American Journal of Psychiatric Rehabilitation

Expert Opinion Article

Machine Learning Based Classification on Factors Used for Identifying Competency Gap In Engineering Students In Thanjavur District

Aishwariyashindhe S[1], Sathyapriya J[2*]

[1]Research Scholar, Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, India, aishwariyashindhe243@gmail.com
[2*]Professor, Department of Management Studies, Periyar Maniammai Institute of Science & Technology (Deemed to be University), Thanjavur, India, sathyapriya@pmu.edu

ABSTRACT
The quantity of graduates generated annually by education institutions is on the rise. The prediction of graduates' employability is of significant importance to industries as it facilitates effective talent acquisition and use. Additionally, it assists students in recognizing the qualifications and abilities they need to enhance prior to completing their degrees in order to secure desired employment opportunities. During the current era of the Digital Revolution, there is a notable occurrence of informal learning with skill development taking place in an unrestricted manner. However, a significant challenge is in effectively connecting and aligning these acquired knowledge and skills with the overall employability rate. The primary aim is to effectively tackle this matter by employing machine learning algorithms to continuously predict and forecast the ongoing skill acquisition and align it with the demands of the industry. The study included various machine learning methods, including Logistic Regression (LR), Decision tree (DT), k-nearest neighbor (K-NN), Support Vector Machine (SVM), and Naïve Bayes (NB), to construct the model. This research holds potential benefits for various entities, encompassing governmental bodies, private enterprises, and businesses, as well as individuals such as students and educators, with the aim of enhancing employability.

Keywords: Machine Learning, Skill acquisition, informal learning, skill enhancement, revolution

1. INTRODUCTION

In the context of digitalization, the focus of education is increasingly shifting towards job outcomes. The reputation of an educational institution is contingent upon the employment success of its students, thus making it a significant area of concern (1). The predictive model, which offers insights into the employment outcomes of students, can assist in identifying those who may require more support. Instead of conventional analytics, the utilization of advanced machine learning techniques, which are a component of Artificial Intelligence, is employed to derive predictive insights on future events.

As a result of the continuously evolving industrial market and the swift progressions in technology. The global need for specialists in the field of Information Technology (IT) is now seeing significant growth (2). Human capital is widely recognized as a crucial economic asset in the production process, serving as a fundamental pillar for enhancing living standards and fostering the development of human resources. Human capital plays a crucial role in the strategic planning endeavors of nations as they strive to attain sustainable development. This is because human capital encompasses the workforce involved in various societal activities, including service provision, production, and consumption. Consequently, Higher educational institutions (HEIs) are generating a growing quantity of graduates annually. The incongruity between the outcomes of higher education and the requirements of the labor market is widely recognized as a significant challenge to economic progress, resulting in elevated levels of unemployment and issues related to job placement.

The discrepancy arises from inadequate collaboration between the skill sets of industry professionals and students. The absence of effective communication leads to an inadequate workforce, thereby resulting in costly production errors (3)(4). In order to mitigate this discrepancy, HEIs must take measures to assure the employability of their graduates.

Machine learning (ML) methodologies have the potential to forecast the employability indicators of information technology (IT) graduates and ascertain the primary factors that influence their employability at the earliest possible stage. This enables the implementation of suitable measures to augment their employability, thereby equipping them with the requisite knowledge and skills prior to their entry into the ever-changing job market.

There has been a growing inclination towards the utilization of machine learning in the realm of higher education, as indicated by previous research. These studies have focused on the prediction of graduates' employability. However, the application of automated ML for the purpose of predicting students' employability is still in its early stages. ML is a specific branch of AI that involves the utilization of computer systems to analyze extensive datasets. The objective is to identify patterns within the data, enabling the systems to generate accurate predictions for fresh data. This approach differs from conventional computer approaches. In the context of conventional reasoning, algorithms refer to a collection of precisely articulated instructions that computers employ for the purpose of delineating or resolving issues. Consequently, within the context of the recruitment procedure, individuals who possess prior experience are highly sought after owing to their heightened production levels and reduced training expenses, in comparison to their inexperienced counterparts. HEIs are required to undertake regular evaluations in order to ensure that they are equipping future engineering graduates with the necessary skills. This is regarded a crucial aspect in producing a competent and qualified engineering workforce.

The objective of this research endeavor is to develop a machine learning framework capable of accurately forecasting the employability outcomes of individuals who have completed their education. This model will involve the assessment of skill levels among students and the industry. Additionally, the study seeks to optimize the utilization of the collected dataset, which will provide valuable insights into the level of preparedness of engineering graduates for entering highly technical careers in the workforce.

- One potential area of improvement lies in narrowing the disparity between the requirements of the industry and the skill set possessed by engineering students
- One potential avenue for enhancing the qualifications of engineering graduates is to align their skillsets more closely with the demands of the industry
- This study aims to offer significant insights that will assist engineering universities in formulating more effective long-term strategies for creating graduates who possess both knowledge and skills, by predicting their employability status
- Make a substantial contribution to the recruitment process for businesses
- One potential solution to mitigate the issue of elevated unemployment rates among engineering graduates is to implement measures aimed at reducing this phenomenon

The subsequent sections of this work are structured in the following manner: Section 2 provides a comprehensive review of the existing literature pertaining to employability prediction within the discipline. Section 3 provides a comprehensive exposition of the proposed methodology. Section 4 presents the outcomes of the employed algorithms and provides a comprehensive examination of the utilized features. The outcomes of this study, together with several limitations and potential areas for development, are presented in Section 5.

2. LITERATURE SURVEY

Z. Othman et al., (5) employed data analytics and ML methodologies, including SVM, LR, Artificial Neural Networks (ANN), DT and discriminant analysis, to make predictions on the employment prospects of graduating students. Moreover, the employed characteristics encompass hard talents, demographic attributes, extracurricular activities, and internships. The data were obtained through the administration of surveys to students and the retrieval of information from institutional databases. The SVM classification technique demonstrated a classification accuracy of 87.26%.

The researchers in (6) sought to determine the primary elements that have a substantial impact on the employment of graduate students. To achieve this objective, they employed three classification algorithms, namely DT, ANN, and SVM. The features included in this study encompass hard talents, soft skills, demographic characteristics, extra/co-curricular involvements, university attributes, and internships. The research data were obtained from databases maintained by academic institutions. The SVM algorithm demonstrates an accuracy rate of 66.096%.

M. Alghamlas and R. Alabduljabbar in (6) developed a web-based application utilizing machine learning algorithms, specifically DT, NB, and NN, to forecast the environmental sustainability of IT students' skills for recruitment purposes, with a focus on both hard skills and soft skills. The data used for this prediction was collected through surveys administered to both students and recruiters. Among the algorithms tested, Naive Bayes attained the highest accuracy rate of 69% .

A. Dubey and M. Mani in (7) presents supervised machine learning methods, including LR, DT, RF, KNN, and SVM, were employed to forecast the employability of high school students for part-time positions within local enterprises. The predictive models were constructed using a combination of hard skills, demographic characteristics, and extra/co-curricular activities data, which were obtained through surveys administered to the students. The logistic regression technique demonstrated a classification accuracy of 93%.

S. R. Rahman et al., (8) conducted an analysis of data obtained from educational institutions. Their objective was to make predictions regarding students' employability and identify the various factors that influence it. To achieve this, they considered a range of variables including hard skills, soft skills, demographic characteristics, extra/co-curricular activities, and university attributes. Subsequently, they applied four machine learning

algorithms, namely DT, NB, SVM, and KNN. The accuracy of the results obtained using the DT and SVM algorithms was found to be 98%.
C. D. Casuat and E. D. Festijo (9) conducted an analysis to identify the attributes that had the highest predictive power in determining the employability of students. This analysis encompassed hard skills, soft skills, and demographic features. The authors utilized graduates surveys and institutional databases as data sources. Additionally, they implemented and analyzed three different approaches, namely SVM, RF, and DT. The SVM demonstrated the highest level of accuracy, reaching 91.22%.
K. B. Aviso et al., in (10) explored the influence of different institutional characteristics on the employability of graduates. They employed a hyperbox-based machine learning model, which yielded a notable accuracy rate of 78%.
A. Bai and S. Hira in (11) put forth a hybrid model known as the deep belief network and Softmax regression (DBN-SR) for the purpose of predicting student employability. The dataset utilized in this study was derived from questionnaires administered to students. The features selected for analysis encompassed hard and soft skills, demographic characteristics, and university-related attributes. Notably, the outcomes of the analysis exhibited a remarkable level of accuracy, reaching 98%.
M. D. Laddha et al., in (12) explains the employability of students was predicted in a study "Performance Analysis of the Impact of Technical Skills on Employability, by utilizing technical skills data acquired from institutional databases. Subsequently, various algorithms were employed for analysis. Among the classification algorithms employed, namely SVM, LR, DT, RF, AdaBoost, and NB, the RF approach has demonstrated the highest accuracy, reaching a maximum of 70%.
D. Jagan Mohan Reddy et al., in (13) have proposed a model that utilizes multiple machine learning techniques, including DT, RF, NN, and NB to predict candidate hiring. They have employed various statistical measures for feature selection, such as hard skills, demographic features, and professional experience. The results indicate that Gaussian NB achieved the highest accuracy of 99%.

3. PROPOSED SYSTEM

Data set

The dataset included in this study was acquired through a survey administered to Engineering graduates residing in the Tanjavur District of Tamil Nadu. An online survey was developed, comprising relevant inquiries, which was subsequently sent to Engineering graduates across several industries in order to obtain the intended research outcomes. The graduates were categorized into five distinct groups based on their skills: Communication, Problem Solving, Soft, Decision Making, and Analytical skills. Each category encompasses the most relevant attributes. The values assigned to the first three categories, denoted as "0" or "1," indicate whether the graduates received any training or specific coursework in these areas during their college years. The numerical value "1" indicates that the graduate has received training or completed a specific course related to those skills. Within the fourth category, the numerical value ranging from 0 to 7 represents the quantity of courses or trainings that the graduate has undertaken in order to meet the industry's prerequisites.

Machine Learning Algorithm

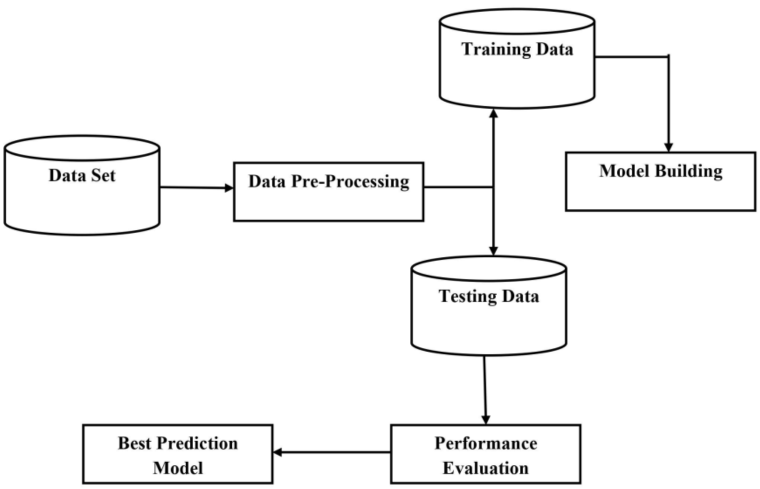

Figure 1: Proposed System Model

Data Pre-processing

The process of data preparation is a crucial step in the development of a ML model due to the inherent challenge faced by machines in interpreting raw datasets in order to generate the desired outcomes (13). Data preprocessing is a crucial step in preparing data to be compatible with a machine learning model. Initially, the process involves the removal of extraneous elements, addressing any instances of incomplete data, and ensuring the uniformity of the dataset. Next, feature selection is employed to discover the pertinent properties that enable classifiers to achieve optimal performance, hence significantly influencing the employability of engineering graduates in alignment with labor market requirements. Ultimately, the dataset was divided into two distinct sets: one including 80% of the data, which was allocated for training the model, and the other consisting of 20% of the data, designated for assessing the model's correctness and optimizing the performance of our ML models.

Prediction Model

Five distinct binary classification techniques are employed to forecast the outcomes of graduates utilizing the dataset that has been gathered. Due of its ability to classify novel findings into any of five distinct categories. The dataset contains a binary class with two distinct values: (0) representing graduates who are not qualified and do not meet the expectations of the labor market, and (1) representing graduates who are qualified. The sample size for this investigation comprises 296 records. The Python programming language was utilized, employing with different library function. There are five classification algorithms:

- Decision Tree
- Navie Bayes
- Random Forest
- Logistic Regression
- Support Vector Machine algorithm

i. Decision Tree Algorithm (DT)

The DT Algorithm can be classified as a type of supervised learning methodology, which can be conceptualized as a series of conditional statements in the form of IF-THEN rules. A hierarchical structure is created by organizing branches and nodes based on the data collected for each attribute throughout the learning process (14). The DT algorithm is utilized to construct decision trees for the purpose of addressing regression and classification issues by utilizing training data. The Gini approach was employed in our suggested model to generate split points. The task at hand is the development of a decision rule that aims to optimize the reduction in impurity inside a certain node.

$$G(x) = 1 - \sum_{m=1}^{n} P_i^2 \qquad (1)$$

In the given context, G(x) represents the Gini impurity at a certain node x, while Pi denotes the proportion of observations belonging to class n at the same node x. This decision-making procedure is repeated until either a predetermined cutoff is reached or all leaf nodes are pure.

ii. Navie-Bayes (NB)

NB, also known as Naive Bayes, is a probabilistic ML method commonly employed for various classification tasks. The methodology is based on the foundational principles of the Bayes theorem and functions on the underlying premise that the predictors exhibit independence from each other (14). The assumed probability of the attributes in our suggested model is

$$P(i|j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(x_i-\mu_j)^2}{2\sigma_j^2}\right. \qquad (2)$$

The parameters σ_j and μ_j are calculated by the utilization of greatest likelihood.

iii. Random Forest Classification (RF)

The RF algorithm can be classified as a form of supervised learning technique. The tool has the capability to be utilized for both classification and regression tasks. The initial step of this model involves the generation of a random forest consisting of several trees. The objective of conducting a vote process to merge randomly selected trees within a forest is to eliminate the tree with the highest level of predictability. In the case where a dataset comprises a collection of x features, the initial step involves the selection of a random feature, denoted as y. The algorithm subsequently endeavors to combine trees by considering the anticipated result and employing a voting mechanism (15).

iv. Logistic Regression (LR)

Logistic regression (LR) employs regression analysis, wherein a logistic regression model necessitates the inclusion of a binary classified class variable (16). In this dataset, the target column, referred to as the employability class, contains binary values indicating the employability status of Engineering graduates. A value of "0" represents graduates who are deemed unqualified and unlikely to meet the demands of the labor market, while a value of "1" indicates graduates who have been predicted to possess the necessary qualifications and meet industry requirements. In the approach proposed, a linear model is incorporated within a logistic function in the following manner:

$$P(j|i) = \frac{1}{1+e^{-(\beta_0+\beta_1 i)}} \qquad (3)$$

The determination of the probability, represented as P(j|i), that the target value j corresponds to class 1 for the i^{th} observation is based on the training data. The parameters to be learned, β0 and β1, are involved in this calculation. It is worth noting that e represents Euler's number. The primary aim of the logistic function is to assess its output as probabilities by imposing a constraint on its range, limiting it to values between 0 and 1.

v. Support Vector Machine (SVM)

The SVM classes inside the dataset should be predetermined in this model. The classification of items in a given dataset is achieved by the utilization of preexisting classes. The process involves the classification of transactions by assigning one or more categories with the aim of enhancing the precision of performance measurement.

4. RESULT AND ANALYSIS

Quality Parameters

To evaluate the effectiveness of the model, a confusion matrix is generated for the expected data, encompassing the metrics of true positive (TP), false positive (FP), true negative (TN), and false negative (FN). The assessment of the study's performance is carried out utilizing criteria such as accuracy, precision, recall, and F1 score. In the following section, a succinct depiction of each will be presented:

Accuracy (Acc): The evaluation of classifier performance is commonly measured using this metric. The computation involves determining the proportion of instances that are correctly categorised in relation to the total number of instances [17]. The formula can be expressed as follows:

$$Acc = \frac{No\ of\ instances\ classified\ correctly}{Total\ no\ of\ instances} \quad (4)$$

Precision: The calculation of this ratio involves the division of the aggregate count of positive forecasts by the count of actual positive instances.

$$Precision = \frac{No\ of\ positive\ instances}{Total\ no\ of\ predicted\ positive\ instances} \quad (5)$$

Recall: The concept of recall pertains to the quantification of accurately predicted positive experiences. The primary emphasis is placed on the model's capacity to accurately identify and encompass all occurrences that exhibit good attributes. The calculation involves determining the proportion of genuine positives in relation to the combined total of TP and FN.

$$Recall = \frac{No\ of\ positive\ instances\ predicted\ correctly}{Total\ no\ of\ predicted\ positive\ instances} \quad (6)$$

F1 Score: The average precision and recall values are obtained by combining them.

$$F1\ Score = 2 * \frac{Precision * recall}{Recision + recall} \quad (7)$$

Performance Analysis

Following the completion of data pre-processing, in accordance with the used technique, a total of 517 data points were collected. The data was divided into two sets: 80% for training and 20% for testing. The results pertaining to this investigation are given in the subsequent manner. Figure 2 displays the correlation matrix pertaining to the utilized features.

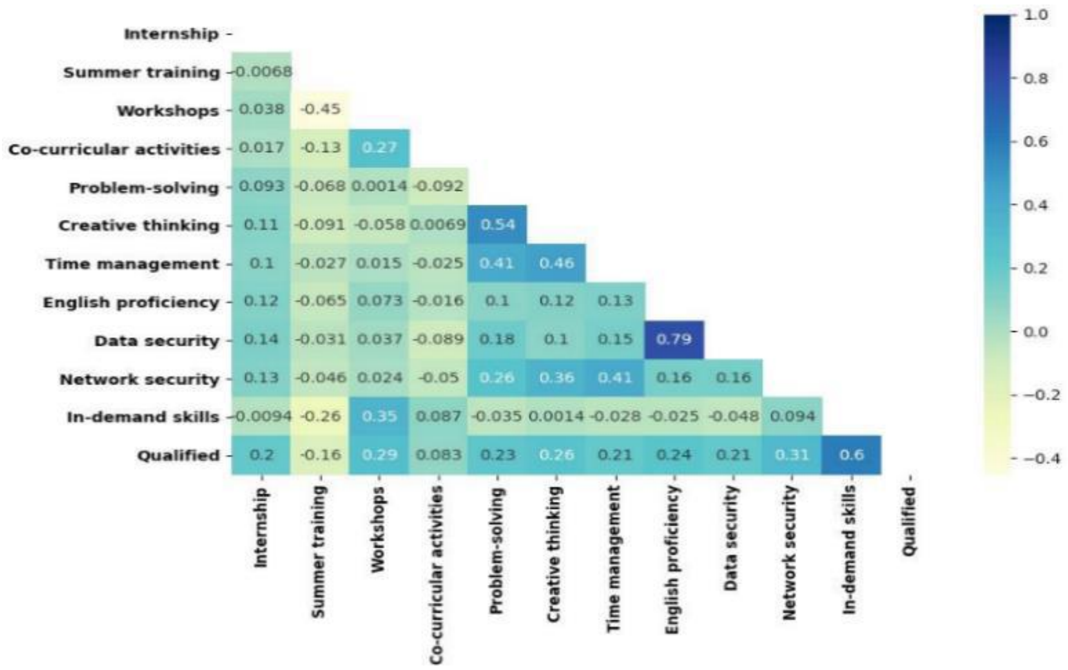

Figure 2: Correlation matrix of proposed model

The classification of the competency gap among various parameters is accomplished through the utilization of proposed classifiers. Table 1 presents the various parameters that are taken into account when assessing the competency gap across student.

Table 1: Factors for competency measurement

Factors/Algorithm	Decision tree (DT)	Navie Bayes (NB)	Random Forest (RF)	Logistic Regression (LR)	Support Vector Machine (SVM)
General competency	65	72	66	82	85
Specific competency	79	81	61	75	83
Technical competency	72	69	75	69	77
Behavioral competency	62	78	72	71	76
Learning competency	76	79	79	80	87

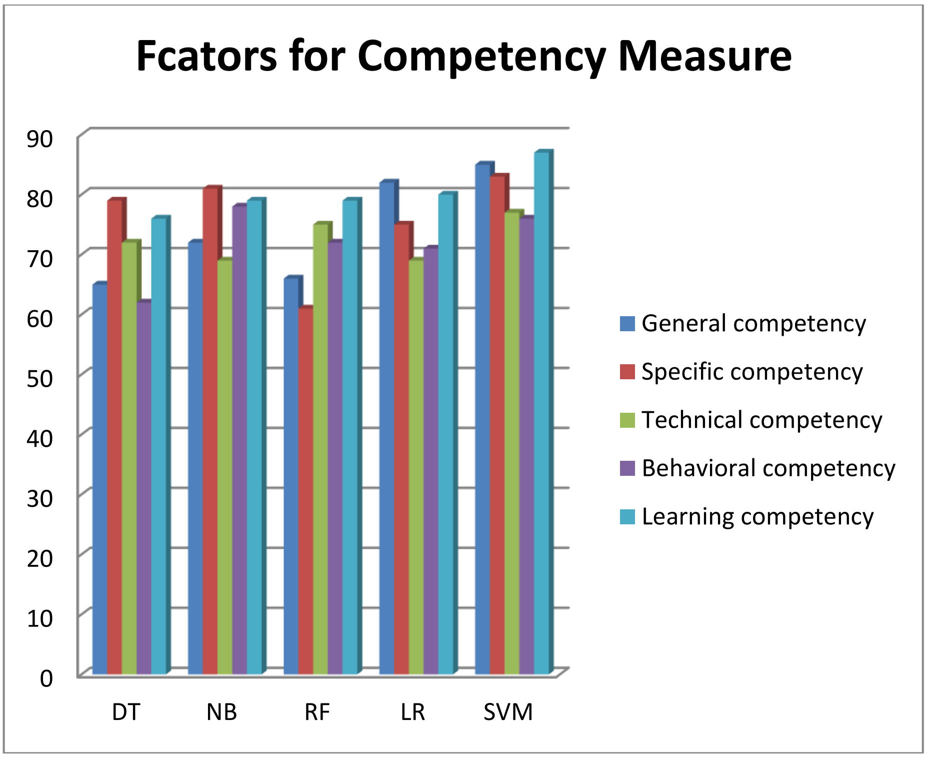

Figure 3: Classification of competency factors

Figure 3 displays the performance of the five classifiers on the competency factors. The performance of the SVM classifier exhibits superior classification capabilities compared to other classifiers.

Table 2 and Figure 4 present a comprehensive compilation of the factors that exert influence on graduates. According to the classifier result, possessing teamwork, self-motivation, and leadership skills are identified as the highest priority characteristics for students to mitigate the competency gap with the industry.

Table 2: **Factors Influencing the Competency of Engineering Graduates**

Factors/Algorithm	DT	NB	RF	LR	SVM
Team Management	68	71	79	76	78
Team Work	72	66	68	69	71
Leadership	75	78	75	81	67
Self motivation	79	73	72	73	74
Learning Competency	66	67	65	80	84

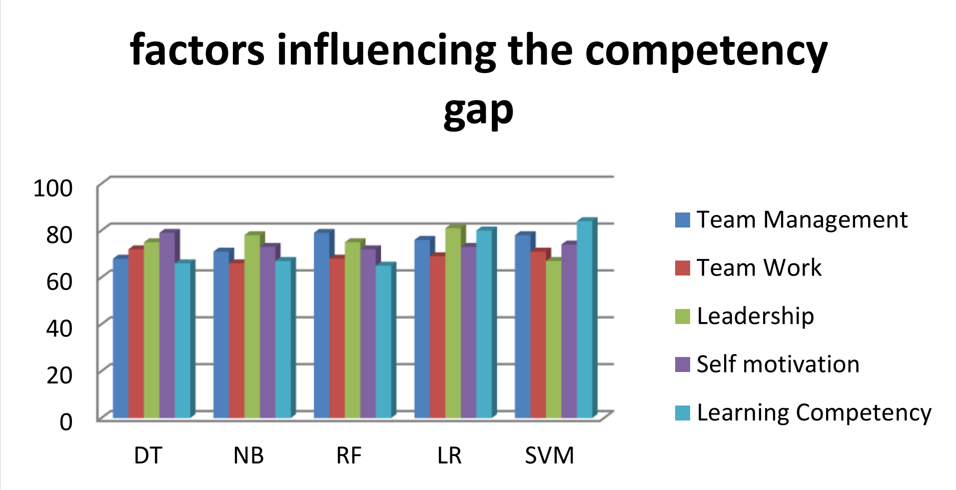

Figure 4: Factors influencing the competency gap

Table 3: Factors that expect by the Industries

Factors/Algorithm	DT	NB	RF	LR	SVM
Communication skills	79	81	69	79	82
Problem solving skills	71	72	80	71	75
Soft skills	84	76	70	74	70
Decision making skills	75	68	77	83	81
Analytical skills	70	83	73	68	69

The classification results presented in Table 3 and Figure 5 are derived from the factors anticipated by the industry. The findings indicate that problem-solving, technical thinking, and decision-making skills are the primary attributes that the industry anticipates from engineering graduates.

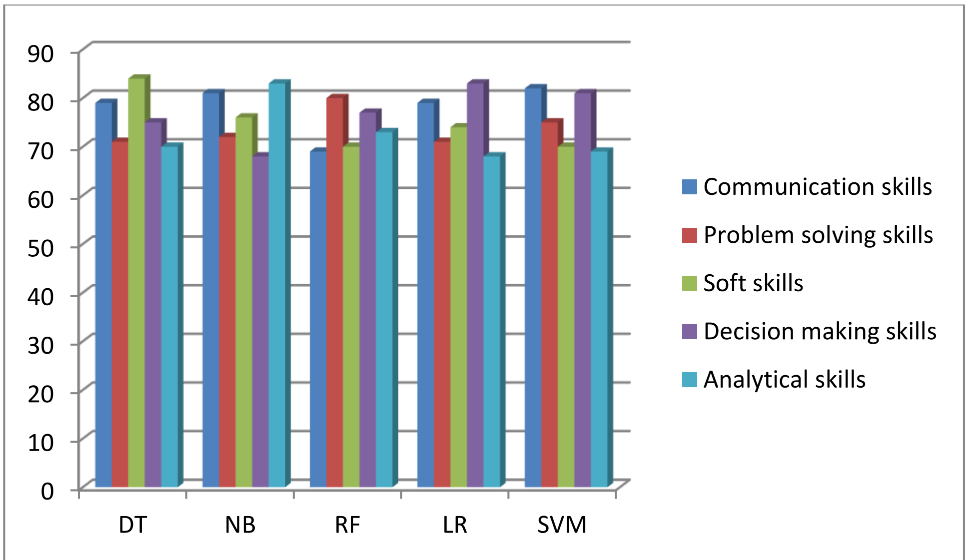

Figure 5: Classification based on industry expectation factors

Table 4: Performance Comparison among classifiers

Quality parameters	DT	NB	RF	LR	SVM
Accuracy	82	78	75	79	84
Sensitivity	75	72	69	72	79
Specificity	87	82	76	83	88
F 1-score	79	75	82	75	85

Table 4 and Figure 6 present an overview of the performance of the classifiers proposed in this study. Based on the data presented in the chart, it can be observed that the SVM exhibits a somewhat superior performance in comparison to the other categorization methods. The SVM model demonstrates an accuracy of 84%. Additionally, it exhibits sensitivity, specificity, and f-score values of 79%, 88%, and 85%, respectively.

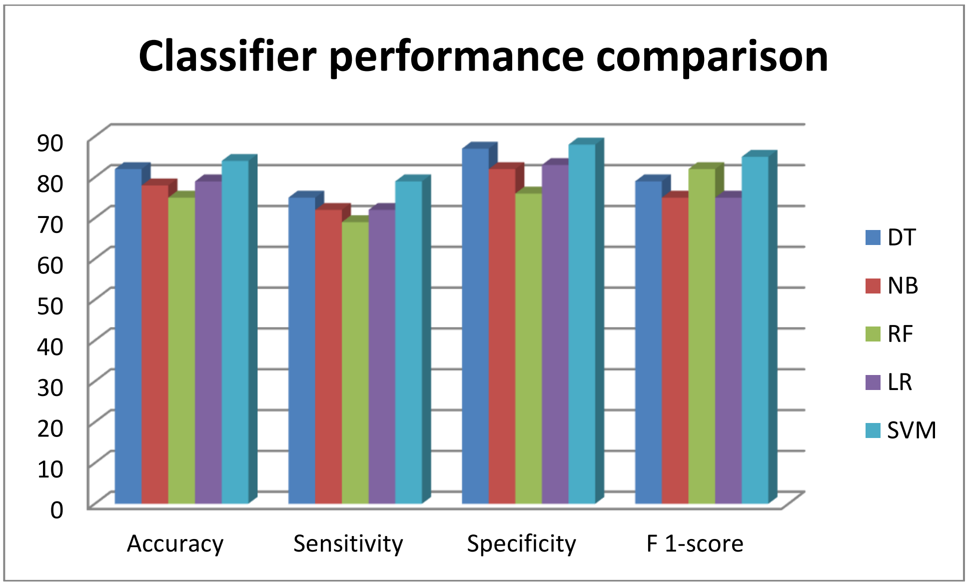

Figure 6: Performance comparison of classifiers

5. CONCLUSION AND FUTURE WORK

The annual production of graduates by educational institutions has been steadily increasing. In order to address the issue of unemployment and the discrepancy between industry expectations and demands, it is imperative to develop a predictive model that utilizes ML techniques to assess the employability of graduates and align it with the objectives of corporations. This study presents, examines, and applies five ML classification methods, specifically DT, NB, LR, RF, and SVM.

This study demonstrated a higher level of accuracy compared to previous research endeavors. The SVM algorithm demonstrates the highest level of accuracy, reaching 84%. Following closely behind is the DT algorithm, which achieves an accuracy of 82%. Conversely, the RF algorithm exhibits the lowest accuracy, achieving a rate of 75%. The primary constraint of this study is the limited size of the dataset. Based on the empirical evidence presented in the research, it can be deduced that machine learning approaches has the capacity to effectively predict the employability prospects of engineering graduates. This is achieved through the examination of diverse aspects that contribute to the existence of a competency gap.

The proposed model holds potential utility for educational institutions in enhancing their capacity to formulate more effective long-term strategies aimed at cultivating graduates who possess a comprehensive understanding, advanced skills, and the ability to meet the demands of the business. The results of the characteristics analysis revealed that there is a need to modify the curriculum to incorporate the abilities that are now in demand in the business. Additionally, enhancing the teaching and learning methods through increased training opportunities is recommended in order to generate high-quality graduates in the future. Moreover, the suggested approach is anticipated to provide substantial assistance to employers in making major contributions to the placement process.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to everyone who contributed to the completion of this research. We are deeply appreciative of the guidance and encouragement from our mentors and colleagues, whose insights were invaluable throughout the research process. We also extend our heartfelt thanks to our families and friends for their unwavering support. Finally, we are grateful to the editorial team and reviewers for considering our submission and providing constructive feedback to further improve our work.

FUNDING

No Funding Available for this work

CONFLICT OF INTEREST

All the authors are having no conflict of interest in publishing this article in this journal.

AUHOR CONRIBUTIONS

All authors are having equal contributions towards this research contribution

ETHICS APPROVAL
All procedures followed the guidelines laid out in the Declaration of journal and applicable national or institutional ethical standards. Informed consent was obtained from all participants, and their anonymity and confidentiality were ensured throughout the study.

DATA AVAILABILITY
The data used in this study are available from the corresponding author upon reasonable request. Due to reasons such as confidentiality, privacy concerns, or legal restrictions, some portions of the data may not be publicly available. However, anonymized datasets or specific data related to the findings can be shared with qualified researchers, subject to approval and adherence to any necessary data-sharing agreements.

ABBREVIATION
LR-Logistic Regression
DT- Decision tree
K-NN- k-nearest neighbor
SVM- Support Vector Machine
NB- Naïve Bayes
HELs- Higher Educational Institutions
ML-Machine Learning
IT- Information Technology
ANN- Artificial Neural Networks
DBN-SR- Deep Belief Network and Softmax Regression
TN-True Negative
FN-False Negative

6. REFERENCES

1. L. S. Hugo, "Predicting Employment Through Machine Learning." https://www.naceweb.org/career-development/trends-and-predictions/predicting-employment-through-machine-learning/ (accessed Oct. 18, 2022).
2. H. B. Kenayathulla, N. A. Ahmad, and A. R. Idris, "Gaps between competence and importance of employability skills: evidence from Malaysia," Higher Education Evaluation and Development, vol. 13, no. 2. pp. 97–112, 2019. doi: 10.1108/heed-08-2019-0039.
3. F. Biagi, J. Castaño Muñoz, and G. Di Pietro, "Mismatch Between Demand and Supply Among Higher Education Graduates in the EU," JRC Tech. Rep., pp. 1–21, 2020, doi: 10.2760/003134
4. M. I. Hossain, K. S. A. Yagamaran, T. Afrin, N. Limon, M. Nasiruzzaman, and A. M. Karim, "Factors Influencing Unemployment among Fresh Graduates: A Case Study in Klang Valley, Malaysia," Int. J. Acad. Res. Bus. Soc. Sci., vol. 8, no. 9, Oct. 2018, doi: 10.6007/IJARBSS/v8-i9/4859.
5. Z. Othman, S. W. Shan, I. Yusoff, and C. P. Kee, "Classification Techniques for Predicting Graduate Employability," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 8, no. 4–2, p. 1712, Sep. 2018, doi: 10.18517/ijaseit.8.4-2.6832.
6. M. Alghamlas and R. Alabduljabbar, "Predicting the Suitability of IT Students' Skills for the Recruitment in Saudi Labor Market," in 2019 2nd, International Conference on Computer Applications & Information Security (ICCAIS), May 2019, pp. 1–5. doi: 10.1109/CAIS.2019.8769577.
7. A. Dubey and M. Mani, "Using Machine Learning to Predict High School Student Employability – A Case Study," in 2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Oct. 2019, pp. 604–605. doi: 10.1109/DSAA.2019.00078.
8. S. R. Rahman, M. A. Islam, P. P. Akash, M. Parvin, N. N. Moon, and F. N. Nur, "Effects of co-curricular activities on student's academic performance by machine learning," Curr. Res. Behav. Sci., vol. 2, no. May, p. 100057, 2021, doi: 10.1016/j.crbeha.2021.100057.
9. C. D. Casuat and E. D. Festijo, "Identifying the Most Predictive Attributes among Employability Signals of Undergraduate Students," Proc. - 2020 16th IEEE Int. Colloq. Signal Process. its Appl. CSPA 2020, no. May, pp. 203–206, 2020, doi: 10.1109/CSPA48992.2020.9068681.
10. K. B. Aviso, J. I. B. Janairo, R. I. G. Lucas, M. A. B. Promentilla, D. E. C. Yu, and R. R. Tan, "Predicting higher education outcomes with hyperbox machine learning: what factors influence graduate employability?," Chem. Eng. Trans., vol. 81, no. 2019, pp. 679–684, 2020, doi: 10.3303/CET2081114.
11. A. Bai and S. Hira, "An intelligent hybrid deep belief network model for predicting students employability," Soft Comput., vol. 25, no. 14, pp. 9241–9254, Jul. 2021, doi: 10.1007/s00500-021-05850-x.
12. D. Jagan Mohan Reddy, S. Regella, and S. R. Seelam, "Recruitment Prediction using Machine Learning," in 2020 5th International Conference on Computing, Communication and Security (ICCCS), Oct. 2020, pp. 1–4. doi: 10.1109/ICCCS49678.2020.9276955.
13. A. Alhassan, B. Zafar, and A. Mueen, "Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 4, 2020, doi: 10.14569/IJACSA.2020.0110425.
14. C. D. Casuat, "Predicting Students' Employability using Support Vector Machine: A SMOTE-Optimized Machine Learning System," Int. J. Emerg. Trends Eng. Res., vol. 8, no. 5, pp. 2101–2106, May 2020, doi: 10.30534/ijeter/2020/102852020.

15. P. Thakar, P. Dr., and D. Manisha, "Role of Secondary Attributes to Boost the Prediction Accuracy of Students' Employability Via Data Mining," Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 11, 2015, doi: 10.14569/IJACSA.2015.061112.
16. HUGO, LINSEY S., M.S., "Analytics A Comparison of Machine Learning Models Predicting Student Employment", Individual Interdisciplinary Program, Master of Science in Higher Education and Business Ohio University, December 2018
17. Sui, F., J. C., Chang, Hsiao, H., Chen, S. Chen. "A Study Regarding the Gap Between the Industry and Academia Expectations for College Student's Employability", IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), December 2018