e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

STARTUP-INVESTOR SUGGESTIONS THROUGH MACHINE LEARNING

Sakshi Tambe*1, Priti Sukale*2, Kharmale Pooja*3, Pansare Siddhi*4, V.V. Jadhav*5

*1,2,3,4,5Department Of Computer Engineering, Jaihind Polytechnic, Kuran, Pune, Maharashtra, India.

ABSTRACT

Investment strategies are successful in helping struggling businesses meet their desired financial goals. However, implementing these strategies poses challenges for startups and other early-stage organizations. This predicament could lead to various issues in achieving investment objectives, which could be highly detrimental. Furthermore, this situation could worsen if potential investors overlook excellent investment opportunities in innovative companies. Businesses heavily rely on initial funding, during which investors make significant investments crucial for future expansion. There's a notable lack of valuable research guiding investors on making precise investment recommendations. To address this gap, an interactive application has been developed, utilizing K Nearest Neighbor, Linear Regression, Artificial Neural Network, and Fuzzy Classification algorithms for investment-related suggestions. These suggestions have proven effective for both investors and registered startups, demonstrating satisfactory performance in facilitating informed decision-making.

Keywords: K Nearest Neighbors, Linear Regression, Artificial Neural Networks, And Fuzzy Classification.

I. INTRODUCTION

The Capitalism is one of the driving factors for catalyzing the growth of the human species and the main ingredient behind the success of many large corporations and other firms. The capitalist effectiveness can be useful in achieving immense growth which will be supported by the investors and other main players in this system. The rate of increase in the number of small scale businesses and startups has shot up exponentially. There are a number of different and innovative campaigns that have been useful in providing the much needed support and effective push towards these firms achieving their goals.

The investment to these companies and startups are crucial as their survival and success depends upon securing enough funding for their firm. This can be quite difficult to manage and secure without effective source for providing the funding. The governments also assist these individuals for the purpose of achieving success and improvement in the economies of the nation as a whole. If such businesses are assisted in their time of need, then there is an increased incidence of such businesses supporting the economy in the times of distress has been well documented. Such startups also are in need for an effective mechanism for the purpose of being listed on a platform for securing the investments.

The improvement in the startups have been affected due to the lack of an effective mechanism to secure the funding. The lack of financial resources can be very problematic as it can be difficult to run a small scale business or a startup without the financial means. These businesses can also use some credit to their advantage, but doing that can be detrimental as it induces unnecessary risk towards the users which can be difficult to manage with the already complex matters of handling and managing the startup. The lack of the funds can be quite difficult aspect that can induce unnecessary stress and other problems which can be quite problematic and can have a negative impact on the business.

If applied to infrastructure improvements, this prevailing thought laid the framework for a minimalist view of financing, in which the firm's rational expectations performance problem could be addressed without regard to investment considerations. Organizations were meant to face financial expenses established by centralized financial markets that were unaffected by a firm's institutional framework. Since the inception of classical concept, significant analytical research has been devoted to assessing the comparable effectiveness of various investment projections, using both a cooperative and mandatory framework, usually lacking regard to the likely causes of financial concerns.

Most of this research is founded on the concept of representative firms, which implies that the very same empirical results apply to all organizations, regardless of structure, with the distinction of economic obstacles. As a consequence, investigators were unsuccessful to identify whether the claimed numerical vulnerability of investment to economic metrics differed depending on the kind of organization. As a corollary, the financial

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

arguments that might have been presented were limited by the representative corporate perspective. This actually indicates that an effective application for investment suggestion is the need of the hour.

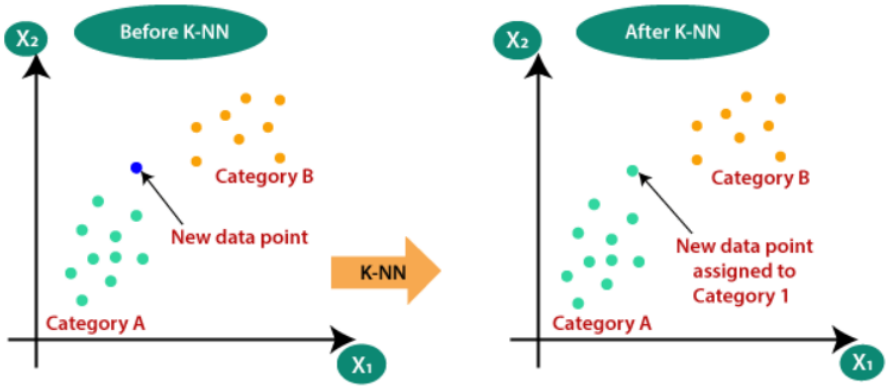

Figure 1: KNN

K Nearest Neighbor (KNN) is a simple yet effective machine learning algorithm used to suggest investments for startup investors. It analyzes factors like financial performance and market trends to find similarities between different startups. By looking at historical data of similar successful startups, KNN recommends investments based on past successes. It's user-friendly, requiring minimal training, making it ideal for investors seeking straightforward recommendations based on existing data. In summary, KNN provides quick and reliable suggestions to help startup investors make informed decisions in the dynamic startup landscape.

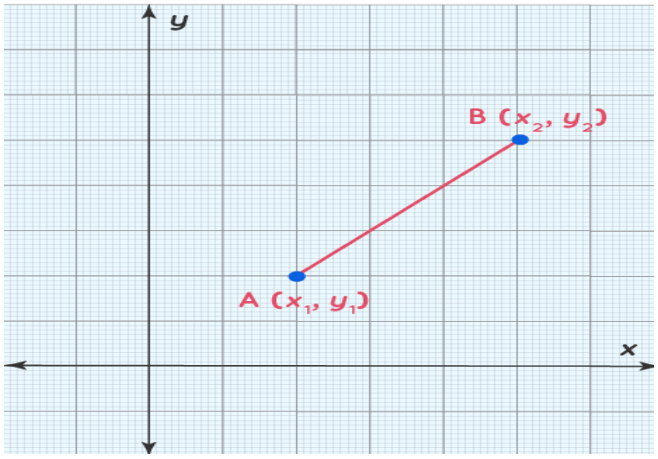

Figure 2: Euclidean Distance

Euclidean distance is a mathematical concept that measures the straight-line distance between two points in space. It's calculated by finding the square root of the sum of squared differences between corresponding coordinates of the points. This distance metric is crucial in various fields, including machine learning, where it helps determine similarity between data points in algorithms like K Nearest Neighbor (KNN) and K-means clustering. Overall, Euclidean distance provides a simple and intuitive way to quantify the separation between points in multi-dimensional space, making it fundamental in data analysis and pattern recognition.

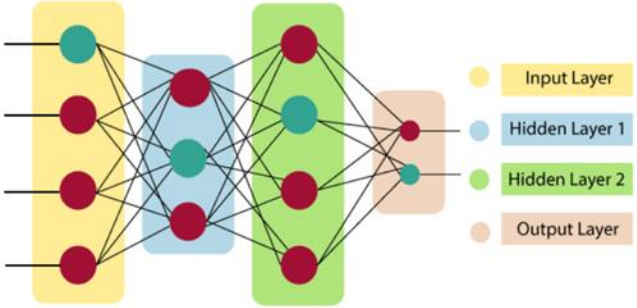

Figure 3: ANN

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Artificial Neural Networks (ANNs) are like smart computers that learn from data. They analyze lots of information about startups, like financial statements and market trends. By doing this, they can find patterns and make predictions about which startups might be good investments. ANNs help startup investors by giving them personalized advice based on what they want and how much risk they're willing to take. So, ANNs are really helpful tools for investors to make smart decisions in today's fast-changing market.

II. METHODOLOGY

Problem Statement

In today's fast-paced market, it can be challenging to connect the right startups with the right investors. However, our solution leverages advanced machine learning technology to streamline this matchmaking process. Think of it as a highly intelligent matchmaker for startups and investors, facilitating smooth and efficient connections in the business world. With the help of Euclidean distance KNN, along with ANN we can search startups and Investors.

Motivation

The motivation for using machine learning to suggest investments for startups is to assist investors in making better decisions. Machine learning analyzes market and startup data to identify potential success, helping investors choose the most promising opportunities and steer clear of risky ones. As machine learning continues to improve, it acts as a smart assistant that learns and enhances its suggestions over time. Ultimately, it's about using technology to simplify investing and increase profitability for all parties involved.

III. MODELING AND ANALYSIS

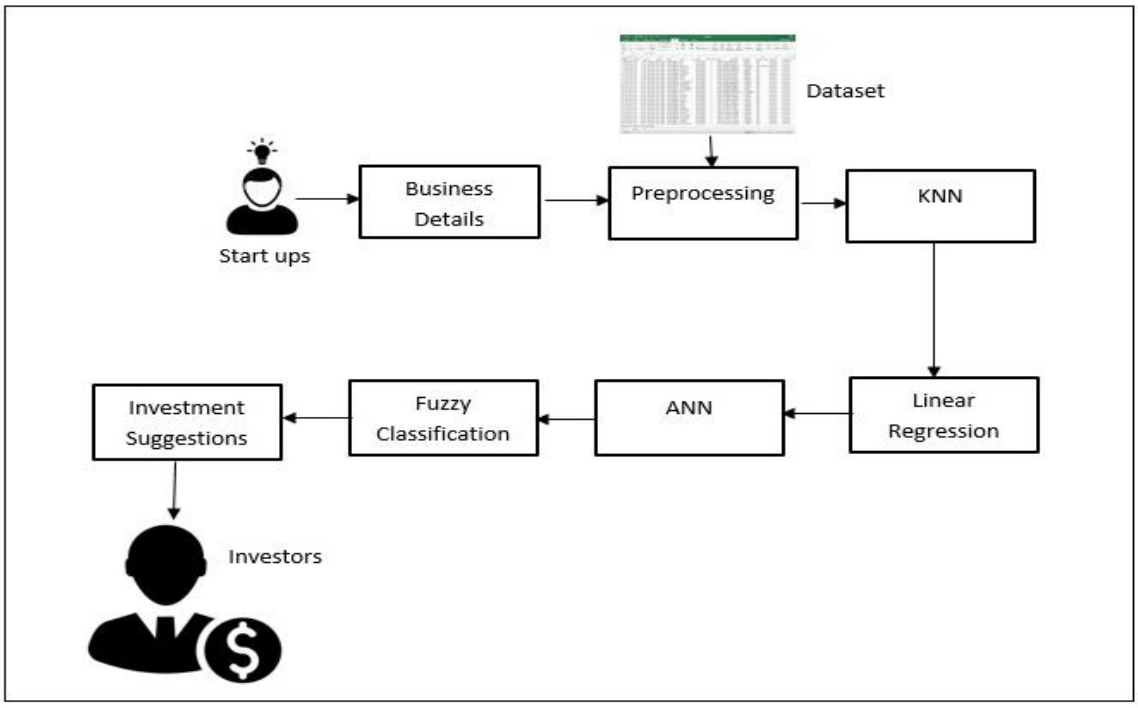

Figure 4: System Overview Design

A dataset containing company information from numerous investors and startups serves as the system's first source. This dataset may come from prepared synthetic datasets, causal datasets, or platforms such as UCI. The input dataset is preprocessed before analysis. To prepare the data for input into the system, this include cleaning it, dealing with missing numbers, and normalising or changing its features. Preprocessed data is used to extract business details like team expertise, project approach, earnings, and assets. These specifics function as the system's input parameters. The ANN is the system's central component. It employs its learning capabilities to analyse patterns and relationships within the data after receiving the business specifics as input.

The output from the ANN consists of predictions or scores for each startup based on the provided business details. These predictions serve as investment suggestions, ranking the startups in terms of their suitability for

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

investment. The system incorporates decision-making logic to interpret the ANN output. It may involve setting a threshold for acceptance or considering additional criteria. The goal is to convert the predictions into actionable investment decisions. Based on the given business details, the ANN generates predictions or ratings for each firm. These forecasts rate the startups according to investment suitability, making recommendations for potential investments. The system interprets the ANN output by including decision-making logic. It can entail establishing an acceptability threshold or taking into account other requirements. The objective is to transform the forecasts into practical investing choices.

Dataset:

The designed module is based on the official dataset from Kaggle, which is readily available.

KNN algorithm:

//Input : Sorted Distance List $SD_{L,}$

//Output: Cluster List K_{CL}

Start

$I_L = \emptyset$ [Inner Layer] $O_L = \emptyset$ [Outer Layer] , $K_{CL}=\emptyset$

MIN= 0 , MAX=SDL_{SIZE-1}

K= (MAX-MIN) /2

K=MIN+K

for i=0 to Size of SD_L

R = $SD_{L\,[i]}$

if(i<=K), then

$I_{L=}\ I_{L+R}$

else

$O_{L\,=}\ O_{L\,+R}$

end for

$K_{CL[0]} = I_L$

$K_{CL[1]} = O_L$

return K_{CL}

Stop

Hidden Layer Estimation

//Input: Cluster List CL, Weight set WS= { }

//Output: Hidden Layer value list HVAL

hiddenLayerEstimation (CL, WS)

Start

HVAL =∅, index=0

for i=0 to size of CL

ROW=CL[i]

for j=0 to N

X=0

for k=0 to size of ROW

ATR=ROW[k]

X=X+ (ATR* WS [index])

index++

end for

HVAL =1/ (1+e-X)

end for

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

end for

return HVAL

Stop

Module Description:

1) **Module A: Pre-processing**
- Input: Search Attributes by Investors ,Start-ups and Dataset
- Process: Attribute identification and Selection
- Output: Pre-processed list

2) **Module B: Labelling**
- Input: Pre-processed list
- Process: unique list and index allocation
- Output: Labelled list

3) **Module C: K-NN**
- Input: Labelled List
- Process: Euclidean Distance and Centroid Estimation
- Output: Classified nearest neighbour list

4) **Module D: Linear Regression**
- Input: User Attributes and Nearest List Row attributes
- Process: Intercept estimation
- Output: Regression List

5) **Module E: ANN**
- Input: Regression List
- Process: Neuron formation and Hidden layer
- Output: Probability list

6) **Module E: Fuzzy Classification**
- Input: Probability List
- Process: Probability Factor estimation through List sorting and IF then Rules
- Output: Investment Suggestion List

IV. RESULTS AND DISCUSSION

Companies mostly depend on seed money, where investors put up large sums of money that are essential for growth in the future. Important research that would advise investors on how to make exact investment recommendations is conspicuously lacking. An interactive application that uses K Nearest Neighbor, Linear Regression, Artificial Neural Network, and Fuzzy Classification algorithms for investment-related recommendations has been created to fill this gap. These recommendations have shown to be successful in promoting well-informed decision-making for investors and registered startups alike.

Gender	Name	Investment Avenues	Stock Market	Factor	Objective	Purpose	Duration	Invest Monitor	Expect
Female	Gunjal	Yes	Yes	Risk	Growth	Savings for Future	3-5 years	Monthly	20%-30%
Female	Dani	Yes	Yes	Risk	Growth	Wealth Creation	1-3 years	Monthly	20%-30%
Female	Tambe	Yes	Yes	Returns	Capital Appreciation	Wealth Creation	1-3 years	Monthly	20%-30%
Male	Adani	Yes	Yes	Returns	Capital Appreciation	Wealth Creation	3-5 years	Monthly	30%-40%
Female	Haware	Yes	Yes	Returns	Income	Returns	3-5 years	Monthly	10%-20%
Male	Oberoi	Yes	Yes	Risk	Capital Appreciation	Wealth Creation	1-3 years	Monthly	20%-30%
Male	Poonawalla	Yes	Yes	Risk	Capital Appreciation	Wealth Creation	3-5 years	Monthly	20%-30%

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Figure 5: Find Investors

Startup Name	Startup Age	Number of Partners	Annual Turnover	Investment Required	Return Type	Return Percentage
RailYatri	3	7	6.0	7.0	Monthly	5.0
5th Vital	5	5	2.0	4.0	Yearly	11.0
Ofbusiness	4	6	5.5	4.5	Monthly	7.0
CreditMantri	6	5	7.5	14.0	Yearly	19.5
HomeLane	2	7	5.0	14.0	Yearly	19.3
The Ken	5	5	4.0	15.0	Quarterly	2.5
The Porter	4	5	4.5	10.0	Quarterly	2.9
Vidgyor	5	3	5.0	9.0	Monthly	6.0

Figure 6: Find Startups

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

V. CONCLUSION

This research has expanded on the strategy that was previously offered in order to facilitate a practical and successful investment proposal approach. These investments could be extra purchases or goods that are primarily made with the goal of creating cash flow or appreciating in value over time. Purchasing goods that are not needed now but could increase profits later on is considered an investment. A financial asset purchased with the intention of making money or appreciating in value and then selling it is called an investment. The current financial arrangement with the hope of a positive improvement in the near future is referred to as an investment. The investment alternatives available nowadays are quite varied. Businesses greatly rely on starting capital to meet their goals. Additionally, innovators make large investments at this time that will be vital as the company grows. Thorough study on the process of providing investors with specific and definitive suggestions and answers is conspicuously lacking. Using K Nearest Neighbor, Linear Regression, and Artificial Neural Networks, this study offers a valid and advantageous approach for financial suggestion. The method's performance parameters, precision and recall, have been thoroughly assessed, yielding incredibly positive results. By using a much larger and more comprehensive dataset, the precision can be further increased.

VI. REFERENCES

[1] M. S. Oduro, H. Yu, and H. Huang," Predicting the Entrepreneurial Success of Crowdfunding Campaigns Using Model-Based Machine Learning Methods," in International Journal of Crowd Science, vol. 6, no. 1, pp. 7-16, April 2022, DOI: 10.26599/IJCS.2022.9100003.

[2] L. Wang, W. Duan, B. Yu, Q. Ying, H. Yang, and Y. Lei," Exploring the Response Mechanism of Remote Sensing Images in Monitoring Fixed Assets Investment Project in Terms of Building Detection," in IEEE Access, vol. 7, pp. 167919-167929, 2019, DOI: 10.1109/ACCESS.2019.2948060.

[3] Z. Wang et al., "A Research on Effect of Response to Internet Financing Reputation Evaluation on Achievement - From the Perspective of Social Network Theory," in IEEE Access, vol. 7, pp. 39352-39361, 2019, DOI: 10.1109/ACCESS.2019.2903681.

[4] Y. Wu, W. W. Y. Hsu, M. Wu, and C. Lu, "Optimal investments strategy for whole life insurance policies," 2014 International Conference on Information Science, Electronics and Electrical Engineering, 2014, pp. 1951-1955, DOI: 10.1109/InfoSEEE.2014.6946264.

[5] Yu Qian Ang, Andrew Chia, and Soroush Saghafian, "Using Machine Learn-ing to Demystify Startups Funding, Post-Money Valuation, and Success," In Babich, V., Birge, J.R., Hilary, G. (eds) Innovative Technology at the Inter-face of Finance and Operations. Springer Series in Supply Chain Management, vol 11. Springer, Cham. https://doi.org/10.1007/978-3-030-75729-8 10.

[6] Y. Zhan, X. Li, Y. Dou, and J. Chen, "A OLS Model for analysis of Institutional Distance of Agricultural Investment and Performance between China and Myanmar," 2020 2nd International Conference on Economic Management and Model Engineering (ICEMME), 2020, pp. 356-361,
DOI: 10.1109/ ICEMME51517.2020.00075.

[7] S. Zhang and J. Li, "KNN Classification with One-step Computation," in IEEE Transactions on Knowledge and Data Engineering, DOI: 10.1109/TKDE. 2021.3119140.

[8] X. Yang et al., "Piecewise Linear Regression Based on Plane Clustering," in IEEE Access, vol. 7, pp. 29845-29855, 2019, DOI: 10.1109/ACCESS.2019.2902 620.

[9] G. Qiu, X. Gui and Y. Zhao, "Privacy-Preserving Linear Regression on Distributed Data by Homomorphic Encryption and Data Masking," in IEEE Access, vol. 8, pp. 107601-107613, 2020,
DOI: 10.1109/ACCESS.2020.3000764.

[10] B. E. Yellakuor, A. A. Moses, Q. Zhen, O. E. Olaosebikan and Z. Qin, "A Multi-Spiking Neural Network Learning Model for Data Classification," in IEEE Access, vol. 8, pp. 72360-72371, 2020,
DOI: 10.1109/ACCESS.2020.298 5257