International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 03 | March - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Startup Investment Suggestion System Using ML

A.A. Shaikh[1], Shaikh Bushra Dilawar[2], Patel Suhana Razzakali[3], Lahare Anushka Sunil[4], Zinjal Pranjal Narendra[5]

[1]*Professor, Dept. of Computer Technology, P.Dr.V.V.P. Institute of Technology and Engineering, Loni, Maharashtra, India*

[2,3,4,5] *Final year Diploma Student, P.Dr.V.V.P. Institute of Technology and Engineering, Loni, Maharashtra, India*

ABSTRACT - For a failing company, investment techniques are a sure-fire way to get the returns they need. For businesses in the early stages of their development, such as startups, the anticipated execution of the plan is still a challenge. This is a problematic scenario since it could lead to many mistakes in achieving the investment potential, which could have devastating consequences. A prospective investor may lose out on a great chance to put money into a ground-breaking business, which can make the problem much worse. Seed financing is crucial for startups to reach their goals. That was also the moment when investors started putting a lot of money into the business, which would be crucial for its future growth. Indeed, there is a dearth of relevant literature on the methods for providing investors with thorough and accurate investment suggestions. To that end, we have developed an interactive software that makes use of K-Nearest Neighbors and Decision Making to provide recommendations for investments. Investors and companies using this platform have benefited from the proposals' ability to bring their ideas to fruition. The provided approach has produced quite satisfactory results, so the findings must be good.

Keywords: Investment Opportunity, K-Nearest Neighbors, and Decision Making.

1.INTRODUCTION

With the startup ecosystem evolving at a quick pace and investment decision-making become more complex, machine learning-based startup investment suggestion systems are becoming more significant in the current context. It is becoming increasingly difficult for investors to sift through the thousands of new companies appearing in fields as varied as sustainability, healthcare, fintech, and technology in search of opportunities with both high growth potential and reasonable risk. Conventional approaches to evaluating investments largely depend on labor-intensive, subjective, and bias-prone manual analysis, expert intuition, and sparse historical comparisons. Financial metrics, market trends, business models, founder histories, consumer traction, funding history, and machine learning-based startup investment suggestion systems all contribute to data-driven decision-making. This aids in the early detection of promising businesses in extremely competitive and unpredictable markets, helps investors make consistent, well-informed investment decisions, and lowers financial risk.

To find promising new businesses to back, this approach employs a number of machine learning algorithms. First, data is gathered from a variety of sources, including startup databases, financial documents, platforms for market research, and records of past investments. To make sure the data is accurate and reliable, it is preprocessed through stages like data cleansing, normalization, outlier elimination, and addressing missing information. Revenue growth rate, burn rate, scalability, industry sector, funding stage, and management team experience are among the main indicators that are identified using feature extraction and selection methodologies. Investment results, success probabilities, and expected returns can be predicted using supervised learning methods like classification and regression. On the other hand, clustering and other unsupervised learning approaches can be used to group startups with similar traits and risk profiles. Prioritized investment recommendations based on investor tastes and comfort level with risk are then generated by applying ranking and recommendation algorithms.

Because of its capacity to represent complicated and non-linear correlations between input characteristics and investment results, the Artificial Neural Network model is pivotal to the startup investment recommendation system. A normalized set of starting qualities is sent into the ANN's input layer, which then uses activation

functions to conduct weighted computations. Finally, the output layer generates predictions, including investment appropriateness scores or success likelihood. In order to train the model to make accurate predictions, it is fed past investment data and its weights are adjusted using backpropagation. In contrast to more conventional forms of analysis, ANNs are able to learn and discover previously unseen patterns and complex interplay between operational, financial, and market variables. A strong instrument for contemporary startup investment analysis, the ANN-based model offers accurate, dependable, and dynamic investment recommendations because to its adaptability, scalability, and high predictive capability.

[1] Three improvements are displayed by the Improved Artificial Immune Algorithm (IAIA) developed by Mingan Wu et al. There is a 38.5% decrease in risk due to the convergence in the 50-generation. Compared to the normal AIA, the precision is improved with a variation that is 5.2% lower. With an out-of-sample Sharpe ratio of 0.82, the robust generalization is demonstrated. Crucial constraints on the key innovation stack. With an efficiency rate ranging from 0.05 to 0.3, adaptive mutation is able to dynamically balance exploitation and exploration. The elite memory bank only stores solutions of the highest quality. Historical patterns, such as templates for asset rotation, are introduced via dynamic vaccinations. The following ways in which the market demonstrates its adaptability: A maximum drawdown of -12.3% was allowed during periods of volatility. For 30 assets, the optimization time was 5.2 seconds, which was 39% faster. In addition, the search space was reduced by 40% by vaccination guidance.

[2] A quantitative way to identify successful investors is given to us by Srishti Gupta et al. there are presently two identifying heuristics that the author is employing. The author will flag an investor if their Invertor Rank does one of the following: (i) remains continuously below 100 and can only go beyond that figure twice; or (ii) shows a general upward tendency, going down or up by no more than eight times from the previous snapshot. Out of 1,524 unlabeled investors in the author's Crunch Base data, the second rule effectively highlights those with a good chance of success; all of these investors had ranks below 200, and in most cases, under 50. As seen in Figure 1, there are a number of effective implantors. Paul Buchheit of General Catalyst Partners is just one example of how Investor Rank unearths promising but lesser-known investors like these. The next stage for the author is to confirm their results by consulting with experts. additionally, intend to incorporate the impacts of investment duration and amount into Investor Rank in the company life cycle.

[3] The Tadawl platform, developed by Hend Al-Khalifa et al., is a marketplace where students may showcase and invest in their startup ideas. The author's experience with gamification in the classroom, where students may build their interests while learning about investing and giving each other feedback, is also detailed in the article. The next logical step, following Tadawl's successful launch, is to expand the platform's capacity to accommodate a greater number of users and to incorporate other interactive features, including a digital wallet.
An examination of earlier research that was deemed a Litereature Survey is presented in the second part of this publication. Section 3 provides a comprehensive description of the proposed methodology, outlining the path of action. The experimental evaluation is covered in Part 4, possible modifications are discussed in Section 5, and the essay concludes with a conclusion on the existing plan.

2. LITERATURE SURVEY

[[4] It is helpful for investment statisticians that the RPA robot given by Minjie Zhu et al. can be developed using simple drag and drop. It is also very straightforward to start using, so even operators who can't program can do it independently. From a managerial perspective, RPA robots can record the entire operation, take screenshots during operations, and produce investment statistics reports with more accuracy thanks to their support for system logs, process execution, process key points, and video recording. This helps managers avoid risks and ensures data is traceable and verified across different levels and departments. From an efficiency standpoint, RPA has been implemented into investment statistics operations such as basic data verification, report generation, data verification of field consistency between tables, data verification of monitoring indicators, finding the causes of early warning, and investment completion risk assessment. This has improved the quality and efficiency of investment statistics.

[5] Machine learning, network analysis, and social media analysis are some of the latest business analytics tools that Sumy National Agrarian University et al. suggest can improve the accuracy of startup prospects assessments. Machine learning, for instance, enables a default prediction accuracy of 98.6% (Ciampi and Gordini,

2012), while network analysis, with centrality coefficients as high as 0.995, finds important actors and relationships. These methods show great efficiency. When comparing traditional startup valuation methodologies to new analytical methods, how do the two stack up? By combining financial measures with digital indicators (such as social media activity or network centrality), hybrid models are able to reduce subjectivity and improve accuracy, surpassing traditional methods (such as SWOT or expert assessments).Which of the following (financial, social, and technological) aspects are most important for Baltic startups to succeed? Factor study showed that social engagement (online activity, sentiment) and economic potential (prof itability, investor availability) are the main determinants, accounting for 88.5% of the variation in startup rankings. High correlation coefficients (> 0.98) for important variables show that data-driven strategies improve decision-making accuracy, which is supported by the results (H1). Accounting for 88.5% of the variation in startup rankings, factor analysis found that social engagement (online activity, sentiment) and economic potential (profitability, investor availability) are the key drivers. By combining taxonomy with machine learning and social media analysis, the created startup rating model achieves better results than conventional methods while simultaneously decreasing subjectivity and increasing dependability. Startups' chances of success in the Baltic states are heavily influenced by economic stability, technological innovation, social activity, and the power of the media. The level of social media involvement (coefficient0.990) and financial and resource potential (fac tor loading coefficient 0.986) were found to have the highest correlation with startup success, according to factor analysis. Significant variations in startup evaluation methods were found in a comparative study of the Baltic nations. Machine learning algorithms and Monte Carlo simulations are among the modern methods utilized in Estonia, accounting for 45% of the total approaches. This highlights Estonia's leadership in digitalization. Lithuania and Latvia are less competitive in the global startup environment because they adopt traditional methods in 50% of cases and 60% of cases, respectively. We were able to classify businesses according to their social and economic potential using taxonomic and cluster analysis. Leading the pack are companies that demonstrate strong social activity and economic stability, with average taxonomy coefficients of 0.784 and 0.86, respectively. In contrast, businesses that have poor indicators, like a taxonomy coefficient below 0.5 (Bi omatter, BoBo), should reconsider their company strategies and look for funding. Adaptable to different geographies and economic sectors, the created startup evaluation model has already demonstrated its usefulness for investment decision-making through its integration of taxonomy, machine learning, and social media analysis. Investors should incorporate analytical tools into decision-making to minimize risks and maximize rewards, while start-ups should concentrate on improving financial stability, social engagement, and reputation for successful development. Implications for theory. A hybrid multifactor approach that incorporates financial, technological, and social factors is proposed in this study, which adds to the existing literature on startup evaluation. It adds to the existing literature by showing how an integrated model including conventional financial measures and digital signals like network centrality and social media activity might be beneficial. Relevance to real life. In order to increase the quality of decision-making and decrease risks, the findings equip investors with evidence-based tools for faster and more accurate startup appraisal. The results can be used by policymakers and startup support groups to create initiatives that improve the startup environment in the Baltic States by increasing financial stability, promoting social engagement, and encouraging the use of advanced analytics. Limitations. The results may not be applicable outside of the Baltic States because the study only included 20 startups in its sample. Furthermore, the research relies on past data and chosen indicators; so, more comprehensive insights could be obtained by including longitudinal data or a wider variety of factors. Future research. The impact of Baltic startups' resiliency and transformation on investment strategies for the future should be investigated further (LSM, 2025; Stats and Market Insights, 2025). The suggested model might be further validated and improved by expanding the dataset to incorporate other regions and additional variables, like ESG measures or consumer sentiment.

[6] To investigate BIV's user experience, Shuqing Liu et al. (UXBIV) presented a UX framework. Thorough literature review and analysis formed the basis of this paradigm. The BIV evaluation process involves the use of independent and dependent variables as well as a research design. Important new features of this framework include the categorization of independent variables, task designs, method-paradigm combinations, and additional dependent variable elements. In addition, the author assessed three BIV designs ranging in complexity as part of a case study to back up this methodology. Authors think UXBIV to be an acceptable framework for

evaluating BIV user experience based on case study analysis and literature survey. There is a notable difference between the three BIV designs in terms of the user experience. Outperforming the other two designs on all task-based ratings and overall UX is Design 3, the most complex of the three. Consequently, this paradigm is a great resource for designers looking to assess the BIV designs' impact on users' experiences. The goal of UXBIV is to help business intelligence (BI) development firms become more competitive and influential while simultaneously improving decision-making performance and customer happiness.

[7] In order to investigate the user experience of BIV (UXBIV), Matteo Castiglioni et al. suggested a new approach, while Shuqing Liu et al. put forward a user experience framework. Thorough literature review and analysis formed the basis of this paradigm. The BIV evaluation process involves the use of independent and dependent variables as well as a research design. Important new features of this framework include the categorization of independent variables, task designs, method-paradigm combinations, and additional dependent variable elements. In addition, the author assessed three BIV designs ranging in complexity as part of a case study to back up this methodology. Authors think UXBIV to be an acceptable framework for evaluating BIV user experience based on case study analysis and literature survey. There is a notable difference between the three BIV designs in terms of the user experience. Outperforming the other two designs on all task-based ratings and overall UX is Design 3, the most complex of the three. Consequently, this paradigm is a great resource for designers looking to assess the BIV designs' impact on users' experiences. The goal of UXBIV is to help business intelligence (BI) development firms become more competitive and influential while simultaneously improving decision-making performance and customer happiness.platform for online advertising initiatives. While most of the existing research is concerned with how to maximize a campaign's revenue, this work introduces the idea of safety for the algorithms that choose the bid allocation daily. The author is keen on making sure that the bids meet the ROI and budget limits set by the firms' business divisions as often as possible. With the unpredictable limits (also known as safety) in mind, the author aims to maximize revenue. We prove that this setting is inapproximable within any strictly positive factor unless P = NP by modeling it as a combinatorial optimization problem. However, we also show that this problem admits an exact pseudo-polynomial-time solution. Most intriguingly, the author establishes that no system for online learning can ensure a sublinear number of violations of uncertain constraints while providing sublinear faux regret. Although the GCB algorithm may break the constraints a linear number of times, the author demonstrates that it suffers from a sublinear pseudo-regret. Consequently, the authors devised GCBsafe, an innovative method that ensures safety in exchange for a linear pseudo-regret. Incredibly, GCBsafe(ς,ϕ) is a version of GCB safe that ensures sublinear pseudo-regret and safety while incurring ψ and ϕ tolerances on the ROI and budget limitations, respectively. Lastly, the author verifies the theoretical results by evaluating the actual effectiveness of their algorithms on synthetic advertising situations. An intriguing area for future research is the development of algorithms that can adapt to changing constraints as they are learned; specifically, algorithms that can detect which constraints are active and loosen those that aren't. Another intriguing area of study is the correlation between loosening a restriction and a corresponding rise in revenue.

[8] An auto-scaler that is ROI-optimal for serverless edge contexts, RIA, was introduced by Huadong Li et al. Before developing a ROI economic model, the author examines current problems with quality of service (QoS) and cost modeling in the literature. To tackle the issues of high dynamic function calls and heterogeneous application loads in edge environments, the author then suggests the SHAP-based Deep Q-Network (SDQN) algorithm. This algorithm combines reactive methods based on thresholds with proactive methods of time-series prediction. To optimize DQN in state space, the author makes use of SHAP values. At last, the author compares state-of-the-art approaches with an implementation of RIA on the open-source serverless platform Open FaaS. RIA was the most cost-effective, had the second-lowest ROI, and the best quality-of-service scores.

[9] The assessment issue is currently dominating concerns, according to Craig A. Stewart et al., who stated, "The basic problems surrounding the measurement of organizational effectiveness are criteria problems." Cyberinfrastructure return on investment studies still lack a comprehensive evaluation of results. Nevertheless, a lot of ground has been covered, and the author can conclude from the cited materials that ROI assessments of cyberinfrastructure have yielded the following results: • Higher education institutions see an increase in academic output and grant receipts when they invest in

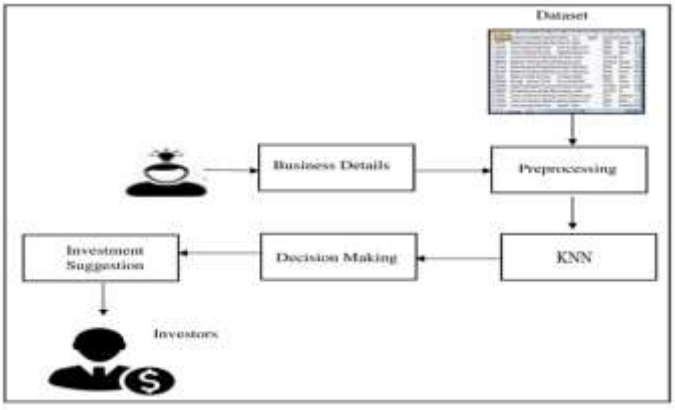
cyberinfrastructure. Investment in (and utilization of) modern computing facilities is positively correlated with both the overall number of publications produced and the effect of those articles. • XSEDE, one of the biggest US cyber infrastructure programs, offers services at a cost savings (ROI > 1.5) compared to the market value of those services, according to an analysis of the program. • Compared to commercial cloud services, on-premise resources are typically more cost-effective. Financial economies are rarely the motivation for universities to use commercial services for research. Cyberinfrastructure return on investment (ROI) evaluations can be conducted using published approaches by any institution that gathers enough data about its own system usage. Research institutions and society as a whole stand to gain monetarily and scientifically from investments in research cyberinfrastructure. What the author knows and what the author and funding agencies want to know are not always the same. The most pressing issues are around the lack of adequate assessment tools. Specifically, how can the value of cyberinfrastructure be determined when it is just one component among numerous that contribute to long-awaited scientific breakthroughs that have far-reaching cultural and social implications? And when you have both quantitative and qualitative measures of return on investment (ROI), what do you do in the near run? Both of these problems are crucial to ROI analysis of innovation and research cyberinfrastructure, but they have so far been unresolved.

[10] A machine learning ensemble prediction model was created by Yaohu Lin et al., which chooses the best prediction methods for each daily k-line pattern automatically. The paper's forecasting approach is predictive, and an investing strategy grounded on the model can produce better returns, according to the empirical results. There are four areas that this study adds to. To begin, this article enhances stock market forecasting research by integrating traditional candlestick charting with cutting-edge AI technologies. The author integrates AI with conventional techniques of technical analysis by testing the predictive power of thirteen distinct one-day candlestick patterns using various machine learning algorithms. Additionally, the author came to the conclusion that there are specific candlestick patterns that seem to have predictive power in the stock market. These patterns include pattern 4 and pattern 5. In addition to the thirteen patterns of daily k-line patterns and volume change features, feature engineering develops an eight-trigram classification of two-day k-line patterns. The Bagua, or eight-trigram scheme, is an important idea in Taoist cosmology, and the simple eight-trigram classification follows it. Using the high and low prices from two days of trading as well as the opening and closing prices, the eight-trigram categorization offers a basic set of attributes. The author included four groups of technical indicators—overlap, momentum, volume, and volatility—as supplementary feature variables to raise the quality of the forecast. When it comes to short-term forecasting, the author finds that momentum indicators are far superior to other indicators through empirical testing. Forecasting may be improved in most circumstances with the addition of additional technical indicators. There are fewer momentum or volatility indications that need to be considered in the short-term prediction in order to successfully predict certain patterns. To choose the best prediction approach for various feature modes, the authors thirdly present a framework for creating numerous machine prediction models. The ensemble model optimizes the parameters of six widely-used effective prediction models: RF, GBDT, LR, KNN, SVM, and LSTM. Generally speaking, the authors of the empirical investigation found that RF and GBDT performed well when it came to short-term prediction. Adding features will increase LR's prediction level. Patterns are the only ones that KNN and SVM can fit. In this case, the LSTM deep learning model's benefits are underappreciated. Lastly, the author has developed an investment plan based on the paper's predictions. Both individual stocks and portfolios can theoretically benefit economically from the approach outlined in this study, according to the empirical data. This further demonstrates the efficacy of result prediction through the use of huge data, iterative training, feature standardization, etc. Investment strategy's expected maximum drawdown, Sharpe Ratio, and Sortino Ratio outperforms purchasing and holding the original stock. But real transactions are heavily influenced by the transaction costs. To get outsized profits from real investments, further considerations are required. While this paper's forecasting framework is predictive, the stop-trading regulations of the Chinese market make it difficult to profit from specific patterns.A machine learning technique called support vector machine (SVM) isn't made for forecasting stock data on a massive scale. The author plans to update the ensemble model by including more appropriate machine learning prediction methods, like reinforcement learning approaches. In addition, the author intends to enhance future predicting outcomes by incorporating more predictive indicators, such as market mood and significant news events.

[11] The challenge of forecasting the early-stage success of startup businesses was taken up by Boris Sharchilev et al. Crunchbase is the biggest open-access startup database, and we used data from that and a crawl of web-based open sources to create a diverse and deep set of signals. In addition, the author conducted the largest trials on the topic to date and shown that our method significantly outperforms the existing state-of-the-art. WBSSP, a robust and diversified prediction pipeline, was built using a combination of many machine learning models. The author's contribution went beyond just creating a prediction model; they also included a comprehensive examination of the model and its outcomes. As one might expect, forecasts rely heavily on organized corporate data, such as category and investor data. A notable discovery from the author's work is the value of considering a company's online mentions across various websites. While these mentions may not have much weight on their own, when added together they paint a picture of how a company is perceived by its target audience and greatly enhance the accuracy of predictions. The author has acknowledged the shortcomings of earlier studies on startup success prediction, but their work also identifies areas where more development is needed. To begin, it's not enough to simply monitor where mentions of startups are coming from; future efforts should also incorporate the substance of these mention pages, perhaps through sentiment analysis. Secondly, as the amount of included mentions approaches the author's whole dataset, prediction quality does not saturate, as shown in Section 6.4's trials. Future research might apply Named Entity Recognition methods to find indirect references, such as those using firm names, since the author's study only used direct mentions in the form of links. Lastly, the author has solely taken domain-level mentions into account, meaning that distinct web pages or second-level domains inside a larger domain were deemed to be equivalent. This makes sense for smaller online resources, but major domains like reddit.com or forbes.com have a plethora of sections covering a wide range of subjects. If we can further differentiate between these parts, we might get a more nuanced signal for our predictive models.

[12] By utilizing a variety of machine learning techniques, Jiyoung Park et al. demonstrated how ESG-related parameters improve the forecasting accuracy of crowd financing success. A number of algorithms, including XGBoost, LightGBM, AdaBoost, CatBoost, and NGBoost, are employed to predict the results of crowdfunding campaigns, with a focus on the environmental, social, and governance aspects of ESG. Predicting crowdfunding success while also incorporating ESG information about the fund increased the prediction rate, according to the study. When comparing the outcomes of each ESG element, it was found that the best performance was achieved when environmental information was used to predict crowdfunding success, followed by social activity information. Lastly, the performance of the crowdfunding projection was quite poor when considering the fund's governance information. It appears that ESG plays a substantial role in the success of crowdfunding campaigns, since the prediction performance improved when ESG elements were included compared to when they were excluded. Both the theoretical and applied dimensions of this paper's contributions can be considered independently. First, this study added to the growing body of academic literature on the topic of non-financial factors influencing crowdfunding success or failure by demonstrating the significance of an ESG factor-based machine learning model for success prediction in the field. J. Park et al.: Development of a Success Prediction Model for Crowdfunding (VOLUME 12, 2024) is a significant academic contribution to the fields of crowd funding and environmental, social, and governance (ESG) research because it investigates how ESG elements can be incorporated into prediction models based on machine learning and lead to improved performance. Secondly, by presenting and investigating machine learning approaches that are not extensively utilized academically in the field of finance and accounting, this work is anticipated to add to the breadth of research in these areas. The second practical application of this work is a success prediction model that takes ESG aspects into account; this allows crowdfunding organizers and investors to make better strategic decisions. Specifically, in light of the growing interest among investors in environmental, social, and governance (ESG) factors, this article offers practitioners some concrete suggestions for how to increase the project's chances of success by highlighting its social responsibility and sustainability. The second practical takeaway from this study is the need of controlling and revealing ESG-related information correctly during crowd funding recruiting. Site administrators, project creators, and investors are the target audience for this paper's crowdfunding success prediction model, which should help them make better decisions. To start, platform operators might employ models that take ESG factors into account and employ techniques that reveal them to choose projects that have a good chance of succeeding. The success rate of the platform, its trustworthiness, and the

experience of investors and project creators will all be positively affected by this. Secondly, by incorporating ESG factors into project design and marketing methods, project creators can pique investors' interest and boost success rates. The model's ESG data is crucial, thus it's in the best interest of the project to make environmental, social, and ethical considerations prominent. Third, using predictive models that incorporate ESG data allows investors to invest in projects with a better success probability, reducing risk and increasing profitability. From a sustainable investment standpoint, it is worthwhile to consider ESG aspects as significant criterion when making investment decisions. Hence, platform operators, project makers, and investors can utilize the prediction model presented in this work as a valuable resource to make decisions based on more data. There are a few caveats to this study. The first limitation is that the study only included data from three Korean crowdfunding platforms, so it can't be applied to other types of crowdsourcing too. Strengthening resilience may be possible with a larger dataset. The second issue is the potential subjectivity in the use of dummy variables to quantify ESG factors. For more unbiased analysis, future studies could utilize advanced natural language processing techniques. Finally, it's possible that seasonal or long-term trends in crowdfunding won't be captured by statistics from December 2023. This matter could be resolved using longitudinal data. Fourth, previous research suggests that deep learning models and other variables (such as creator reputation and investor demographics) could enhance forecast accuracy. Fifth,



we corrected the arithmetic inequality between successful and failed campaigns; nevertheless, we may improve the model even more by including techniques like ensemble learning. Sexiest point number six: these results may only work on Korean platforms and not in other areas. In order to make it more applicable, global datasets are required. Future research and methodological development should focus on addressing these constraints. In order to create a more objective and generalizable forecasting model, the author plans to gather data from a wider variety of crowdfunding platforms in future studies. The author also plans to study the effects of using deeper learning or more sophisticated machine learning techniques to enhance prediction outcomes.

[13] Zeinab Shahbazi et al. conducted research on the application of Hierarchical Risk Parity (HRP) and Reinforcement Learning (RL) techniques to the analysis of cryptocurrency network risk management. When compared to other machine learning methods that have been applied to this domain, reinforcement learning yields very good performance evaluation results. The learning-based nature of RL makes it a natural fit for this process, as it allows system structures to achieve high accuracy in providing the correct information to the system. In addition, the HRP possesses the best qualities and the most coveted variety. After re-balancing the chosen period, we examined the findings using several estimating windows and approaches. By enhancing risk management, the implemented HRP provides transitional asset allocations with significant alternatives. To improve the technique's efficacy in risk management, future studies will expand its use of out-of-sample testing to more assets and classes and employ optimization approaches.

[14] The new dataset that Abhinav Nadh Thirupathi et al. examined included 3,160 distinct items, 32 of which were features at the individual level and 16 at the business level. A binary outcome—whether a firm had a liquidation event in the form of an IPO or M&A—was predicted using the collected data by training an XGBoost classifier. We used leave-one-out cross validation (LOOCV) to validate the model, and we used AUC, accuracy, precision, recall, TPR for different FPR thresholds, and a reliability diagram for statistical calibration to evaluate its performance. According to these criteria, the author's model performed admirably (0.91 AUC).

3. METHODOLOGY

Figure 1: Proposed model System Overview

In figure 1 above, we can see a system overview of the proposed methodology for investment advice. Below, we have detailed the sequential procedures to implement the proposed approach.

Step 1: Data Gathering – An interactive online tool has been developed to gather information from the startups, and here is where the provided investment suggestion approach begins. With the help of a JSP page, the java code is created to create the web application, and the Glassfish server is hosted on the local host.

Startups use this interactive website to share information about their company, including: stock market (if listed), factors (such as returns, risk, and locking period), duration, objective (capital appreciation, income, or growth), invest monitor (daily, weekly, or monthly), and expected return percentage. After collecting these attributes during setup, they are successfully stored in the database.

The website also allows investors to interact by entering details that are used to find firms that are relevant to them. Information such as the amount needed, the kind of return, and the percentage of return are supplied by the investor. In addition, the database stores these details after successfully reading them from the website.

Step 2: Preprocessing – While we await the data to be added to the database, a python thread is being started. Activating the python thread and retrieving the required data from the database follows the data being uploaded to the database. In this stage, the input dataset is also retrieved from the following URL: https://www.kaggle.com/datasets/nitindatta/finance-data. What about the startups' dataset? It's a synthetic dataset that was created with this goal in mind; it's what the investors utilize.

The dataset is in a worksheet format, and the python code uses the pandas library to read it. A list containing the extracted dataset data is subsequently preprocessed. To achieve this goal, we extract the necessary attributes from the dataset and remove the unnecessary ones.

These characteristics are essential for a startup and include things like stock market, element, objective, duration, invest, monitor, and expect. Alternatively, investor-required features include things like quantity, returntype, and return. After the user inputs their values in the previous step using the interactive web site, the next phase uses these attributes and their values to classify the data.

Step 3: K-nearest Neighbor classification – Here the user's input and the preprocessed dataset from earlier steps of the approach are received. We are estimating the distance between the user input and each row of the preprocessed dataset using this data. Use the following equation [1] to determine the distance.

ED=pP(AT i − AT j) 2 ----------[1]

Where, ED=Euclidian Distance

ATi=Attribute at index i

ATj= Attribute at index j

The complete list is effectively sorted into ascending order using the bubble sort approach once the distance from each row has been computed and attached to the corresponding rows. Two clusters are produced by setting the k parameter to 2 in this implementation. To go on to the next level of decision-making, the first cluster containing high-quality data is supplied.

Step 4: Decision Making – In this step of the decision-making process, the probability list that was created in the previous step is used as input. After sorting through the possibilities, the most appropriate suggestion conclusion is shown to the investor via the user-friendly interface.

4. RESULTS AND DISCUSSIONS

Using the NetBeans and Spyder IDE, a machine learning–based investment suggestion approach has been developed. Python and Java have been chosen as the programming languages for this method. A standard configuration in the implementation laptop is an Intel i5 CPU, 1 TB hard drive, and 8 GB of RAM. The storage needs were handled by means of the MySQL database system.

A thorough evaluation of the proposed methodology's efficacy has been conducted using this strategy. We used the precision and recall approach to look at the evaluation parameters.

Performance Evaluation based on Precision and Recall

Two very helpful metrics for gauging the exactness of a specific component's execution in our system are recall and precision. The component's precision encompasses a broad range of dependability and determines its relative precision.

The method's precision measure was calculated as the ratio of the number of trials completed to the number of correct Investment Suggestions.

In contrast, the recall requirements supplement the precision measure and help decide the Artificial Neural Network component's absolute correctness.

This approach calculates recall as the proportion of right investment

Advice on the total number of erroneous investment recommendations. The calculations that follow provide a quantitative expansion of this.

Precision and Recall can be depicted as below:
X = The number of accurate Investment Suggestions
Y= The number of inaccurate Investment suggestions
Z = The number of accurate investment suggestions not done
So, precision can be defined as
Precision = (X / (X+ Y)) *100
Recall = (X / (X+ Z)) *100

Shown below are the results of the experiments that were conducted utilizing the formula that was discussed earlier. As seen in figure, these statistical variables are used to create a visual representation.

No of Trials	Accurate Investment Suggestions(X)	Inaccurate Investment Suggestions (Y)	Accurate Investment Suggestions Not Done	Precision	Recall
1	1	0	0	100	100
4	3	0	0	100	75
7	6	1	0	85.71429	100
10	8	2	0	80	100
13	10	1	0	90.90909	83.33333
16	13	2	0	86.66667	92.85714
19	17	1	0	94.44444	94.44444
22	17	3	0	85	89.47368
25	20	3	0	86.95652	90.90909
28	24	2	2	92.30769	92.30769

Table 1: Precision and Recall Measurement Table

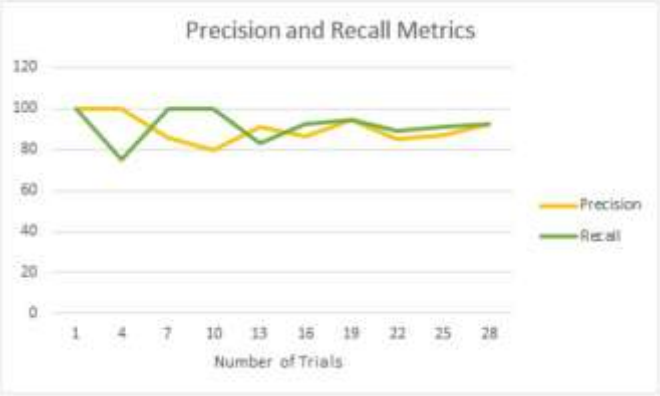


Figure 7.2: Comparison of Precision and Recall

For different numbers of trials, the graph shows how well the Artificial Network performed and what suggestions were most appropriate given the input data. The method's remarkable dependability is demonstrated by its recall rate of 91.83% and precision rate of 90.19%. These figures are noteworthy because the first time this technique was put into action, and they have yielded positive results.

5. CONCLUSION AND FUTURESCOPE

To conclude, this study put out a machine learning-based intelligent startup investment suggestion system to help investors evaluate and choose promising businesses in today's data-rich, competitive market. Biased conclusions, higher risk, and lost opportunities are common outcomes of the traditional investment decision-making procedures that depend on human judgment, limited historical insights, and manual analysis. By analyzing huge and varied statistics pertaining to startup performance, financial health, market trends, scalability, and management competence, the suggested method is able to overcome these restrictions. Better investment decisions may be made with the help of this data-driven method, which allows for a more efficient, consistent, and impartial assessment of businesses.

By successfully learning complicated and non-linear correlations among numerous startup features and investment outcomes, the system's prediction potential is greatly enhanced by including the Artificial Neural Network model. The ANN model is able to detect patterns that would otherwise be impossible to find using traditional analytical techniques because of its methodical training on past data. Consequently, investors are able to reduce risk and maximize return with the help of the system's accurate investment appropriateness scores and dependable suggestions. The system's continued usefulness in ever-changing market conditions is assured by the model's adaptability to fresh data.

In sum, the findings show that investors, VC companies, and banks can greatly benefit from a startup investment suggestion system that is based on machine learning. Better investment strategies and more effective allocation of resources are outcomes of the suggested method, which improves prediction accuracy, decreases human bias, and allows for scalable analysis. Intelligent and sustainable investment ecosystems could be greatly improved with future updates to the system, such as the ability to integrate data in real-time, use explainable AI approaches, and create tailored recommendation frameworks.

The startup investment suggestion system that utilizes

machine learning has a vast potential for growth and improvement in the future. In order to make better, more current investment suggestions, the system might use real-time data from social media trends, startup performance dashboards, and financial markets. Incorporating explainable AI techniques can enhance decision-making transparency and trust by providing investors with a better understanding of the reasons behind model projections.

In order to make predictions that are more accurate and resilient across many industries and market conditions, future progress may involve using ensemble models and sophisticated deep learning. It is also possible to implement personalized recommendation frameworks that take into account investors' risk tolerance, investing objectives, and personal preferences. This platform is an intelligent and all-inclusive investment decision-support tool, and it can be extended to accommodate startup ecosystems around the world.

REFERENCES

[1] M. Wu, H. Yang, and X. Xu, "Research Regarding the Financial Intelligent Investment Approach Grounded in the Enhanced Artificial Immune Algorithm," in ICCSIE '25: Proceedings of the 10th International Conference on Cyber Security and Information Engineering, Xining, China, 2025, pp. 126–131. [Online]. Available: https://doi.org/10.1145/3759179.3759198

[2] S. Gupta, R. Pienta, A. Tamersoy, D. H. Chau, and R. C. Basole, "Identifying Successful Investors in the Startup Ecosystem," in WWW '15 Companion: Proceedings of the 24th International Conference on World Wide Web, Florence, Italy, 2015, pp. 39–40. [Online]. Available: https://doi.org/10.1145/2740908.2742743.

[3] H. Yu and W. Lu, "Dynamic Resource Allocation for Cloud Computing Based on an Improved Genetic Algorithm," in ICCSIE '25: Proceedings of the 10th International Conference on Cyber Security and Information Engineering, Xining, China, 2025, pp. 132–137. [Online]. Available: https://doi.org/10.1145/3759179.3759199

[4] Z. Li, "An Intelligent Financial Investment Model Grounded in the Improved Genetic Algorithm," in ICCSIE '25: Proceedings of the 10th International Conference on Cyber Security and Information Engineering, Xining, China, 2025, pp. 138–143. [Online]. Available: https://doi.org/10.1145/3759179.3759200

[5] V. Shcherbak, O. Dorokhov, K. Ukrainski, D. Djakons, O. Kovalyova, and L. Dorokhova, "Business Analytics and Digitalization as Drivers of Startup Evaluation: The Experience of the Baltic States," Organizacija, vol. 58, no. 4, pp. 353–368, Nov. 2025. [Online]. Available: https://doi.org/10.2478/orga-2025-0022

[6] S. Liu, H. Zhang, Z. Yang, J. Kong, L. Zhang, and C. Gao, "UXBIV: An Evaluation Framework for Business Intelligence Visualization," IEEE Access, vol. 11, pp. 92403–92415, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3300418

[7] M. Castiglioni, A. Nuara, G. Romano, G. Spadaro, F. Trovò, and N. Gatti, "Safe Online Bid Optimization with Return on Investment and Budget Constraints," in KDD '25: Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining, vol. 1, Toronto, ON, Canada, 2025, pp. 1–11. [Online]. Available: https://doi.org/10.1145/3690624.3709288

[8] H. Li, H. Liu, A. Chen, X. Ma, Q. Liu, and J. Du, "RIA: Return on Investment Auto-scaler for Serverless Edge Functions," in ICPP '24: Proceedings of the 53rd International Conference on Parallel Processing, Gotland, Sweden, 2024, pp. 1–11. [Online]. Available: https://doi.org/10.1145/3673038.3673099

[9] C. A. Stewart, J. A. Wernert, C. M. Costa, D. Hancock, R. Knepper, and W. G. Snapp-Childs, "Return on Investment in Research Cyberinfrastructure: State of the Art," in PEARC '22: Practice and Experience in Advanced Research Computing, Boston, MA, USA, 2022, pp. 1–8. [Online]. Available: https://doi.org/10.1145/3491418.3535131

[10] Y. Lin, S. Liu, H. Yang, and H. Wu, "Stock Trend Prediction Using Candlestick Charting and Ensemble Machine Learning Techniques With a Novelty Feature Engineering Scheme," IEEE Access, vol. 9, pp. 101433–101446, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3096825

[11] B. Sharchilev, M. Roizner, A. Rumyantsev, D. Ozornin, P. Serdyukov, and M. de Rijke, "Web-based Startup Success Prediction," in CIKM '18: Proceedings of the 27th ACM International Conference on Information

and Knowledge Management, Torino, Italy, 2018, pp. 2283–2291. [Online]. Available: https://doi.org/10.1145/3269206.3272011.

[12] J. Park, H. J. Na, and H. Kim, "Development of a Success Prediction Model for Crowdfunding Based on Machine Learning Reflecting ESG Information," IEEE Access, vol. 12, pp. 197274–197289, 2024. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3519219

[13] Z. Shahbazi and Y.-C. Byun, "Machine Learning-Based Analysis of Cryptocurrency Market Financial Risk Management," IEEE Access, vol. 10, pp. 37842–37856, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3162858

[14] A. N. Thirupathi, T. Alhanai, and M. M. Ghassemi, "A Machine Learning Approach to Detect Early Signs of Startup Success," in ICAIF '21: Proceedings of the Second ACM International Conference on AI in Finance, Virtual Event, 2021, pp. 1–8. [Online]. Available: https://doi.org/10.1145/3490354.3494374