Industrial Engineering Journal
ISSN: 0970-2555
Volume: 56, Issue 04, No. 4, April, 2024

SHARK TANK – REALTIME STARTUP SUCCESS AND FAILURE PREDICTION USING MACHINE LEARNING ALGORITHMS

Mrs.P.Satyavathi[1], Chaganti Saraswathi Satya Swetha[2], Devisetti Sreeja[3], Kotha Naga Sri Lakshmi[4], Charishma Sama[5], Uppuluri Sai Nikitha Devi[6]

[1] Assistant Professor, Department of AI & ML in Pragati Engineering College, Surampalem, Andhra Pradesh, India.
[2,3,4,5,6] UG Students Department of AI & ML in Pragati Engineering College, Surampalem, Andhra Pradesh, India.

Abstract: In the dynamic and competitive milieu of startup enterprises, the ability to accurately prognosticate success transcends mere desirability to assume a pivotal role in strategic decision-making. Statistical data from 2019 accentuates the formidable challenge, with an overwhelming 90% of startups encountering adversities, thereby emphasizing the imperative of employing rigorous methodologies for foretelling a startup's trajectory towards success. This study adopts a comprehensive approach, harnessing the formidable analytical capabilities of Machine Learning (ML) algorithms such as AdaBoost, Gradient Boosting, and Random Forest to conduct an exhaustive evaluation of diverse key performance indicators (KPIs) germane to startup enterprises. The parameters subjected to scrutiny extend beyond conventional metrics such as funding rounds and investor demographics to encompass more nuanced variables, including but not limited to, the composition of founding members, industry classification, and participant demographics. Notably, AdaBoost emerges as the preeminent algorithm, boasting a commendable predictive accuracy rate of 78%, thereby attesting to its efficacy in discerning patterns indicative of startup success. Drawing upon a rich corpus of historical data pertaining to startup enterprises, the models developed in this study furnish nuanced insights into the probabilistic landscape of startup success. By elucidating the interplay between diverse factors and the likelihood of achieving desired outcomes, our findings furnish invaluable guidance to stakeholders and prospective investors navigating the labyrinthine pathways of the startup ecosystem. Consequently, the insights gleaned from this research serve to inform strategic decision-making processes, thereby fostering more informed and judicious allocation of resources within the volatile and uncertain milieu of startup enterprises.

Keywords: Startup Success Prediction, Machine Learning, AdaBoost, Gradient Boosting, Random Forest, Predictive Modeling, Startup Metrics.

1. Introduction

In recent years, the global landscape has witnessed an unprecedented surge in the proliferation of startups, marking a significant departure from traditional economic models and signaling a paradigm shift towards entrepreneurship-driven growth. This surge has been particularly pronounced in the aftermath of the COVID-19 pandemic, which has not only accelerated the adoption of digital technologies but also spurred a remarkable surge in entrepreneurial activity worldwide. In the United States alone, the number of applications for new business formations

reached a record high of 551,657 in July 2020, representing a staggering 95% increase compared to the same period in 2019.

The success of a startup is multifaceted, encompassing various dimensions that extend beyond mere profitability. While financial viability is undoubtedly a crucial aspect, the success of a startup is also contingent upon factors such as innovation, market traction, scalability, and the ability to adapt to changing market dynamics. Successful startups are characterized by their ability to identify unmet needs or pain points in the market and develop innovative solutions that resonate with customers. Moreover, scalability is essential for startups to capitalize on growth opportunities and achieve sustainable long-term success.

However, navigating the path to success is fraught with challenges, and not all startups succeed in realizing their aspirations. The failure rate of startups in 2019 was alarmingly high, standing at around 90%. Research indicates that the failure rate varies across different stages of a startup's lifecycle, with 21.5% failing in the first year, 30% in the second year, 50% in the fifth year, and 70% by their tenth year.

The determinants of startup success are manifold and can vary depending on factors such as industry, market conditions, and the competitive landscape. Common challenges faced by startups include limited access to capital, intense competition, regulatory hurdles, and the risk of market saturation. Predicting success enables entrepreneurs, investors, and stakeholders to make informed decisions regarding resource allocation, strategic planning, and risk management.

By leveraging predictive analytics and machine learning algorithms, stakeholders can gain valuable insights into the factors that drive startup success and identify potential pitfalls that may impede growth. Moreover, predictive models can help identify promising startups early on, thereby enabling investors to allocate capital more efficiently and mitigate investment risks. Successful startups not only generate employment opportunities but also stimulate innovation, foster competition, and contribute to overall economic growth.

Furthermore, successful startups have the potential to address pressing societal challenges, ranging from healthcare and education to environmental sustainability and social inequality. By understanding the determinants of startup success and leveraging predictive analytics, stakeholders can enhance their ability to identify promising ventures, allocate resources effectively, and contribute to economic development and societal welfare.

2. Literature review

The contemporary entrepreneurial landscape has been teeming with dynamic activity, particularly underscored by the burgeoning global startup ecosystem. This phenomenon has experienced an exponential surge, especially in the aftermath of the seismic disruptions wrought by the COVID-19 pandemic, which served as a catalyst for a pronounced proliferation of novel enterprises. Notably, amidst the socioeconomic turbulence precipitated by the pandemic, a substantial cohort of individuals, compelled by circumstances such as job displacement and economic uncertainty, opted to pivot towards entrepreneurial pursuits as a means of economic sustenance and self-actualization. Evidentially, official records and

empirical data gleaned from reputable sources attest to an unprecedented proliferation of new businesses throughout the tumultuous year of 2020, eclipsing antecedent annual metrics and evincing an unmistakable entrepreneurial fervor that permeated various sectors and geographies.

In the Indian context, which serves as a focal point of burgeoning entrepreneurial endeavors, the trajectory of startup proliferation has been particularly noteworthy, underpinned by a confluence of factors including favorable regulatory frameworks, access to capital, burgeoning technological infrastructure, and an increasingly robust support ecosystem comprising incubators, accelerators, and venture capital firms. The ascendancy of India's startup landscape assumes paramount significance not only as a testament to the nation's burgeoning entrepreneurial spirit but also as a barometer of its economic dynamism and innovation prowess on the global stage. Understanding the nuanced contours of the Indian startup milieu is thus imperative, affording insights into the underlying drivers propelling founders, the multifarious challenges besieging them, and the requisite support mechanisms indispensable for their sustenance and growth trajectory.

An incisive examination of data gleaned from a myriad of sources, including qualitative interviews, quantitative surveys, and extant scholarly discourse, constitutes a foundational pillar in elucidating the intricacies characterizing the Indian entrepreneurial landscape. Such empirical analyses serve not only to discern prevailing trends and patterns but also to identify emergent phenomena and nascent opportunities that may have hitherto eluded scholarly scrutiny. Moreover, they provide a robust evidentiary basis for policymakers, practitioners, and stakeholders to formulate informed strategies aimed at fostering an enabling environment conducive to entrepreneurial dynamism and innovation.

Of particular intrigue within the realm of contemporary entrepreneurship is the burgeoning prevalence of startups helmed by individuals hailing from diverse sociocultural milieus. These nascent enterprises, often characterized by their agility, innovation, and propensity for disruption, epitomize the democratization of entrepreneurship, transcending traditional barriers and hierarchies to empower individuals from marginalized or underrepresented backgrounds. The motivational substrates underpinning the entrepreneurial aspirations of such individuals, as well as the attendant vicissitudes and challenges they confront along their entrepreneurial journey, constitute fertile terrain for scholarly inquiry and empirical investigation.

Notwithstanding the palpable exuberance surrounding entrepreneurial ventures, the harsh veracity of startup failure looms large as a sobering reminder of the inherent risks and uncertainties endemic to entrepreneurial endeavors. Scrutiny into the determinants underpinning startup demise not only sheds light on the pitfalls and challenges confronting nascent ventures but also underscores the imperative for a nuanced comprehension of the elements engendering entrepreneurial success. Leveraging methodological tools such as Machine Learning algorithms, coupled with an exhaustive examination of diverse facets of startup gestation, engenders a repository of discernments pivotal for aspiring entrepreneurs, investors, policymakers, and other stakeholders alike. By distilling empirical findings into actionable insights, scholars and practitioners can inform evidence-based interventions aimed at mitigating risk, enhancing resilience, and catalyzing success within the entrepreneurial

ecosystem. Ultimately, a nuanced understanding of the dynamics undergirding entrepreneurial success or failure furnishes a foundational substrate for the cultivation of a nurturing ecosystem conducive to entrepreneurial endeavor, thereby fostering innovation, economic growth, and societal prosperity.

3. Existing System

The prior research landscape in the field of startup success prediction underwent extensive examination by scholars such as Pan et al. and Arroyo et al., who scrutinized various aspects of startup outcomes using diverse methodological approaches. Pan et al. focused on forecasting specific events within startup trajectories, such as mergers, acquisitions, and initial public offerings, employing algorithms like Logistic Regression, Random Forests, and K Nearest Neighbors. Notably, K Nearest Neighbors emerged as the best performer, displaying superior F1 scores.

In contrast, Arroyo et al. expanded their investigation to include a broader range of startup outcomes, including subsequent funding rounds and closures. They employed algorithms such as Support Vector Machines, Decision Trees, Random Forests, Extremely Randomized Trees, and Gradient Tree Boosting. Remarkably, Gradient Tree Boosting achieved an accuracy level of approximately 82%.

However, the prevailing trend in startup success prediction leaned heavily on conventional statistical models and a limited set of predictors. This reliance is evident in the repeated use of methodologies like K-nearest neighbors (KNN) to forecast success rates among established firms. While effective for mature organizations, these conventional approaches often struggle with the complexities of the dynamic and heterogeneous early-stage startup ecosystems. The limitations of these methodologies include difficulties in addressing the inherent uncertainty and ambiguity characteristic of the initial stages of venture development. Consequently, such constraints lead to inaccurate predictions and suboptimal decision-making processes.

4. Proposed System

Our proposed system represents a significant advancement over existing methodologies, offering several advantages that enhance its effectiveness in predicting startup success. By incorporating AdaBoost, Gradient Boosting, and Random Forest algorithms, our system leverages the strengths of each approach. These ensemble methods are known for their ability to handle complex relationships within data and mitigate overfitting, resulting in more accurate predictions compared to single algorithms. AdaBoost, one of the key components of our system, has demonstrated the highest accuracy among the employed algorithms, achieving a notable 78%. This superior performance is indicative of its robustness in capturing the intricate patterns and dynamics inherent in startup data.

Unlike previous models that relied on a limited set of predictors, our approach considers a wide range of parameters such as investor count, founder demographics, funding details, industry type, and investor profiles. By integrating diverse factors that influence startup success, our model provides a holistic view of the startup ecosystem, leading to more reliable predictions. The output format of our system is designed to be clear and straightforward, providing

stakeholders and investors with concise information about the likelihood of a startup's success or failure. This transparency enables informed decision-making and facilitates proactive measures to mitigate risks or capitalize on opportunities.

Leveraging extensive historical data, our models offer nuanced insights into the potential success of startups. By analyzing past trends and patterns, our system identifies key drivers of success and highlights areas of concern, enabling stakeholders to make strategic adjustments and optimize their investment strategies. The dynamic nature of the startup landscape demands adaptive prediction models capable of accommodating evolving trends and market conditions. Our system's flexibility makes it well-suited to navigate the complexities of the ever-changing startup ecosystem, ensuring relevance and reliability over time.

5. Software Environment

Software Environment	**Technology**
Operating System	Windows 10/11
Development Language	Python 3.10
IDE	Visual Studio Code
Front-End Technologies	HTML5, CSS3, JavaScript
Back-End Framework	Django
Database Language	SQL
Database Management System	MySQL
Local Development Server	XAMPP
Development Web Server	Django Development Server
Design & Modelling Tool	Rational Rose

Table1: Software Environment

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 56, Issue 04, No. 4, April, 2024

6. System Architecture

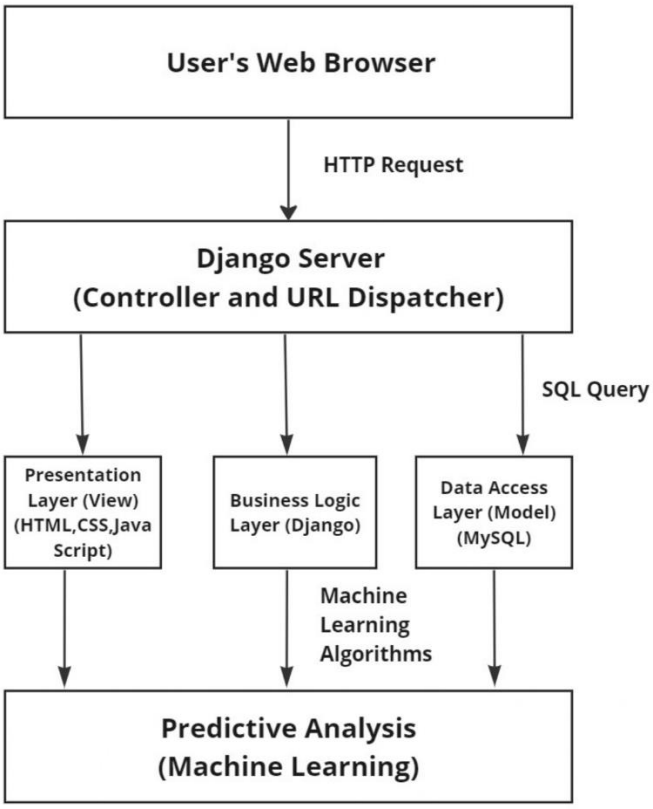
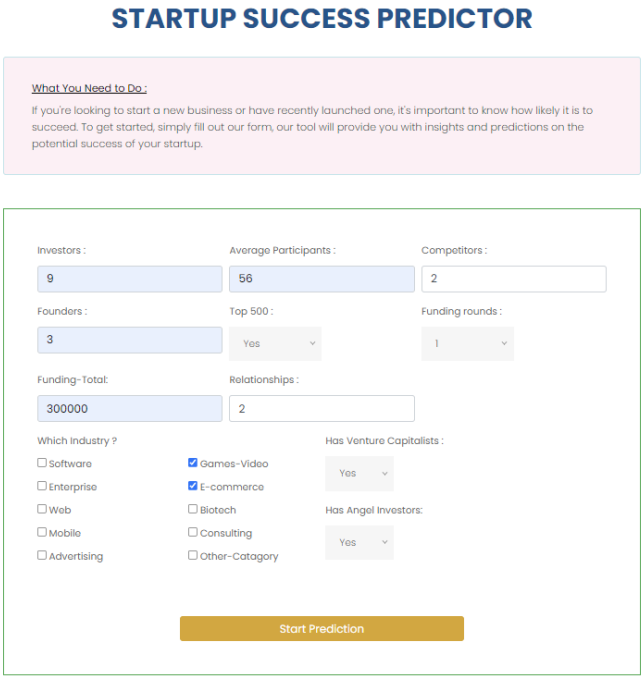


Figure 1: System architecture

7. Results



Industrial Engineering Journal
ISSN: 0970-2555
Volume: 56, Issue 04, No. 4, April, 2024

Figure 2: Predicting Startup Company – 1

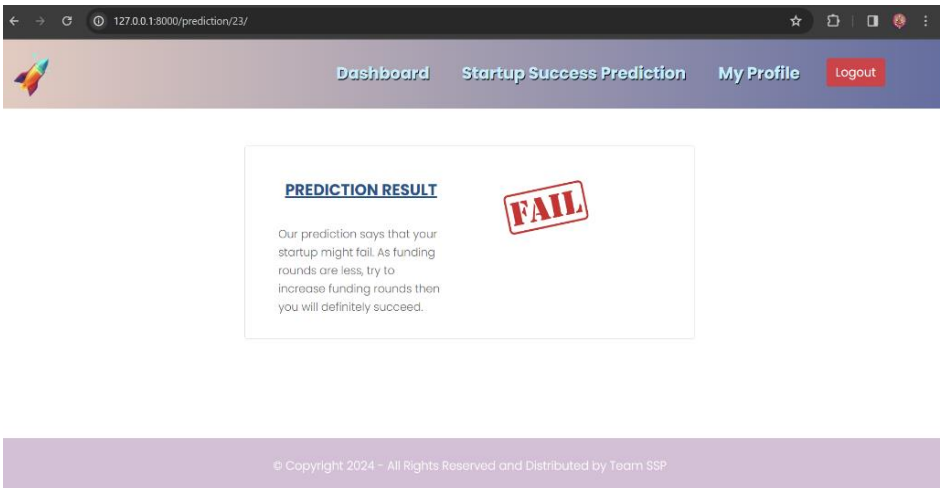


Figure 3: Startup Company – 1 Result Predicted as Fail

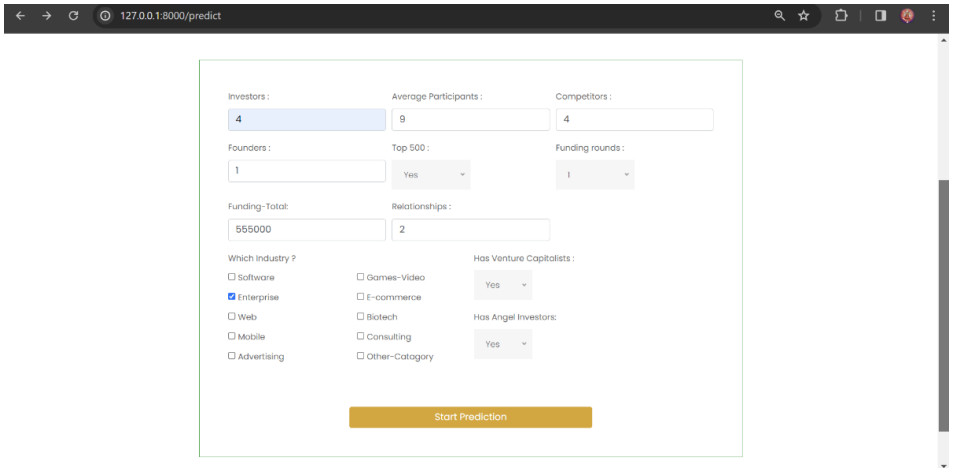


Figure 4: Predicting Startup Company – 2

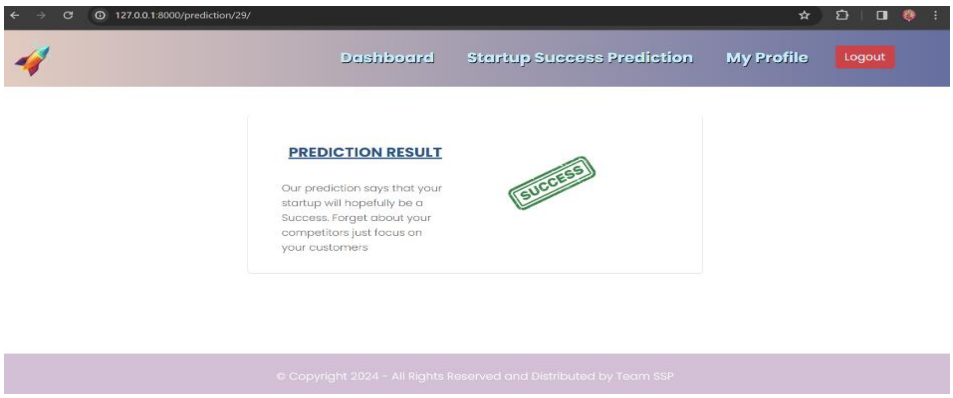


Figure 5: Startup Company – 2 Result Predicted as Success

Industrial Engineering Journal
ISSN: 0970-2555
Volume: 56, Issue 04, No. 4, April, 2024

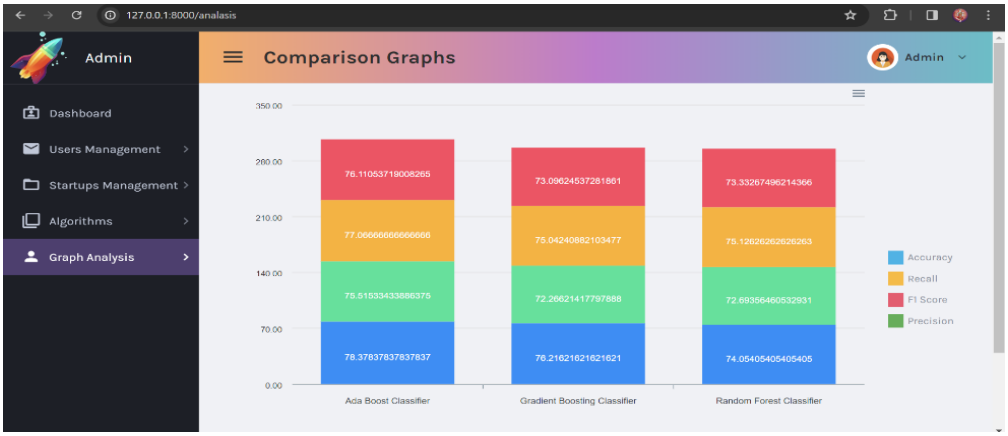


Figure 6: Comparison Graph

8. Future Directions

The pursuit of leveraging machine learning algorithms to forecast startup success is an ongoing endeavor marked by continuous refinement and innovation. While our current research has established a solid groundwork, the horizon is brimming with enticing opportunities, each poised to elevate our models to greater heights of potency and adaptability. Real-time Data Integration is a focal point, exploring the integration of dynamic data streams like market trends and user feedback to enhance model adaptability. Incorporating User Feedback from social media or customer reviews holds promise in capturing market dynamics and detecting early signs of traction or challenges.

Additionally, prioritizing Explainable AI (XAI) techniques ensures transparency in predictions, aiding stakeholders' understanding of underlying factors contributing to success probabilities. Investigating Cross-border Startup Analysis aims to extend model applicability across diverse regions, while Temporal Analysis offers insights into startup trajectory evolution over time. Integration of Unstructured Data sources like sentiment analysis and news articles enriches predictive models with deeper market insights. Ethical Considerations remain paramount, necessitating the development of frameworks to ensure responsible and equitable model application in the startup ecosystem.

9. Conclusion

In conclusion, this study presents a significant contribution to the domain of startup success prediction by leveraging the power of machine learning algorithms. Our exploration, which employed a combination of AdaBoost, Gradient Boosting, and Random Forest algorithms, achieved a commendable accuracy of 78% and demonstrated the effectiveness of these models in identifying patterns that signal a startup's potential trajectory. Furthermore, by incorporating a broader range of success indicators compared to existing research, our approach offers a more comprehensive framework for startup viability assessment.

This study not only advances the field of predictive modelling within the startup ecosystem but also empowers decision-makers with valuable information. By enabling stakeholders to make more informed choices regarding resource allocation and investment opportunities, this study

has the potential to fuel economic growth and foster a more vibrant environment for entrepreneurship.

Our study serves as a stepping stone for future advancements in startup success prediction as the field continues to evolve. By incorporating the proposed avenues for future exploration, such as real-time data integration, and Explainable AI, this study can be further extended to develop even more comprehensive and adaptable models. These enhanced models have the potential to provide stakeholders with unparalleled insights into the complex world of startups, ultimately contributing to a more successful and sustainable startup ecosystem.

References

[1] Y.B. Altun ”Pandemic Fuels Global Growth Of Entrepreneurship And Startup Frenzy”forbes.com.https://www.forbes.com/sites/forbestechcouncil/2021/04/09/pandemicfuels-global-growth-of-entrepreneurship-and startup frenzy/?sh=292d4e3f7308 (accessed: May 7,2022)

[2] S. Korreck. The Indian Startup Ecosystem: Drivers, Challenges and Pillars of Support. ORF Occasional Paper #210

[3] S. Torkington ”How the Great Resignation is driving a boom in startups from more diverse founders” weforum.org. https://www.weforum.org/agenda/2022/02/the-great-resignation-boomin-startups-from-more-diverse-founders/ (accessed: May 7,2022)

[4] T. Eschberger ”5 TOP reasons why startups fail” lead-innovation.com. https://www.lead-innovation.com/english-blog/reasons-startups-fail (accessed: May 7,2022)

[5] M. Van Gelderen, R. Thurik, and N. Bosma. Success and risk factors in the pre-startup phase. Small Business Economics, 24(4):365–380, 2005.

[6] T. M. Begley and W.-L. Tan. The socio-cultural environment for entrepreneurship: A comparison between east asian and anglo-saxon countries. Journal of international business studies, pages 537–553, 2001.

[7] R. Dickinson. Business failure rate. American Journal of Small Business, 6(2):17–25, 1981.

[8] W. B. Gartner. Who is an entrepreneur? is the wrong question. American journal of small business, 12(4):11–32, 1988

[9] N. K. Malhotra. Marketing Research: An Applied Orientation, 5/E. Pearson Education India, 2008.

[10] D. C. McClelland. Characteristics of successful entrepreneurs*. The journal of creative behavior, 21(3):219–233, 1987.

[11] Marco Felgueiras, Fernando Batista, Joao Paulo Carvalho, Creating Classification Models from Textual Descriptions of Companies Using Crunchbase”, IPMU 2020, CCIS 1237, pp. 695–707.

[12] Chenchen Pan, Yuan Gao, Yuzi Luo,” Machine Learning Prediction of Companies’ Business Success”, CS229: Machine Learning, Fall 2018, Stanford University, CA,2018.

[13] Javier Arroyo , Francesco Corea, Guillermo Jimenez-D ´ ´ıaz, Juan A. Recio-Garc´ıa," Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments", IEEE Access 7,2019, pp. 124233-124243

[14] Amar Krishna, Ankit Agrawal, Alok Choudhary, "Predicting the Outcome of Startups: Less Failure, More Success", IEEE, 2016.