

Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling

Renato R. Maaliw III

*College of Engineering
Southern Luzon State University
Lucban, Quezon, Philippines
rmaaliw@slsu.edu.ph*

Bernard H. Ugalde

*Information Technology Department
University of Technology and Applied Sciences
Salalah, Dhofar, Sultanate of Oman
bernard.u@sct.edu.om*

Karen Anne C. Quing

*College of Arts and Sciences
Southern Luzon State University
Lucban, Quezon, Philippines
kquing@slsu.edu.ph*

Ace C. Lagman

*Department of Information Technology
FEU Institute of Technology
Manila, Philippines
aclagman@feutech.edu.ph*

Michael Angelo D. Ligayo

*Department of Electronics Engineering
Quezon City University
Quezon City, Philippines
michael.angelo.ligayo@qcu.edu.ph*

Abstract—Higher educational institutions have a responsibility and commitment to deliver employable graduates as it impacts their well-being and the economy. This study compared the accuracy of several classification algorithms to build an ensemble prediction model capable of forecasting graduates' employability using extensive data mining techniques. Based on the evaluation metrics, an ensemble model composed of Random Forest (RF), Support Vector Machines (SVM), and Naïve Bayes (NB) achieved the highest cross-validated accuracy score of 93.33%. Association rule mining and permutation feature importance analysis from 500 graduates of the electronics engineering program of a university revealed that grit is firmly attributed to employability, including the capabilities to acquire technical skills and professional certifications. Thus, the knowledge gained can be used to develop a range of policies, initiatives, and strategies to increase students' employment prospects.

Keywords—association rule mining, classification, clustering, feature importance, grit, non-cognitive skills

I. INTRODUCTION

Producing employable graduates is an essential component of the educational process encompassing the complete instructional spectrum objectives, from knowledge transmission, cognition, development of skills, and positive personal qualities. Employability refers to the probability of an individual getting hired based on their ability to adapt with the changing employer and client needs to fulfill their potential profession [1]. In recent decades, its conception has gained attention, becoming the principal focus and priority of all higher educational institutions (HEIs).

According to a published document by the United Nations, Educational, Scientific, and Cultural Organization (UNESCO), global enrollment in HEIs is projected to quadruple from 100 million in 2000 to 472 million in 2035. The current report highlights that the global unemployment rate was 6.30%, compared to 5.37% in 2019 – this is a considerable increase. Meanwhile, the Department of Labor and Employment (DOLE) in the Philippines pointed job mismatch as one of the top factors causing Filipinos to lose out of work. In an article written by [2], the country registered an unemployment rate of 8.90% while underemployment was

at 14.70%, the highest among Southeast Asian nations. These numbers are accentuated further by a report from [3] that even college graduates (21.60%) are experiencing unemployment. These stark figures signal enormous issues for the future generation if no intervention is done to halt this trend in the way economies are creating jobs. Thus, the graduates need to strengthen their abilities to meet specific expectations to improve the employment prospects. It starts with their initial education to provide them adequate skills for any profession to overcome barriers, whether seeking a secure career or creating their employment through entrepreneurship. Numerous studies exist to determine the factors affecting graduates' employability, but their prediction is still in infancy due to the term's multiple contextual definitions. Experts proposed different frameworks and models to address these challenges.

The authors [4] use psychological methodologies to investigate the performance score of entry-level engineers centered on skill weights via questionnaires. Their studies infer that soft skills are more significant than technical skills. It is also reflective of [5] regarding the job-readiness of information technology (IT) professionals. Myriad published papers utilized a cross-sectional retrospective method survey based on students' tracer of business, economics, education, and technology [6 - 9]. It concluded that curriculum-related issues were the most significant employability factors in an educational perspective. Regarding the methods applied in the studies, it is worth noting that they primarily rely on descriptive statistics and basic correlation criterion for their analysis that are weak in revealing hidden patterns. The evolution of data science prompted researchers to implement machine learning algorithms. Among the strategies by [10] to rank key variables of getting hired using actual data are the Bayesian network and decision tree. In a recent attempt, scientists employed multiple approaches to forecast the job status of new graduates [11 – 13], including one-dependence estimators (AODE), Naïve Bayes, Logistic Regression, ID3, C4.5, and NBTree. On the other hand, researchers used hierarchical clustering and random forest to assess how students acquire work after finishing their degrees, and they recommended validation strategies to refine the models. The majority of approaches cited in the literature are exclusively

concerned with algorithmic accuracy evaluation rather than interpreting the knowledge extracted. In addition, the datasets are imbalanced with limited features incapable of capturing the overall picture, which can result in overfitting and instability. Furthermore, there are few studies about the engineering graduates' employment prediction.

To address these gaps, our main objective is to discover unknown patterns that could predict graduate's employability from cognitive and non-cognitive data from numerous sources. Our paper diverges from prior publications as we provide interpretations on uncovered knowledge using intuitive ensemble machine learning-based approaches to generate predictive models on employment, thereby, creating metric insights to benefit tertiary graduates.

II. METHODOLOGY

We utilized knowledge discovery in databases (KDD) framework to extract undiscovered knowledge from several repositories. It entails various steps like selection, preprocessing, transformation, feature engineering, evaluation, and interpretations [14]. The analysis was carried out using a number of statistical and machine learning packages such as *numpy*, *pandas*, *sklearn*, *matplotlib*, *seaborn*, and *eli5* using the Python programming language. The following subsections cover each implementation.

A. Data Acquisition, Features, and Preprocessing

We collected the information of 500 electronics engineering graduates from a university in the Philippines who earned their degrees from 2014 to 2019. They were purposively chosen due to their identical curriculums. Moreover, their profession has consistently been in the top 30 of the most sought occupations for the last ten years, according to statistics from Jobstreet.com, Asia's leading online job market [15]. These circumstances established an unbiased baseline as employability is influenced by external factors such as market demands and economic conditions.

Unlike other studies, we extensively consulted different sources (graduate tracer) to acquire their profiles, scholastic ratings (cognitive), on-the-job training assessments from direct supervisors (non-cognitive), and behaviors in a learning management system for selected online courses (C++, Data Communications, and Signals & Spectra). Although not part of a graduate's internship evaluation, we also included the concept of grit (passion & perseverance) as pertinent information. Current articles insinuate its importance on an individual's success. Table 1 presents the category, features, values, and encoding, including their employability classification.

The bulk of the attributes are self-explanatory; yet, it is our obligation to the readers to articulate our employability context clearly. Graduates who have secured a job or started a business relevant to their academic degrees are employable. Meanwhile, underemployed individuals work in low-skilled, low-wage occupations (e.g., engineers working as factory line workers or bartenders). Finally, less employable are those actively seeking opportunities but cannot find work. Our typology is the proper levels of employability not explored in

previous researches. We purposefully conceptualized the classification with a threshold of one year based on the available data provided by [17]. Results from the said study showed that most graduates, on average, spend six months in preparing for the electronics engineering board examinations before seeking employment. Furthermore, the researchers are not dismissing the ability of the graduates to seek employment beyond the set threshold. In fact, it is considered as one this study's limitations.

TABLE 1
EMPLOYABILITY FEATURES, VALUES, AND ENCODING

Category	Features	Values & Encoding
Profile	Sex	Male (1), Female (2)
	Family income ^a	Low (1), Middle (2), Upper (3)
Cognitive	Technical skills ^b	Pass (1), Fair (2), Good (3), Very Good (4), Excellent (5)
	Academic distinctions ^c	No (1), Yes (2)
Non-Cognitive	Professional certifications ^d	No (1), Yes (2)
	Attendance & punctuality	
	Work habits	
	Quality of work	
	Judgment	
	Cooperation	
	Honesty & dependability	
	Communication skills	
	Safety consciousness	
	Relationship with superior	
	Relationship with co-workers	
	Emotional stability & maturity	
	Leadership	
	Grit	Low (1), High (2)
	Class Labels	Employability
		Less employable (1), Underemployed (2), Employable (3)

a. Social classes based on monthly household income in the Philippines [16]

b. Excellent (1.00 – 1.24), Very Good (1.25 - 1.74), Good (1.75 - 2.49), Fair (2.50 – 2.99), Pass (3.00)

c. Graduated with cum laude, magna cum laude, summa cum laude or equivalent

d. Electronics engineering (ECE) board examinations given by Professional Regulation Commission (PRC) [17]

B. Grit Levels Identification

Grit is a psychological construct drawn from positive psychology that emphasizes perseverance as a determinant of long-term success and is related to attaining aspirational objectives over long periods. According to [18], grit's ingredients are the consistency of interests that exhibit an individual's ability to stay resolute and fixated on long-term tasks and goals; persistence of effort portrays the power to keep going even when there are hindrances and mishaps. The characteristic can be measured using the grit scale questionnaire [19]. In reality, most people desire to create a favorable image of themselves when partaking in surveys and may distort the truth – this is the social desirability concept. Instead, we quantify the attributes using clustering based on their learning management system (LMS) behaviors. Table 2 shows the variables we formulated for grit level identification.

TABLE 2
GRIT RELATED ATTRIBUTES IN LMS

Relevant Behavior	Value
Repetitively attempting advanced self-assessment tests until achieving a passing score.	Total number of completed attempts
Logs to the supplementary online courses beyond designated class hours	Total number of hours

A self-assessment test is a kind of examination where the results are not graded. Attempting to complete such activity requires a high level of intrinsic motivation and persistence to

achieve a passing score [20]. Grittier persons based on [21] spend more time studying beyond the required class hours, indicating consistent effort towards a long-term goal. We identified grit levels using a non-parametric density-based clustering algorithm (DBSCAN) illustrated in Figure 1.

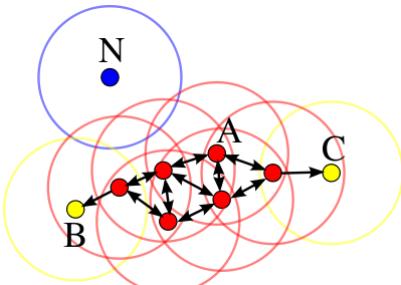


Fig. 1. Illustration of a DBSCAN clustering algorithm

It utilizes a straightforward estimation of the minimal density level, based on a specified value for the number of neighbors (*minPts*) with parameter of 4, within a specified radius (*epsilon*), and indicated by circles. Inside this radius, objects having more than the *minPts* neighbors are density reachable (arrows) *corepoint* A, while non-core points are called *border points* (B & C) because these points are density connected. N is described as a noise point because it is not density accessible. This is a proven strategy to efficiently reduce noise and detect outliers (for removal) producing good results than K-means clustering [22].

We calculated the optimized epsilon parameter of 1.5 (maximum) using a curvature graph via the K-Nearest Neighbor (KNN) shown in Figure 2, *minPts* values of 4 via a ruling (the number of features multiplied by 2), and the derived grit clusters depicted in Figure 3.

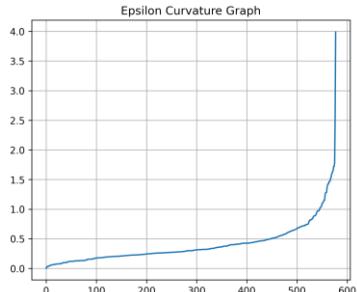


Fig. 2. Epsilon curvature graph via KNN

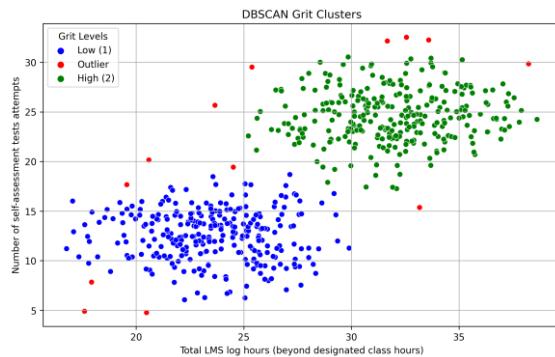


Fig. 3. Derived grit clusters via DBSCAN algorithm

C. Data Transformation

As a general rule, feature scaling is ubiquitous in boosting variable subset selection and machine learning performance. Using the original data with different ranges puts a lot of weight on larger variable values resulting to convoluted models. We used a *min-max* scaling technique to solve this problem as our data do not follow a Gaussian distribution to transform them to a scale from 0 to 1. Equation 1 displays the mathematical formula:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Where x' is the scaled data, x_{\max} and x_{\min} are the largest and smallest values of the features, respectively.

D. Two-Stage Feature Selection

The side-effects of high dimensional data (18 features) are prediction complexity and diminished generalizability. As a solution, feature elimination is necessary to create smaller datasets with fewer variables. The tradeoff is to sacrifice a loss of information in favor of model simplicity. The research adopted a two-stage feature selection approach based on unsupervised principal component analysis (PCA) and supervised linear discriminant analysis (LDA) to curtail the size of the features while keeping maximum amounts of entropy. We utilized PCA's explained variance ratio for the first stage – a measure indicating a component's relevance before building a model. A 75% aggregated value or better is ideal for preserving certainty. According to Figure 4, the first two components account for a total of 90.80% of variation – PC1 (79.52%) and PC2 (11.28%).

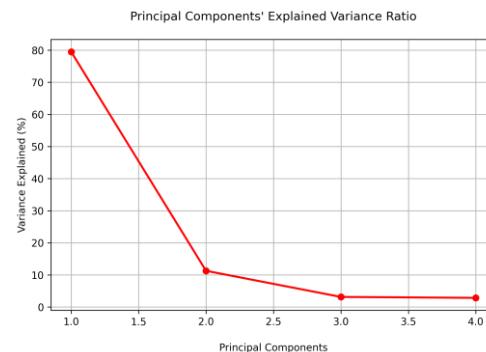


Fig. 4. Explained variance ratios of principal components

Table 3's feature loadings indicated that grit (0.414), professional certification (0.412), technical skills (0.395), communication skills (0.386), work habits & attitude (0.281), and leadership (0.279) are the contributors for each principal component.

Attribute	FEATURE LOADINGS BASED ON PRINCIPAL COMPONENTS			
	PC1	PC2	PC3	PC4
Grit	0.414	0.134	0.031	0.019
Professional certification	0.412	0.128	0.082	0.023
Technical skills	0.395	0.123	0.147	0.024
Communication skills	0.386	0.118	0.131	0.013
Work habits & attitude	0.146	0.281	0.121	0.021
Leadership	0.141	0.279	0.133	0.018

We further refined the subset using LDA. This multi-stage strategy according to the work of [23] is an efficient way to reduced dimensionality with accuracy. Based on the graph illustrated in Figure 5, removal of the variable ‘Leadership’ provides an even better class boundary with only a 2.18% information loss.

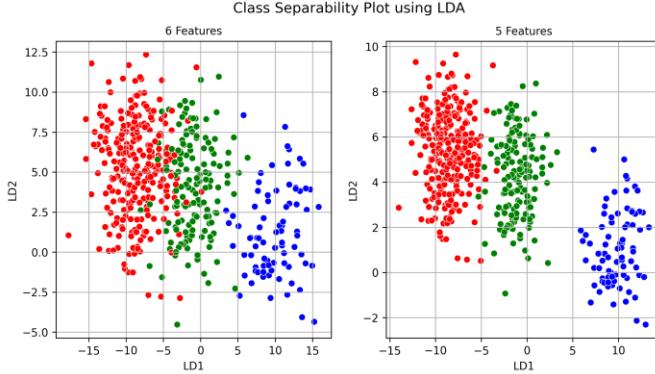


Fig. 5. Comparative LDA class separability plots with 6 and 5 features

E. Ensemble Classification and Hyperparameter Tuning

We tested our final dataset to various classification models using a 70/30 training and testing split, spread into two phases. The training portion contains the data that the algorithms can learn from, while the test set was used to measure its prediction capabilities.

During the first phase, we benchmarked the accuracies of different classifiers such as Artificial Neural Network (ANN), Decision Trees (DT), K-Nearest Neighbors (KNN), Logistic Regression (LogReg), Naïve Bayes (NB), Random Forest (RF), and Support Vector Machines (SVM) using a 10-fold cross-validation. This method allows each classifier to train on multiple equal subsets of classes on train-test samples. Furthermore, we implemented a GridSearch technique to find the optimized hyperparameters by observing weighted accuracy changes for each model via automatic tuning versus manual configuration.

At the last phase, we opted to combine the prediction power of the top three performing classifiers using *max-voting*. It produces a final prediction based on each base model’s majority votes (prediction). These mechanisms allow us to pre-analyze and gauge each classification better.

F. Evaluation and Interpretation

We used standard measurements such as accuracy, precision, recall, F1-score, cross-validation, confusion matrix, receiver operating characteristics (ROC), and area under the curve (AUC) to estimate the machine learning algorithms’ prediction quality and reliability. Interpretation of knowledge was handled through permutation feature importance to evaluate the impact of an attribute’s presence or absence on a model, and association rule mining.

III. RESULTS

This section details the outcomes of hyperparameter configurations, evaluation metrics, and the knowledge or patterns learned from the data.

A. Hyperparameter Optimization

Hyperparameters are user-specified settings in contrast to a model’s learned parameters. Supplying appropriate values enhance a machine learning performance, yet it is one of the most laborious and neglected task in the data mining process. It is a computationally expensive process but the benefits outweigh its disadvantages. Table 4 shows the optimized configurations for each classifier before the validation testing.

TABLE 4
OPTIMIZED HYPERPARAMETER CONFIGURATIONS USING GRIDSEARCH

Classifier	Settings
ANN	solver = ‘adam’, batch_size = 30, activation = ‘relu’
DT	max_depth = 5, max_leaf_nodes = 10, min_samples_leaf = 1, splitter = ‘best’
KNN	n_neighbors = 9, weights = ‘uniform’, metric = ‘manhattan’
LogReg	solver = ‘newton-cg’, penalty = ‘l2’
NB	var_smoothing = 0.0314
RF	max_depth = 5, max_features = ‘log2’, n_estimators = 20
SVM	C = 100, gamma = 0.01, kernel = ‘rbf’

B. Individual Model’s Accuracy Scores

Table 5 shows the evaluation metrics results of different classification algorithm for employability prediction. It discloses that the RF achieves the highest accuracy with 94% followed by SVM (93%), and ANN (92%). DT, NB, and KNN yielded 90%, 89%, 89%, while LogReg performs the least with 87% due to its deficiency in multi-classifications [24].

TABLE 5
SUMMARY OF INDIVIDUAL MODEL’S EVALUATION

Classifier	Metrics	Precision	Recall	F1-Score
Random Forest	0.94	0.91	0.93	0.92
Support Vector Machine	0.93	0.86	0.88	0.87
Artificial Neural Network	0.92	0.89	0.91	0.90
Decision Tree	0.90	0.86	0.86	0.86
K-Nearest Neighbor	0.89	0.82	0.83	0.83
Naïve Bayes	0.89	0.84	0.84	0.84
Logistic Regression	0.87	0.78	0.79	0.79

C. Individual Model’s Cross-Validated Accuracy Scores

Table 6 demonstrates the RF’s superiority to other classification approaches with an accuracy of 91.36% because of its intrinsic ability to handle multiclass problems [25], followed by the SVM and NB with 90.83% and 89.37%. The top three performing classifiers were subjected to an ensemble algorithm to create the final prediction model.

TABLE 6
10-FOLD CROSS-VALIDATED MEAN ACCURACY SCORES

Classifier	Accuracy
Random Forest	91.36%
Support Vector Machine	90.83%
Naïve Bayes	89.37%
Artificial Neural Network	88.53%
Decision Tree	88.12%
K-Nearest Neighbor	87.63%
Logistic Regression	86.50%

D. Ensemble Model’s Evaluation

We combined the prediction capabilities of RF, SVM, and NB to construct the ensemble model to surpass the accuracies

of individual classification through the (hard) majority of prediction votes, implemented with *sklearn's* '*VotingClassifier*'. The results fetched a cross-validated accuracy of 93.33%, a mean increase of 2.73% compared to standalone models. Table 7 shows the distribution of correctly classified and wrong predictions that the ensemble has made. It signifies that it was easier to discern graduates who found actual and missed out of work (employable & less employable) than the underemployed ones.

TABLE 7
ENSEMBLE MODEL'S CONFUSION MATRIX

Test Size, N = 150 (30%)	Predicted: Employable	Predicted: Underemployed	Predicted: Less Employable
Actual: Employable	76	2	0
Actual: Underemployed	1	18	4
Actual: Less Employable	2	1	46

We also generated ROC and AUC curve plots to measure the ensemble model's prediction quality and generalization capabilities. The graph empirically defines the capacity to distinguish and identify classes' degrees of separability. Figure 6 reveals AUC scores of 0.98, 0.90, and 0.84 explaining substantial prediction ratings beyond random guessing (red line). Likewise, micro and macro averages of 0.91 reinforced its robustness.

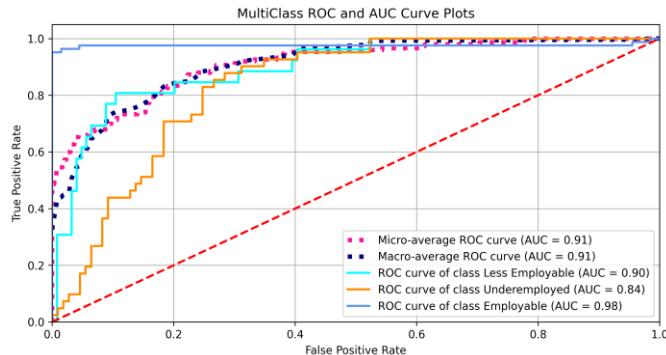


Fig. 6. ROC and AUC plots of the ensemble model's prediction quality

E. Model Interpretation

'Black box' models such as ensembles produce higher prediction accuracy than straightforward frameworks (e.g., linear regression, decision trees) with fewer parameters. However, they are exceedingly difficult to interpret because the knowledge is concealed inside its stochastic structure, making the relationship between predictors and predictions extremely challenging to be mapped out. To unravel this dilemma, we calculated mean decrease accuracy (MDA) regularities using permutation feature importance.

It is a scheme to solve the comprehensibility crisis by randomly shuffling attribute values and quantifying their impact on the model's predictive power. A decline in the average accuracy score is indicative of the feature's relevance. Table 8 shows the mean feature importance using ten arbitrary exchanges.

TABLE 8
ATTRIBUTE'S IMPORTANCE BY RANK

Weight	Feature
0.2446 ± 0.0678	Grit
0.1413 ± 0.0331	Professional certifications
0.1181 ± 0.0441	Technical skills
0.0842 ± 0.0275	Communication skills
0.0531 ± 0.0231	Work habits

The analysis confirmed that grit (0.2446) was the strongest influential predictor in the model's prediction accuracy, followed by professional certifications (0.1413), technical skills (0.1181), and communication skills (0.0842). Work habits were the weakest contributor.

In addition, we observed common patterns and co-occurrences using association rule mining (apriori) – a rule-based machine learning process for discovering interesting correlations between variables [26]. Table 9 captures the results with a minimum support configuration of 0.5.

TABLE 9
INTERESTING PATTERNS ON EMPLOYABILITY

Category	Frequent Patterns	Support
Employable	(High grit)	0.952
	(Has professional certifications)	0.879
	(High grit, Has professional certifications)	0.860
	(High grit, Satisfactory technical skills)	0.534
Underemployed	(Has professional certifications, Satisfactory technical skills)	0.523
	(Low grit)	0.912
	(No professional certifications)	0.809
	(Low grit, No professional certifications)	0.802
	(Low grit, Fair technical skills)	0.732
Less Employable	(Fair technical skills, No professional certifications)	0.625
	(No professional certifications)	0.942
	(Low grit)	0.917
	(Fair technical skills)	0.887
	(Low grit, Fair technical skills, No professional certifications)	0.875
	(Fair communications skills, No professional certifications, Fair technical skills)	0.712

It exposed that the graduate's level of grit is the primary determinant with support of 0.952 (employable) and 0.912 (underemployed). The second indicator was the professional certifications (0.879 & 0.809). For the less employable individuals, it warranted that the lack of professional certifications (0.942), low grit (0.917), and fair technical skills (0.887) affected their employability.

IV. DISCUSSIONS

An accurate ensemble model was built to predict the employment status of electronics engineering graduates through extensive optimization and evaluation procedures [27] [28]. The research discovered that grit has a profound effect on employability as it is the top-ranked and dominant attribute appearing on patterns.

Our findings indicated that gritty graduates are highly technically skilled and passed the board examinations, which leads them to employment aligned with their specializations. It is a critical trait described by [29] in earning a demanding engineering degree. In addition, the investigation of [30 - 34] is in unity with our results, citing that grit drives a person to persevere and persists even in adversities. One possible explanation for this is that grit has been found to have a direct relationship and is considered a fundamental component of a

growth mindset [35]. Individuals with growth mindset tend to see their intelligence and abilities as malleable and can be improved with consistent effort [36]. Thus, gritty graduates often have the confidence to explore their cognitive development and learn from their difficulties. They often persevere until they have reached their goals. It is the complete opposite for underemployed and less employable graduates who have low grit levels, lack of professional certifications, and technical skills. Relatively, low levels or lack of grit are often associated with a fixed mindset, which leads to the belief that one's intelligence and abilities are inborn and cannot be enhanced [35]. Moreover, our secondary results differ from [37] [38], stressing that employers preferred communication over technical skills. In the case of highly specialized professions such as engineering, it was noted otherwise [39] [40].

These new perspectives have significant implications for colleges and universities traditionally seen as a training ground for developing technical skills in response to the workforce needs. A comprehensive strategy must be put in place by HEIs to teach and produce graduates with grit in preparation for their professional future. This paper is limited only to the characteristics of individuals that do not capture other significant data such as curriculum design, school reputation, ranking, and prestige that also influenced employability.

V. CONCLUSIONS AND RECOMMENDATIONS

Employability is a criterion inextricably linked to the quality of education given by tertiary institutions since they play a principal role in their graduates' academic and professional growth – including non-cognitive skills. They must prepare their students for professional careers in the face of increasing competition and demands for a competent workforce. This research analyzes student-related variables utilizing rigorous data mining approaches to construct an ensemble prediction model for employability outcomes.

The main contribution of our work is the discovery of a unique viewpoint on the importance of developing and cultivating students' grit during their academic life to help them overcome challenges and become persevering to achieve their long-term goals. Academic institutions have always focused on developing cognitive skills, but it is also worthy of teaching students other psychological qualities like grit. Conversely, it was inferred that technical skills and professional certifications are rooted on grit, and combinations of these attributes are significant predictors of graduate employment. Identifying additional intrinsic and extrinsic attributes and implementing other machine learning techniques to create an even better model will be explored in the following stages of the research.

REFERENCES

- [1] L. Small, K. Shacklock and T. Marchant, "Employability: A contemporary review for higher education stakeholders", *Journal of Vocational Education & Training*, 70(1), pp. 148-166, 2018.
- [2] PSA, "Unemployment rate in September 2021 is estimated at 8.9 percent, Philippine Statistic Authority, Available at: <https://psa.gov.ph/content/unemployment-rate-september-2021-estimated-89-percent>, 2021.
- [3] E. Rosero, "Why many fresh college grads don't get hired, according to survey of managers", GMA News, Available at: <https://www.gmanetwork.com/news/money/economy/250239/why-many-fresh-college-grads-don-t-get-hired-according-to-survey-of-managers/story/>, 2021.
- [4] Y. Yusoff, M. Omar, A. Zaharim, A. Mohamed and N. Muhamad, "Employability skills, performance score, for fresh engineering graduates in Malaysian industry", *Asian Social Science*, 18(16), pp. 140-145, 2012.
- [5] A. Albina and L. Sumagaysay, "Employability tracer study of information technology education graduates from a state university in the Philippines", *Social Sciences & Humanities Open*, 100055, pp. 1-6, 2020.
- [6] L. Pinto and D. Ramalheira, "Perceived employability of business graduates: The effect of academic performance and extracurricular activities", *Journal of Vocational Behavior*, vol. 99, pp. 165-178, 2017.
- [7] F. Suleman, "The employability skills of higher education graduates: Insights into conceptual frameworks and methodological options", *Higher Education*, 76(2), pp. 263-278, 2018.
- [8] S. Monteiro, L. Almeida, C. Gomes, and J. Sinval, "Employability profiles of higher education graduates: A person-oriented approach", *Studies in Higher Education*, pp. 1-14, 2020.
- [9] S. McKenzie, J. Coldwell-Neilson and S. Palmer, "Understanding the career development and employability of information technology students", *Journal of Applied Research in Higher Education*, 10(4), pp. 456-468, 2018.
- [10] Z. Othman, S. Shan, I. Yusoff and C. Kee, "Classification techniques for predicting graduate employability", *International Journal on Advanced Science, Engineering and Information Technology*, 8(4-2), pp. 1712-1720, 2018.
- [11] A. Dubey and M. Mani, "Using machine learning to predict high school student employability – A case study", *2019 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 604-605, 2019.
- [12] C. Casuat and E. Festijo, "Predicting students' employability using machine learning approach", *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pp. 1-5, 2019.
- [13] K. Aviso, J. Janairo, R. Lucas, M. Promentilla, D. Yu and R. Tan, "Predicting higher education outcomes with hyperbox machine learning: What factors influence graduate employability?", *Chemical Engineering Transactions*, vol. 81, pp. 679-684, 2020.
- [14] N. Oliver, F. Pérez-Cruz, S. Kramer, J. Read and J. Lozano, "Machine learning and knowledge discovery in databases", *Research Track*, 12975, Springer Nature, 2021.
- [15] JobStreet, Accessed: November 1, 2021, [Online], Available at: <https://www.jobstreet.com.ph>
- [16] K. Domingo, "Who are identified rich, poor? Gov't shows income class brackets in PH", ABS-CBN News, Available at: <https://news.abs-cbn.com/news/09/17/20/who-are-identified-rich-poor-govt-shows-income-class-brackets-in-ph>, 2020.
- [17] R. Maaliw, "Early prediction of electronics engineering licensure examination performance using random forest," *2021 IEEE World AI IoT Congress (AIoT)*, pp. 41-47, 2021.
- [18] A. Duckworth, "Grit: Why passion and resilience are the secrets to success", *NIDA Journal of Language and Communication*, 24(35), pp. 105-106, 2019.
- [19] K. Quing and J. Baudin, "The roles of internal locus of control and grit in managing the academic stress of senior high school students in Southern Luzon State University", *Journal of Education, Management and Development Studies*, 1(1), pp. 1-9, 2021.
- [20] M. Ryan and E. Deci, "Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions", *Contemporary Educational Psychology*, vol. 61, 101860, 2020.

- [21] E. Huéscar Hernández, J. Moreno-Murcia, L. Cid, D. Monteiro and F. Rodrigues, "Passion or perseverance? The effect of perceived autonomy support and grit on academic performance in college students", International Journal of Environmental Research and Public health, 17(6), 2143, 2020.
- [22] E. Schubert, J. Sander, M. Ester, H. Kriegel and X. Xu, "DBSCAN revisited: Why and how you should (still) use DBSCAN", ACM Transactions on Database Systems (TODS), 42(3), pp. 1-21, 2017.
- [23] M. Li, H. Wang, L. Yang, Y. Liang, Z. Shang and H. Wan, "Fast hybrid dimensionality reduction method for classification based on feature selection and grouped feature extraction", Expert Systems with Applications, vol. 150, 113277, 2020.
- [24] R. Couronné, P. Probst and A. Boulesteix, "Random forest versus logistic regression: A large-scale benchmark experiment", BMC bioinformatics, 19(1), pp. 1-14, 2018.
- [25] R. Primartha and B. Tama, "Anomaly detection using random forest: A performance revisited", 2017 International Conference on Data and Software Engineering (ICoDSE), pp. 1-6, 2017.
- [26] A. Telikani, A. Gandomi and A. Shahbahrami, "A survey of evolutionary computation for association rule mining", Information Sciences, vol. 524, pp. 318-352, 2020.
- [27] R. Maaliw, "A personalized virtual learning environment using multiple modeling techniques", 2021 Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), pp. 8-15, 2021.
- [28] R. Maaliw, M. Ballera, Z. Mabunga, A. Mahusay, D. Dejelo and M. Seño, "An ensemble machine learning approach for time series forecasting of COVID-19 cases", 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 633-640, 2021.
- [29] G. Zamarro, M. Nichols, A. Duckworth and S. D'Mello, "Validation of survey effort measures of grit and self-control in a sample of high school students", PloS One, 15(7), e0235396, 2020.
- [30] W. Widodo and C. Chandrawaty, "Assessing the effect of grit and employability on organizational commitment mediating by job involvement", Management Science Letters, 10(11), pp. 2403-2410, 2020.
- [31] N. Butz, R. Stratton, M. Trzebiatowski and T. Hillery, "Inside the hiring process: How managers assess employability based on grit, the big five, and other factors", International Journal of Business Environment, 10(4), pp. 306-328, 2019.
- [32] H. Mashod, H. Kura and K. Ringim, "Theorising the boundary condition between psychological capital and graduate employability: Does grit matter?", ETH Learning and Teaching Journal, 2(2), pp. 78-83, 2020.
- [33] J. Kasler, L. Zysberg and N. Harel, "Hopes for the future: demographic and personal resources associated with self-perceived employability and actual employment among senior year students", Journal of Education and Work, 30(8), pp. 881-892, 2017.
- [34] S. McGinley, A. Mattila and T. Self, "Deciding to stay: A study in hospitality managerial grit", Journal of Hospitality & Tourism Research, 44(5), pp. 858-869, 2020.
- [35] A. Hochanadel and D. Finomore, "Fixed and growth mindset in education and how grit helps students persist in the face of adversity", Journal of International Education Research – 1st Quarter, 11(1) pp. 47-50, 2015.
- [36] S. Keesey, A. Schaefer, M. Loy and C. Allen, "Developing growth mindset and grit in preservice teachers", Kentucky Teacher Education Journal: The Journal of the Teacher Education Division of the Kentucky Council for Exceptional Children, 5(1), pp. 1-7, 2018.
- [37] P. Sotiriadou, D. Logan, A. Daly and R. Guest, "The role of authentic assessment to preserve academic integrity and promote skill development and employability", Studies in Higher Education, 45(11), pp. 2132-2148, 2020.
- [38] A. Moumen, E. Bouchama and Y. El Bouzekri El Idirissi, "Data mining techniques for employability: Systematic literature review", 2020 IEEE 2nd International Conference On Electronics, Control, Optimization and Computer Science (ICECOCS), pp. 1-5, 2020.
- [39] A. Fadhilah, M. Noraini, A. Noreliza and K. Chua, "Analysis on the employability skills of UNITEN graduates in engineering program", 2018 IEEE 10th International Conference on Engineering Education (ICEED), pp. 34-38, 2018.
- [40] M. Mutambala, G. Sheikheldin, B. Diyamett and B. Nyichomba, "Student industrial secondments in East Africa: Improving employability in engineering", 2020 IFEES World Engineering Education Forum - Global Engineering Deans Council (WEF-GEDC), pp. 1-5, 2020.



1 of 1

 [Download](#)
 [Print](#)
 [Save to PDF](#)
 [Add to List](#)
 [Create bibliography](#)

2022 IEEE 12th Annual Computing and Communication Workshop and Conference, CCWC 2022 • Pages 288 - 294 •
 2022 • 12th IEEE Annual Computing and Communication Workshop and Conference, CCWC 2022 • Virtual, Las Vegas • 26
 January 2022 through 29 January 2022 • Code 17778

Document type

Conference Paper

Source type

Conference Proceedings

ISBN

978-166548303-2

DOI

10.1109/CCWC54503.2022.9720783

[View more](#)

Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling

Maaliw, Renato R.^a ; Quing, Karen Anne C.^b ;
 Lagman, Ace C.^c ; Ugalde, Bernard H.^d ; Ballera, Melvin A.^e ;
 Ligayo, Michael Angelo D.^f
 [Save all to author list](#)

^a College of Engineering, Southern Luzon State University, Quezon, Lucban, Philippines^b College of Arts and Sciences, Southern Luzon State University, Quezon, Lucban, Philippines^c Feu Institute of Technology, Department of Information Technology, Manila, Philippines^d University of Technology and Applied Sciences, Information Technology Department, Dhofar, Salalah, Oman[View additional affiliations](#) 3 98th percentile
Citations in Scopus6.45
FWCI 7
Views count [View all metrics](#) [Full text options](#) [Export](#) **Abstract**

Author keywords

Indexed keywords

Sustainable Development Goals
2023

SciVal Topics

Metrics

Abstract

Higher educational institutions have a responsibility and commitment to deliver employable graduates as it impacts their well-being and the economy. This study compared the accuracy of several classification algorithms to build an ensemble prediction model capable of forecasting graduates' employability using extensive data mining techniques. Based on the evaluation metrics, an ensemble model composed of Random Forest (RF), Support Vector Machines (SVM), and Naïve Bayes (NB) achieved the highest cross-validated accuracy score of 93.33%. Association rule mining and permutation feature importance analysis from 500 graduates of the electronics engineering program of a university revealed that grit is firmly attributed to employability, including the capabilities to acquire technical skills and professional certifications. Thus, the knowledge gained can be used to develop a range of policies, initiatives, and strategies to increase students' employment prospects. © 2022 IEEE.

Author keywords

Association rule mining; Classification; Clustering; Feature importance; Grit; Non-cognitive skills

Indexed keywords

Sustainable Development Goals 2023

New

SciVal Topics

Metrics

References (40)[View in search results format](#)

All [Export](#) [Print](#) [E-mail](#) [Save to PDF](#) [Create bibliography](#)

1 Small, L., Shacklock, K., Marchant, T.

[Employability: a contemporary review for higher education stakeholders](#)

Cited by 3 documents

Barrier-Free Routes in a Geographic Information System for Mobility Impaired People

Ugalde, B.H. , Maaliw, R.R. , Palarimath, S. (2022) 2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2022

A Deep Learning Approach for Automatic Scoliosis Cobb Angle Identification

Maaliw, R.R. , Susa, J.A.B. , Alon, A.S. (2022) 2022 IEEE World AI IoT Congress, AlloT 2022

Clustering and Classification Models for Student's Grit Detection in E-Learning

Maaliw, R.R. , Quing, K.A.C. , Susa, J.A.B. (2022) 2022 IEEE World AI IoT Congress, AlloT 2022

[View all 3 citing documents](#)

Inform me when this document is cited in Scopus:

[Set citation alert](#) **Related documents**

Employability Prediction of Information Technology Graduates using Machine Learning Algorithms

ElSharkawy, G. , Helmy, Y. , Yehia, E. (2022) International Journal of Advanced Computer Science and Applications

Applying Data Mining in Graduates' Employability: A Systematic Literature Review

Mpia, H.N. , Mburu, L.W. , Mwendia, S.N. (2023) International Journal of Engineering Pedagogy

A Deep Learning Approach for Automatic Scoliosis Cobb Angle Identification

Maaliw, R.R. , Susa, J.A.B. , Alon, A.S. (2022) 2022 IEEE World AI IoT Congress, AlloT 2022

[View all related documents based on references](#)

Find more related documents in Scopus based on:

[Authors](#) > [Keywords](#) >

All



ADVANCED SEARCH

Conferences > 2022 IEEE 12th Annual Computi... ?

Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling

Publisher: IEEE

Cite This

PDF

Renato R. Maaliw ; Karen Anne C. Quing ; Ace C. Lagman ; Bernard H. Ugalde ; Melvin A. Ballera ; Micha... All Authors

5
Paper
Citations148
Full
Text Views

< Previous | Back to Results | Next >

**Need
Full-Text**

access to IEEE Xplore
for your organization?

CONTACT IEEE TO SUBSCRIBE ➤

Abstract

Document Sections

I. Introduction

II. Methodology

III. Results

IV. Discussions

V. Conclusions and
Recommendations**Authors**

Figures

References

Citations

Keywords

Metrics

Footnotes

Abstract:

Higher educational institutions have a responsibility and commitment to deliver employable graduates as it impacts their well-being and the economy. This study compared the accuracy of several classification algorithms to build an ensemble prediction model capable of forecasting graduates' employability using extensive data mining techniques. Based on the evaluation metrics, an ensemble model composed of Random Forest (RF), Support Vector Machines (SVM), and Naïve Bayes (NB) achieved the highest cross-validated accuracy score of 93.33%. Association rule mining and permutation feature importance analysis from 500 graduates of the electronics engineering program of a university revealed that grit is firmly attributed to employability, including the capabilities to acquire technical skills and professional certifications. Thus, the knowledge gained can be used to develop a range of policies, initiatives, and strategies to increase students' employment prospects.

Published in: 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)

Date of Conference: 26-29 January 2022**INSPEC Accession Number:** 21743454**Date Added to IEEE Xplore:** 04 March 2022**DOI:** 10.1109/CCWC54503.2022.9720783**ISBN Information:****Publisher:** IEEE**Conference Location:** Las Vegas, NV, USA**I. Introduction**

Producing employable graduates is an essential component of the educational process encompassing the complete instructional spectrum objectives, from knowledge transmission, cognition, development of skills, and positive personal qualities. Employability refers to the probability of an individual getting a job with the changing employer and client needs to fulfill their potential profession [1]. In recent decades, its conception has gained attention, becoming the principal focus and priority of all higher educational institutions.

Authors

Renato R. Maaliw

College of Engineering Southern Luzon State University, Lucban, Quezon, Philippines

Karen Anne C. Quing

College of Arts and Sciences Southern Luzon State University, Lucban, Quezon, Philippines

Ace C. Lagman

More Like This

Financial Data Mining Based on
Support Vector Machines and
Ensemble Learning

2010 International Conference on
Intelligent Computation Technology and
Automation
Published: 2010

Parallel framework based
modified fuzzy support vector
machine model for data mining
with prior knowledge

2016 IEEE International Conference on
Advances in Computer Applications
(ICACA)
Published: 2016

Show More


Get Published in the
**IEEE Open Journal of
Industry Applications**

**Learn More**

Bernard H. Ugalde

Information Technology Department, University of Technology and Applied Sciences,
Salalah, Dhofar, Sultanate of Oman

Melvin A. Ballera

Graduate Programs Technological Institute of the Philippines, Manila, Philippines

Michael Angelo D. Ligayo

Department of Electronics Engineering, Quezon City University, Quezon City, Philippines

Figures



References



Citations



Keywords



Metrics



Footnotes



IEEE Personal Account

[CHANGE USERNAME/PASSWORD](#)

Purchase Details

[PAYMENT OPTIONS](#)

[VIEW PURCHASED DOCUMENTS](#)

Profile Information

[COMMUNICATIONS PREFERENCES](#)

[PROFESSION AND EDUCATION](#)

[TECHNICAL INTERESTS](#)

Need Help?

US & CANADA: +1 800 678 4333

WORLDWIDE: +1 732 981 0060

[CONTACT & SUPPORT](#)

Follow



[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [IEEE Ethics Reporting](#) | [Sitemap](#) | [IEEE Privacy Policy](#)
A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

© Copyright 2023 IEEE - All rights reserved.

[Feedback](#)



My... » My email messages »

#169 (1570784099): Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling

#169 (1570784099): Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling



Property	Change Add	Value																																				
Conference and track		2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC) - CCWC Paper - Regular Research Paper																																				
Authors		<table><thead><tr><th colspan="6">Affiliation</th></tr><tr><th>Name</th><th>ID</th><th>Edit</th><th>Flag</th><th>(edit for paper)</th><th>Email</th></tr></thead><tbody><tr><td>Renato R. Maaliw III</td><td>1841045</td><td></td><td></td><td>Southern Luzon State University, Philippines</td><td>renatomaaliw3@gmail.com</td></tr><tr><td>Karen Anne C. Quing</td><td>1925228</td><td></td><td>not creator</td><td>Southern Luzon State University, Philippines</td><td>karenannequing@gmail.com</td></tr><tr><td>Ace C. Lagman</td><td>1925232</td><td></td><td></td><td>FEU Institute of Technology, Philippines</td><td>acelagman69@gmail.com</td></tr><tr><td>Melvin A. Ballera</td><td>1893153</td><td></td><td></td><td>Technological Institute of the Philippines, Philippines</td><td>mballera27@gmail.com</td></tr></tbody></table>	Affiliation						Name	ID	Edit	Flag	(edit for paper)	Email	Renato R. Maaliw III	1841045			Southern Luzon State University, Philippines	renatomaaliw3@gmail.com	Karen Anne C. Quing	1925228		not creator	Southern Luzon State University, Philippines	karenannequing@gmail.com	Ace C. Lagman	1925232			FEU Institute of Technology, Philippines	acelagman69@gmail.com	Melvin A. Ballera	1893153			Technological Institute of the Philippines, Philippines	mballera27@gmail.com
Affiliation																																						
Name	ID	Edit	Flag	(edit for paper)	Email																																	
Renato R. Maaliw III	1841045			Southern Luzon State University, Philippines	renatomaaliw3@gmail.com																																	
Karen Anne C. Quing	1925228		not creator	Southern Luzon State University, Philippines	karenannequing@gmail.com																																	
Ace C. Lagman	1925232			FEU Institute of Technology, Philippines	acelagman69@gmail.com																																	
Melvin A. Ballera	1893153			Technological Institute of the Philippines, Philippines	mballera27@gmail.com																																	
Title	Only the chairs (ieeccwc2022- chairs@edas.info) can edit	Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling																																				
Abstract	Only the chairs (ieeccwc2022- chairs@edas.info) can edit	Higher educational institutions have a responsibility and commitment to deliver employable graduates as it impacts their well-being and the economy. This study compared the accuracy of several classification algorithms to build an ensemble prediction model capable of forecasting graduates' employability using extensive data mining techniques. Based on the evaluation metrics, an ensemble model composed of Random Forest (RF), Support Vector Machines (SVM), and Naïve Bayes (NB) achieved the highest cross-validated accuracy score of 93.33%. Association rule mining and permutation feature importance analysis from 500 graduates of the electronics engineering program of a university revealed that grit is firmly attributed to employability, including the capabilities to acquire technical skills and professional certifications. Thus, the knowledge gained can be used to develop a range of policies, initiatives, and strategies to increase students' employment prospects.																																				
Keywords	Only the chairs (ieeccwc2022- chairs@edas.info) can edit	association rule mining; classification; clustering; feature importance; grit; soft skills																																				
Topics	Only the chairs (ieeccwc2022- chairs@edas.info) can edit	Big Data, Data Management and Analytics																																				
Presenter(s)		Renato R. Maaliw III (bio) and Karen Anne C. Quing																																				
Registration		Renato R. Maaliw III has registered and paid for T2:IEEEENM																																				
DOI	Only the chairs (ieeccwc2022- chairs@edas.info) can edit																																					
URL																																						
Status		Accepted																																				
Copyright form		IEEE; IEEE: Jan 12, 2022 03:55 Asia/Hong_Kong																																				
Review		However, authors cannot upload: paper status																																				

manuscript		Document (show)	Pages	File size	Changed
			7	515,584	Jan 5, 2022 10:50 Asia/Hong_Kong
Final manuscript	Could upload until Jan 27, 2022 07:59 Asia/Hong_Kong.	However, authors cannot upload: final deadline			
Presentation		Document (show)	Pages	File size	Changed
	Could upload until Jan 28, 2022 06:23 Asia/Hong_Kong.		54,059,823 (13 min 11 sec)		Jan 24, 2022 11:00 Asia/Hong_Kong

Personal notes



You are the creator, an author and a reviewer for this paper. You have authored an accepted paper.

Reviews

3 IEEE CCWC 2022s

Review 1

None of the scores are visible for your roles.

Comments for reviewers (write at least 100 words regarding this review)

This paper compared the accuracy of several classification algorithms to build an prediction model capable of forecasting graduates' employability using data mining techniques.

It is a good idea but the reliability of the data set is very poor.

Instead of a direct student survey, they consulted different tracer records to acquire current student profiles, scholastic ratings, on-the-job training assessments from direct supervisors, and behaviors in a learning management system for online courses.

Data collection from various websites is highly unreliable. Profiles may be incorrect or not updated. They failed to discuss such issues in the paper.

Review 2

None of the scores are visible for your roles.

Comments for reviewers (write at least 100 words regarding this review)

The machine learning algorithms used are pretty basic: Random Forest (RF), Support Vector Machines (SVM), and Naive Bayes (NB). It would be nice to have RNN or CNN to yield the results.

TABLE 1 EMPLOYABILITY FEATURES, VALUES, AND ENCODING talk about Latin Honors which I guess are not valid worldwide. Could you please explain the term a little bit more?

Random Forrest yields the best results, which is again not a very advanced algorithm for machine learning. More intensified algorithm should have used

Review 3

None of the scores are visible for your roles.

Comments for reviewers (write at least 100 words regarding this review)

In the paper, the results of comparing the accuracy of several classification algorithms to build an ensemble prediction model capable of forecasting graduates' employability using extensive data mining techniques are presented.

Based on the evaluation metrics, an ensemble model composed of Random Forest, Support Vector Machines, and Naive Bayes achieved the highest cross-validated accuracy score above 93%.

The paper demonstrates an adequate understanding of the relevant literature in the field and cites an appropriate range of literature sources. The literature covers publications of recent years (2017-2021). The paper provides a detailed analysis of the theoretical background of prediction models.

The paper contains many practical results.

The paper is well prepared.

FACULTY POSITION RECLASSIFICATION FOR SUCS

(DBM-CHED Joint Circular No. 3, series of 2022)

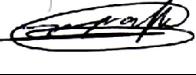
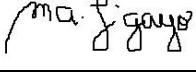
CERTIFICATION OF PERCENTAGE CONTRIBUTION

(Research Output with Multiple Authors)

Title of Research: Employability Prediction of Engineering Graduates Using Ensemble Classification Modeling

Type of Research Output: Conference Paper (Scopus-Indexed, IEEE Paper)

Instruction: Supply ALL the names of the authors involved in the publication of Research output and indicate the contribution of each author in percentage. Each author shall sign the Conforme column if he/she agrees with the distribution. The Conforme should be signed by all the authors in order to be considered. Please prepare separate Certification for each output.

	Name of Authors	% Contribution	Conforme (Sign if you agree with the % distribution)
1	Renato R. Maaliw III	40%	
2	Karen Anne C. Quing	40%	
3	Ace C. Lagman	5%	
4	Bernard H. Ugalde	5%	
5	Melvin A. Ballera	5%	
6	Michael Angelo D. Ligayo	5%	
* Should have a total of 100%		100.00%	

Prepared by:


Renato R. Maaliw III, DIT
(Name and Signature)
Faculty

Certified by:


Nicanor L. Guinto, Ph.D.
(Name and Signature)
Director, Office of Research Services