KDU Journal of Multidisciplinary Studies (KJMS)
Volume 6. Issue 2 Nov 2024
DOI: https://doi.org/10.4038/kjms.v6i2.136

KJMS 2024 VOL.6 (2): 151-156

A CLUSTERING APPROACH TO DETECT IMPOSTER SYNDROME AMONG SRI LANKAN UNDERGRADUATES

A Brahmana

Department of Information and Communication Technology, Faculty of Technology,
University of Ruhuna, Matara, Sri Lanka

ABSTRACT

Imposter Syndrome is another name for perceived fraudulence, which is characterized by feelings of personal inadequacy and self-doubt that endure despite education, achievement, experience and success. This is not a disease or abnormality, so there is no obvious reason to imposter emotions. Therefore, even if they suffer from imposter syndrome, they are not able to know this. The results of an undergraduate with imposter syndrome may be inappropriate academic choices, the impact on mental health and social isolation. The aim of the present study is to develop a computerized framework based on a data mining strategy to identify the Severity Level of imposter syndrome for Sri Lankan undergraduates. Thus, this research shows whether the person suffers from imposter syndrome as Low or Moderate or High in level. During the model development, a formal questionnaire was developed examining different influencing factors like depression, anxiety, parentification, family expectations, perfectionism, and low trait self-esteem that can affect the imposter syndrome of an undergraduate and was used to collect data from Sri Lankan undergraduates. In this study, five different unsupervised machine learning techniques, namely K-means, K-medoids, Spectral Clustering, Hierarchical Clustering and Gaussian Mixture Model Clustering were used. Clustering was selected as the best approach as it allows to detect patterns and similarities associated with undergraduates linked to imposter syndrome. To calculate the goodness of the clustering algorithms, the Silhouette index and the Calinski-Harabasz index were used. Among these five clustering algorithms, the best result was shown in the three clusters of K-means Hence, the finalized method helps to predict and classify severity levels of imposter syndrome among Sri Lankan Undergraduates into three groups as low, moderate or high. The research found that among 316 data points, 32.28% showed a low level of imposter syndrome, 16.77% displayed a moderate level, and 50.95% exhibited a high level.

KEYWORDS: *Imposter Syndrome, Sri Lankan Undergraduates, Clustering, Severity Level*

Corresponding Author: A Brahmana, Email: akila@ictec.ruh.ac.lk

https://orcid.org/ 0009-0000-5357-9106

1. INTRODUCTION

Psychologist Audrey Ervin says imposter syndrome can apply to anyone who is "unable to internalize and own their successes". Imposter Syndrome (IS) or perceived fraudulence or impostor phenomenon is the belief that, despite education, experience, success, and accomplishments, they are underestimated by others. Psychology researchers Pauline Clance and Suzanne Imes, introduced the phrase "Imposter Syndrome" for the first time in 1978. People with Imposter Syndrome believe that their achievement is not a result of their abilities or intelligence, but rather only pure luck, external fault, or hard work (Maqsood *et al.*, 2018). Those who have this syndrome worry that others will recognize them as frauds. Students are one of the high-risk groups for this syndrome. In modern society, there is huge competition among children in the field of education. Among them, the competition between undergraduates is high. Therefore, undergraduates may unknowingly suffer from imposter syndrome. There is no obvious reason to imposter emotions because this is neither a disease nor an abnormality. Instead, several factors may combine to trigger them, such as the pressures of perfectionism, increasing social comparisons, overworking, undermining one's own achievements, fear of failure and discounting praise (Maqsood *et al.*, 2018). Therefore, even if they suffer from imposter syndrome, they are not able to know this. Considering the effects of IS on undergraduates, it is important to investigate this in advance as the students can then take necessary steps to avoid it. (Maqsood *et al.*, 2018; Bravata *et al.*, 2020)

Considering prior studies in foreign countries, these studies have proven that impostor syndrome is a serious public health problem for undergraduates. No research has yet been published on impostor syndrome among Sri Lankan undergraduates. That means there is no source to know whether there is impostor syndrome among Sri Lankan undergraduates. When compared to western countries, Sri Lanaka is having different cultural aspects and educational settings. Hence this research covers various aspects by analyzing hidden patterns associated with imposter syndrome in Sri Lankan context. Additionally, this study fills important gaps related to mental health and machine learning while considering evidence-based insights. While drawing attention to it, it is very important to find out whether undergraduates in Sri Lanka have imposter syndrome or not.

No sufficient mechanism exists to determine the IS severity level, and no data mining framework has been applied to predict the level. Hence, the present research study's main aim is to implement a model to predict the severity level of imposter syndrome among Sri Lankan undergraduates. This study incorporates five unsupervised machine-learning techniques, such as K-means, K-medoids, Spectral Clustering, Hierarchical Clustering and Gaussian Mixture Model Clustering. such a strategy has not been explored. The Silhouette index and the Calinski-Harabasz index, two widely used performance metrics, were employed in the results to confirm the suggested technique's enhanced performance.

2. LITERETURE REVIEW

This research has mainly investigated the IP level of Sri Lankan undergraduates who are studying in state and private universities in Sri Lanka. Many studies on IP have been done with undergraduate students in foreign countries. But currently, no any research has been conducted to develop a computer based model to determine the level of IP in Sri Lanka.

Some studies have been done to determine the extent to which imposter syndrome affects university students (Qureshi *et al.*, 2017). Similarly, several studies have been conducted on how imposter syndrome affects university students based on factors, such as perfectionism, psychological distress, mental health and personality traits related to imposter syndrome (Wang, Sheveleva and Permyakova, 2019). Many researchers have collected data from a variety of questionnaires. Some researchers used the Clance Imposter Phenomenon Scale Questionnaire (Maqsood *et al.*, 2018) and Young Imposter Scale Questionnaire based on random sampling. The collected data were analyzed using Preliminary analyses, Mediation analyses, Moderation analyses, Correlational patterns and Regression analyses. From the final results of these studies, statistical findings (results are shown

quantitatively) show how imposter syndrome affects undergraduates.

(Sullivan and Ryba, 2020) conducted a study and revealed that 57% of pharmacy and 15% of medical students in the United States, exhibit imposter syndrome, according to the most recent IS research. Indeed, on a worldwide and regional scale, IS growing to be a substantial public health risk. For instance, Pakistan and Malaysia both had prevalence rates of 47% and 45.7%, respectively, for IS among medical students. (Qureshi *et al.*, 2017)

Only one research has been published on IS regarding Sri Lanka. The number shows that the topic of Impostor syndrome is new to Sri Lanka. This research was based on MBA students. The impact of context-dependent factors were examined, such as job fit career stage and organizational tenure, on imposter fear using accomplishment goal theory as a foundation. To learn more about the causes of imposter fears, factors like self-efficacy , gender and achievement-related qualities were investigated. Key findings indicate that according to both studies and job fit impostor fears are frequently predicted by organizational tenure and career stage. Similar levels of imposter fears were expressed by men and women, and impostor fear was predicted by self-efficacy and locus of control. The findings on Sri Lankan personality predictions are also same with those from studies with a North American focus, showing the concept's cross-cultural applicability. These research results aid in the development of focused management strategies that support the creation of treatments like orientation programs that improve socialization processes and lessen concerns about impostor fears. (Kumar, Kailasapathy and Sedari Mudiyanselage, 2022)

Furthermore, one study has been conducted to prove the use of computational techniques such as data mining. In this study, the prediction of medical students' (Holliday *et al.*, 2020) IS was carried out using the YIS scale and three distinct machine learning approaches: neural network, ensemble learning and random forest (Khan *et al.*, 2022).

3. METHODOLOGY

Conduct a Comprehensive Literature Review

A Comprehensive analysis has been carried out to identify the factors of IS and to gather the required background knowledge from existing studies.

Data Collection

In order to identify the factors influencing IS for undergraduate students, a literature review has been conducted. Based on previous studies, seven factors have been selected as the causal factors of impostor syndrome in an undergraduate. These are depression, parentification and family expectations, anxiety, perfectionism, low trait self-esteem, fear of failure and comparison. (Fassl, Yanagida and Kollmayer, 2020)

After selecting the factors, a structured questionnaire of twenty-eight questions was created and prepared in both Sinhala and English languages as the data set was collected from the undergraduate community in Sri Lanka. This questionnaire includes twenty-one questions based on seven factors (Yaffe, 2023) with three questions per factor and seven demographic features, including gender, field, level, GPA, platform, accommodation and hours. This “factors twenty-one item questionnaire” is in Likert type.

Data were collected from 18th February 2022 to 6th June 2022 from public and private undergraduate students in Sri Lanka from the first to the fourth years. It was assured the survey participants that all information they provided would be treated in strict confidence and used only for research. A Google form was created and distributed it among undergraduates from various universities. Ultimately, 350 successful data was collected for analysis.

Data Preprocessing

The quality data was obtained after applying several preprocessing techniques. The pre-processing helped extract the required data from the entire data set to perform proper data mining. Out of the twenty-eight features from gathered data, twenty-one are survey entries based on a Likert scale, and except for two numerical columns (GPA and hours), the rest are entries based on various categorical values, such as

gender, year, field of study, etc. First, One-Hot Encoding was used to convert the categorical values into numerical values that could be given to the machine learning algorithms to better predict. Next, outliers were detected in the numerical column (hours) using a box plot and handled them using the IQR method. Then, the noise or non-linear values in the hours column were divided into three bins using the smoothing by bin boundary method. After data visualization, some unnecessary data were removed by identifying data that were not important for future processes. Finally, missing values were filled in with the most frequent values in that column. 316 proper observations remained for the research analysis after cleaning the raw data.

Algorithm Implementation

Unsupervised learning was used for prediction since there was not a labeled dataset. The clustering approach allows to detect patterns and similarities among undergraduates who may exhibit imposter syndrome characteristics. These methods can reveal student groups who share similar physiological and behavioral features linked to imposter syndrome. The data was analyzed using five unsupervised machine learning methods: K-means, K-medoids, Spectral Clustering, Hierarchical Clustering and Gaussian Mixture Model Clustering. Python was used as the programming language for algorithm implementation.

Validation indices describe how efficiently an algorithm partitions data into clusters. To select the most suitable and precise algorithm, validation indices named the Silhouette index and the Calinski-Harabasz index were used in this study.

Final Prediction

The finalized model was used to test with five clustering algorithms with two indices from three to ten clusters. The most suitable data mining model was chosen upon in accordance with its performance. According to the highest performance gained from the Silhouette index and Calinski-Harabasz index, to implement the model in this study the K-means approach was chosen as the data mining model.

Jupyter 6.4.5, a web-based interactive computing notebook environment, was used to perform the main data preprocessing tasks, including outlier treatment, filling in missing values and smoothing data, and selecting the most suitable and accurate algorithm. Once the model is developed, an undergraduate can identify the level at which he or she suffers from imposter syndrome. If the results are moderate or high, the undergraduate can take the necessary steps to suppress them.

4. RESULTS ANS DISCUSSION

Selection of the model

After gathering the data from undergraduates in several Sri Lankan universities, pre-processing techniques were applied to the data to obtain suitable data for the model development. Then the application of unsupervised learning techniques in machine learning was carried out. As shown in TABLE 1, the five clustering algorithms, with three to ten clusters, were compared with two validation indices, namely the Silhouette index and the Calinski-Harabasz index. Based on the results of the validation indices, the total number of clusters was finalized, and the best efficient cluster could be selected accordingly. From Table 1, it is clearly observed that all the indices decrease or increase after the third cluster and in terms of indices, we can confirm that the third cluster is the best as compared to other clusters.

Table 1: Overall performance of the algorithm

Clustering Algorithm	Validity Indices	Total Number of Clusters		
		3	4	5
K-means Clustering	Silhouette Index	0.1765	0.1443	0.1464
	Calinski-Harabasz Index	59.9749	50.8048	44.9126
K-medoids Clustering	Silhouette Index	0.0467	0.0689	0.0443
	Calinski-Harabasz Index	22.3996	29.3210	24.3665
Spectral Clustering	Silhouette Index	0.1685	0.1367	0.1354
	Calinski-Harabasz Index	58.7778	49.7090	44.2493
Hierarchical Clustering	Silhouette Index	0.1594	0.1483	0.1378
	Calinski-Harabasz Index	56.8646	46.2755	41.0024
Gaussian Mixture Models	Silhouette Index	0.1756	0.1588	0.1512
	Calinski-Harabasz Index	59.9072	50.2033	44.7649

This table illustrates that the Silhouette index and Calinski-Harabasz index in the three clusters are higher in the K-means algorithm compared to other clustering algorithms. Therefore, we have concluded that K-means clustering produces better results.

Model Results

After the model development research revealed that, higher levels of imposter syndrome were associated with higher GPAs. Based on this finding, the cluster labels of the three clusters in the K-means are mapped to the levels as Low, Medium and High.

Based on the calculation of the K-means method with the help of Jupyter software, as a result, out of 316 undergraduates, The study has identified 102(32.28%) undergraduates as "low" level, 53(16.77%) undergraduates as "moderate" level and 161(50.95%) undergraduates as "high" level.

Research findings

After reviewing the experimental results, it was found that a significant percentage of undergraduates in Sri Lanka suffer from impostor syndrome, as most undergraduates belong to moderate and high levels. Previous studies have revealed that third-year (Maqsood *et al.*, 2018) and fourth-year (Villwock *et al.*, 2016) students have a higher prevalence and severity of impostor syndrome. Reflecting research findings, these indicate that third and fourth years undergraduates suffer from moderate/high categories of IS. From that point of view, the academic year affects on impostor syndrome in the Sri Lankan undergraduate context. According to some studies, women were twice as affected as men (Fraenza, 2016). Similarly, female respondents had more impostor syndrome than male respondents in our observation.

Among the seven selected factors, some factors represent the most influential factors on undergraduates with a high level of IS, while others have the least influence on undergraduates with a low level of IS. According to these findings, for undergraduates in Sri Lanka depression and anxiety are the most influencing factors, while parentification and family expectations, perfectionism and low trait self-esteem are the least influencing factors.

The goal of the study was to develop a more accurate predictive model for impostor syndrome among Sri Lankan undergraduates. Also, this study attempts to address a relatively new research area.

This study was able to conclude four major findings. It revealed that there are significant undergraduate students in the high and moderate categories, which are severity levels. The severity level of IS was higher in female respondents than in male respondents. IS was not limited to a particular academic year and this study proved that impostor syndrome is a significant public health problem for university students in Sri Lanka.

Nowadays, some undergraduates study in a physical learning environment while others study in a virtual learning environment. Therefore, the special point to be mentioned here is that when the undergraduates answer the questionnaire from the corresponding learning environment, the data they provide may have less impact on the accuracy of the proposed algorithm.

5. CHALLENGES AND LIMITATIONS

With the current data set, Severity level of imposter syndrome was divided into three main levels as low, medium and high. But time to time these influential factors can be changed with various economic and social factors. Data was collected from both private and state university students. State university students might have different schedules, learning environments or accessibility issues compared to private university students. It might lead to an uneven sample. It will lead to limit the generalizability of the research findings. There is a dynamic and evolving nature of imposter syndrome. Questionnaire always may not capture this dynamic nature of the syndrome.
Sometimes students may not reveal their genuine ideas for these kinds of questionnaires and it may affect to make correct assessments on the dataset.

6. CONCLUSION

This is the first study that focus on impostor syndrome based on Sri Lankan undergraduates. Identifying a well-fit model for predicting the severity level of impostor syndrome is the main goal of this study. To determine whether an undergraduate has IS, a Likert-type questionnaire was developed based on demographic factors and the factors selected after conducting the literature survey. Through the use of a Google form, data was collected from various universities in Sri Lanka for the research. Unlabeled dataset led to follow clustering approach in this study. Using two validation indices, the outcomes was compared of five machine-learning clustering algorithms. When comparing the performances, the three clusters of the K-means algorithm scored higher on the validation indices. Finally, the three cluster labels in K-means are mapped as low, moderate, and high. The present study confirmed that the severity of IS was higher in female undergraduates than in males. Also, IS was limited to a specific academic year, and third-year and fourth-year undergraduates suffered from moderate/high categories of IS. Based on the findings of the study, respective institutions and governing authorities should take necessary actions and implement policies to assist students experiencing imposter syndrome.

7. REFERENCES

Bravata, D.M. *et al.* (2020). Prevalence, Predictors, and Treatment of Impostor Syndrome: a Systematic Review, *J. of General Internal Medicine*, 35(4), pp. 1252–1275. Available at: https://doi.org/10.1007/s11606-019-05364-1.

Fassl, F., Yanagida, T. and Kollmayer, M. (2020).Impostors Dare to Compare: Associations Between the Impostor Phenomenon, Gender Typing, and Social Comparison Orientation in University Students, *Frontiers in Psychology*, 11, p. 1225. Available at: https://doi.org/10.3389/fpsyg.2020.01225.

Holliday, A.M. *et al.* (2020).High Prevalence of Imposterism Among Female Harvard Medical and Dental Students, *J. of General Internal Medicine*, 35(8), pp. 2499–2501. Available at: https://doi.org/10.1007/s11606-019-05441-5.

Khan, Md.N.A. *et al.* (2022).Predicting Young Imposter Syndrome Using Ensemble Learning, *Complexity*. Edited by L. Guo, 2022(1), p. 8306473. Available at: https://doi.org/10.1155/2022/8306473.

Kumar, S., Kailasapathy, P. and Sedari Mudiyanselage, A. (2022).It's my luck: impostor fears, the context, gender and achievement-related traits, *Personnel Rev.*, 51(9), pp. 2222–2238. Available at: https://doi.org/10.1108/PR-03-2021-0149.

Maqsood, H. *et al.* (2018).The descriptive study of imposter syndrome in medical students, *International J. of Research in Medical Sc.*, 6(10), p. 3431. Available at: https://doi.org/10.18203/2320-6012.ijrms20184031.

Qureshi, M.A. *et al.* (2017).Imposter Syndrome among Pakistani Medical Students, *Annals of King Edward Medical Univ.*, 23(2). Available at: https://doi.org/10.21649/akemu.v23i2.1647.

Villwock, J.A. *et al.* (2016).Impostor syndrome and burnout among American medical students: a pilot study, *Int. J. of Medical Education*, 7, pp. 364–369. Available at: https://doi.org/10.5116/ijme.5801.eac4.

Wang, K.T., Sheveleva, M.S. and Permyakova, T.M. (2019).Imposter syndrome among Russian students: The link between perfectionism and psychological distress, *Personality and Individual Differences*, 143, pp. 1–6. Available at: https://doi.org/10.1016/j.paid.2019.02.005.

Yaffe, Y. (2023).The Association between Familial and Parental Factors and the Impostor Phenomenon—A Systematic Review, *The American J. of Family Therapy*, 51(5), pp. 527–545. Available at: https://doi.org/10.1080/01926187.2021.2019140.