# An Heuristic Feature Selection Algorithm to Evaluate Academic Performance of Students

Samuel-Soma M. Ajibade
*Computer Science Department, UTM*
Johor Bahru, Malaysia
*samuel.soma@yahoo.com*

Nor Bahiah Ahmad
*Computer Science Department UTM*
Johor Bahru, Malaysia
*bahiah@utm.my*

Siti Mariyam Shamsuddin
*Computer Science Department UTM*
Johor Bahru, Malaysia
*Mariyam@utm.my*

*Abstract*—**The value of schooling and academic performance of student is the topmost priority of all academic institutions. Educational Data Mining (EDM) is an evolving area of research which aids academic institutions to enhance their student's performances. Feature Selection algorithms eradicates inapt and unrelated data from the dataset, thereby increasing the classifiers performances that are utilized in EDM. This aim of this paper is to evaluate the performance of students utilizing a heuristic technique known as Differential Evolution for feature selection algorithms on the dataset of students and some other feature selection algorithms have also been used which have never been used before on the dataset. Also, classification techniques such as Naïve Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN) and Discriminant Analysis (DISC) were used to evaluate. The Differential Evolution (DE) algorithm is proposed as a better feature selection algorithm for evaluating the academic performance of students and this gave a better accuracy than other feature selection algorithm that were used. The outcome of the different feature selection algorithms and classification techniques will help researchers to find the finest combinations of the classifiers and feature selection algorithms. This paper is a step towards playing an important role in enhancing the standard of education in academic institutions and also to carefully guide researchers in strategically interfering in academic issues.**

***Keywords—educational data mining, differential evolution, student performance, feature selection algorithm, classification algorithm***

## I. INTRODUCTION

Educational Data Mining, is a field of methodical review and analysis which depends on the enhancement of techniques which is implied not just for information disclosure that is inside the particular sorts of data that is gotten from various educational settings yet in addition for utilizing those techniques afterwards to efficiently comprehend the learners as well as the surroundings that they learn in, has appeared as a free research field as of late [1].

Feature Selection is a functioning and dynamic area of study which comprises of machine learning and data mining. The process of feature selection is performed so as to choose a subclass by removing non prescient information. Likewise, the accuracy of performance prediction is upsurged and diminishes the multifaceted nature of academic outcomes [2, 3, 4]. When the feature techniques are utilized, then the efficiency of prediction model is enhanced.

Feature selection has been effectively connected to numerous fields, for example, text categorization, face recognition, cancer classification, gene classification, recommender system The entire space of exploring consists of all the possible subsets of features and this suggests that the request space measure is 2n where n is the quantity of the real features. Hence, the issue of finding the optimum feature subset is a NP-hard issue [5,6]. In previous works, feature selection algorithms have been applied to predict student performance classifier accuracy, but the use of heuristic algorithm is still relatively low, hence the accuracy performances has been low. In this paper, the DE has been introduced as a better feature selection algorithm for evaluating the performance of students. The DE is best known for reducing computation time and increases accuracy of classifiers. This work attempts to distinguish the best combinations of feature section techniques and classification methods on the dataset of students. Also, the low performances have been accredited to the inadequate use of variables as well as singular use of base classifier and as such the use of heuristic algorithms in prediction of performance has been recommended [11]. Section II below discusses some reviews of literatures, while section III describes the methodology used. In section IV, the results are discussed, and section V draws the conclusion of this work.

## II. LITERATURE REVIEW

The Differential Evolution (DE) is a masses-based algorithm that can be seen as like Genetic Algorithm (GA) since it uses operators such as: crossover, mutation and selection. The rule differentiate among DE and GA is in structuring better goals, where DE relies upon the operation of mutation and GA relies upon the operation of crossover. This DE was founded by Storn and Price in 1997 [7], Who use a certifiable number enhancer and utilizes the operators of DE to the arrangements of the features that makes comparative features to be practiced on different events in the solution vector.

A hybrid approach that links Artificial Bee Colony (ABC) optimization and DE together is suggested by [8] and DE mutation operator is associated with a fated percent of good resolution found by the ABC. In this study, ABC algorithm which is clear and capable is generated, and this is contrasted with customary ABC algorithm. The result demonstrate that the hybrid algorithm is superior to ABC for benchmark purposes.

A feature selection approach that is based on wrapper method that makes use of ABC is proposed by [9]. In this procedure, the ABC created the feature subsets and a classifier such as J48 is used to evaluate the feature subsets. The C4.5 algorithms that is utilised to create decision trees is applied in Weka as a classifier known as J48. This approach is probably assessed with over 10 different datasets that are gotten from the University of California Irvine (UCI) Repository. Therefore, it is found that the precision of the classification vacillates within the scope of 81.26% and 98.55% for the datasets, and the feature measurement is decreased while the classification precision improved.

In [10], a hybrid search approach is recommended by combining Harmony Search Algorithm (HSA) and Stochastic Local Search (SLS) for the selection of feature on classification. Then a probabilistic choice strategy is used to apply stochastic exploitation. Hence, this algorithm is wrapped with Support Vector Machine (SVM) classifier. The exploratory results show that HSA-SLS technique is superior to HSA and GA for selection of features.

## III. METHODOLOGY

This paper aims to analyse the performances of various feature selection techniques that are applied on various methods of classification by the use of educational dataset. The evaluation between various feature selection algorithms gives a better view to data miners concerning the performance of various feature selection strategies on student data. In order to accomplish our aim in this study, a student dataset is gotten from a reliable repository after which differential evolution (DE) and other feature selection algorithms were applied to it, which has never been used before on the dataset. Various classification techniques are thereby applied by utilizing feature selection algorithms, and also it was analysed to check the one with the most accurate performance of all techniques used on the data of students.

The dataset used in this work is gotten from a LMS called Kalboard 360 [11]. The Kalboard 360 is a multi-agent LMS which is established to in order for learning to be improved by the use of leading-edge technology. The data was collected using a learner activity tracker tool known as Experience API (XAPI) [10]. XAPI is a component of the Training and Learning Architecture (TLA) that allows the tracking of learning experiences and actions of learners like reading an article or watching an educational video. In this study, the dataset numbers into 500 students with 16 features.

At the moment, feature selection is a very vital and a frequently used data preprocessing method and it is a very vital factor in machine learning. This study makes emphasis on six (6) vital feature selection algorithms, namely: Correlation feature selection (CFS), Rank Importance of predictors, kullback-leibler divergence, sequential forward selection (SFS), sequential backward selection (SBS) and a newly introduced heuristic technique known as Differential Evolution (DE) which were evaluated. These feature selection algorithms consist of both filter and wrapped methods. Also, classification techniques such as Naïve Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN) and Discriminant Analysis (DISC) were used to evaluate.

The algorithm proficiency is estimated through performance evaluation measures such as: Precision, Recall, F-measure and accuracy of prediction (Correctly Classified Instances). F-measure is the symphonious mean of recall and precision [12].

## IV. RESULT AND DISCUSSION

This paper centers around the feature selection techniques performance alongside different methods of classification. The results of five approaches of feature selection which were used on four (4) classifiers are shown in tables below.

.

### A. Correlation Feature Selection (CFS)

CFS looks out for the subsets of feature as per the level of repetition amongst the features [13]. The assessor means to discover the feature subsets which are independently exceedingly linked with the class however have quite low inter-relationship.

TABLE I. PERFORMANCE EVALUATION OF CFS

| Classifiers | Precision | Recall | F-Measure |
|---|---|---|---|
| NB | 72.16 | 68.25 | 65.85 |
| DT | 80.55 | 78.57 | 75.54 |
| KNN | 77.27 | 74.21 | 73.40 |
| DISC | 75.15 | 71.83 | 70.40 |

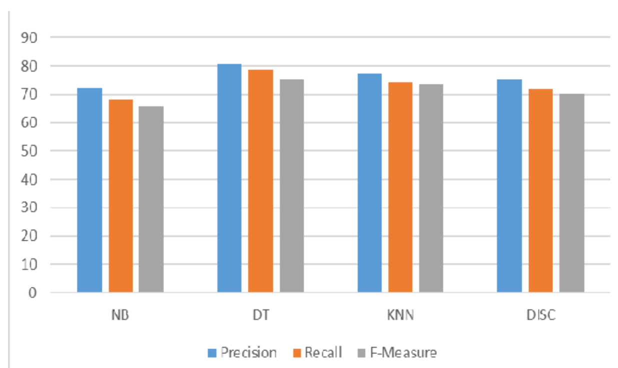


Fig. 1. Precision, Recall and F-Measure values for CFS

Table 1 shows the values of Precision, Recall and F-Measure for the four classifiers using correlation feature selection. Fig 1 displays Table I in graphs. The outcome in Table 1 displays NB as having the minimum performance in terms of the evaluation measure, while DT has the highest performance.

### B. Relief F

The Relief F method is using a method to rank the importance of the features; so the features will be sorted based on their importance rank [14].

Table II shows the values of Precision, Recall and F-Measure for the four classifiers using Relief F. Fig 2 represents Table II in graphs. The result in Table II shows that NB has the lowest performance in terms of the evaluation measure, while KNN has the highest performance.

TABLE II. PERFORMANCE EVALUATION OF RELIEF F

| Classifiers | Precision | Recall | F-Measure |
|---|---|---|---|
| NB | 72.16 | 68.25 | 65.85 |
| DT | 74.14 | 73.81 | 68.90 |
| KNN | 77.27 | 74.21 | 73.40 |
| DISC | 75.15 | 71.83 | 70.40 |

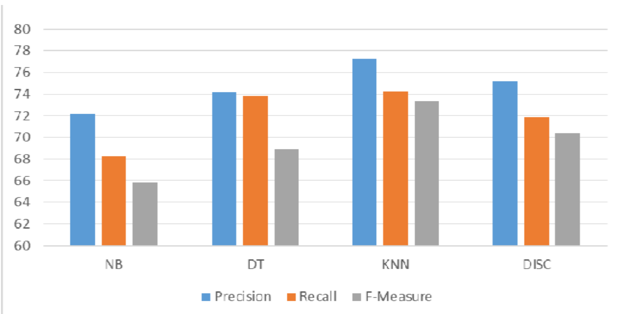


Fig. 2. Precision, Recall and F-Measure values for Relief F

## C. Kullback-Leibler Divergence

The Kullback-Leibler divergence method first normalizes the indexed features and classes. Then calculates some kind of distance measure called Kullback-Leibler divergence; then the features are sorted based on this measure [15].

TABLE III. PERFORMANCE EVALUATION OF KULLBACK-LEIBLER DIVERGENCE

| Classifiers | Precision | Recall | F-Measure |
|---|---|---|---|
| NB | 71.15 | 67.06 | 64.30 |
| DT | 80.63 | 76.19 | 74.91 |
| KNN | 74.35 | 69.84 | 70.86 |
| DISC | 74.14 | 70.63 | 68.90 |

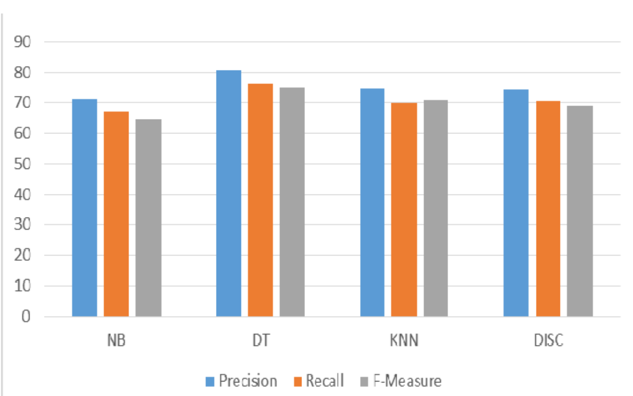


Fig. 3. Precision, Recall and F-Measure values for Kullback-Leibler Divergence

Table III shows the values of Precision, Recall and F-Measure for the four classifiers using kullback-leibler divergence. Fig 3 displays Table III in graphs. The result in Table III shows that NB has the lowest performance in terms of the evaluation measure, while DT has the highest performance.

## D. Sequential Forward Selection (SFS)

The SFS performs better when the ideal subset consists of a set number of features [16]. A huge number of states can be conceivably assessed if the search is close to the unfilled set. Towards the full set, the section analyzed by SFS is smaller since the vast majority of the features have been chosen.

TABLE IV. PERFORMANCE EVALUATION OF SEQUENTIAL FORWARD SELECTION (SFS)

| Classifiers | Precision | Recall | F-Measure |
|---|---|---|---|
| NB | 77.88 | 75.00 | 72.84 |
| DT | 79.37 | 76.98 | 75.93 |
| KNN | 77.27 | 74.21 | 73.40 |
| DISC | 76.19 | 73.02 | 71.90 |

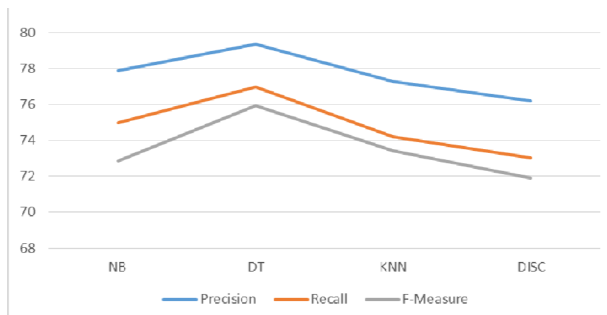


Fig. 4. Precision, Recall and F-Measure values for SFS

Table IV shows the values of Precision, Recall and F-Measure for the four classifiers using Sequential Forward Selection (SFS). Fig 4 displays Table IV in graphs. The result in Table IV displays that DISC has the lowest performance in terms of the evaluation measure, while DT has the highest performance.

## E. Sequential Backward Selection (SBS)

Sequential Backward Selection (SBS) [17] works the other way of SFS. SBS is at its best when the ideal feature subset has an expansive number of features, since SBS invests the greater part of its energy visiting substantial subsets.

TABLE V. PERFORMANCE EVALUATION OF SEQUENTIAL BACKWARD SELECTION (SBS)

| Classifiers | Precision | Recall | F-Measure |
|---|---|---|---|
| NB | 77.88 | 75.00 | 70.84 |
| DT | 82.54 | 80.95 | 78.66 |
| KNN | 81.52 | 79.76 | 77.10 |
| DISC | 82.54 | 80.95 | 78.66 |

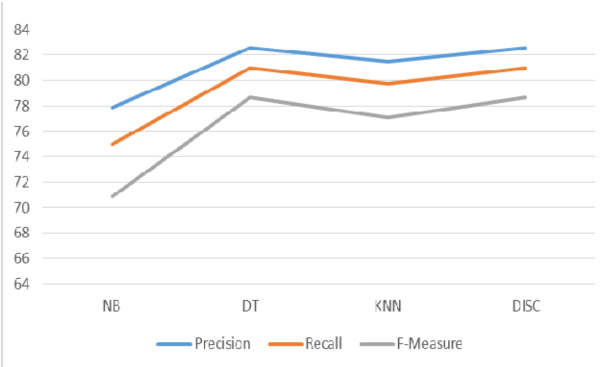


Fig. 5. Precision, Recall and F-Measure values for SBS

Table V shows the values of Precision, Recall and F-Measure for the four classifiers using Sequential Backward Selection (SBS). Fig 5 presents Table V in graphs. The result in Table V displays that NB has the lowest performance in terms of the evaluation measure, while DT and DISC both have the highest performance.

TABLE VI. PERFORMANCE EVALUATION OF DIFFERENTIAL EVOLUTION

| Classifiers | Precision | Recall | F-Measure |
|---|---|---|---|
| NB | 71.15 | 67.05 | 64.30 |
| DT | 80.45 | 80.44 | 78.60 |
| KNN | 84.35 | 82.52 | 82.01 |
| DISC | 81.04 | 80.95 | 79.66 |

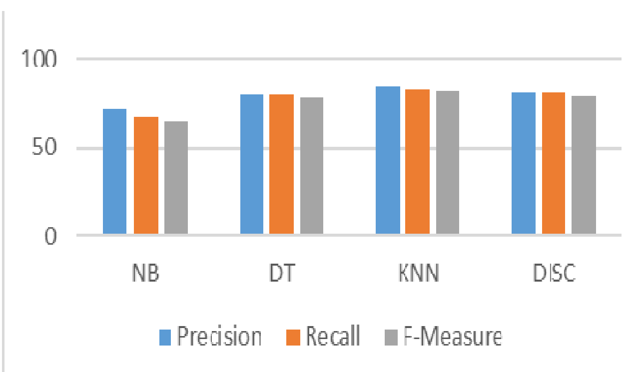


Fig. 6. Precision, Recall and F-Measure values for DE

Table VI demonstrates the values of Precision, Recall and F-Measure for the four classifiers using DE. Fig 6 displays the graph of Table VI. The result in Table VI displays that NB has the minimum performance in terms of the evaluation measure, while KNN have the highest performance.

TABLE VII. PERFORMANCE EVALUATION OF FS ALGORITHMS WITH CCI

| | **Prediction Accuracy (CCI)** | | | | | |
|---|---|---|---|---|---|---|
| **FS Algorithms** | **NB** | **DT** | **KNN** | **DISC** | **Mean** | **Variance** |
| CFS | 73.61 | 81.94 | 80.56 | 77.78 | **78.5** | **10.13** |
| Rank Importance of Predictors | 73.61 | 76.39 | 80.56 | 77.78 | **77.1** | **6.28** |
| Kullback-Leibler Divergence | 63.89 | 83.33 | 79.17 | 76.39 | **75.7** | **52.55** |
| SFS | 77.78 | 83.33 | 80.56 | 79.17 | **80.2** | **4.21** |
| SBS | 77.78 | 84.72 | 83.33 | 84.72 | **82.6** | **8.18** |
| DE | 79.13 | 84.94 | 85.21 | 83.09 | **83.09** | **7.87** |

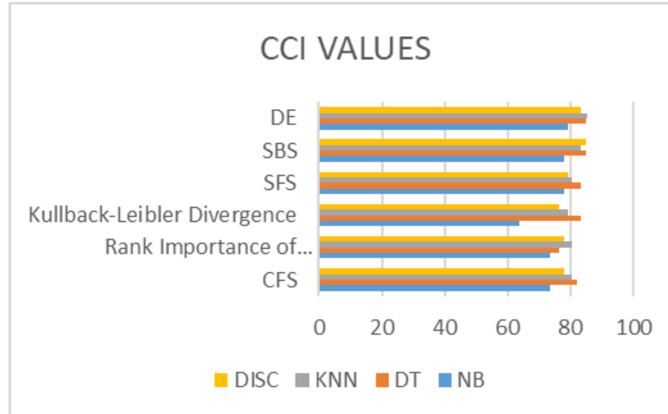


Fig. 7. CCI Values for FS Algorithms

The result in Table VII shows the values of each Feature selection algorithm with various classifiers. The mean and variance of the Feature selection algorithm was gotten so as to check the variations in the FS algorithm performance with the classification techniques. KNN displays better performance that the other classifiers and the DE also shows better results as shown in Fig 8 and Fig. 9. The results in Fig 8 and Fig 9 shows the mean and variance of the FS algorithms.

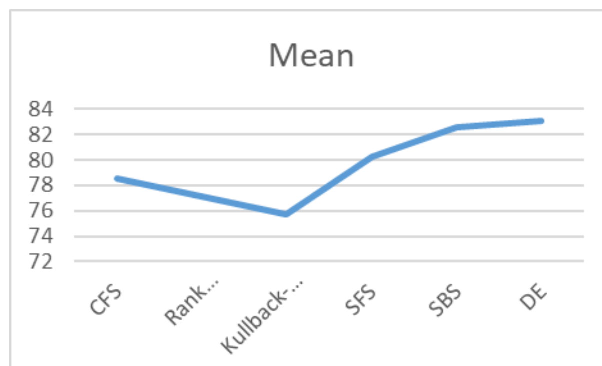


Fig. 8. Mean Values for FS Algorithms

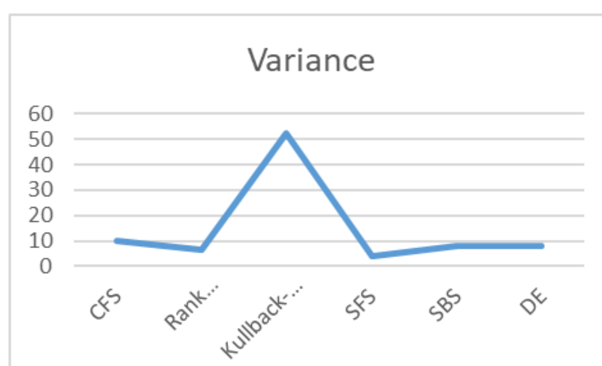


Fig. 9. Variance Values for FS Algorithms

## CONCLUSION AND FUTURE WORKS

In this work, various feature selection techniques have been evaluated in performance. The outcomes demonstrate that there is a noteworthy change in the performances. The DE exhibits preferred outcomes over other feature selection algorithms. This investigation likewise demonstrates that KNN classifier performed superior to other various classifiers on the student data. So as to be able to predict the performance of students that have expansive number of features, we analyzed the utilization of wrapper and filter methods. Later on, more understudy datasets of different estimated can be utilized for assessments. In future study, analysis can be performed with the use of larger datasets. Furthermore, this study can be enhanced by the use of hybrid metaheuristics algorithms for feature selection on the student data in order to predict efficiently the performance of students.

## REFERENCES


[1] E. Osmanbegović, M. Suljić, and H. Agić, "Determining dominant factor for students' performance prediction by using data mining classification algorithms" *Tranzicija,* vol. 16, pp. 147-158, 2015.

[2] M. Ramaswami and R. Bhaskaran, "A study on feature selection techniques in educational data mining," *arXiv preprint arXiv:0912.3924,* 2009.

[3] D. Koller and M. Sahami, "Toward optimal feature selection," Stanford InfoLab1996.

[4] P. Mitra, C. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE transactions on pattern analysis and machine intelligence,* vol. 24, pp. 301-312, 2002.

[5] I.A. Gheyas, L.S. Smith, Feature subset selection in large dimensionality domains, Pattern Recogn. 43 (2010) 5–13.

[6] H. Liu, H. Motoda, Computational Methods of Feature Selection, Chapman & Hall/CRC, 2007.

[7] R. Storn, K. Price Differential Evolution—A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. Journal of Global Optimization. 1997: 341–359

[8] Y. Xu, P. Fan, & L. Yuan. A simple and efficient artificial bee colony algorithm. Mathematical Problems in Engineering, 2013.

[9] S. Palanisamy & S. Kanmani. Artificial bee colony approach for optimizing feature selection. *International Journal of Science Issues, 9*(3), 432–438, 2012.

[10] M. Nekkaa & D. Boughaci. Hybrid harmoy search combined with stochastic local search for feature selection. *Neural Processing Letters*, pg. 1–2, 2015.

[11] S. S. M. Ajibade, N. B. Ahmad & S. M. Shamsuddin, "A Data Mining Approach to Predict Academic Performance of Students Using Ensemble Techniques," International Conference on Intelligent Systems Design and Applications. Springer, Cham, pp. 749-760, 2018.

[12] T. Velmurugan and C. Anuradha, "Performance Evaluation of Feature Selection Algorithms in Educational Data Mining," Performance Evaluation, vol. 5, 2016.

[13] M. Doshi. "Correlation based feature selection (CFS) technique to predict student performance," *International Journal of Computer Networks & Communications*, *6*(3), 197, 2014

[14] S. Dahiya, S. Handa & N. Singh. "A rank aggregation algorithm for ensemble of multiple feature selection techniques in credit risk evaluation," *International Journal of Advanced Research in Artificial Intelligence*, *5*(9), 1-8, 2016.

[15] Z. Lian, A. Godil, B. Bustos, M. Daoudi, J. Hermans, S. Kawamura & Y. Ohkita. "A comparison of methods for non-rigid 3D shape retrieval," *Pattern Recognition*, *46*(1), 449-461, 2013.

[16] D. Ververidis & C. Kotropoulos, "Sequential forward feature selection with low computational cost," In *2005 13th European Signal Processing Conference* (pp. 1-4). IEEE, 2005.

[17] S. Mustafa, "Feature selection using sequential backward method in melanoma recognition," *13th International Conference on Electronics, Computer and Computation (ICECCO)* (pp. 1-4). IEEE, 2017.