***Bridge damage prediction post earthquake using machine learning algorithms***

# Yash Garg

B. Tech CSE

ASET, Amity University

Noida, India

[ygarg704@gmail.com](mailto:skhiearth@gmail.com)

# Arpit Masih

B. Tech CSE

ASET, Amity University

Noida, India

[glitcheritzu@gm](mailto:naman.mohan09@outlook.com)ail.com

***Abstract*— The Indian Government has been promoting entrepreneurship on a nation-wide scale for many years, yet a majority of the Indian youth doesn’t prefer to start their venture. Our objective is to predict the cause behind the lack of Entrepreneurial Competency in university students and suggest potential measures to improve the same. We performed an analysis to identify a correlation between the different personality traits associated with Entrepreneurship and also cluster students into different groups and extract information from this analysis using data collected from 198 university students from across India. We have used several Machine Learning algorithms like k-NN, Logistic Regression, Naïve Bayes, Support Vector Machine, Decision Trees, Random Forests, and K-Means Clustering.**

***Keywords—Entrepreneurship, Machine Learning, Clustering Algorithms, Classification Algorithms, Personality Traits***

1. INTRODUCTION

Entrepreneurship provides thousands of opportunities both for the entrepreneur and the society. Startups are revolutionizing the world, and a world-changing startup requires a supporting eco-system to thrive. Although there have been significant improvements in the last decade, India hasn't made a mark in the startup eco-system despite the government's efforts.

Our inspiration for this paper is a report [1] from the Global Entrepreneurship Monitor (GEM), a global consortium researching entrepreneurship that reports that few people in India are motivated to improve their lives by pursuing entrepreneurial opportunities.

The lack of entrepreneurial competence and interest in entrepreneurship among Indian University students presents itself as an alarming situation because a lack of entrepreneurs can relate to a direct negative impact on the economy and further slow-down the startup ecosystem in the country. Entrepreneurship opens up countless possibilities and doors, both for the startup owner and the common masses, and we want to solve the problem of the lack of entrepreneurial competency by using Machine Learning algorithms.

The following characteristics, traits, and factors of an individual are considered for building the machine learning model:

1. *Personality and Entrepreneurial Competency*

It is believed that certain characteristics and personality traits make an individual more likely to succeed as an entrepreneur than others. According to a study [2], the following personality traits are highly linked with entrepreneurial competency: Perseverance, the desire and

willingness to take the initiative, Competitiveness, Self- Confidence, Good Physical Health, a strong need to achieve and Self-Reliance.

1. *Geographical, Educational and Mental Traits’ correlation with Entrepreneurial Competency*

Apart from personality traits, an array of other factors is also key in making an individual much more suited to succeed as an entrepreneur than others. The city [3], mental health [4], education [5], gender [6], age [7] and the existence of certain traits [8] like Vision, Work Ethic, Resilience, Positivity, and Passion are also believed to be linked to the entrepreneurial competency of an individual.

This paper is divided into the following sections: Related Work, Proposed Approach, Experimental Set-Up, Results and Discussions, Conclusion and Future Scope, and References.

1. RELATED WORK

In paper [9], the authors used Classification algorithms to detect mental stress in University Students. In paper [10], the researchers predicted the entrepreneurial behavior of an individual by applying the Theory of Planned Behavior. This research was in the domain of social psychology.

J. Agarwal et. al. [11] used K-Means Clustering for Crime Analysis on a dataset of offenses recorded by the police in England and Wales. The working paper [12] talks about the personality traits of an entrepreneur and reviews other literature in the field.

The paper [13] uses cluster analysis to map the pattern of growth of entrepreneurial ventures. In the research paper [14], Melissa S. et. al. uses cluster analysis to explore a categorization pattern that best describes the differences among entrepreneurs from different minority groups.

1. PROPOSED APPROACH

In this paper, first we cleaned the data by removing the missing values and unwanted rows. Then we did classification for the prediction of the bridge damage prediction.

Next, we found the feature on the basis of which we would classify our dataset. Then, after we decided to classify our dataset on the basis of whether the bridge is damaged or not and if it is damaged then what is the damaged level. The damage level was divided into 4 parts , which are ‘no damage’, ‘moderate/severe damage’, ‘partial/overall damage’ and finally ‘total collapse’.

Then we repeated the following steps for decision tree and random forest classification so that we can get a higher accuracy score.

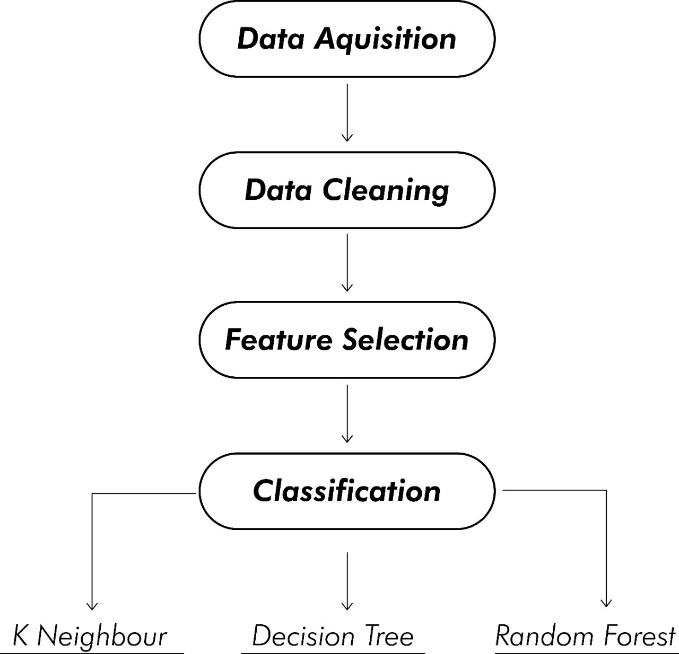


Figure 1. Generalized Framework.

1. EXPERIMENTAL SET UP, RESULTS AND DISCUSSIONS
2. *Dataset, Pre-processing, and Analysis*

We collected the dataset with 250 entries of “San Francisco-Oakland Bay Bridge”, “Interstate 5 (Golden State Freeway), Gavin Canyon”, “Interstate 10 (over La Cienega Boulevard)”, “Interstate 405 (over Jefferson Boulevard)” and many more bridges. We used many attributes for classification like material\_type of which the bridge is made of. The type I material is steel and the type II material is concrete as shown in fig(2).

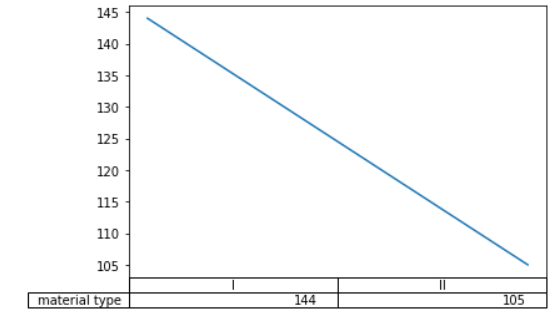


Figure 2. No. of bridges based on Material type.

Other features consist of the age of the bridge and the type of the bridge which is classified into 4 types which are beam, truss, cable, arch bridge which is shown in fig(3). We also used the magnitude of the earthquake which is the most crucial of all the attributes as shown in fig(4).

Furthermore, we used the damage level, which is divided into 4 levels which are no damage, moderate damage, partial damage and total collapse depicted in fig(5), but first we checked whether the bridge was damaged or not, 1 indicated damaged and 0 indicates not damaged which is represented in fig(6).

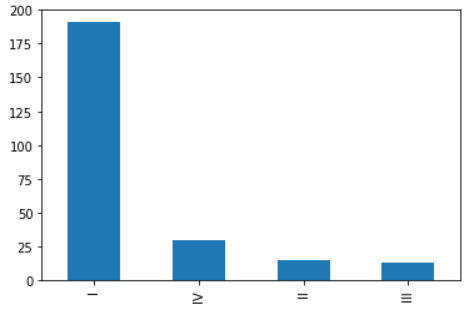


Figure 3. Bridge types.

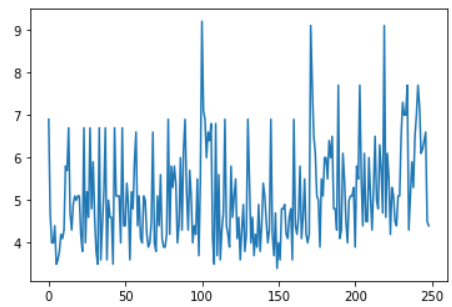


Figure 4. Magnitude scale

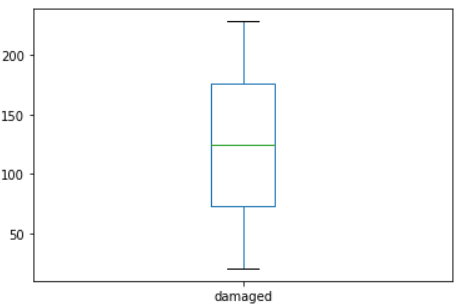


Figure 5. Damaged or not.

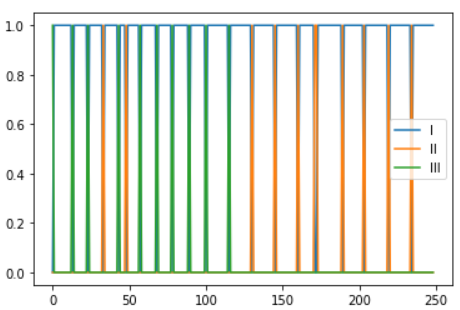


Figure 6. Damage level.

1. EXPERIMENTAL SET UP, RESULTS AND DISCUSSIONS
2. *Dataset, Pre-processing, and Analysis*

We collected the dataset with 250 entries of “San Francisco-Oakland Bay Bridge”, “Interstate 5 (Golden State Freeway), Gavin Canyon”, “Interstate 10 (over La Cienega Boulevard)”, “Interstate 405 (over Jefferson Boulevard)” and many more bridges. We used many attributes for classification like material\_type of which the bridge is made of. The type I material is steel and the type II material is concrete as shown in fig(2).

Figure 2. No. of bridges based on Material type.

Other features consist of the age of the bridge and the type of the bridge which is classified into 4 types which are beam, truss, cable, arch bridge which is shown in fig(3). We also used the magnitude of the earthquake which is the most crucial of all the attributes.

Furthermore, we used the damage level, which is divided into 4 levels which are no damage, moderate damage, partial damage and total collapse but first we checked whether the bridge was damaged or not, 1 indicated damaged and 0 indicates not

F

Apa

TABLE I

COMPARISION OF TRAITS IN THE DATASET

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Dominant Gender | Mental Disorder | Dominant Key Trait |
| All | Male (73.7%) | No (69.1%) | Positivity (34.8%) |
| Entrepreneurs | Male (80.4%) | No (69.5%) | Positivity (37.8%) |
| Non- Entrepreneurs | Male (68.9%) | No (68.9%) | Positivity (32.7%) |

It is observed that Entrepreneurs have a better state of mental health and are much more positive than non-entrepreneurs. For closer analysis, we would use k-Means clustering to divide the dataset into disjoint groups and study the features of each group.

1. *Tools used*

R (dplyr, caTools, e1071, factoextra and ggplot2) and Python (Pandas, Scikit and matplotlib) are used for Data Analysis, Visualization, Clustering, and Classification.

R and Python were selected due to the availability of well- maintained popular libraries.

1. *Clustering*

We use k-Means Clustering to arrange the survey respondents into a set number of related clusters, or groups, and analyze these clusters to help improve entrepreneurial competency by detecting clusters with a large percentage of students possessing the same. K-Means Clustering is a very popular algorithm used for partitioning a set of patterns into disjoint clusters. This method is efficient in producing respectable results for many applications [15].

We manually fixed the number of clusters, say k, to initialize the algorithm. We used two popular methods to determine the number of clusters: Elbow Method and Silhouette method [16].

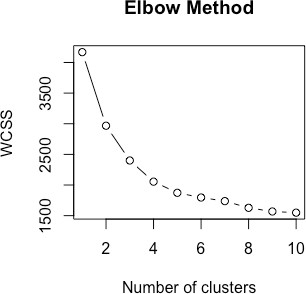


Fig. 2 Elbow Method used to initialize the number of centroids

In the elbow method, a sharp drop in the WCSS, or the so-called ‘Elbow Point’ is expected to show up. In our case, we don’t see a very distinct drop, although the Elbow Method does suggest that the ideal number of clusters would be either two or three. To verify this, we use the Silhouette method, another popular way to determine the appropriate number of clusters for the k-Means algorithm.

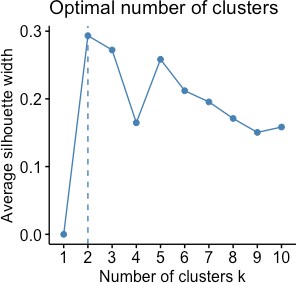


Fig. 3 Silhouette Method used to initialize the number of centroids

The Silhouette method confirms that two would be the ideal number of clusters for this particular dataset. So, we proceed with k-Means clustering intending to divide the students into two disjoint clusters. However, only two clusters predicted by Elbow and Silhouette methods suggest that the statistical power of the results would be low because it is known that statistical power surges with an escalating number of clusters and one can estimate the inter-cluster disparity more accurately [17].

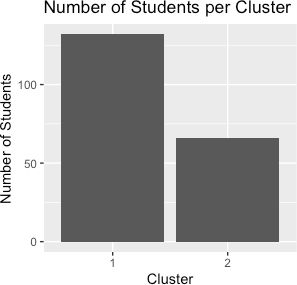


Fig. 4 Number of Students per Cluster

After analyzing the clusters, it is seen that the majority of students lie in Cluster 1. Further, we check for the number and percentage of students possessing Entrepreneurial Competency in these clusters. Checking for the percentage is important because it might be possible that a particular cluster outputs a greater number of entrepreneurs, and also covers a large number of students.

TABLE II

COMPETENCY OF STUDENTS PER CLUSTER

|  |  |  |
| --- | --- | --- |
| Cluster | No. of Entrepreneurs | Percentage of Entrepreneurs |
| Cluster 1 | 57 | 43.181% |
| Cluster 2 | 25 | 37.878% |

A low percentage of entrepreneurs in either cluster reflects the lack of motivation in Indian University Students to pursue a career as an entrepreneur. Additionally, the second cluster has a lower percentage of entrepreneurs out of the two, and to improve entrepreneurial competency, we would analyze this cluster further and compare the results to the slightly ‘more competent’ cluster.

TABLE III

COMPARISION OF TRAITS IN CLUSTERS

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Dominant Gender | Mental Disorder | Dominant Key Trait |
| Cluster 1 | Male (78.7%) | No (72.7%) | Positivity (37.1%) |
| Cluster 2 | Male (63.6%) | No (62.1%) | Passion (37.8%) |

Analysis of the clusters reveals that students in Cluster 1 pick Positivity as their dominant key trait, whereas students in Cluster 2 pick Passion as the key trait. Also, students in Cluster 1 generally haven’t suffered from any mental disorder in the past, a finding in line with the reported Positivity in these students.

Students in Cluster 2, however, comparatively have a lower score in terms of Mental Health. In terms of Gender, male students are dominant in both clusters, more so in Cluster 1, but this finding was expected provided that the maximum number of survey respondents were male.

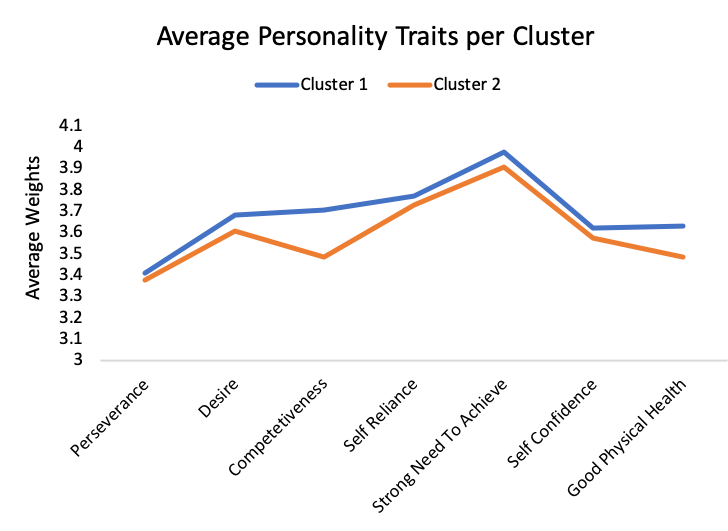


Fig. 5 Average Personality Traits per Cluster

A comparative study regarding the personality traits of students in both clusters reveals that students in the first cluster score higher in all personality traits. The maximum difference in the average score is in Competitiveness, followed by Good Physical Health.

1. *Classification*

We use Machine Learning Classification algorithms, namely k-NN, Logistic Regression, Naïve Bayes, Support Vector Machine, Decision Trees and Random Forest to try and classify students and predict if a student with the given set of attributes possesses entrepreneurial traits or not. All the features collected in the survey were used as independent variables for the classification algorithms.

A training-test dataset split of 72:25 is used for feeding the classification algorithms. The classifiers are developed on the training set and then applied to the test set samples [18].

We used Repeated CV in k-NN to find the appropriate number of ‘k’ in the algorithm, the number of closest neighbors we would be inspecting.

TABLE IV

COMPARISION OF ACCURACY REPORTED BY DIFFERENT CLASSIFICATION ALGORITHMS

|  |  |  |
| --- | --- | --- |
| S. No. | Algorithm | Accuracy |
| 1 | Random Forests (500 Trees) | 40.81% |
| 2 | k-Nearest Neighbors (k = 5) | 53.06% |
| 3 | Support Vector Machine | 59.18% |
| 4 | Logistic Regression | 57.14% |
| 5 | Naïve Bayes | 48.97% |
| 6 | Decision Tree | 57.14% |

SVM is comparatively the best performing algorithm out of the six but since the accuracy reported isn’t very high, this indicates that the dataset in question isn’t easily classifiable and that the traits reported independently to be linked to entrepreneurship don’t have a strong correlation that can be used to accurately predict the entrepreneurial capabilities of a student given his/her interests, age, gender and other personality traits in question.

1. CONCLUSION AND FUTURE SCOPE

Embracing the startup ecosystem and promoting entrepreneurship is the need of the hour to strengthen the economy, open countless new opportunities for entrepreneurs and provide a platform for innovation.

The satisfactory results of cluster analysis can be used by policymakers and other people working high in the chain in the entrepreneurship ecosystem to make clever decisions that help boost the interests of the Indian youth in entrepreneurship.

Good mental health, stress-free life, good physical health, high levels of competitiveness and a positive lifestyle are some of the insights that can be drawn from the cluster analysis to increase entrepreneurship competence in university students. The survey responses also highlight that a very large percentage of students who aren’t entrepreneurs claim that it is due to the lack of knowledge and awareness. This is something alarming, given the recent awareness campaigns and financial programs launched by the government. Academic pressure is also considered by a large percentage of students as a barrier to entrepreneurship, and academic leaders should look into this matter by making university programs more adaptable to those who want to pursue entrepreneurship, offer courses in the field, organize workshops, seminars, and events.

The low accuracy of the classification models can be improved by future researchers possessing business acumen or knowledge in the domain of psychology by using feature engineering and/or feature scaling to eliminate certain traits and features from the classification algorithm to make the model more robust and report a higher accuracy score. Also, analyzing clusters and performing classification on a much larger dataset can lead to clearer statistics and trends.

The processed and clean datasets generated from the survey are open-sourced and future researchers are free to experiment on the data [19].

REFERENCES

1. N. Bosma and D. Kelly, Global Entrepreneurship Monitor 2018/2019 Global Report
2. W.E. Jennings, “A Profile of the Entrepreneur” in Entrepreneurship: A Primer for Canadians
3. Acs, Zoltan J. and Glaeser, Edward L. and Litan, Robert E. and Fleming, Lee and Goetz, Stephan J. and Kerr, William R. and Klepper, Steven and Rosenthal, Stuart S. and Sorenson, Olav and Strange, William C., Entrepreneurship and Urban Success: Toward a Policy Consensus, January 2008.
4. U. Stephan, “Entrepreneurs’ Mental Health and Well-Being: A Review and Research Agenda” in Academy of Management Executive, vol. 32, p. 290-322, January 2018.
5. HULUGAPPA, H. Ramakrishna H. Entrepreneurship Education in India: Emerging Trends and Concerns. Journal of Entrepreneurship & Management, [S.l.], pp. 28-41, December 2013. ISSN 2277-6850.
6. B. Kar, R. N. Subudhi, and N. Kar, “Gender-Gap in Entrepreneurship: A Study on Ideation, Efficacy, Planning Differentiation Measures”, Amity Journal of Entrepreneurship, vol. 1, pp. 49-61, May 2016.
7. Azoulay, Pierre and Jones, Benjamin F. and Kim, J. Daniel and Miranda, Javier, Age and High-Growth Entrepreneurship (April 2018). NBER Working Paper No. w24489. Available at SSRN: https://ssrn.com/abstract=3158929
8. R. Westwood, (January 2017) “The Traits Entrepreneurs Need To Succeed”. The Forbes Website. [Online]. Available: [https://www.forbes.com/sites/ryanwestwood/2017/01/09/the-traits-](http://www.forbes.com/sites/ryanwestwood/2017/01/09/the-traits-) entrepreneurs-need-to-succeed/#46bcdc55c10f
9. R. Ahuja, and A. Banga, “Mental Stress Detection in University Students using Machine Learning Algorithms”. International Conference on Pervasive Computing Advances and Applications – PerCAA 2019. Procedia Computer Science 152 (2019) 349–353.
10. Kautonen, Teemu & Gelderen, Marco & Tornikoski, Erno. (2011). Predicting Entrepreneurial Behaviour: A Test of the Theory of Planned Behaviour. Applied Economics. 45. 697-707.
11. J. Agarwal, R. Nagpal, and R. Sehgal, “Crime Analysis using K-Means Clustering”, International Journal of Computer Applications (0975 – 8887), vol. 83, December 2013
12. S. P. Kerr, W. R. Kerr, and T. Xu, “Personality Traits of Entrepreneurs: A Review of Recent Literature”, Working Paper 18- 047, Harvard Business School. Available: <http://www.hbs.edu/faculty/publication%20files/18-047_b0074a64-> 5428-479b-8c83-16f2a0e97eb6.pdf
13. A. S. Setyaningsih, Using Cluster Analysis Study to Examine the Successful Performance Entrepreneur in Indonesia, Procedia Economics and Finance, vol. 4, 2012, pp. 286-298, ISSN 2212-5671. Available: https://doi.org/10.1016/S2212-5671(12)00343-7.
14. M. S. Cardon, R. S Shinnkar, M. Eisenman, and E. G. Rogoff, “Segmenting the Population of Entrepreneurs: A Cluster Analysis Study”, Journal of Development Entrepreneurship, vol. 13, pp. 293- 314, 2008.
15. S. Singh, and N.S Gill, “Analysis and Study Of K-Means Clustering Algorithm”, International Journal of Engineering Research & Technology (IJERT), vol. 2, pp. 2546-2551, July 2013. ISSN 2278- 0181.
16. P.J Rousseuw (1987). "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics. 20: 53–65.
17. Peter C. Austin & George Leckie (2018) The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models, Journal of Statistical Computation and Simulation, 88:16, 3151-3163
18. Dobbin and Simon: Optimally splitting cases for training and testing high dimensional classifiers. BMC Medical Genomics 2011 4:31.
19. https://github.com/skhiearth/Statistical-Research-Entrepreneurial- Competency-Datasets