WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Intelligence for Startup Profit Prediction System

Aaryan P Shiurkar, Atharva Kankate, Pratap Joshi, Sanskar Tundurwar, Prof . Vidya Deshmukh

Department of Computer Engineering, P E S Modern college of engineering, Pune

ABSTRACT—

Machine learning techniques are used to discover hidden patterns of application-centric data analysis. Many applications can be developed to support different businesses using this technology. This model will be used by business leaders and managers to create a sustainable business, grow and solve global business problems. Against this backdrop, the use of machine learning in business analytics can be a valuable tool that can help new businesses and entrepreneurs grow and expand over time. Therefore, in this article, we provide a review of existing machine learning methods that have recently helped understand startups' needs, business trends, and can provide ideas to inform future ideas to solve business problems. Second, based on the analysis, we plan a future-proof way to create and develop information systems that will support new entrepreneurs in India.

Keywords—Machine learning, recommendation system, business data analytics, algorithm design, support system, handholding of startups.

I. INTRODUCTION

Startup India is the Government of India's initiative to create a sustainable system that will facilitate the growth of new companies, promote sustainable financial development and increase capacity. Through this initiative, government officials plan to attract new businesses to increase prosperity. Many projects have been tested since the Prime Minister announced it in January 2016 [1]. Thanks to this program, many new ventures start their business life without any business background or experience. But not all businesses are successful and some new ventures fail to survive due to various internal and external business and business challenges. On the other hand, with the emergence of Industry 4.0, the economy is also affected by the digitalization of the world, and the manufacturing industry is also in the development phase of the transition from the physical framework to the digital framework. Industry 4.0 is the new achievement of computing and information printing. It combines digital technology, the Internet of Things and cloud computing. Changing the assembly area by keeping it to very high standards is very important to prepare us for the future. To seize this opportunity, FICCI has launched an Industry 4.0 Board. The committee is intended to focus on this new trend and support the government in creating business guidelines [2].

Therefore, in this article, we propose a smart strategy using machine learning (ML) technology to support startups and entrepreneurs in solving the next generation of business technology competition. The aim is to support and advise startups to help them develop a successful personal business. Machine learning (ML) can process large amounts of business data and retrieve many models. ML is a process that uses mathematical algorithms to discover underlying patterns from raw data [3].

Collections of algorithms are used to explore and understand the relationship between two or more features that focus on work. Therefore, we may share, distribute and/or make predictions using information we collect or create using this technology. This process includes classification, clustering, association rule mining, etc. It may include different types of algorithms such as Techniques or algorithms are applied to data based on the use of data models. In this section, we provide a brief overview of the recommended AI products.

II. LITERATURE REVIEW

The article presents different recently developed methods to predict startup success and provide a framework to support and grow new businesses using machine learning techniques. With all the company's data, forward-thinking models can be developed and validated as a critical measure of the true model. K. Å bikowski et al. [4] Uses data from Crunchbase. The important thing is to build a good model based on machine learning to predict the success of the organization. They organize the review in a way that prevents information from leaking at the last minute. Developers analyze logistic regression (LR), support vector machines (SVM), and gradient boosting (GB) classifiers. The model achieved good results in terms of accuracy, analysis, and F1 scores (57%, 34%, and 43%). The best results were obtained with the GB classifier. They provide information about what is important. Sharchilevet al.

They stated that the task is to predict whether the organization that has already started granting will participate in another grant. They use a rich combination of indicators to evaluate the model's predictive performance and potential for electronic access points. Pre-prepared documents for the initialization of biological systems, containing information from work and planning, personal distance communication and networks. Using these, they

trained a machine learning tube with various underlying models. They suggest that using the organization's reports on the web can create rich presentations. They also examined the incoming samples. A. Krishna et al [6] plan to create a forward-thinking model for a new business based on various requirements at different levels. The speed of success in measuring new business is attractive. Seed revenue, time, subsidies, etc. They proposed a strategy to predict new business outcomes based on factors. Crunchbase, Tech Crunch etc. They make some models based on data from sources. Many data mining techniques have been used, including optimization and application from previous data. They provide analysis using Random Forest (RF), ADTree and Bayesian Networks (BN). They measured the model's accuracy based on the area under the ROC curve, precision, and recall .

H. Janáková et al. They also record current and future events. Work on the latest AeroMobile and eSense. Developers have focused on the advancement of SR since its initial inception. They prepare the necessary opportunities for new companies. The development of new IT businesses is strictly business-centric, while the regional tone is local. A new business is created that is financed after the sale to financial support. Evaluation of results is a matter set for evaluation. The goal is to review and explain how to evaluate and re-evaluate the estimation process using data and writing to learn and evaluate skill. They identify and legitimize the owner. A machine learning model called CapitalVX is used to predict the outcome of new ventures; for example, will they go public, be acquired, go neutral, or remain private. Usage of various features, output results and future contributions are predicted with 80-89% accuracy. They say VC/PE firms can choose to leverage machine learning to analyze predictions using valid data. Kaiser et al. They think about patience, business development, return on investment, new patents and donations. The main points they consider include credentials, location, property and job. Add address company name, business description, launcher and startup properties to different elements. To make an estimate, we simply multiply the number of items in the BPS table. accurate measurement of economic development requires a combination of BPS and organizations . All predictions need to be accurate and precise as basic information is required. Crowdfunding offers another way to connect new companies with many financial backers. S. Zhang et al. Track organizations that successfully raise money over a 7-month period and document their community commitments. Through various measures of peer engagement analysis shows that regular promises from online advertising are associated with more funding. Sometimes the level of commitment is higher in successful organizations.

In addition to tree selection, SVM, KNN, etc. are used to predict the organization's ability to improve its financial performance. In addition, despite class irregularity, the model can predict performance with 84% accuracy. The company receives funds from financial backers. Founders of 40 new companies are discussed to identify differences involving the importance of forecasting. Scholars consider the policy and financial reforms of the two collectives from an economic perspective. Results show some differences between organizations' and funders' views on progress important .

It is difficult to invest in early stage companies when there is no information. Venture capitalists often rely on their instincts when making decisions. F. Corea et al. They conduct research and develop with the analysis of more than 600,000 companies. The framework is a smart list of 21 features that will help investors. Machine learning can close this gap. This process is made possible thanks to the information provided by thousands of companies from Crunchbase. Previous efforts have focused on predicting two categories, purchasing or distribution.

J. Arroyo and colleagues [13] will attempt to predict possible outcomes, including future financing or closure. This approach will lead venture capitalists to achieve lower risk and higher returns. They will analyze the effectiveness of various machine learning methods on data from more than 120,000 companies and try to predict their success. T. Antretter et al. They used machine learning techniques to analyze more than 187,000 tweets from 253 business accounts. The results showed that they were able to accurately identify up to 76% of failed businesses. They contribute to the debate about the importance of machine learning in business research. McKenzie et al. I) positive scores are not associated with energy, performance, activity or health; iii) Machine learning techniques are not developed today; Guerzoni et al.

As an experiment, machine learning allows another part of development to take place, continuing Italian policies to make it easier to start new companies. They study the impact of thought on the population in general. First, they examined seven auditors of Chinese education to show creative companies according to 2013 company statistics. Second, use the latest 2008 data set and predict which companies will be labeled as innovators. Finally, this new indicator is considered as a regressor in the resilience model that reveals the company's capability.

S.P. Jun et al. First, they analyze the existing language and help R&D partners work with SMEs. Machine learning and discriminant analysis are then used to identify companies that will achieve success through four R&D partners. Finally, they run tests using new information forgotten by model changes. Therefore, the accuracy of the proposed model is as high as 91%. Using this model, companies are expected to choose the type of R&D partner more effectively and increase success.

J. C. Kaminski et al. Semantic style in crowdfunding efforts Prospects that spark interest or clarify understanding are a good indicator of role fulfillment. A negative development in the product may reduce the evaluation of the new product and reduce the purchase target. These findings suggest that positive psychological messages are important when goals are rare and risk-based, such as pleasure. Using happy quotes or sharing something in real life will attract people's attention.

J. R. Saura et al.'s [19] idea is to diversify Twitter content to create new companies and new businesses. They suggest identifying important elements. First of all, Latent Dirichlet Allocation (LDA) is used, which determines the storage content. Additionally, sensitivity analysis was performed using SVM to classify threshold values as negative, positive, and irrelevant. Third, analysis of the text was carried out using text search techniques. They believe that the factors that have a positive impact on the success of a new business are innovation, individual creators and processes. Negative content patterns and expressions are a prerequisite for types of work and holy missionaries. Controversial topics are Optimization of field trial strategies, releases and Hatcheries and geographical regions. The limits are the number of tweets to check and the time limit..

III. PRELIMINARIES

The business environment is important for the entrepreneur. The business is in the best possible position; T. Bilen et al. The system collects job characteristics and learns from future data. It estimates the features and shows the cluster in the well-located region. The system is evaluated against a reference that predicts the restaurant's best location. The first stage of decision making is to estimate the mentioned features. They determine the best regression model for each feature. In the second stage, a hierarchical tree is used to group regions based on these predictions. It is interesting to note that while writing about new companies is useful, analyzing most promotion decisions is generally not. M. Cantamessa et al.

Another related undertaking is to complete the research on large amounts of data. The agreement shows that the main reason for the initial disappointment was the lack of economic development. Jealousy of this organization affects the development of his business. I. Afolabi et al. The goal is to propose a plan that identifies, predicts and confirms. Their predictions are based on analysis of data connections and connection between Naive Bayes and J48.

The heuristic for this finding is drawn from a review of existing counseling methods, practical strategies, and humanities experts. The framework will work on the speed of business execution and enable the business development phase. M. Bauer et al. They found that machine learning is being adopted by businesses, but small and medium-sized businesses are struggling compared to large businesses. To identify effective and efficient employees, the authors conducted a qualitative study. The results show that SMEs do not use machine learning technology due to insufficient machine knowledge. Location and Estimation of Corporate Defaults and Chapter 11, Basic Operations. These questions are important for financial managers to articulate their expectations, given the results of forecast models.

IV. SYSTEM MODEL

1. Data collection: Collect data on financial initiatives, business, business trends, and other changes that may affect results.

2. Preliminary data: Clean and process preliminary data to resolve missing values and anomalies and ensure that its type is suitable for machine learning algorithms.

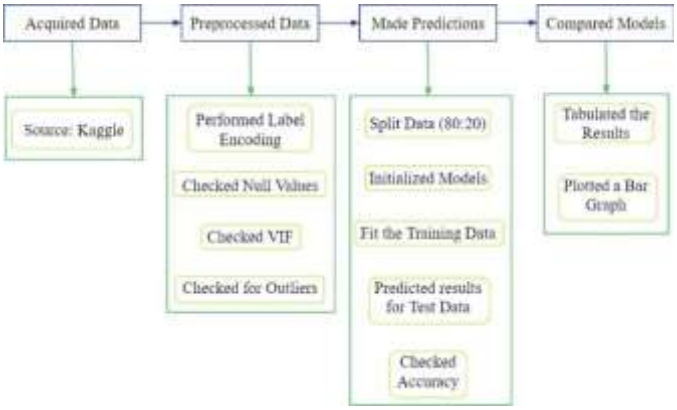
3. Feature Engineering: Identify key features that may affect predictive results and design new features as needed to improve model performance.

4. **Model Selection**: Explore a variety of machine learning algorithms such as linear regression, decision trees, random forests, or neural networks to determine the best model for your prediction.

5. Model training: Split the data into training and testing, train the selected model on the training data and use the test data to evaluate its performance.

6. Model evaluation and improvement: Evaluate the model's accuracy, precision, and recall to correct its flaws and improve its predictive capabilities.

7. Export: Once you have a successful prototype, send it to the manufacturer to generate instant profits for your startup.



V. METHODLOGY

1.Problem definition: Clarify the problem that needs to be solved based on the betting results. Determine what factors impact new business profitability and how machine intelligence can help predict future profits.

2.Data collection: Collect important data such as financial data, customer data, business models and other data that may be relevant to benefits.

3.Data processing: Clean data for analysis by handling missing values and errors and creating appropriate data.

4.Feature Engineering: Identify key features that may affect predictive results and design new features as needed to improve model accuracy.

5.Model selection: Choose the appropriate learning machine such as linear regression, decision trees, random forests or neural networks according to the nature of the data and the problem at hand.

6. Model training: Split the data into training set and test set, train the selected model on the training data and adjust the parameters for a good run.

7.Model evaluation: Evaluate your model's performance using metrics such as mean square error (MSE), root mean square error (RMSE), or undermean error (MAE) to evaluate its accuracy in predicting results.

8.Deployment: Deploying the training model to the production environment can provide rapid results for startups.

9.Monitoring and control: Regularly monitor your model's performance, regularly retraining it with new data and updating it as necessary to ensure its predictions remain timely.

VI. SECURITY ANALYSIS

1.Data Security: Use strong data encryption technology to protect important data such as financial information and customer details. Use secure data storage and access control procedures to prevent unauthorized access.

2.Model security: Protect machine learning models from attacks using techniques such as model encryption, input sanitization, and monitoring for anomalous behavior that may indicate security.

3.System architecture: Create a secure system architecture, isolate components, use secure communication protocols, and restrict user access responsibilities and permissions.

4.Regular audits: Perform regular security audits and vulnerability assessments to identify and resolve system security issues.

5.User Authentication and Authorization:** Use strong user authentication methods such as multi-factor authentication and role control to ensure that only authorized users have access to the system.

6.Monitoring and logging: Create monitoring tools to monitor system activities, identify anomalies, and create logs for audits. Monitor performance and security measures to identify and resolve security issues.

7.Emergency response plan: Create an emergency response plan that outlines the procedures to be followed if a security breach occurs. Make sure all team members are trained on how to properly respond to a security incident.

8.Compliance: Ensure revenue forecasting systems comply with data protection laws such as GDPR, HIPAA, or specific business standards to protect data privacy and security.

VII. PERFORMANCE ANALYSIS

The performance analysis of the machine learning model is crucial to understanding its accuracy, reliability, and overall effectiveness. For the Startup Profit Prediction System, we conducted a comprehensive evaluation using various metrics and techniques.

This section discusses the key aspects of the performance analysis, including the dataset, evaluation metrics, cross-validation results, and comparative analysis of different algorithms.

Evaluation Metrics

To assess the performance of the machine learning models, we used several evaluation metrics:

1.Mean Absolute Error (MAE): Measures the average magnitude of errors in predictions, without considering their direction.

2.Mean Squared Error (MSE): Provides a measure of the average squared difference between actual and predicted values.

3.Root Mean Squared Error (RMSE): The square root of MSE, providing error magnitude in the same unit as the predicted values.

4.R-squared (RÂ²): Indicates the proportion of variance in the dependent variable that is predictable from the independent variables.

5.Adjusted R-squared: Adjusts the RÂ² value based on the number of predictors in the model, providing a more accurate measure in multiple regression.

CONCLUSIONS

Machine learning is used to analyze various business processes and other relevant factors. This improves business efficiency, manages large amounts of information, reduces work and errors, increases productivity and growth. In this context, the sustainable development of new businesses is a major challenge for the world. This topic has started to attract attention recently due to new business challenges and issues (e.g. covid-19). To reduce risk and understand the key points of successful trading, in this article we plan to review the latest studies and collaborations on this subject. Understanding and predicting the success of a new business can depend on many factors. It is important for investors and founders to consider, risk assess and plan for seed financing. But starting to forecast can reduce investors' risk and increase future profits. On the other hand, using machine learning methods to predict

patterns of successful startups is not only beneficial for investors, but may also help address high inequality in the new economy as a development proposition.

By using different characteristics and measurements of success, we can not only predict success or failure, but also provide startups with recommendations for their own development. In addition to determining success or failure, different data can also be used. This will depend on the type and scope of publication. Finally, we can see that there are many supervised and unsupervised learning methods used to create prediction and recommendation models. Many of these ideas are influenced by regression analysis and classical machine learning algorithms.

In some cases, educational monitoring and NLP-based techniques can also help determine the success and failure of online initiatives. Finally, by reflecting on the facts and methods, we arrive at some important facts that are useful for our future research, which will be presented in the following section.

We also provide future plans for research requests by aggregating content to identify and recommend important features that increase the sustainability of plans of a new nature.

REFERENCES

[1] https://www.startupindia.gov.in/

[2] https://ficci.in/

[3] Pujari, Arun K., "Data Mining Techniques", Universities press, 2001.

[4] K. Żbikowski, P. Antosiuk, "A machine learning, bias-free approach for predicting business success using Crunch base data", Information Processing and Management 58 (2021) 102555

[5] B. Sharchilev, M. Roizner, A. Rumyantsev, D. Ozornin, P. Serdyukov, M. de Rijke, "Web-based Startup Success Prediction", CIKM 2018, October 22–26, Lingotto, Turin, Italy © Copyright held by the owner/author(s). ACM ISBN 123-4567-24-567/08/06

[6] A. Krishna, A. Agrawal, A. Choudhary, "Predicting the Outcome of Startups: Less Failure, More Success", IEEE 16th International Conference on Data Mining Workshops, 23759259/16© 2016 IEEE

[7] H. Janáková, "The Success Prediction of the Technological Start–up Projects in Slovak Conditions", Procedia Economics and Finance 34 (2015) 73 – 80

[8] G. Ross, S. Das, D. Sciro, H. Raza, "CapitalVX: A machine learning model for startup selection and exit prediction", The Journal of Finance and Data Science 7 (2021) 94-114

[9] U. Kaiser, J. M. Kuhn, "The value of publicly available, textual and non-textual information for startup performance prediction", Journal of Business Venturing Insights 14 (2020) e00179

[10] Q. Zhang, T. Ye, M. Essaidi, S. Agarwal, V. Liu, B. T. Loo, "Predicting Startup Crowd-funding Success through Longitudinal Social Engagement Analysis", CIKM'17, November 6–10, 2017, Singapore, ACM ISBN 978-1-4503-4918-5/17/11

[11] A. Prohorovs, J. Bistrova, D. Ten, "Startup Success Factors in the Capital Attraction Stage: Founders' Perspective", Journal of East-West Business, DOI: 10.1080/10669868.2018.1503211

[12] F. Corea, G. Bertinetti, E. M. Cervellati, "Hacking the venture industry: An Early-stage Startups Investment framework for data-driven investors", Machine Learning with Applications 5 (2021) 100062

[13] J. Arroyo, F. Corea, G. J. Diaz, J. A. Recio-Garica, "Assessment of Machine Learning Performance for Decision Support in Venture Capital Investments", Volume 7, 2019, IEEE

[14] T. Antretter, I. Blohm, D. Grichnik, J. Wincent, "Predicting new venture survival: A Twitter-based machine learning approach to measuring online legitimacy", Journal of Business Venturing Insights 11 (2019) e00109

[15] D. McKenzie, D. Sansone, "Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria", Journal of Development Economics 141 (2019) 102369

[16] M. Guerzoni, C. R. Nava, M. Nuccio, "Start-ups survival through a crisis Combining machine learning with econometrics to measure innovation", Economics of Innovation and New Technology, DOI: 10.1080/10438599.2020.1769810.

[17] S. P. Jun, H. S. Yoo, J. Hwang, "A hybrid recommendation model for successful R&D collaboration: Mixing machine learning and discriminant analysis", Technological Forecasting & Social Change 170 (2021) 120871

[18] J. C. Kaminski, C. Hopp, "Predicting outcomes in crowdfunding campaigns with textual, visual, and linguistic signals", Small Bus Econ (2020) 55:627–649

[19] J. R. Saura, P. P. Sanchez, A. Grilo, "Detecting Indicators for Startup Business Success: Sentiment Analysis Using Text Data Mining", Sustainability 2019, 11, 917; doi:10.3390/su11030917

[20] T. Bilen, M. E. Ozcevik, Y. Yaslan, S. F. Oktug, "A Smart City Application: Business Location Estimator using Machine Learning Techniques", 978-1-5386-6614-2/18/$31.00 ©2018 IEEE, DOI 10.1109/HPCC/SmartCity/DSS.2018.00219