World Academy of Science, Engineering and Technology
International Journal of Educational and Pedagogical Sciences
Vol:16, No:11, 2022

Predictive Analytics of Student Performance Determinants in Education

Mahtab Davari, Charles Edward Okon, Somayeh Aghanavesi

***Abstract*—**Every institute of learning is usually interested in the performance of enrolled students. The level of these performances determines the approach an institute of study may adopt in rendering academic services. The focus of this paper is to evaluate students' academic performance in given courses of study using machine learning methods. This study evaluated various supervised machine learning classification algorithms such as Logistic Regression (LR), Support Vector Machine (SVM), Random Forest, Decision Tree, K-Nearest Neighbors, Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis, using selected features to predict study performance. The accuracy, precision, recall, and F1 score obtained from a 5-Fold Cross-Validation were used to determine the best classification algorithm to predict students' performances. SVM (using a linear kernel), LDA, and LR were identified as the best-performing machine learning methods. Also, using the LR model, this study identified students' educational habits such as reading and paying attention in class as strong determinants for a student to have an above-average performance. Other important features include the academic history of the student and work. Demographic factors such as age, gender, high school graduation, etc., had no significant effect on a student's performance.

***Keywords*—**Student performance, supervised machine learning, prediction, classification, cross-validation.

I. Introduction

PERFORMANCE in studying is a phenomenon that can be affected by different factors: from family background to social interactions and style of preparation for examinations. Researchers have made several attempts to provide more clarity on these important factors. Benavides [1] attempted to use behavioral attributes to identify at-risk students. Findings from such studies have particularly interested higher institutions of learning. As a result, educational authorities leverage these findings to periodically review and update academic administrative plans towards achieving the improved performance of students during studies [2]. In line with evaluating student academic performance, machine learning (ML) approaches have gained wide popularity and acceptance in solving regression and classification problems as well as determinations of feature importance [3]. Classification techniques are learning methods used for problems that require determining the classes or categories of given observations. With the number of existing classification techniques available in the literature, determining a superior method from another may require more studies. To do this, several factors such as application, nature of the problem or available data could be considered.

Mahtab Davari is with Dalarna University, Sweden (e-mail: mahtab.davari@gmail.com).

Some ML methods are identified to be appropriate in solving classification problems [2]. These techniques include K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), SVM, LDA, Quadratic Discriminant Analysis (QDA) and LR. To provide a summary of mentioned methods in similar domains, the KNN algorithm is a non-parametric supervised learning method, and its output is a class membership. It is an effective classification method. The K value in the KNN algorithm represents how many neighbors need to be checked to classify a specific entry based on the distance between the data points [3]. According to Naghesh et al. [4], using the KNN classification technique to predict student performance would provide educational institutions with sufficient information to take appropriate measures towards improving the institute's quality.

The DT algorithm is used in classification problems and can be performed on categorical and numerical data. DT can manage problems with multiple outputs, such as multiple classifications [5]. Awad et al. [6] evaluated the performance of DT in predicting student performance using three DT algorithms. In comparison, the vJ48 algorithm was identified as the best performing algorithm compared with the Random Tree and RepTree algorithms. Despite this, Ghosh and Janan [7] concluded that building a model with RF proved effective and pragmatic for predicting student progress compared to using other traditional ML algorithms. In a similar study, Katarya [8] showed the best prediction accuracy among other methods such as SVM and KNN.

Similarly, other related studies [9]-[11] mentioned SVM as showing high performance in classification and prediction when working with high-dimensional data. LDA is another robust classification method used to classify patterns between two or more classes [12], [13]. The LDA method presumes that each class's observations are extracted from a multivariate Gaussian distribution with a class-specific mean vector with a covariance matrix common to all K classes. LDA is generally used to classify the categorical output variable. Aye et al. recommended this method for both binary classifications [14]. QDA is an extension of LDA and presumes that each class has its covariance matrix [15]. QDA keeps the same property as LDA. QDA grants more flexibility to the covariance matrix and can fit the data better than LDA [16], [17]. Therefore, it might show significant results compared with LDA. LDA and QDA were not found to be widely applied for this typical problem. LR can be used, as an algorithm, to model the probability of belonging to a specific class. It is helpful for linearly separable data when the outcome is binary [18]. According to Abraham and Das [19], LR had the most accurate prediction for the final grades of students when compared to other supervised ML algorithms, such as DT, Naïve Bayes, SVM, KNN, Sequential Minimal Optimization and Neural Networks. A similar outcome is observed when the application of LR is compared with Neural Network on a closely related subject [14].

Given the known facts of the various ML methods, their usability, interpretability, and flexibility are usually based on some trade-off considerations [4]. Determining the best ML method for predicting students' performance could be a challenge with the number of existing alternatives [4]. Consequently, based on all reviewed articles, there is still a need to investigate the predictive performances of the identified methods and key determinants of above-average student performance in any given course. This study attempts to address this research gap using the 5-fold cross-validation and investigate metrics of accuracy, precision, recall, and F1 scores. In addition, this study will examine which specific features (i.e., demographic, academic and behavioral) influence a student's academic performance in any given course. In other words, the answers to the following research questions are attempted to be researched in detail:

1. What classification method best predicts the expected student academic performance?
2. What variables would be best for predicting the expected performance of students?

II. Data

A. Source of Data

The data used in this study were obtained from an open access survey data source [20] published in 2019. It contains the survey data from students of two faculties (a faculty of engineering and the faculty of educational sciences in a higher institute of learning). This dataset provides features about the respondent's socioeconomic, family characteristics and educational habits including 33 variables which consist of numerical and categorical types. A total of 145 records of students' participation was obtained.

III. Methodology

In this study, prior to training and evaluating the performance of the seven ML methods (LR, SVM, DT, RF, KNN, LDA, and QDA), we performed a descriptive analysis to evaluate the balance of the dataset by its distribution as well as evaluating the outliers and missing values. It was also done to identify if there are similarities among variables. For descriptive analysis, we also produced cross-tabulation showing frequencies and percentages on each variable against student performance (binary response: 1 – "Above average performance", 0 – "Below average performance").

All data processing and analytics were done using Jupyter notebook (python) with Panda, Numpy, Scikit-learn (sklearn) libraries and machine learning extensions (mlxtend).

A. Feature Selection

A key step in data analytics of this study for model specification in training a ML method is feature selection [21], [22]. Determining and collecting the correct features required in training a method have been known to be a major determinant of how well such a method may perform. A challenge with implementing a ML method and obtaining correct prediction accuracy is to collect the right features [23]. The ineffectual features may decrease the performance of the learning (classification or regression) algorithms or crash the model without producing any realistic result. This makes optimizing the selection of features crucial, especially when the "size of observation" to "number of features" ratio is very low [24]. As the principle of parsimony (Occam's Razor) states, the best reasoning for a problem is the one with the fewest assumptions [25]. Feature selection in ML attempts to find out the best set of predictors that permits one to build the best possible models for the case study [26]-[28]. Some popular feature selection techniques in ML are forward selection, backward selection (Elimination), and exhaustive selection; suggested for studies dealing with classification problems [3].

Backward elimination is also an iterative approach. This technique begins the process by considering all the features and removing the least significant feature. This elimination process continues until the elimination of any features does not show any improvement in the model's performance [2].

The backward elimination has the advantage of assessing the joint predictive ability of variables since the process starts by including all variables in the model. Therefore, this method was adopted for feature selection. The list of selected features (13 variables) can be found in Appendix I.

B. ML Algorithms

1) Logistic Regression

LR is a classification model derived by transforming the linear regression cost function using the sigmoid cost function. In this study, we will focus on the binary LR model, which is used when there are only two possible discrete outcomes. The function returns a probability score ranging from 0 to 1. If the function is greater or equal to the threshold, it concludes that the data point belongs to the response feature class 1—otherwise, 0.

2) Support Vector Machine

SVM is an ML algorithm for classification and Regression. This algorithm draws all data points in the n-dimensional space (n features as our predictors in the data set). It finds an n-dimensional hyperplane that divides all the features into two distinct classes using three kernel functions (i.e., Linear, Non-linear, and Radial Basis Function).

3) Random Forest

RF creates different training subsets from the training data and makes the DT for the classification with the majority vote. One of the properties of the RF algorithms is the fact that can handle continuous and discrete data sets, for Regression and categorical problems and performs better.

4) Decision Tree

A DT is a ML algorithm for classification systems that can be used for developing prediction algorithms for a target variable. This method classifies a dataset into branch-like segments that construct an upside-down tree with a root node, internal nodes, and leaf nodes. The internal nodes represent the features of a dataset, branches represent the decision rules, and each leaf node denotes the response. The DT algorithm is non-parametric.

5) K-Nearest Neighbor

The KNN approach is a ML algorithm that estimates the conditional distribution of the response variable given a certain explanatory variable and then classifies a certain observation to a class for which the estimated probability is highest. Using this algorithm, the classifier spots the K points in the train data nearest to the test observation and then estimates the conditional probability.

6) LDA & QDA

It is supervised ML and is used to find a linear combination of features that separates two or more classes of objects. Its result can be used as a linear classifier. LDA assumes a shared covariance matrix for all classes instead of one for each specific class. QDA is a variant of LDA, but an individual covariance matrix is calculated for every class of observations.

C. Evaluation of ML Methods

After implementation of ML methods, the features that were selected for each method was identified. ML methods were tested with 5-Fold cross-validation (CV). The CV results (Accuracy, Precision, Recall, and F1) were compared per ML method. Based on the evaluation the best ML algorithm for predicting student performance was determined.

IV. Results

The backward feature selection algorithm using the linear regression approach selected thirteen features (number of features corresponding to the lowest MSE [25]) to obtain the best features to train our ML methods (see Fig. 1). These features are *age*, *work*, *cumulative_gpa*, *highs-school_type_state*, *scholarship*, *read_frequency_sometimes*, *read_frequency_often*, *prepare_exam*, *listens_sometimes*, *listens_often*, *course_id_5, course_id_6, course_id_8*.

This study's findings provide insights into the predictive capacity of various ML algorithms in determining students' academic performance in any enrolled course. The frequency distribution in Appendix 1 gives an overview of all selected features used in training our ML algorithms for prediction, estimation, and attribution. Table IV in the Appendix shows the distribution of students' socioeconomic characteristics and educational habits by academic performance. The results form descriptive analysis [30] show that larger proportion of males tends to have better performance whereas for females there is a balance between the two groups of below or above average performances. Across gender, male students are seen to have better above-average performance compared to female students. The results in Table I show the various performance levels of all methods in predicting student performance.

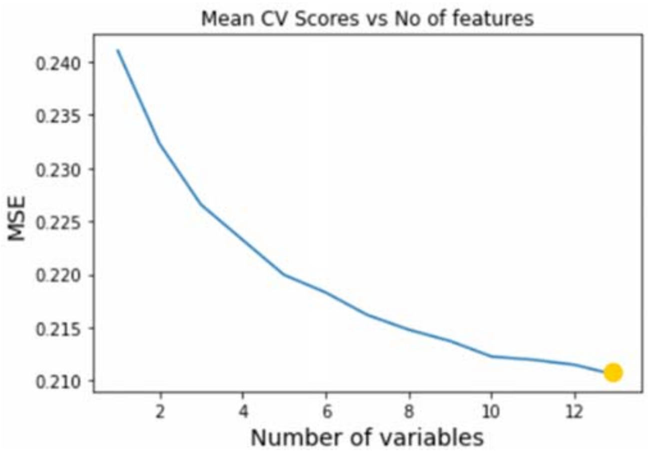

Fig. 1 Validation Score

TABLE I
5-Fold Cross-Validation Results

Method	Accuracy	Precision	Recall	F1 Score
LR	0.71	0.73	0.77	0.74
SVM (Linear Kernel)	0.73	0.76	0.75	0.75
SVM (Non-linear)	0.68	0.71	0.72	0.71
SVM (Radial Basis Function)	0.64	0.67	0.67	0.66
RF	0.66	0.68	0.71	0.69
DT	0.59	0.66	0.55	0.59
KNN	0.66	0.69	0.65	0.66
LDA	0.72	0.76	0.74	0.74
QDA	0.46	0.00	0.00	0.00

Using the accuracy, precision, recall and F1 score, the best-performing methods having above 70% in all test scores are SVM (using a Linear Kernel), LDA, and LR, respectively. Amongst other methods, LR model was used further to understand the existing relationship between significant predictors and response variables for obvious reasons (i.e., high performing prediction rate and easy interpretation of existing relationships). Table II shows the odds ratio (OR) results from the logistic model with significant predictors (P-value < 0.05). This algorithm suggests that the major significant predictors of students' performance at a 5% level of significance are reading habits, listening in class, work, and the cumulative grade point average in the last semester. Table III shows the feature importance of the SVM, RF and DT.

The top five important features from the top selected models are shown. The two common features among all models (SVM, LDA, LR) were the reading and paying attention in class. Such educational habits play a pivotal role in the student's performance.

TABLE II
LR RESULTS

Variables	Logistic Model			
	OR	P-Value	Confidence interval	
			Lower	Upper
Reading (Often vs. Never)	11.004	0.007	1.927	62.855
Listening (Often v.s Never)	7.475	0.009	1.669	33.485
Reading (Sometimes vs. Never)	3.648	0.046	1.024	12.997
Work (Yes vs. No)	2.968	0.045	1.026	8.589
Cumulative GPA	2.509	0.000	1.594	3.951

TABLE III
TOP FIVE IMPORTANT FEATURES FOR THE OUTPERFORMING METHODS

Features	Models		
	SVM	LDA	LR
Reading (often vs never)	X	X	X
Course_id_6	X	X	
Listen (often vs Never)	X	X	X
Course_id_5	X	X	
Work (Yes vs no)		X	X
Cumulative GPA			X
Prepare Exam (often vs never)			
High school (Others vs Private)			
Reading (sometimes vs Never)	X		X
Listen (often vs sometimes)			
Scholarship			
Age			
Course_id_8			

V. DISCUSSION

The best ML algorithms to use in most cases are determined by the nature of the problem. The results from this study showed significantly varying outcomes amongst all investigated methods. The reason for this might be related to the properties of each method and the nature of the relationship existing amongst variables in the dataset used in the data analytics process. Overall, SVM (using a linear kernel) LDA, LR showed the high prediction performance when compared to other investigated ML methods, having an approximately 75% for the F1 score. Other methods such as SVM (using non-linear Kernel), RF, DT, and KNN also had good prediction results (F1 => 60%). The 'reading frequency' feature had the highest contribution (like the logistic result). Two course IDs were found as important among the other selected ones. They can refer to the subjects based on which the performance can vary. It would be recommended to exclude the subjective courses and investigate only the education factors in future. From the LR result in Table II it was entailed that a student who studies regularly would be 11 times more likely to perform above average than a student who never studies. Approximately, a student who reads often would most likely perform 100% better than a student that never reads. Paying attention in class contributes greatly to the final performance of a student in the class. Also, this result shows that students who often listen in class are about seven times more likely to have above-average performance in class compared to students who never concentrate in class. This result also suggests that a student who works is about three times more likely to have an above-average performance compared to students who do not work. Finally, the logistic model shows that the cumulative grade point average (CGPA) may also significantly determine a student's performance in a course. As students work hard to improve their CGPA, so do their chances of having an above-average performance improve by more than 100% on average. Using a suitable learning method, evidence in the literature suggests that features such as high school grade, proficiency in English language, class attendance, study effort, academic self-efficacy and family socioeconomic status may be key determinants of a student's performance in school [29]. This study's findings agree that class attendance, study effort and academy self-efficacy are key determinants of an above-average performance for students. However, no evidence was seen in this study to suggest the same for socioeconomic characteristics. In this study, the model selection was inspired by Gareth's advice that if a set of models show equally good performance, then it is better to select the simplest model, the model with the smallest number of predictors [3].

Considering the LR model results, we can conclude that educational habits such as reading and paying attention in class play a pivotal role in determining whether a particular student may perform above or below average in the class. This study also validates the views of Efron [2] with methods such as SVM having high performance in terms of predicting classification problems.

Even though the results are in line with some studies, a large sample size needs to be used to obtain more precise estimates or coefficients for similar studies in the future.

VI. CONCLUSION

This paper contributed in 1) demonstrating the predictive ability of various ML methods (supervised learning) with data from student records, and 2) identifying the level of influence the features had on student performance.

Among the seven ML methods, three methods showed promising results. Those methods were SVM, LDA, and LR. The findings from this study showed that it is feasible to use ML methods to predict the students' performance. The findings suggest optimizing the methods by adjusting the hyperparameters for the attainment of more accurate outcomes. A Monte-Carlos experiment could be a potential approach to be used.

APPENDIX I

The descriptive statistics of all selected variables obtained from the backward selection algorithm (using linear regression) across the overall student performance are shown in Table IV.

TABLE IV
FREQUENCY DISTRIBUTION OF EDUCATIONAL CHARACTERISTICS BY ACADEMIC PERFORMANCE

Features	Student End of Term Performance					
	Below Average		Above Average		Total	
	N	%	N	%	N	%
Age						
18-21	29	44.62	36	55.38	65	100
22-25	34	48.57	36	51.43	70	100
Above 26	4	40.00	6	60.00	10	100
Work						
No	41	42.71	55	57.29	96	100
Yes	26	53.06	23	46.94	49	100
Cumulative grade point average						
2.00-2.49	27	71.05	11	28.95	38	100
2.50-2.99	9	36.00	16	64.00	25	100
3.00-3.49	18	45.00	22	55.00	40	100
< 2.00	12	70.59	5	29.41	17	100
Above 3.49	1	4.00	24	96.00	25	100
High School Type						
Private	15	60.00	10	40.00	25	100
State	44	42.72	59	57.28	103	100
Other	8	47.06	9	52.94	17	100
Reading Frequency (Non-scientific books/journals)						
Never	17	62.96	10	37.04	27	100
Often	7	36.84	12	63.16	19	100
Sometimes	43	43.43	56	56.57	99	100
Preparation for Exams						
Closest Date to the Exam	59	47.97	64	52.03	123	100
Never	0	0.00	2	100.00	2	100
Regularly during the Semester	8	40.00	12	60.00	20	100
Listens in Class						
Always	13	35.14	24	64.86	37	100
Never	16	55.17	13	44.83	29	100
Sometimes	38	48.10	41	51.90	79	100
Course ID						
1	41	62.12	25	37.88	66	100
2	1	50.00	1	50.00	2	100
3	0	0.00	8	100.00	8	100
4	0	0.00	4	100.00	4	100
5	0	0.00	7	100.00	7	100
6	0	0.00	8	100.00	8	100
7	0	0.00	15	100.00	15	100
8	14	100.00	0	0.00	14	100
9	11	52.38	10	47.62	21	100
Grade						
AA	0	0.00	17	100.00	17	100
BA	0	0.00	13	100.00	13	100
BB	0	0.00	17	100.00	17	100
CB	0	0.00	10	100.00	10	100
CC	0	0.00	21	100.00	21	100
DC	24	100.00	0	0.00	24	100
DD	35	100.00	0	0.00	35	100
FAIL	8	100.00	0	0.00	8	100
Total	67	46.21	78	53.79	145	100

REFERENCES

[1] F. Benavides, "Equity and Quality in Education," 2012. https://www.oecd.org/education/school/50293148.pdf (accessed May 20, 2022).

[2] B. Efron, "Prediction, Estimation, and Attribution," Journal of the American Statistical Association, vol. 115, no. 530, pp. 636–655, Apr. 2020, doi: 10.1080/01621459.2020.1762613.

[3] G. James, D. Witten, T. Hastie, and R. Tibshirani, An introduction to statistical learning: with applications in R. Springer, 2013.

[4] A. Seetharam Nagesh, C. Satyamurty and K. Akhila, "Predicting Student Performance Using KNN Classification in Bigdata Environment", CVR Journal of Science and Technology, vol. 13, no. 7, pp. 83-87, 2017. Available: ISSN 2277 – 3916.

[5] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman, "Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique," Decision Analytics, vol. 2, no. 1, Mar. 2015, doi: 10.1186/s40165-014-0010-2

[6] A. K. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 2, p. 26, 2018, doi: 10.9781/ijimai.2018.02.004.

[7] S. K. Ghosh and F. Janan, "Prediction of Student's Performance Using Random Forest Classifier," 11th Annual International Conference on Industrial Engineering and Operations, Sep. 2022.

[8] R. Katarya, "Review: Predicting the Performance of Students Using Machine Learning Classification Techniques," Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE, pp. 36–41, 2019.

[9] K. Sixhaxa, A. Jadhav and R. Ajoodha, "Predicting Students Performance in Exams using Machine Learning Techniques," 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2022, pp. 635-640, doi: 10.1109/Confluence52989.2022.9734218.

[10] H. S. Liu, "Research on the College Students' Physique Test Based on Artificial Neural Network and Support Vector Machines Approach," Advanced Materials Research, vol. 945–949, pp. 3558–3561, Jun. 2014, doi: 10.4028/www.scientific.net/amr.945-949.3558.

[11] H.-H. Huang and T. Zhang, "Robust discriminant analysis using multi-directional projection pursuit," Pattern Recognition Letters, vol. 138, pp. 651–656, Oct. 2020, doi: 10.1016/j.patrec.2020.09.013.

[12] A. Salah Hashim, W. Akeel Awadh, and A. Khalaf Hamoud, "Student Performance Prediction Model based on Supervised Machine Learning Algorithms," IOP Conference Series: Materials Science and Engineering, vol. 928, p. 032019, Nov. 2020, doi: 10.1088/1757-899x/928/3/032019.

[13] N. Mohd Rusli, "Predicting Students' Academic Achievement: Comparison between LogisticRegression, Artificial Neural Network, and Neuro-Fuzzy."

[14] Y.-Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," Shanghai archives of psychiatry, vol. 27, no. 2, pp. 130–5, 2015, doi: 10.11919/j.issn.1002-0829.215044.

[15] P. Aye, "Predicting Students Performance in the University System Using Discriminant Analysis Method," Multidisciplinary International Journal of Research and Development, vol. 01, no. 02, 2021.

[16] K. Elkhalil, A. Kammoun, R. Couillet, T. Y. Al-Naffouri, and M.-S. Alouini, "Asymptotic Performance of Regularized Quadratic Discriminant Analysis Based Classifiers," IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6, 2017.

[17] S. Sothan, "The determinants of academic performance: evidence from a Cambodian University", *Studies in Higher Education*, vol. 44, no. 11, pp. 2096-2111, 2018. Available: 10.1080/03075079.2018.1496408.

[18] R. A. FISHER, "The Use of Multiple Measurements in Taxonomic Problems," Annals of Eugenics, vol. 7, no. 2, pp. 179–188, Sep. 1936, doi: 10.1111/j.1469-1809.1936.tb02137.x.

[19] A. Abraham and Swagatam Das, *Computational Intelligence in Power Engineering*. Berlin Springer Berlin, 201.

[20] "EDA on Higher Education Students Performance," kaggle.com, 2019. https://www.kaggle.com/code/janietran/eda-on-higher-education-students-performance, Accessed: May 20, 2022.

[21] J. Chain, "A Multilevel Analysis of Student, Family, and School Factors Associated with Latino/a Parental Involvement in the Middle School Learning Environment," scholarsbank.uoregon.edu, Nov. 2016, Accessed: May 20, 2022. (Online). Available: http://hdl.handle.net/1794/20664.

[22] M. Saarela and S. Jauhiainen, "Comparison of feature importance measures as explanations for classification models," SN Applied Sciences, vol. 3, no. 2, Feb. 2021, doi: 10.1007/s42452-021-04148-9.

[23] A. J. Ferreira and M. A. T. Figueiredo, "Efficient feature selection filters for high-dimensional data," Pattern Recognition Letters, vol. 33, no. 13,

pp. 1794–1804, Oct. 2012, doi: 10.1016/j.patrec.2012.05.019 .

[24] Z. Zhao and H. Liu, "Spectral Feature Selection for Supervised and Unsupervised Learning," Proceedings of the 24 th International Conference on Machine Learning, 2007. https://www.public.asu.edu/~huanliu/papers/icml07.pdf (accessed May 20, 2022).

[25] J. W. Nagge, "Regarding the Law of Parsimony," The Pedagogical Seminary and Journal of Genetic Psychology, vol. 41, no. 2, pp. 492–494, Dec. 1932, doi: 10.1080/08856559.1932.10533115.

[26] M. Wallowa Mwadulo, "A Review on Feature Selection Methods For Classification Tasks," International Journal of Computer Applications Technology and Research, vol. 5, no. 6, pp. 395–402, Jun. 2016, doi: 10.7753/ijcatr0506.1013.

[27] I. Guyon and A. De, "An Introduction to Variable and Feature Selection André Elisseeff," Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003, (Online). Available: https://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf?ref=driverlayer.com/web. (accessed May 20, 2022).

[28] G. Doquire and M. Verleysen, "An Hybrid Approach to Feature Selection for Mixed Categorical and Continuous Data," Proceedings of the International Conference on Knowledge Discovery and Information Retrieval, 2011, doi: 10.5220/0003634903860393. (accessed May 20, 2022).

[29] M. Behzad, K. Asghari, M. Eazi, and M. Palhang, “Generalization performance of support vector machines and neural networks in runoff modeling,” Expert Systems with Applications, vol. 36, no. 4, pp. 7624–7629, May 2009, doi: 10.1016/j.eswa.2008.09.053.

[30] Descriptive analysis table, 2022. https://www.kaggle.com/datasets/gmasschannel/descriptive-table-of-student-data. (accessed Aug 12, 2022)