS STARTUPS

¹SK. AHMED MOHIDDIN, ²GOPISETTI BHANUVARSHA, ³ERADALA NAGA MAHA JAYA
SRI, ⁴PAVANA SRI RAM BASWANI, ⁵KUMBHAM BABY SWARUPA

¹Associate Professor, ^{2,3,4,5}Students, Department of Computer Science and Engineering, SRI VASAVI INSTITUTE
OF ENGINEERING & TECHNOLOGY, NANDAMURU, ANDHRA PRADESH

ABSTRACT

The rapid growth of the global startup ecosystem has created significant opportunities for innovation, economic development, and technological advancement. However, despite the increasing number of startups emerging each year, a large proportion of them fail during their early operational stages due to factors such as poor market fit, financial instability, inefficient resource management, and strategic misalignment. Identifying the underlying causes of startup failure is therefore essential for investors, founders, and policymakers to make informed decisions and reduce financial risk. This project, titled “Hidden Pattern Mining in Failed Business Startups,” aims to develop an intelligent analytical platform capable of predicting startup failure risk and uncovering hidden patterns associated with unsuccessful ventures. The proposed system, called the Startup Failure Risk Intelligence Platform (S-FRIP), integrates full-stack web technologies with advanced machine learning techniques to analyze structured startup data such as funding history, industry type, team size, operational duration, and burn rate. The platform employs the XGBoost algorithm for predictive classification to estimate the probability of startup success or failure. In addition, K-Means clustering is used to segment startups into meaningful risk categories, while the FP-Growth algorithm is applied to discover frequent patterns and associations that commonly

occur in failed startups. These analytical techniques enable the system to generate both quantitative risk predictions and qualitative insights into contributing factors. The results are presented through interactive dashboards that transform complex analytical outputs into intuitive visualizations for non-technical users. By combining predictive analytics, pattern mining, and visualization within a web-based system, the project demonstrates how data-driven intelligence can enhance investment decision-making, support startup founders in identifying weaknesses, and ultimately contribute to reducing failure rates within the entrepreneurial ecosystem.

I INTRODUCTION

The rapid growth of the global startup ecosystem has significantly influenced modern economies by encouraging innovation, entrepreneurship, and technological advancement [1]. Startups are widely recognized for introducing disruptive solutions that transform traditional industries such as finance, healthcare, and e-commerce [2]. Despite their potential impact, a large percentage of startups fail within the first few years of operation due to various operational and strategic challenges [3]. Studies indicate that poor market understanding, insufficient funding, and weak business models are among the major reasons for startup failure [4]. In many cases, startups struggle to identify sustainable revenue streams and effective growth strategies [5]. Additionally, the lack of experienced leadership

and poor decision-making processes often contribute to organizational instability [6]. Traditional methods used by investors and analysts to evaluate startup performance rely heavily on manual analysis and expert opinions [7]. These approaches are often subjective and may not effectively capture complex relationships within large volumes of business data [8]. With the increasing availability of digital business datasets, data-driven approaches have become an important area of research in entrepreneurial studies [9]. Data mining techniques enable researchers and analysts to extract useful patterns and insights from large and complex datasets [10]. Machine learning algorithms have proven effective in predicting business outcomes and supporting decision-making processes [11]. These algorithms analyze historical data to identify patterns that may indicate the likelihood of success or failure in business ventures [12]. In particular, classification models have been widely applied in financial risk analysis and business performance prediction [13]. Clustering techniques are also used to group organizations with similar characteristics for better analytical understanding [14]. Association rule mining methods help discover relationships between different variables that may influence business performance [15].

In response to the increasing need for data-driven startup evaluation, intelligent analytical systems are being developed to support investors and entrepreneurs in decision-making [16]. These systems integrate machine learning and data mining techniques to analyze startup-related data and generate meaningful insights [17]. Predictive analytics allows organizations to estimate the probability of future outcomes based on historical information [18]. Ensemble learning algorithms have gained popularity due to their ability to improve prediction accuracy by combining

multiple learning models [19]. Among these algorithms, XGBoost has demonstrated strong performance in classification tasks involving complex datasets [20]. XGBoost utilizes gradient boosting techniques to enhance predictive accuracy and reduce model errors [21]. In addition to predictive modeling, clustering algorithms such as K-Means are used to segment startups into different risk categories based on similar characteristics [22]. This segmentation helps analysts understand patterns within groups of startups that share common attributes [23]. Association rule mining algorithms such as FP-Growth are applied to identify frequently occurring patterns among failed startups [24]. These patterns can reveal critical combinations of factors that contribute to business failure [25]. By integrating predictive classification, clustering, and pattern mining, analytical systems can provide a more comprehensive understanding of startup risk factors [26]. Such systems not only predict potential outcomes but also explain the underlying reasons behind those predictions [27]. Visualization tools further enhance these systems by presenting analytical results through interactive dashboards [28]. These dashboards allow users to interpret complex analytical information in a simple and intuitive manner [29]. Consequently, data-driven platforms can significantly improve investment decision-making and help startup founders identify potential weaknesses in their business strategies [30].

II LITERATURE SURVEY

The study of startup success and failure has attracted significant attention from researchers in the fields of entrepreneurship, data analytics, and business intelligence. Early research focused on identifying financial indicators and organizational factors that influence business sustainability and

performance [1]. Traditional statistical methods such as regression analysis and discriminant analysis were initially used to predict business failure based on financial ratios and operational indicators [2]. These approaches helped establish foundational models for risk assessment and bankruptcy prediction in organizations [3]. Over time, researchers recognized that startup failure is influenced by multiple complex and interconnected factors, including market conditions, management quality, funding availability, and innovation capability [4]. Studies on entrepreneurial ecosystems highlight that many startups fail due to the absence of product–market fit and poor strategic planning during early growth stages [5]. Additionally, limited financial resources and inefficient resource allocation often contribute to operational instability in emerging ventures [6]. Researchers have also emphasized the importance of business models and competitive strategies in determining startup performance and long-term survival [7]. As the availability of digital business data increased, the focus gradually shifted toward the use of data mining techniques for extracting hidden patterns from large datasets [8]. Data mining enables analysts to identify relationships among multiple business variables that may not be easily observed through traditional statistical approaches [9]. Techniques such as classification, clustering, and association rule mining have become widely used for discovering useful insights from complex datasets [10]. Machine learning methods have further enhanced analytical capabilities by enabling automated prediction and pattern discovery from historical data [11]. Several studies demonstrate that predictive models can effectively analyze business data to estimate the likelihood of startup success or failure [12]. Ensemble learning techniques and decision tree–based algorithms have also been widely applied in

financial risk prediction and business intelligence systems [13]. These advancements highlight the growing importance of intelligent analytical tools in understanding business performance and supporting data-driven decision making [14][15].

Recent research has increasingly focused on applying advanced machine learning algorithms to analyze startup ecosystems and predict business outcomes more accurately. Predictive analytics techniques allow organizations to evaluate future risks by learning patterns from historical business data [16]. Algorithms such as support vector machines, random forests, and gradient boosting models have demonstrated strong performance in classification problems related to business success prediction [17]. Among these approaches, the XGBoost algorithm has gained considerable attention due to its efficiency, scalability, and ability to handle complex nonlinear relationships within datasets [18]. Researchers have successfully applied XGBoost models to financial forecasting, risk prediction, and startup evaluation tasks [19]. In addition to predictive modeling, unsupervised learning techniques are also widely used to explore underlying structures within startup datasets [20]. Clustering algorithms such as K-Means enable analysts to group startups based on similarities in operational characteristics, allowing the identification of different risk segments [21]. These clusters help researchers understand patterns in startup growth, funding behavior, and failure trends across industries [22]. Association rule mining methods have also been applied to identify relationships between different startup attributes and failure outcomes [23]. Algorithms such as FP-Growth efficiently discover frequent patterns in large transactional datasets and reveal combinations of factors associated with unsuccessful ventures [24]. Such analytical approaches provide deeper insights into how multiple variables interact to

influence startup performance [25]. Furthermore, modern business intelligence platforms integrate predictive analytics with visualization tools to improve interpretability and usability of analytical results [26]. Interactive dashboards allow stakeholders to understand complex machine learning outputs in a simple and intuitive manner [27]. These visualization systems support investors and entrepreneurs in evaluating startup risk and making informed decisions [28]. Overall, the integration of machine learning, clustering, and pattern mining techniques provides a comprehensive framework for analyzing startup ecosystems and uncovering hidden factors that contribute to business failure and success [29][30].

III METHODOLOGY

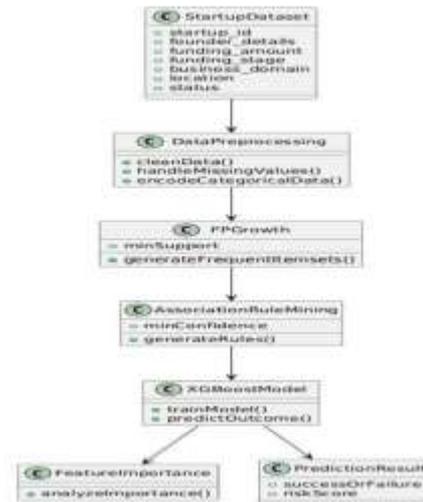
The methodology of the proposed system, Startup Failure Risk Intelligence Platform (S-FRIP), is designed to analyze startup data and identify hidden patterns associated with business failure using a combination of machine learning and data mining techniques. The system begins with the collection of structured startup data that includes attributes such as funding history, industry type, operational duration, team size, burn rate, and geographical information. This data is first processed through a data preprocessing phase where missing values, inconsistencies, and noise are handled to ensure the quality and reliability of the dataset. Feature engineering techniques are then applied to transform raw attributes into meaningful analytical features that improve model performance. After preprocessing, the dataset is divided into training and testing subsets to enable proper evaluation of model generalization. The core predictive component of the system is implemented using the XGBoost classification algorithm, which applies gradient boosting techniques to learn patterns from historical startup outcomes and

predict the probability of startup success or failure. This supervised learning model provides a quantitative risk score that helps stakeholders understand the likelihood of business failure. In addition to prediction, the methodology integrates unsupervised learning techniques to provide deeper analytical insights. K-Means clustering is applied to normalized feature sets to segment startups into different risk categories such as low-risk, moderate-risk, and high-risk groups, enabling comparative analysis between startups with similar characteristics. Furthermore, the FP-Growth algorithm is used for association rule mining to identify frequent combinations of attributes that commonly occur in failed startups, revealing hidden behavioral patterns within the data. These association rules provide interpretable insights that complement the predictive results and help stakeholders understand the contributing factors behind failure risks. The outputs generated by the machine learning engine are then integrated into a web-based interface where results are presented through intuitive dashboards and visualizations. This approach ensures that complex analytical outputs are transformed into understandable insights for investors and startup founders, enabling informed decision-making and proactive risk management.

IV SYSTEM DESIGN

The system design of the Startup Failure Risk Intelligence Platform (S-FRIP) provides a structured blueprint that defines how different components of the system interact to deliver predictive and analytical insights regarding startup failure risk. The design follows a layered architectural approach that separates the system into multiple logical layers, ensuring modularity, scalability, and maintainability. The presentation layer represents the user interface of the system and

is responsible for enabling interaction between users and the platform. This layer allows different stakeholders such as administrators, investors, and startup founders to access the system through a web-based interface. Users can register, log in, submit startup information, and view analytical results through intuitive dashboards and visualizations. The application layer manages the core business logic of the system and processes user requests received from the frontend interface. This layer handles operations such as data validation, authentication, authorization, and communication with backend services. Secure access control mechanisms ensure that different user roles interact with the system according to predefined privileges. The data layer is responsible for storing and managing structured startup data, including user information, startup profiles, and prediction results. A relational database is used to maintain data integrity, enable efficient querying, and support historical tracking of prediction outcomes. The layered architecture ensures separation of concerns so that modifications in one layer do not significantly affect other parts of the system, thereby simplifying maintenance and future enhancements.



In addition to the layered structure, the system design integrates a microservice-inspired architecture to incorporate machine learning capabilities within the platform. Instead of embedding machine learning models directly within the backend service, a dedicated Python-based machine learning service is used to host analytical models and perform predictive computations. Communication between the backend application and the machine learning service occurs through RESTful APIs, allowing the system to exchange data efficiently and maintain flexibility in model deployment. When a user submits startup data through the interface, the backend service validates and forwards the data to the machine learning engine for processing. The machine learning module performs preprocessing operations and applies algorithms such as XGBoost for predicting startup failure probability, K-Means clustering for identifying startup risk categories, and FP-Growth for discovering frequent patterns associated with failure scenarios. The analytical results generated by the machine learning service are then returned to the backend, stored in the database for future reference, and displayed to users through the frontend dashboards. UML diagrams such as the Use Case Diagram, Class

Diagram, and Sequence Diagram are used to visually represent system components, relationships, and workflows, helping developers understand system structure and behavior. These diagrams illustrate how users interact with the system, how data flows between components, and how various modules collaborate to produce meaningful insights. The modular and service-oriented design ensures that S-FRIP remains scalable, secure, and adaptable for future improvements and integration with additional analytical technologies.



V PROPOSED SYSTEM

The proposed system, Startup Failure Risk Intelligence Platform (S-FRIP), is designed to provide a data-driven approach for analyzing startup ecosystems and predicting the likelihood of business failure. The primary objective of the system is to help investors, entrepreneurs, and

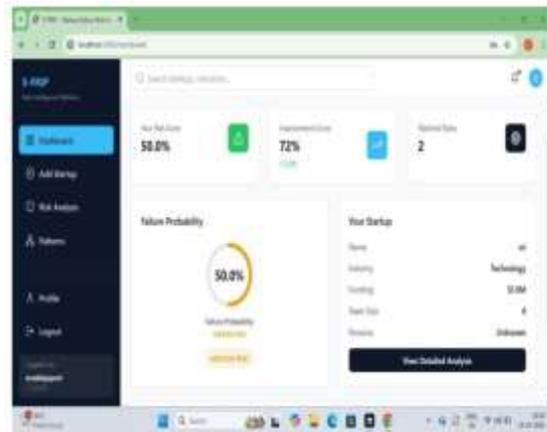
business analysts make informed decisions by utilizing advanced machine learning and data mining techniques. The platform collects and analyzes structured startup data such as funding history, industry type, operational duration, team size, burn rate, and other relevant business attributes. These datasets are processed using data preprocessing techniques to remove inconsistencies, handle missing values, and transform raw information into meaningful analytical features. The processed data is then used to train predictive models that evaluate the probability of startup success or failure. The system employs the XGBoost algorithm as the core predictive model because of its high accuracy, scalability, and ability to handle complex relationships between multiple variables. By analyzing historical startup data, the system can generate risk predictions that indicate the potential chances of business failure, allowing stakeholders to take preventive actions or adjust their strategies accordingly.

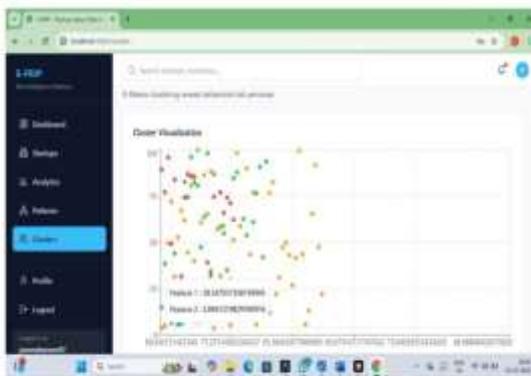
In addition to predictive analytics, the proposed system integrates unsupervised learning techniques to uncover deeper insights into startup failure patterns. The K-Means clustering algorithm is applied to categorize startups into different risk groups such as low-risk, moderate-risk, and high-risk segments based on their operational characteristics. This clustering approach allows users to understand similarities between startups and identify patterns within groups that share common features. Furthermore, the FP-Growth algorithm is utilized for association rule mining to discover frequent combinations of attributes that often appear in failed startups. These association rules help reveal hidden relationships between business factors and failure outcomes, providing valuable insights that go beyond simple prediction results. The system presents the analytical findings

through a user-friendly web interface with interactive dashboards and visualizations that simplify complex data analysis for non-technical users. By combining predictive modeling, clustering analysis, and pattern mining within a unified platform, the proposed system provides a comprehensive solution for startup risk assessment and strategic decision-making.

VI RESULTS & DISCUSSION

The results of the proposed Startup Failure Risk Intelligence Platform (S-FRIP) demonstrate the effectiveness of integrating machine learning and data mining techniques for analyzing startup data and predicting potential business failure risks. The XGBoost classification model successfully analyzes various startup attributes such as funding history, operational duration, industry type, and team size to estimate the probability of startup success or failure with improved predictive accuracy. The K-Means clustering algorithm groups startups into meaningful risk categories, enabling stakeholders to identify patterns among similar startups and understand different levels of risk exposure. Additionally, the FP-Growth algorithm discovers frequent combinations of factors that commonly occur in failed startups, revealing hidden relationships within the data. The analytical results are presented through interactive dashboards, allowing users to easily interpret complex insights. Overall, the system provides valuable decision-support capabilities for investors and entrepreneurs by offering predictive insights, identifying risk patterns, and supporting data-driven strategies that can reduce startup failure rates and improve long-term business sustainability.





In conclusion, the Startup Failure Risk Intelligence Platform (S-FRIP) demonstrates how modern data mining and machine learning techniques can be effectively utilized to analyze startup ecosystems and predict potential business failures. The project focuses on identifying hidden patterns within startup datasets by integrating predictive analytics, clustering methods, and association rule mining. By using the XGBoost classification algorithm, the system is able to evaluate various startup attributes such as funding history, operational duration, industry type, and team characteristics to estimate the likelihood of startup success or failure. This predictive capability provides valuable insights for investors, entrepreneurs, and business analysts who aim to minimize financial risks and improve decision-making processes. In addition to prediction, the implementation of K-Means clustering enables the system to categorize startups into different risk groups, helping stakeholders understand similarities and variations among startup profiles. The FP-Growth algorithm further enhances the analytical process by discovering frequent combinations of attributes that are commonly associated with failed startups, revealing hidden relationships that may not be easily detected through traditional analytical methods. The results generated by the analytical models are presented through interactive dashboards, making complex data insights accessible even to non-technical users. Overall, the integration of predictive modeling, clustering, and pattern mining within a unified platform provides a comprehensive analytical framework for understanding startup risks. The proposed system not only supports data-driven investment decisions but also assists startup founders in identifying weaknesses in their strategies, thereby contributing to improved business sustainability and reducing the overall rate of startup failure in competitive markets.

VII CONCLUSION

REFERENCES

1. Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609.
2. Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: An overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), 63–93.
3. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
4. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
5. Brownlee, J. (2016). *Machine learning mastery with Python*. Machine Learning Mastery.
6. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
7. Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2021). Finding the unicorn: Predicting early-stage startup success through a hybrid intelligence method. *Information Systems Journal*, 31(5), 889–917.
8. Drucker, P. F. (2014). *Innovation and entrepreneurship*. Routledge.
9. Gompers, P., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions? *Journal of Financial Economics*, 135(1), 169–190.
10. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. Springer.
11. Jain, P., Sompura, D. U., & Serene, I. M. (2024). Start-up success prediction analysis using hybrid machine learning techniques. *International Journal of Intelligent Systems and Applications in Engineering*.
12. Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), 995–1003.
13. Kitching, J., & Smallbone, D. (2012). Are freelancers a neglected form of small business? *Journal of Small Business and Enterprise Development*, 19(1), 74–91.
14. Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning*, 30(2–3), 271–274.
15. Li, J. (2020). Prediction of startup company success based on support vector machine and random forest. *Proceedings of the International Workshop on Artificial Intelligence and Education*.
16. Marmer, M., Herrmann, B., Dogrultan, E., & Berman, R. (2011). Startup genome report: A new framework for understanding why startups succeed.
17. Martínez, M. R., Carmona, P., & Vargas, J. M. (2025). Predicting business failure with the XGBoost algorithm: The role of environmental risk. *Sustainability*.
18. McKinney, W. (2018). *Python for data analysis*. O'Reilly Media.
19. Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
20. Ningrum, I. W. K., Ridho, F., & Wijayanto, A. W. (2023). Predicting startup success using a machine learning approach. *Journal of Applied Information and Communication Technologies*.
21. Osterwalder, A., & Pigneur, Y. (2010). *Business model generation*. Wiley.

22. Porter, M. E. (2008). The five competitive forces that shape strategy. *Harvard Business Review*, 86(1), 78–93.
23. Razaghzadeh Bidgoli, M., Vanani, I. R., & Goodarzi, M. (2024). Predicting the success of startups using a machine learning approach. *Journal of Innovation and Entrepreneurship*.
24. Ries, E. (2011). *The lean startup*. Crown Business.
25. Romero Martínez, M., Carmona Ibáñez, P., & Martínez Vargas, J. (2025). Predicting business failure using machine learning models. *Sustainability*.
26. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach*. Pearson.
27. Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to data mining*. Pearson.
28. Witten, I. H., Frank, E., & Hall, M. A. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
29. Zemmouchi-Ghomari, L., & Maroua, M. (2025). Beyond failure rates: Unveiling startup success factors with machine learning. *Journal of Computational Social Science*.
30. Zhou, Z. H. (2012). *Ensemble methods: Foundations and algorithms*. CRC Press.