JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2385 - 2390
https://publishoa.com
ISSN: 1309-3452

Expert Machine Learning Algorithms to Predict Start-Up Unicorns

Dr. Abdul Rasool MD[1], Soban Al Haque[2], Obaid Khan[3], Mohammed Muneebuddin Haqqani[4], MohdNaseemuddin[5]

[1] *Associate Professor, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.*
[2, 3, 4, 5] *Research Scholar, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.*

Email : [1] *rasool.501@gmail.com*

ABSTRACT

Human decision-making will soon be surpassed by artificial intelligence, which is now a hot issue in research. There is still a disagreement as to whether humans or machines are better at complicated and creative tasks like innovation. [2] A machine's failures can be divided into two categories: processing and interpreting "soft" data (information which can be quantified) and predicting the future in situations of "unknowable risk" (severe uncertainty). When this occurs, the computer lacks representative data for a specific outcome. There is still a need for humans to use their intuition to analyse "soft" information. [5] Thus, we use a Hybrid Intelligence approach to combine the complimentary qualities of humans and robots in ability to predict the future of businesses. To achieve this goal, we adopt a design science research methodology to construct a Mixed Intelligent program that uses the strengths of both computer and intellectual capacity to illustrate its applicability for predictions under severe uncertainty.

Index Terms— Hybrid intelligence is a term that refers to a combination of both human and machine intelligence.

I. INTRODUCTION

In the near future, artificial intelligence will be able to make better, faster, and cheaper administrative decisions than humans. Unstructured data can be processed reliably by machines to find patterns and predicting future happenings (e.g. Agrawal and Dhar 2014; Baesens et al. 2016). [1] However, in more sophisticated and creative circumstances such as innovation and entrepreneurship, the question of whether machines are superior to humans remains open.. "Soft" information (information that can't be quantified) and "unknown risk" scenarios require intuitive decision-making are the two categories of situations in which machines fail, according to Petersen (2004) (Petersen 2004). When this occurs, the computer overfits on input samples at the expense of a student's real performance since it lacks representative information for a specific outcome (Attenberg et al. 2015) [10]

The prediction of early-stage business success is an example where both "soft" information signals and "unknown risk" are critical. Investors like angels, who put their own money into startups, must weigh the pros and cons of a firm before committing their resources. Angel investors frequently make decisions before a new product's viability or market existence has been established (Maxwell et al. 2011). In these situations, angel investors are unable to accurately judge a startup's quality and consequently anticipate its continued prospects (Dutta and Folta 2016). Furthermore, in situations of great uncertainty, such knowledge may simply not exist, making the outcome unknown. Since no markets or new technologies exist yet, projections [1] are made about their potential financial returns even while their practicality is unclear (Alvarez and Barney 2007). Recognizing companies that are both inventive and disruptive while also yielding enormous returns on investors' investments is so critical.

The "gold standard" for processing "soft signals" such as originality, innovativeness, etc., that cannot be easily quantified into models is still people (Baer and McKool 2014), who employ an affective judgement tool to discern patterns in prior decisions: intuition (Huang and Pearce 2015). In the face of tremendous ambiguity, intuition proved to be a useful tool. Bounded rationality impairs the forecasts of individual human judges, however, highlighting the fact that instead of optimising every judgement, people tend to rely on heuristics (i.e. mental shortcuts) and hence focus on readily available information (Simon 1955; Kahneman 2011). Unfortunately, this can lead to inaccurate interpretations and, as a result, potentially dangerous forecasts (Busenitz and Barney 1997). Use of "collective intelligence" through crowdsourcing is a valuable answer to this problem, which may be found in human computation research (e.g. Brynjolfsson et al. 2016; Larrick et al. 2011; van Bruggen et al. 2010). [2] The ability of humans to provide subjective evaluation of characteristics that are difficult to quantify objectively through machines (e.g., innovativeness) or use their past domain-specific expertise to make intuitive decisions is a suitable technique to exploit the benefits of humans in prediction tasks (Blattberg and Hoch 1990). Individual human decision makers' statistical errors are eliminated by the accumulation of information and the predictions that result (Larrick et al. 2011) [4] Weargue that a Hybrid Intellectual ability approach combining the additional advantages of humans and machines allows for predictions in conditions of extreme unpredictability such as the terms of early startup success through the application of formal analysis of "hard"

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2385 - 2390
https://publishoa.com
ISSN: 1309-3452

information and intuitive decision-making processing, as well as "soft" information.

Predicting the likelihood of early-stage startup success is the goal of this study. A design science approach (Hevner 2007; Gregor and Hevner 2013) is used to develop a Hybrid Intelligence methodology, which incorporates both machine intelligence (machine learning techniques) and collective intelligence (collective intelligence), which utilises the intuitions and creative potentials of individuals while limiting methodological flaws through statistical methods (e.g. (Shmueli and Koppius 2011). [5] This is why we want to demonstrate that a hybrid strategy outperforms machine or human-only methods when it comes to predicting startup success in highly uncertain environments. For the next phase of our research, we will conduct an experiment to test our suggested strategy for forecasting company success under severe uncertainty.

Using prior research and domain knowledge, we first created a taxonomy of indicators that could serve as possible predictors of early-stage business success (Shmueli and Koppius, 2011).[7]

These predictors were then used as input for machine learning algorithms and collective intelligence, which separately assesses the chance of success and then combines these results to a composite estimate places. We subsequently devised this method." In addition, we offer a glimpse into what our research will look like in the future.

This research will thus have a significant impact on a number of vital areas of information systems and management studies. Before we can model startup success forecasts, we need a taxonomy containing possible determinants that can be generalised. [3] Using a Hybrid Intelligence Method, this research contributes to the literature on predictive research in information systems and data analytics (e.g. Chen and colleagues 2012) with the introduction of a new method for predicting uncertain outcomes under limited knowledge and unknowable risk.[2-7] Formal examination of "hard" data can be combined with intuitive predictions based on "soft" data using this approach. So our research will provide prescriptive information in this regard (Gregor and Jones 2007). A final benefit of our work is that it adds to past work on collective intelligence, such as that done by [7-9] Malone et al. in 2009 and Wooley and colleagues in 2010. Last but not least, we'll present a workable solution that gives angel investors something to rely on when making investment decisions.

II. RELATED WORK

Cognitive decision-making processes can be examined to provide insight into how people make predictions in uncertain situations. Among the theories that can be helpful in this situation is the dual theoretical approach of decision making. It is assumed that people use two cognitive modes, one characterised by intuition (system 1), and one defined by explicit analytical predictions (system 2). (Tversky and Kahneman 1983; Kahneman 2011). [3]

It is highly difficult to predict the performance of early stage initiatives because there are often only hazy ideas, prototypes have not yet been developed, and so the proof of concept [1] has yet to be established. Furthermore, such concepts may not have a current market, but they have enormous growth growth in the field (Alvarez and Barney 2007). Since the probable outcomes and probabilities are unknown, the decision-making context is highly unclear. Two ideas explain this fact: asymmetry of information and unknown risk (Alvarez and Barney 2007; Huang and Pearce 2015) [4]

Situations in which forecasters are given incomplete information are known as information asymmetry (Spence 1974). It is common for decision makers to look for numerous indicators that indicate the likelihood of future events in the absence of perfect knowledge (Morris 1987). There are both hard (e.g. industry, technology) and soft (e.g. team size) types of signals that can be easily quantified and categorised in our setting (e.g. Innovativeness, personality of entrepreneur). [6] People then use formal analysis to collect signals that help them make intentional, rule-based system 2 judgments. 2 (Kahneman 2011). A decision maker can't acquire evidence that indicates a possible outcome or make conclusions depending on rigorous analysis because they don't exist in the case of unknowable risk. Statistical Bayesian model error term might be the most appropriate comparison. The term "unknown risk" refers to events that deviate from the established order (Kaplan and March 1988). A unicorn start-up in our setting, for example, would be a great example of what we're talking about here. Because of the lack of typical cases, formal analyses are not applicable in these situations. [9] The decision-making process is largely focused on intuition (system 1) instead of formal analysis in instances when humans "don't know what they don't know" (Tversky and Kahneman 1983; Huang and Pearce 2015). It's difficult to gauge the future performance of businesses in their early stages, making misclassification costly in terms of bad investment decisions or missed opportunities for profit (Attenberg et al. 2015). In the case of early stage initiatives, previous research strongly suggests that combining analytic (system 2) and instinctive (system 1) predictions delivers the best accuracy in terms of performance (Huang and Pearce 2015).

III. METHODOLOGY

Economic growth is greatly aided by the creation of new businesses. The new ideas, innovations, and jobs they generate help to move the economy forward. There's been a dramatic increase in the number of start-ups in recent years. By accurately predicting a startup's chances of success, investors can uncover companies with the potential for fast expansion and gain an advantage over their competitors.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2385 - 2390
https://publishoa.com
ISSN: 1309-3452

The project's goal is to predict if a current startup will be a positive or a negative. An IPO (Initial Tender Offer) or M&A (Joint venture company and Acquisition) is considered a sign of a company's success if the company's founders get a big sum of money (Initial Public Offering). If a business would have to be shut down, it would be deemed a failure.
We employed powerful machine learning techniques including Gradient Boosting, SVM, Random Forest, and Decision Tree to construct this project and predict the startup's success or failure [8]. Data from the STARTUP dataset is used to train ML algorithms, which may subsequently have been used to predict the successes and failures of future Start-up Test Data.

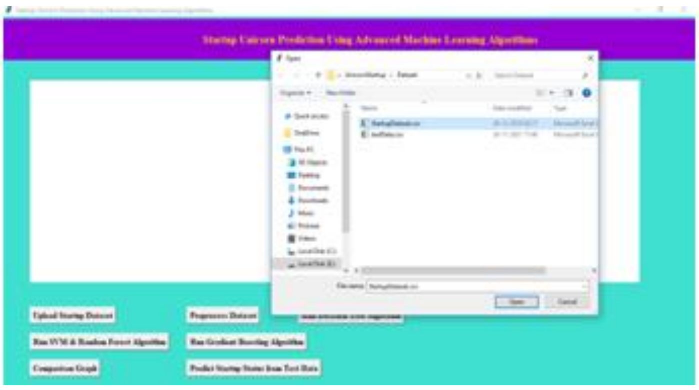
The planning and control are required for project implementation:

Upload Startup Dataset: This module will be used to upload datasets to the application.
Preprocess Dataset: We'll use this module to clean up the data and then split it into train and test datasets, with 80% of the dataset records being used to train the ML algorithms and 20% being used to forecast and calculate accuracy.
Run Decision Tree Algorithm: training a decision tree algorithm with this module
Run SVM & Random Forest Algorithm: We will train Random Forest as well as Svm classifiers using this module.
Run Gradient Boosting Algorithm: The Gradient Boosting method will be trained in this module
Comparison Graph: It is through this module that we will be able to compare the accuracy, precision, recall, and FSCORE performance of various algorithms.
Predict Startup Status from Test Data: the ML algorithms that analyse testing data and forecast the success or failure of a company or startup can be used in this module.

IV. RESULT AND DISCUSSION

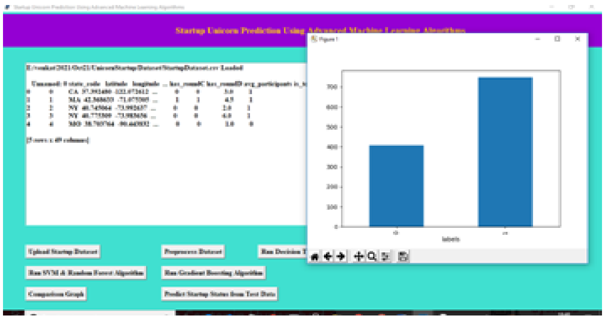
To run the project to get below result

To upload a start-up dataset, click the 'Upload Start-up Dataset' button in the above result.



Selecting and uploading the 'StartupDataset.csv' file in the above result and then clicking on the "Open" button to load the dataset produced the following result.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2385 - 2390
https://publishoa.com
ISSN: 1309-3452



There are a few dataset values shown in the above screen, and I've plotted a graph showing the success or failure of each label on the x-axis and the number of records under each y-axis in the dataset. Simply click the "Preprocess Dataset" option above to eliminate any missing data and then partition the set of data into training and testing sets sections as described previously.

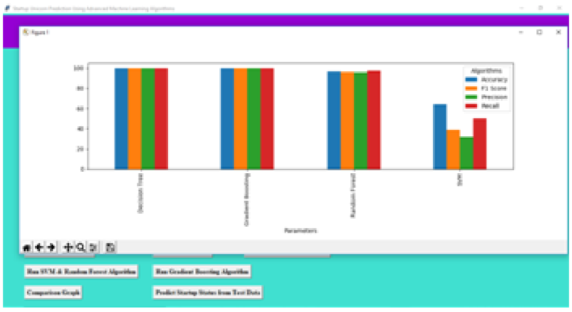
There are 1154 records altogether in the dataset, with 32 features/columns in each record. The programme splits the dataset into 80 percent (923 records) for training, with the remaining 231 records being used to test the correctness of the algorithms learned on those 923 data points. When you've completed the training and testing of your Decision Tree, click the 'Run Decision Tree Algorithm' button to get the results shown below.

Here's what happens when you click on the "Run SVM & Random Forest Algorithms" button after you get the above-mentioned result with decision tree:

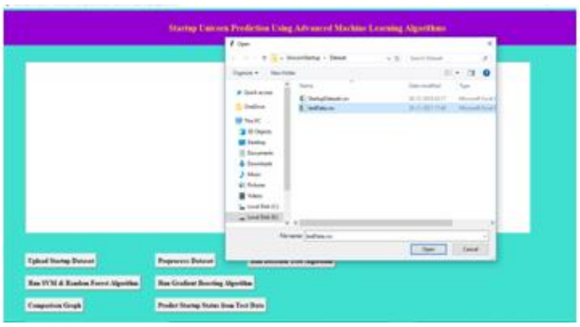
JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2385 - 2390
https://publishoa.com
ISSN: 1309-3452

As shown in the results above, random forest achieved a precision of 96%, while SVM achieved a precision of 64%. Clicking on the 'Run Gradient Boosting Algorithm' button will train gradient boosting on the given dataset and produce the following output:

Also, using Gradient Boosting, the aforementioned result has 100 percent accuracy, and clicking on the 'Comparison Graph' option will bring up the graph below.



Accuracy, precision, recall, and FSCORE are all represented as axes on the y-axis in the graph above, while other metrics for each method are represented as coloured bars. We can see in the graph above that Decision Tree and Gradient Boosting outperform the competition. Once you've closed the above graph, click the 'Predict Startup Progress from Test Data' tab to submit testing data and obtain the output shown below.



Predictions are generated by first loading 'testData.csv' from a file in the upper-right corner of the screen and then clicking the "Open" tab.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2385 - 2390
https://publishoa.com
ISSN: 1309-3452

Predictions are made based on the test values shown in the square brackets on the upper screen. A FAILURE or SUCCESS forecast is shown after each test record's ==== arrow sign.

V. CONCLUSION

Misclassification of early-stage start-ups is costly, as it could lead to poor funding decisions and the omission of worthwhile opportunities for return. Machine and general intelligence should be combined in order to create predictions in these kinds of situations. Even if machines excel at analysing enormous amounts of "hard" signals that predict the effectiveness of a product business, people are better at reading "soft" signals like the temperament of an entrepreneurial or the organizational innovation of a new device. [7] It is also possible to identify profitable start-ups that could be found by depending on existing data, thanks to human intuition. Therefore, we propose that collective intelligence be used to transcend the limitations of individual bounded rationality. Toward that end, we created a preliminary Mixed Intelligence technique, which we will test out before beginning our actual investigation. Further testing will be done in the framework of companies (e.g. growth, survival rate) and other areas of great uncertainty to see if it can be applied (e.g. innovation in general). It is also a goal of ours to see if people are more likely to accept advise from human sources when it is accurate and transparent (e.g., nkal and colleagues 2009). We hope that our work will have a significant impact on academics and practitioners alike. As a starting point, we present [1] a taxonomy of possible determinants that can be generalised to modelling startup success forecasts (e.g. Böhm et al. 2017). Using a Hybrid Intelligence method, this research work is focused on predictive research in information systems and data analytics (e.g. Chen and colleagues 2012) by providing a new way for predicting elements of uncertainty under limited knowledge and unknowable risk. This approach enables the integration of "hard" information analysis with "soft" information predictions. Such a hybrid strategy may be useful in other situations where there is a high degree of uncertainty. [5-10] As a result, we expect that our findings will provide guidance that may be applied to a variety of data science applications (Gregor and Jones 2007). To round out our work on collective intelligence, we propose new uses for machines and crowds that build on past work (e.g. Malone et al. 2009; Wooley et al. 2010). We believe that our proposed method can increase the overall intelligence of the collective. However, future study may look into how machine intelligence may be used as commentary for the crowd and so hint toward more collaborative interactive ways. This paper uses a parallel methodology (e.g. Calma et al. 2016). A realistic prediction challenge that may help angel investors make decisions and decrease the frequency of incorrect investment decisions is finally addressed in this paper.

REFERENCES

1. Afuah, A., and Tucci, C. L. 2003. "A model of the Internet as creative destroyer," IEEE Transactions on Engineering Management (50:4), pp. 395–402.
2. Shepherd, D. A., and Zacharakis, A. 1999. "Conjoint analysis: A new methodological approach for researching the decision policies of venture capitalists," Venture Capital: An International Journal of Entrepreneurial Finance (1:3), pp. 197–217.
3. Shmueli, G., and Koppius, O. R. 2011. "Predictive analytics in information systems research," MIS Quarterly, pp. 553–572.
4. Silver, N. 2012. The signal and the noise: why so many predictions fail--but some don't: Penguin.
5. Simon, H. A. 1955. "A behavioral model of rational choice," The Quarterly Journal of Economics (69:1), pp. 99–118.
6. Simon, H. A. 1982. Models of bounded rationality: Empirically grounded economic reason: MIT Press.
7. Song, M., Podoynitsyna, K., van der Bij, H., and Im Halman, J. 2008. "Success factors in new ventures: A meta-analysis," Journal of Product Innovation Management (25:1), pp. 7–27.
8. Spence, M. 1974. "Competitive and optimal responses to signals: An analysis of efficiency and distribution," Journal of Economic theory (7:3), pp. 296–332.
9. Spiegel, O., Abbassi, P., Zylka, M. P., Schlagwein, D., Fischbach, K., and Schoder, D. 2015. "Business model development, founders' social capital and the success of early stage internet start-ups: a mixed-method study," Information Systems Journal (26:5), pp. 421–449.
10. Sudek, R. 2006. "Angel investment criteria," Journal of Small Business Strategy (17:2), p. 89.