https://doi.org/10.24191/jeesr.v22i1.006

Job classification: The Application of Feature Selection Techniques on Graduates' Data

Muhammad Haziq Haikal Hisham, Mohd Azri Abdul Aziz*, and Ahmad Asari Sulaiman

***Abstract*—In this paper, job classification is viewed as a process to classify or to recommend jobs to the graduates according to the criteria set. The purpose of this study is to compare three feature selection techniques on the graduates' data to determine the relevant features in the job classification process for graduates. The experiment included three different feature selection techniques which are Analysis of Variance (ANOVA), Chi-squared test, and Recursive Feature Elimination (RFE). The dataset used for the experiment covered 12 graduates' feature that are needed to be tested to determine the impact of each graduates' feature on the result. The final feature ranking was listed for each of the feature selection techniques used and two common features among the rank lists had been found out as important features that affect job classification among graduates.**

***Index Terms*—ANOVA, Chi-squared test, Classification, RFE, graduates' data**

I. Introduction

JOB classification serves a big area to be studied for. Hence, it has attracted many researchers to further study about it. There are studies with several specific purposes conducted under this research area, among them are to predict the best fit candidate for a job [1], to recommend jobs to the candidates [2], to predict graduates' job placement according to their study's performance [3], etc. To achieve their own potential, it is important for the employers to have people with the right competences who can fit their culture [4]. Today's organizations are looking for the employees which not just has the basic academic knowledge but also has the ability to link their skills set and the needs of the respective job [5]. Thus, the hiring process play an important role in recruiting the right employees for the suitable job positions. Not only the candidates need to be classified according to their skills and background course, but the jobs need to be classified accordingly too.

Fresh graduates' unemployability rate issue should not be neglected because it will affect the future of a country. Graduate recruitment is different from other contexts in that most of the graduates lack in job-related experience [6]. Other factor such as the preference from the recruiters to recruit the older workers as the long-term investment of the company because the senior workers tend not to quit the job also can contribute towards the unemployability rate [7]. In Malaysia, the statistic shows that the unemployability for the fresh graduates in 2019 reached up to 170,300 graduates which is 5.5% more than previous year which is 161,300 graduates [8]. It is important for the authorities to lessen down the number of the unemployability rate among the fresh graduates to ensure the country's future economic plan will not be affected. Therefore, job classification with specifically focusing on the graduate could be a good alternative and better approach in matching graduate with their job.

From a certain point of view, the 'job classification' term can also be interpreted as a 'job recommender'. Where the main idea is to classify certain group of jobs' candidates according to the feature requirements and recommend jobs that match their features. The recommender systems are being used to figure out the interested feature for a certain user by employing a variety of information resources that is related to the users and features [9]. It was also found that recruiting the appropriate person is a challenge faced by most companies, as well as the unavailability of certain candidates in some skill areas has long been identified as a major obstacle to companies success [10]. Thus, by furthering the study in this area, it will contribute towards the idea to solve the major problem faced by the companies by recommending the most suitable candidates for the job advertised.

The application of feature selection is proposed in this paper to improve the job classification process. To test the robustness of the data, three feature selection techniques are being used in this study such as Analysis of Variance (ANOVA), Chi-squared test, and Recursive Feature Elimination (RFE).

II. Literature Review

Data used in the job classification study in various classification fields have been reviewed. By focussing on the main idea of classifying jobs in this study, reviews have found out that there is still no job classification method has been done on graduates' data with this job classification's concept of recommending a job for the candidates by using several features. Some of the job classification studies that are closely related with this proposed method only covered about the preference of employers and recruiters about the features that they are looking for in the potential candidates such as leadership skill, people skill, etc [11][12]. To improve the quality of the classification, the implementation of feature selection techniques is proposed in this paper.

Feature Selection (FS) algorithm plays a vital role to remove the irrelevant and redundant attributes or features from dataset [13]–[19]. It will give better performance for the classifiers in

This manuscript is submitted on 22 December 2021 and accepted on 30 October 2022. M.A.A.Aziz, and A.A.Sulaiman are with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA, 40450 Shah Alam, Selangor (azriaziz@uitm.edu.my).

Muhammad Haziq Haikal Hisham was with the School of Electrical Engineering, College of Engineering, Universiti Teknologi MARA. (email: 2020402226@isiswa.uitm.edu.my).

*Corresponding author
Email address: azriaziz@uitm.edu.my

terms of accuracy in classifying and predicting the suitable job for graduates. Many feature selection techniques can be applied in the pre-processing phase to improve the results of the experiment. In this paper, we will be focusing on several selected FS techniques like ANOVA, Chi-Squared test, RFE and ET.

ANOVA is a powerful method and very simple FS technique to test the difference and variation in means between groups [20]–[22]. It can tell us whether the means are equal or have slight difference in the means of different populations [21]. According to [20], this FS technique is being used by researchers due to some of it advantages like the robustness to most violations of its assumptions, it is more natural for us to analyse the interaction of the variables, the effectiveness of the algorithm is not affected even the number of observations is different in each group, and it can be easily generalized to more groups (more than two groups) without increasing the Type 1 error.

Chi-Squared test is one of the most common used statistical tests that measures divergence from the distribution expected if there is assumption towards the feature occurrence is actually independent of the class value [23][24]. This FS is used for analysing two different type of comparing which are tests of goodness of fit and tests of independence [25]. Besides, Chi-squared test was developed initially for microarray-based cancer classification where the data used was less than 100 [26], same as this study.

RFE also had been used in various studies and proved its quality such as in analysis of agro-industrial products, RFE had been used as a part of the system that successfully classified data that were categorised as completely independent test sets [27], in manufacturing model study, RFE had proven its effectiveness in both linear and non-linear cases study [26], in high dimensional multi-category data, RFE several variations of RFE had been tested and successfully given a good recognition results in the analysis of the data [28], in cancer classification study, a variation of RFE classifier had able to select better gene subsets and improved the cancer classification accuracy [29], etc.

III. Methodology

This section is parted into several sections, beginning with theoretical background followed by the flow of the process. The intelligent technique was then conducted in Python coding in Anaconda Navigator (Anaconda3) platform.

A. Theoretical Background

1) Analysis of Variance (ANOVA)

ANOVA is the technique used to analyse the experimental data in which one or more response variables are observed under several different conditions identified by one or more classification variables [22]. The statistic for ANOVA is known as F-statistic that can be calculated by using following steps [30]:

(1) The calculation of variation between the group:

Between sum of squares:

$$(BSS) = n_1(.\overline{X}_1 - \overline{X})^2 + n_2(\overline{X}_2 - \overline{X})^2 + \cdots \quad (1)$$

Between mean squares:

$$(BMS) = BSS / d\,f \quad (2)$$

(2) The calculation of variation within the group:

Within sum of squares:

$$(WSS) = (n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2 + \cdots \quad (3)$$

Within mean of squares:

$$(WMS) = WSS / d\,f_\omega \quad (4)$$

Denote that d f = degree of freedom, d f_ω = (N − k), σ = standard deviation N = Number of samples, k = Number of groups, and n_k = no. of samples in group k.

(3) The calculation of F-statistic:

$$F = BMS / WMS \quad (5)$$

The input to the algorithm is in a matrix form of *N* x *M,* where:
N = total number of feature sets
M = the number of samples in the dataset.

2) Chi-squared Test

Chi-squared test attribute evaluation observed and analysed the worth of each feature by computing the value of the chi-squared statistic with respect to the class [24]. It is to estimate whether the class label is independent of a feature or not. Chi-squared score with C class and r values is defined as follows [25]:

$$x^2 = -\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(n_{ij}-\mu_{ij})}{\mu_{ij}} \quad (6)$$

Denote that n_{ij} is the amount of samples value with i^{th} value of the feature.

$$\mu_{ij} = \frac{(n_{*j}n_{i*})}{n} \quad (7)$$

Where:
n_{i*} is the amount of samples with the i^{th} the feature value.
n_{*j} is the amount of samples in class j.
n is the number for samples.

3) Recursive Feature Elimination (RFE)

As RFE is closely related to Support Vector Machine (SVM) [26], we used linear-based RFE kernel of SVM as a supervised learning estimator. The SVM aims to find the hyperplane that part the classes the most [31]. RFE works as a recursive cycle as follows [28]: pre-defined feature measure rules for the current dataset are used to calculate the ranking positions for all features and each of the ranking values indicate the feature's classification ability. Then, the feature with the lowest value will be removed. The ranking value for the remaining feature will be recalculated and feature with the smallest value will be removed. This process is repeated until only one feature is left. It will indirectly uses feature ranking heuristic criteria to make sure the algorithm performance performs at its best.

Suppose *D(x)* indicates the decision function for the hyperplane and c represents the number of the classes. If the data has multiple classes that is more than two, *q*, where the entire number of the hyperplane is calculated according to the equation *q=c(c-1)/2.* Equation (8) and equation (9) show the decision function for the binary class case and for multi-class dataset respectively [31].

$$D(x) = sign(x * w) \tag{8}$$

$$D(x) = sign(x * w_j), j = 1,2,4, \dots, q \tag{9}$$

In the linear decision function, x indicates a vector with the components of a given spectrum, and w is a vector perpendicular to the hyperplane providing a linear decision function [31]. Equation (10) is used to get the weight value for the evaluation of variable importance according to SVM-RFE.

$$W_s = \frac{1}{q}\sum_{i=1}^{q} W_i \tag{10}$$

B. Data Collection

The graduates' performance data was collected from a survey conducted for Universiti Teknologi MARA graduates from School of Mechanical Engineering, School of Electrical Engineering, and Faculty of Science and Mathematics. While the job preference data was collected from a Chief Executive Officer (CEO) of a private Software Engineering/IT company. There are 12 different features (refer Table 1) that were included in the survey where the graduates need to rate themselves accordingly. All the features included were narrowed down from estimately 449 features reviewed from previous researches observed in 'student classification', 'job classification', and 'job recommender' keywords.

For 'CGPA' feature, the graduates had to include their CGPA and we pre-processed them into several levels to be converted as a usable data for this study. For example, if the CGPA is equal or bigger than 3.00, then it was graded as number '4', if the CGPA is equal or bigger than 2.00 and smaller than 3.00, then it was graded as number '3', if the CGPA is equal or bigger than 1.00 and smaller than 2.00, then it was graded as number '2', and if the CGPA is lower than 1.00, then it was graded as number '1'. For 'Prior experience' feature, if the graduates have prior experience, then it was graded as number '1' and if they do not have any prior experience, then it was graded number '0'. For the rest of the features, the graduates had to rate themselves on a scale of 1 to 5 for each feature.

TABLE I
GRADUATES' FEATURES

Graduates' features	**References**
CGPA	[32]–[37]
Prior experience	[17][18][23]
Teamwork	[18][19]
Leadership	[39]–[41]
Motivation	[42]
Flexibility	[43]
Dependability	[43]
Problem solving	[17][18][28]
Creative thinking	[18][24][29][30]
Communication	[18][31]–[33]
Analytical	[18][24][29][30]
Observing	[42]

C. Flowchart of the job classification

The experiment started by the process of data collection. The details of this sub-process as in methodology part B above. The data collected during this sub-process were obtained in raw data the need to be pre-processed for the experiment. During the pre-processing phase, this is where three selected feature selection techniques (ANOVA, Chi-squared test, RFE) are implemented to the graduates' data to eliminate the irrelevant features. Then, the Support Vector Machine (SVM) – Radial Basis Function (RBF) algorithm was applied to the pre-processed data to classify it. Only then the output was generated and the data were classified. The output data that did not meet the criteria set will go through the classifying process again until it met the criteria. The flow of the system is summarized as shown as in Figure 1 below.

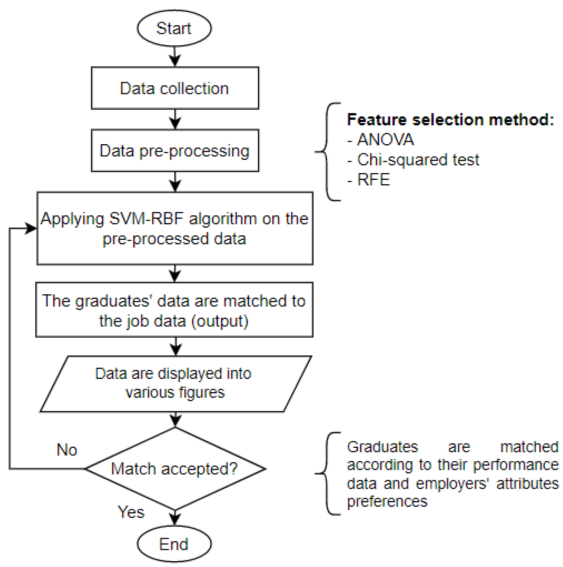

Fig. 1. Flowchart of the classfication

IV. RESULT AND DISCUSSION

The results of the three methods of feature selection test on the dataset are discussed below.

1) Analysis of Variance (ANOVA)

Figure 2 and Table 2 show the scatter plot and the feature ranking of the dataset after ANOVA test respectively. In ascending order sequence, the tested features are ranked as follow; Prior experience, CGPA, Dependability, Flexibility, Leadership, Communication, Motivation, Teamwork, Observing, Problem solving, Creative thinking, and Analytical. By referring to [30], it is said that two main hypothesis affects the results, i.e., Null hypothesis and alternate hypothesis. Where the null hypothesis denote that the classes are same (no difference between the properties of the group) and the alternate hypothesis assumes that there exists some significant difference between the groups. The null hypothesis means that the features do not affect the result and features can be discarded, while on the other hand, the alternate hypothesis means that the features have significant difference between their properties. Thus, the features are accepted.

The higher the F-value, the higher the ranking of the features. To be concluded, top four features chosen by ANOVA test according to the highest F-value are ‘Analytical’, ‘Creative thinking’, ‘Problem solving’, and ‘Observing’.

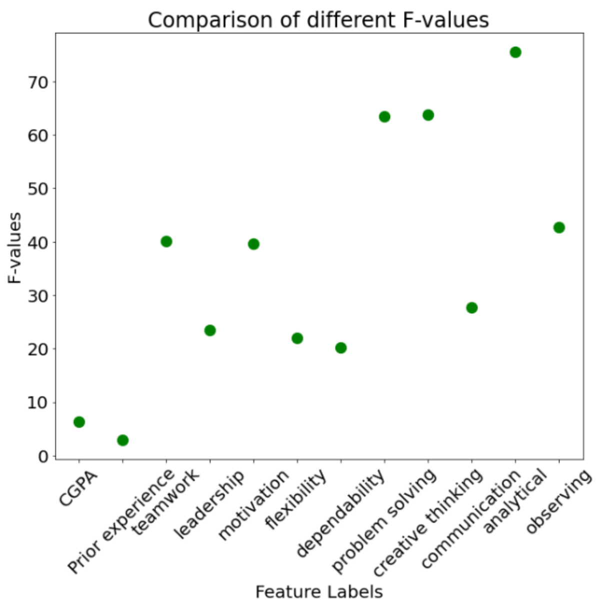

Fig. 2. The scatter plot of ANOVA test on graduates’ data

TABLE II
THE F-VALUE AND FEATURE RANK BASED ON ANOVA TEST

Graduates’ features	F-value	Feature ranking
Analytical	75.4875	1
Creative thinking	63.8122	2
Problem solving	63.3971	3
Observing	42.7909	4
Teamwork	40.1381	5
Motivation	39.5828	6
Communication	27.7952	7
Leadership	23.4453	8
Flexibility	22.0305	9
Dependability	20.2117	10
CGPA	6.2947	11
Prior experience	3.0202	12

2) Chi-Squared Test

Figure 3 and Table 3 show the bar chart and feature rankings of the graduates’ dataset after Chi-squared test had been applied respectively. In ascending order sequence, the tested features are ranked as follow; Prior experience, CGPA, Flexibility, Dependability, Observing, Leadership, Teamwork, Communication, Motivation, Analytical, Problem solving, and Creative thinking. It can be said that this feature selection technique evaluates the value of a feature by computing the value of the Chi-squared statistic with respect to the class or group [24]. After calculating the values of all features, the features’ rank was determined. Where, the bigger the calculated value, the more important the feature is as in sequence above.

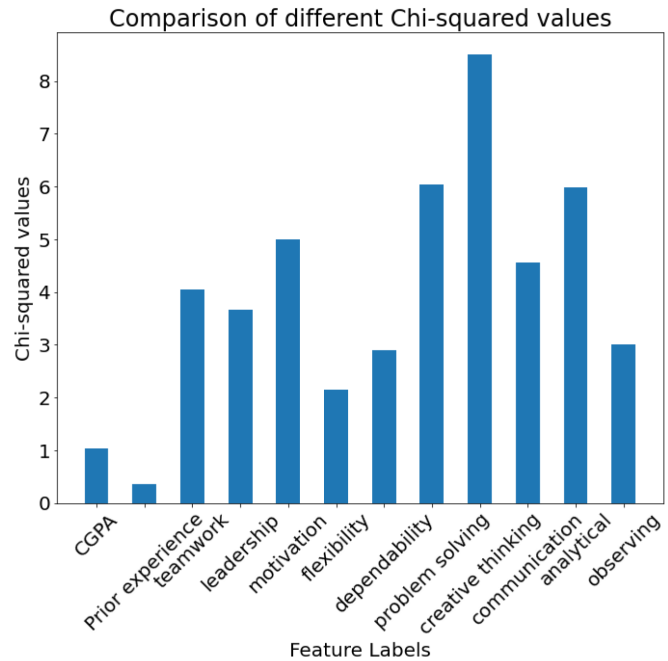

Fig. 3. The bar chart of Chi-squared test on graduates’ data

TABLE III
THE FEATURE SCORE AND FEATURE RANK BASED ON CHI-SQUARED TEST

Graduates’ features	Feature score	Feature ranking
Creative thinking	8.4976	1
Problem solving	6.0373	2
Analytical	5.9796	3
Motivation	4.9963	4
Communication	4.5626	5
Teamwork	4.0542	6
Leadership	3.6726	7
Observing	3.0076	8
Dependability	2.9045	9
Flexibility	2.1508	10
CGPA	1.0354	11
Prior experience	0.3576	12

3) Recursive Feature Elimination (RFE)

Table 4 below shows the rank of each feature after tested by RFE technique. In ascending order sequence, the tested features are ranked as follow; Analytical, Prior experience, Teamwork, Creative thinking, CGPA, Motivation, Flexibility, Dependability, Observing, Communication, Problem solving, and Leadership. As RFE evaluates and observes features rank by their weights in SVM solution, it eliminates and rank each of the features tested accordingly. RFE results are affected by its looping process or recursion, where it is needed because for some measures of the features’ value can change substantially when it is tested over a different subset of features [27]. Each of the features are ranked by using numeric numbers. The highest weighted feature rank is ranked as number '1', while the bigger the rank number indicates the more ineffectiveness of the feature in the dataset.

TABLE IV
THE FEATURE RANK BASED ON RFE TEST

Graduates' features	Feature ranking
Analytical	1
Prior experience	2
Teamwork	3
Creative thinking	4
CGPA	5
Motivation	6
Flexibility	7
Dependability	8
Observing	9
Communication	10
Problem solving	11
Leadership	12

As a summary, each of the FS techniques used have their own ability and way of working. Where the ANOVA FS works by analyzing the data in the response variables that are observed under several different conditions identified by the classification variables, Chi-squared test works by analyzing whether the class label is independent of a feature or not, and RFE works as a recursive cycle where it indirectly uses the feature ranking heuristic criteria to sort for the algorithm output. Thus, the output produced for the same dataset will probably have differences, even a slight difference.

By referring to the Table 5 below, the top four features with the highest rated score from each FS were summarized. Which showing that 'Analytical' and 'Creative thinking' are the features that are ranked in the top four of each FS techniques. By being rated among the top four position in each FS techniques, it shows that these two features have significant value in the dataset compared to the other features.

TABLE V
THE TOP FOUR FEATURES WITH THE HIGHEST SCORE IN EACH OF THE FS TECHNIQUE

FS technique	Features chosen (Ascending order)
ANOVA	Analytical, Creative thinking, Problem solving, Observing
Chi-squared test	Creative thinking, Problem solving, Analytical, Motivation
RFE	Analytical, Prior experience, Teamwork, Creative thinking

V. CONCLUSION

In this paper, the application of three selected feature selection techniques which are ANOVA, Chi-squared test, and RFE were tested on the graduates' data. Each of the feature selection technique was tested to rank and evaluate 12 graduates' features. The results were observed and analised with the different approaches and algorithms behind each of the feature selection technique.

According to the output of each feature selection techniques, only 'Analytical' and 'Creative thinking' are ranked in the top four with the highest score in each of the feature selection technique. Hence, the comparison of the feature selection techniques used in this study had determined and obtained 'Analytical' and 'Creative thinking' as the final features that play the important role when classifying jobs among graduates in this study. Thus, as an idea for future work, a research by using these feature selection techniques on any data is recommended to determine the relevant features in the dataset used.

REFERENCES

[1] K. Appadoo, S. Muhammad Bilaal, and M.-D. Zahra, "JobFit : Job Recommendation using Machine Learning and Recommendation Engine," *IEEE Asia-Pacific Conf. Comput. Sci. Data Eng.*, pp. 1–6, 2020.

[2] G. Özcan and S. G. Öguducu, "Applying different classification techniques in reciprocal job recommender system for considering job candidate preferences," *2016 11th Int. Conf. Internet Technol. Secur. Trans. ICITST 2016*, pp. 235–240, 2017.

[3] S. Venkatachalam, "Data Mining Classification and analytical model of prediction for Job Placements using Fuzzy Logic," pp. 765–770, 2021.

[4] R. M. Lima, D. Mesquita, C. Rocha, and M. Rabelo, "Defining the Industrial and Engineering Management Professional Profile: A longitudinal study based on job advertisements," *Producao*, vol. 27, no. Specialissue, pp. 1–15, 2017.

[5] V. Ajit and P. B. Deshmukh, "Factors Impacting Employability Skills of Engineers," *Int. J. Sci. Res.*, vol. 2, no. 4, pp. 30–32, 2013.

[6] A. Walker and C. L. Caballero, "Work readiness in graduate recruitment and selection: A review of current assessment methods," *J. Teach. Learn. Grad. Employab.*, vol. 1, no. 1, pp. 13–25, 2010.

[7] D. Abrams, H. J. Swift, and L. Drury, "Old and Unemployable? How Age-Based Stereotypes Affect Willingness to Hire Job Candidates," *J. Soc. Issues*, vol. 72, no. 1, pp. 105–121, 2016.

[8] D. F. Kamarudin, "Pandemik tingkat pengangguran graduan," Malaysia, 2021.

[9] S. T. Al-Otaibi, "A survey of job recommender systems," *Int. J. Phys. Sci.*, vol. 7, no. 29, pp. 5127–5142, 2012.

[10] S. T. Al-Otaibi and M. Ykhlef, "Job Recommendation Systems for Enhancing E-recruitment Process," *Proc. Int. Conf. Inf. Knowl. Eng.*, pp. 1–7, 2012.

[11] N. S. Ismail, N. Ismail, M. H. F. Rahiman, M. N. Taib, N. A. M. Ali, and S. N. Tajuddin, "Polynomial tuned kernel parameter in SVM of agarwood oil for quality classification," *Proc. - 2018 IEEE Int. Conf. Autom. Control Intell. Syst. I2CACIS 2018*, no. October, pp. 77–82, 2019.

[12] M. W. B. Trotter, B. F. Buxton, and S. B. Holden, "Support vector machines in combinatorial chemistry," *Meas. Control*, vol. 34, no. 8, pp. 235–239, 2001.

[13] M. Zaffar, M. A. Hashmani, and K. S. Savita, "Performance analysis of feature selection algorithm for educational data mining," *2017 IEEE Conf. Big Data Anal. ICBDA 2017*, vol. 2018-Janua, pp. 7–12, 2018.

[14] M. Zaffar, M. A. Hashmani, K. S. Savita, and S. S. H. Rizvi, "A study of feature selection algorithms for predicting students academic performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 541–549, 2018.

[15] I. Anuradha, "Feature Selection Techniques To Analyse Student Acadamic Performance Using Naive Bayes Classifier," 2016.

[16] M. Doshi and S. K. Chaturvedi, "Correlation Based Feature Selection (CFS) Technique to Predict Student Perfromance," *Int. J. Comput. Networks Commun.*, vol. 6, no. 3, pp. 197–206, 2014.

[17] C. Anuradha and T. Velmurugan, "Performance Evaluation of Feature Selection Algorithms in Educational Data Mining," *Int. J. Data Min. Tech. Appl.*, vol. 5, no. 2, pp. 131–139, 2016.

[18] S. S. M. Ajibade, N. B. Ahmad, and S. M. Shamsuddin, "An Heuristic Feature Selection Algorithm to Evaluate Academic Performance of

Students," *ICSGRC 2019 - 2019 IEEE 10th Control Syst. Grad. Res. Colloquium, Proceeding*, no. August, pp. 110–114, 2019.

[19] C. Jalota and R. Agrawal, *Feature Selection Algorithms and Student Academic Performance: A Study*, vol. 1165. Springer Singapore, 2021.

[20] H. Ding, P. M. Feng, W. Chen, and H. Lin, "Identification of bacteriophage virion proteins by the ANOVA feature selection and analysis," *Mol. Biosyst.*, vol. 10, no. 8, pp. 2229–2235, 2014.

[21] M. Bejani, D. Gharavian, and N. M. Charkari, "Audiovisual emotion recognition using ANOVA feature selection method and multi-classifier neural networks," *Neural Comput. Appl.*, vol. 24, no. 2, pp. 399–412, 2014.

[22] M. Sheikhan, M. Bejani, and D. Gharavian, "Modular neural-SVM scheme for speech emotion recognition using ANOVA feature selection method," *Neural Comput. Appl.*, vol. 23, no. 1, pp. 215–227, 2013.

[23] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *CrossRef List. Deleted DOIs*, vol. 1, pp. 1289–1305, 2000.

[24] A. Bidgoli and M. Naseriparsa, "A Hybrid Feature Selection by Resampling, Chi-squared abd Consistency Evaluation Techniques Flexible Interactive Querying View project," *Researchgate.Net*, no. August, 2018.

[25] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," *Proc. - 2015 7th Int. Conf. Inf. Technol. Electr. Eng. Envisioning Trend Comput. Inf. Eng. ICITEE 2015*, pp. 420–424, 2015.

[26] J. A. Ramírez-Hernández and E. Fernandez, "Control of a re-entrant line manufacturing model with a reinforcement learning approach," *Proc. - 6th Int. Conf. Mach. Learn. Appl. ICMLA 2007*, pp. 330–335, 2007.

[27] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemom. Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, 2006.

[28] W. You, Z. Yang, and G. Ji, "Feature selection for high-dimensional multi-category data using PLS-based local recursive feature elimination," *Expert Syst. Appl.*, vol. 41, no. 4 PART 1, pp. 1463–1475, 2014.

[29] K. Duan, J. C. Rajapakse, S. Member, H. Wang, F. Azuaje, and S. Member, "Multiple SVM-RFE for Gene Selection in Cancer Classification With Expression Data," vol. 4, no. 3, pp. 228–234, 2005.

[30] M. Kumar, N. K. Rath, A. Swain, and S. K. Rath, "Feature Selection and Classification of Microarray Data using MapReduce based ANOVA and K-Nearest Neighbor," *Procedia Comput. Sci.*, vol. 54, pp. 301–310, 2015.

[31] H. Jeon and S. Oh, "Hybrid-recursive feature elimination for efficient feature selection," *Appl. Sci.*, vol. 10, no. 9, pp. 1–8, 2020.

[32] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," *Comput. Educ.*, vol. 113, pp. 177–194, 2017.

[33] 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, pp. 1712–1720, 2018.

[34] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The Prediction of Students ' Academic Performance Using Classification Data Mining Techniques," vol. 9, no. 129, pp. 6415–6426, 2015.

[35] N. Buniyamin, U. Mat, and P. M. Arshad, "Educational Data Mining for Prediction and Classification of Engineering Students Achievement," pp. 49–53, 2015.

[36] I. Singh, A. S. Sabitha, and A. Bansal, "Student performance analysis using clustering algorithm," *Proc. 2016 6th Int. Conf. - Cloud Syst. Big Data Eng. Conflu. 2016*, pp. 294–299, 2016.

[37] N. Elassy, "Quality Assurance in Education Article information :," 2015.

[38] D. M. Costello-Nichitas, "from the SAGE Social Science Collections . All Rights," *Hisp. J. Behav. Sci.*, vol. 9, no. 2, pp. 183–205, 1987.

[39] M. Abdulwahed, W. Balid, M. O. Hasna, and S. Pokharel, "Skills of engineers in knowledge based economies: A comprehensive literature review, and model development," *Proc. 2013 IEEE Int. Conf. Teaching, Assess. Learn. Eng. TALE 2013*, pp. 759–765, 2013.

[40] B. L. Hartmann and C. T. Jahren, "Leadership: Industry needs for entry-level engineering positions," *IEEE Eng. Manag. Rev.*, vol. 44, no. 3, pp. 76–85, 2016.

[41] A. Prof and D. Mishra, "Engineering Employability Skills Required By Employers in India," pp. 961–964, 2016.

[42] P. Protsch and H. Solga, "How Employers Use Signals of Cognitive and Noncognitive Skills at Labour Market Entry: Insights from Field Experiments," *Eur. Sociol. Rev.*, vol. 31, no. 5, pp. 521–532, 2014.

[43] V. Wickramasinghe and L. Perera, "Graduates', university lecturers' and employers' perceptions towards employability skills," *Educ. Train.*, vol. 52, no. 3, pp. 226–244, 2010.

[44] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, "A Comparative Study of Classification and Regression Algorithms for Modelling Students' Academic Performance," *Proc. 8th Int. Conf. Educ. Data Min.*, pp. 392–395, 2015.

[45] M. Y. Yuzainee, A. Zaharim, and M. Z. Omar, "Employability skills for an entry-level engineer as seen by Malaysian employers," *2011 IEEE Glob. Eng. Educ. Conf. EDUCON 2011*, pp. 80–85, 2011.

[46] A. Zaharim, M. Z. Omar, Y. M. Yusoff, N. Muhamad, A. Mohamed, and R. Mustapha, "Practical framework of employability skills for engineering graduate in Malaysia," *2010 IEEE Educ. Eng. Conf. EDUCON 2010*, pp. 921–927, 2010.

[47] J. Moody and C. Bolt-lee, "Showcasing Expanding," pp. 21–33.

[48] A. C. Teo, "Perceptual Differences Between Recruiters and Students on the Importance of Applicant and Job Characteristics: A Research Note Based on Evidence From Singapore," *Int. J. Hum. Resour. Manag.*, vol. 4, no. 1, pp. 231–239, 1993.

[49] N. Ahadiat and K. J. Smith, "A Factor-Analytic Investigation of Employee Selection Factors of Significance to Recruiters of Entry-Level Accountants.," *Issues Account. Educ.*, vol. 9, no. 1, p. 59, 1994.