

# An Efficient Clustering Based Feature Selection for Predicting Student Performance

R. Sasi regha M.Sc., M.Phil.<sup>\*1</sup>, Dr R. Uma rani Ph.d.<sup>2</sup>

<sup>1</sup>Assistant professor, Department of computer science, SSM College of Arts & Science, Tamilnadu, India

<sup>2</sup>Associate professor, Department of computer science, Sri Sarada College for Women, Tamilnadu, India

<sup>1</sup>sasireghaphd123@gmail.com

**Abstract**– The student failure prediction at school has turn into a very complicated challenge owing to both the large number of factors which can influence the high performance of students and the balanced nature of student’s databases which are maintaining by Educational Data Mining (EDM) techniques. The main goal of this study is to detect and remove the both irrelevant and redundant features that can be used to enhance the classification accuracy in predicting the student performance. This goal is achieved by establishing the novel technique for feature or attributes selection process by hybrid of Artificial fish swarm-Cuckoo Search optimization algorithm to remove the irrelevant features or obtaining relevant features. Also, Non-negative Matrix Factorization Clustering algorithm (NMFC) performs the removal of redundant feature or attributes which are presented in the relevant features. The performance of this technique is analyzed by using the student database which comprises the gathering of student’s information from different colleges. For analyzing the performance of this technique, the comparative evaluation is carried out between the classifiers used in this research such as Prism and J48 without the feature selection and classifiers with our proposed technique. The experimental consequences illustrate that hybrid of artificial fish swarm-cuckoo search optimization feature selection along with NMFC approach is accomplishing high accuracy rate than other techniques. This study facilitates us to enhance the performance of the student’s failure and dropout prediction. In other words, this helps to increase the accuracy of the classification result.

**Keywords**– Educational Data Mining (EDM), Feature Selection, Non-negative Matrix Factorization Clustering (NMFC), Artificial fish swarm, Cuckoo search optimization

## I. INTRODUCTION

Educational Data Mining (EDM) refers the research field which is concerned with the applications that are developed by means of using data mining, machine learning and statistics for generation of information from educational settings and by using these methods the student’s performance in the settings where they learn the knowledge. This EDM interests on the development of advanced tools and algorithms to realize the data patterns. EDM investigates the learning theories and updates the educational practice.

Educational Data Mining (EDM) is appearing as research area with computational and psychological methods to understand the students learning performance. In modern world, EDM has the ability to log fine-grained data about student’s computer-based learning atmosphere which increases the significance in developing techniques to analyse the high amounts of data generated in educational settings.

Generally, EDM consists of four processes [1] such as pre-processing, validation, prediction and decision-making. Pre-processing involves classification, regression, clustering and etc. The classified features should valid to avoid overfitting. The valid features are predicted to know about the student’s learning atmosphere. This prediction process helps us to make the decision in order to avoid student’s failure and dropout. In EDM, the data composed from the students are separated for feature classification process which is used to develop the prediction model. In EDM techniques, numbers of feature classification methods are considered without feature selection so that these methods provide less accuracy in data prediction process. Therefore, the feature classification process includes the feature selection method to provide high accuracy in order to predicting the data accurately.

The goal of this paper is to propose a novel technique to select the feature for predicting the student performance in educational field. In this paper, the information in database which is composed from the students is processed by the feature selection technique based on our proposed scheme called as Artificial fish swarm-Cuckoo search optimization to select the features. This proposed optimization technique eliminates the irrelevant features or attributes in order to acquire the relevant features. Then the redundant feature in the relevant features is also removed by the proposed technique named as Non-negative Matrix Factorization

(NMF) which is based on clustering method. Hence, this proposed feature selection technique helps to identify the relevant features and also increases the accuracy rate in classification process with feature selection.

The remaining part of this paper is organized as follows: Section 2 describes works that are related to our proposed methodology. Section 3 explains the concept of our proposed feature selection technique for classification of student's performance. Section 4 presents the performance evaluation results of the proposed technique. Section 5 concludes the research work and also provides the directions for future research.

## II. RELATED WORKS

Baradwaj, B. K., & Pal, S., [2] proposed the analysis of educational data to observe the student's performance. This research utilizes the decision tree approach for classification process which is provided to evaluate the performance of the students. Here, the knowledge which describes the student's performance in end semester examination is extracted to identify the dropouts and the special students who needs appropriate counselling. The accuracy of the decision tree classification technique is investigated to predict the student performance. But this technique has limitations such as the pruning process is necessary which makes more complexity.

Delibasic, B., et al [3] proposed block-box and white-box decision tree algorithms to know about the student's acceptance of data mining algorithms. The black-box algorithm is utilized for permitting the users to utilize predefined algorithms to set parameters. The white-box algorithm is utilized for letting the users to assemble algorithms from algorithm building blocks. This research is utilized to analyse the acceptance of the students for using these algorithms in education data mining technology. The limitation is the white-box algorithm is more complex.

Kabakchieva, D., [4] proposed the data mining classifications algorithms to predict the student's performance. This research uses the various data mining classification algorithms such as Bayes classifiers, decision tree classifiers and also Nearest Neighbour classifier. This research allows forecasting the student's performance at university range according to the student's personal and their pre-university characteristics. But the limitations in this work are that the determination of k value for nearest neighbour classifiers and Bayes classifier has less accuracy.

Nikam, S. S., [5] presented the comparative study of data mining classification algorithms to know about their features and limitations. This research compares the different classification algorithms such as C4.5, ID3, k-nearest neighbour classifier, Naïve Bayes, SVM and ANN in data mining for the purpose of feature classification. By using this research, the features and limitations for corresponding classification techniques are identified. Also, this work shows that determination and grouping of data for each classification technique.

ElGamal, A. F., [6] proposed a data mining classification model to predict the performance of the student in programming course. This research absorbs the student's mathematical skills, programming aptitude, problem solving skills, computer programming experience and e-learning usage, etc. This proposed work involves the three processes such as data pre-processing, attribute selection and rule extraction algorithm. In pre-processing, fuzzification process is carried out to transform the attributes into linguistic form. For classification process rule extraction is performed based on decision tree classification. However, the limitations in this work are decision tree classifies the features by rectangular partitioning and also suffer from over-fitting.

Singh, B., et al [7] proposed a feature selection technique based on symmetric uncertainty for high dimensional data. In this approach, the sorted features are initially partitioned and the essential features are searched in both forward and backward fashion. In this paper, correlation based feature ranking method called as symmetric uncertainty is used to select the features. This work efficiently identifies the major features and eliminates the redundant features. The limitations in this algorithm are it is well for only numerical data not for mixed type of data.

Singh, M., et al [8] proposed the model for feature extraction for identifying the student's risk level in academic activities. This research uses Naïve Bayes classifier to predict the features which are used for recognizing the performance of the second year students in their computer and application course by taking partially relevant or fully relevant features. This model has ability for extracting the fitness procedure sequences by each student who is predicted in at-risk group. This identification improves the student's performance in academia. However, the classification accuracy is still poor.

Mustafa, M. N., et al [9] presented the dynamic prediction model to identify the student's dropout in developing country. This research applies chi square test for feature selection from the separating factors such as gender, financial conditions, and dropping year, etc. In this work, Classification and Regression Tree (CART) and CHAID tree is used as data mining techniques to predict the features. However, the classification accuracy is low. The limitations in this research are it is only based on background information.

Rachburee, N., & Punlumjeak, W., [10] presented the comparison between feature selection techniques in data mining. This research compares the feature selection techniques such as Greedy, IG-ratio, Chi-square and

mRMR to identify the efficiency of student's performance prediction methods. The classification models which are used in this work are such as k-nearest neighbour, neural network, naïve bayes and decision tree. This work uses best couple validation to get high accuracy. However, there are various feature selection and classification techniques are also available to improve the efficiency of student's performance.

Kaur, P., et al [11] proposed the data mining algorithms for classification and prediction for prediction of slow learners in education field. This research tests the database of the student's academic records and applies several classification techniques such as Naive Bayes, SMO, J48 and REPTree. This paper also used to know about the importance of the prediction and classification algorithms in educational data mining. This work is further investigated for other fields such as medicine, sports and etc.

Shahiri, A. M., et al [12] presented the data mining techniques for predicting the student's performance. This research is used to study about the prediction methods for identifying student's performance in Malaysian institutions. This work mainly focused on the prediction algorithm which is used for identifying the most popular attributes in database of the students. Also, this work studies about the variables which are used for analysis of the student's performance. However, the meta-analysis for predicting student's performance helps for monitoring by systematic way.

Sen, B., et al [13] proposed data mining method for predicting and analyzing student's placement-test scores. This research uses large and feature rich database which are collected from the secondary education transition system in Turkey. This work uses sensitivity analysis for prediction methods to identify the most important predictors. However, the prediction models are depends on the most priority factors to analyse the student's performance.

Emary, E., et al [14] proposed the feature selection approach based on the binary Gray Wolf Optimization (bGWO). This research uses the bGWO for selecting the features in order to maximize the classification accuracy by selecting relevant features. This bGWO is used to find the optimal regions in the complex search space. This work is compared with the particle swarm optimization and genetic algorithms. The limitations in this work are, to avoid repeatability and robustness this method converges to similar solutions.

### III. ARTIFICIAL FISH SWARM-CUCKOO SEARCH OPTIMIZATION BASED FEATURE SELECTION

Our proposed system introduces a novel technique for the purpose of feature or attributes selection which is called as hybrid of Artificial fish swarm-Cuckoo search optimization. The effectiveness of feature selection is achieved by our proposed technique which incorporated the two coupled components of irrelevant and redundant feature elimination. In our existing method NMFC uses the Symmetric Uncertainty (SU) to remove the irrelevant features. The SU should maintain the mutual information and entropy. To avoid this limitation, in this paper, Artificial Fish Swarm-Cuckoo Search Optimization to remove the irrelevant features or to obtain the relevant features rapidly. After selecting the relevant features, we have to find the redundant features which are presented in the relevant features. For removal of redundant features, we are employing a novel Non-negative Matrix Factorization based Clustering technique. After feature selection procedure, two classification methods such as Prism and J48 are used to predict the student's performance.

#### A. Removal of Irrelevant Feature

Irrelevant features are eliminated by using Hybridization of Artificial Fish Swarm algorithm (AFS) and Cuckoo Search Optimization (CSO) algorithm. The AFS algorithm is a type of random search algorithm which has characteristics of parallelism, simplicity and tracking but it has low speed of convergence rate. Thus, the hybridization of AFS algorithm with CSO is used to achieve high convergence rate when their search area is limited by levy flight optimization.

This algorithm initialized including the number of features, number of iterations, number of fishes, number of selected features, etc. In initial iteration, all fish will arbitrarily pick a feature subset from  $m$  features. Only the best subsets ( $k < nf$ ) can be utilized to bring up to date the visual position and manipulate the feature subsets of the next iteration. In further iterations, all fish can begin with  $m-p$  features which are arbitrarily chosen from the previously chosen  $k$  best subsets wherein  $p$  is an integer that values from 1 to  $m-1$ . This progression of feature selection may interruption the entire system to pick the relevant features. Therefore, Cuckoo search algorithm is utilized to pick the optimal feature. The feature selected by the cuckoo is carried out for next generation. By using this, the features which characterize the best  $k$  subsets have more probability to present in the features of the next iteration. However it will achievable for all fish to consider other features. For a known fish  $j$ , those features are the ones which accomplish the best cooperation between the previous understanding such as visual position and the current best of cuckoo search.

The feature selection exploited by artificial fish involves the following:

- $n$  features which consist of the original set,  $W = \{f_1, \dots, f_n\}$
- $nf$ , the amount of artificial fishes to search via the feature space

- $V_i$ , the visual position related with the feature  $f_i$
- For every fish  $j$ , a list which encloses the certain feature subset,  $S_j = \{s_1, \dots, s_m\}$

*Algorithm 1:*

//Removal of irrelevant features

//Initialization

1. Set  $v_i = c$ , where  $c$  is a constant and  $i = 1, \dots, n$
2. Define maximum number of iterations
3. Define  $k$ , where  $k$ -best subsets can influence the subsets of next iteration
4. Define number of features to be select ( $m$ )
5. Define number of artificial fishes ( $nf$ )
6. Define  $p$ , where  $m-p$  is the number of features

//each artificial fish can select initial feature subset

7. For  $j = 1$  to  $nf$ , arbitrarily consign the subset of  $m$  features to  $S_j$
8. Evaluate the fitness by means of SVM classification of the features of  $S_j$
9. Sort the subsets according to fitness and update the maximum fitness and store the corresponding subset of features.
10. Using the feature subsets of the best  $k$  fish: for  $j = 1$  to  $k$ ,

$$\Delta V_i = \left\{ \frac{\max_{g-1:k}(Acc_g) - Acc_j}{\max_{h-1:k}(\max_{g-1:k}(Acc_g)) - Acc_j} \right\} \text{ if } f_i \in S_j$$

$$V_i = \rho \cdot V_i + \Delta V_i, \text{ where } \rho \text{ is a constant}$$

//Select the remaining  $p$  features for each fish

11. For  $j = 1$  to  $nf$ , choose feature  $f_i$  from given subset  $S_j$  by cuckoo search algorithm such as follows:
12. Initialize a population of  $n$  host nests and desire the current best nest through evaluating the fitness,

$$\sigma_u = \left\{ \frac{\Gamma(1+\beta) \sin(\pi\beta/2)}{\Gamma[1+\beta/2] \Gamma[\beta/2] \beta_2^{(\beta-1)/2}} \right\}^{1/\beta}, \sigma_v = 1, \beta = 1.5$$

13. While ( $t < \text{MaxGeneration}$ )
14. Search out a cuckoo arbitrarily by Levy flights
 
$$x_i^{(t+1)} = x_i^{(t)} + \alpha t^{-\lambda}$$
15. Evaluate its fitness,  $a_j$  and pick a nest among  $n$  arbitrarily.
16. If ( $a_i > a_j$ ), change  $j$  by the new solution and abandon the worst nests.
17. Remain best solutions and order the solutions.
18. Discover the current best feature  $f_j$ .
19.  $S_j = S_j \cup \{f_j\}$
20. Replace the duplicated subset with randomly chosen subsets and go to step 11.

### B. Removal of Redundant Features

After obtaining the relevant features, the redundant features are also appeared in the relevant features. Therefore, the NMF approaches are used to cluster the features. The redundant features are eliminated from the relevant features by selecting representative feature in the clusters and the final feature subset is acquired.

NMF defines a matrix factorization method which decides the positive factorization for a particular positive matrix. The collected feature set represented as  $W = \{f_1, \dots, f_n\}$  and  $X$  is the weighted frequency vector. Assume that the given feature set includes  $k$  clusters. The major aim is to factorize the  $m \times n$  matrix  $X$  into the two non-negative matrices such as  $m \times k$  matrix  $A$  and  $k \times n$  matrix  $B^T$  that reduces the objective function as follows:

$$F = \frac{1}{2} \|X - AB^T\| \tag{3}$$

In above equation (3),  $\|\cdot\|$  refers the squared sum of all elements in the matrix. This objective function also rewritten as:

$$F = \frac{1}{2} \{\text{tr}(X - AB^T)(X - AB^T)^T\} \tag{4}$$

$$= \frac{1}{2} \{\text{tr}(XX^T) - 2\text{tr}(XBA^T) + \text{tr}(AB^TBA^T)\} \tag{5}$$

Let  $A = [a_{ij}]$ ,  $B = [b_{ij}]$ , and  $A = [A_1, A_2, \dots, A_k]$ . The above minimization problem can be restated as follows: reduce  $F$  with respect to  $A$  and  $B$  under the restraints of  $a_{ij} \geq 0$ ,  $b_{xy} \geq 0$ , where  $0 \leq i \leq m$ ,  $0 \leq j \leq k$ ,  $0 \leq x \leq n$ , and  $0 \leq y \leq k$ . This is the characteristic constrained optimization problem, and it can be explained by using the Lagrange multiplier process. Let  $\alpha_{ij}$  and  $\beta_{ij}$  be the Lagrange multiplier for restraint  $a_{ij} \geq 0$ ,  $b_{ij} \geq 0$  respectively and  $\alpha = [\alpha_{ij}]$ ,  $\beta = [\beta_{ij}]$ , the Lagrange  $L$  is described as:

$$L = F + \text{tr}(\alpha A^T) + \text{tr}(\beta B^T) \tag{6}$$

The derivatives of L in respect of A and B are given as:

$$\frac{\partial L}{\partial A} = -XB + AB^T B + \alpha \quad (7)$$

$$\frac{\partial L}{\partial B} = -X^T A + BA^T A + \beta \quad (8)$$

By using the Kuhn-Tucker condition  $\alpha_{ij}a_{ij} = 0$  and  $\beta_{ij}b_{ij} = 0$ , obtain the following equations for  $a_{ij}$  and  $b_{ij}$  correspondingly:

$$(XB)_{ij}a_{ij} - (AB^T B)_{ij}a_{ij} = 0 \quad (9)$$

$$(X^T A)_{ij}b_{ij} - (BA^T A)_{ij}b_{ij} = 0 \quad (10)$$

The above equations help to the following updating formulae:

$$a_{ij} \leftarrow a_{ij} \frac{(XB)_{ij}}{(AB^T B)_{ij}} \quad (11)$$

$$b_{ij} \leftarrow b_{ij} \frac{(X^T A)_{ij}}{(BA^T A)_{ij}} \quad (12)$$

To create the unique solution, the Euclidean length of the column vector in matrix A is required as one. This requirement of normalizing A can be accomplished by:

$$b_{ij} \leftarrow b_{ij} \sqrt{\sum_i a_{ij}^2} \quad (13)$$

$$a_{ij} \leftarrow \frac{a_{ij}}{\sqrt{\sum_i a_{ij}^2}} \quad (14)$$

Redundant features are composed in each cluster are correlated with every other features. The correlation coefficients are ranked and the top k coefficient can be selected as an exemplary feature and other redundant features can be eliminating from the cluster. These accomplished features are the final feature set which is called as effective feature in the data.

*Algorithm 2:*

//removal of redundant features

1. Factorize X into the non-negative  $m \times k$  matrix A and the non-negative  $k \times n$  matrix  $B^T$  that diminish the objective function Equation 3,
2. Kuhn-Tucker condition  $\alpha_{ij}a_{ij} = 0$  and  $\beta_{ij}b_{ij} = 0$ ,
3. Find the non-negative matrices A & B by using Equation 11 & 12 correspondingly,
4. Normalize the non-negative matrices A & B by using Equation 13 & 14 respectively,
5. Obtain the matrix  $B^T$  by taking transpose of the matrix B i.e.,  $B^T_{k \times n} = B_{n \times k}$
6. Form the clusters, compute correlation between each features in the cluster
7. Rank the correlation coefficients and select the top k feature as an exemplary feature from each cluster and include in  $S_j$
8. Return  $S_j$

From the above algorithm, initially the irrelevant features are eliminated from the database. For this purpose, the hybrid of AFS and CS optimization technique is used. Here, the artificial fish are move to the next state which is better than the current state in terms of measuring crowd factor, step size and Levy distribution function. Levy distribution function helps to optimize the better next state which is selected as the relevant state. The variables in the relevant state are represented as relevant features in the data. From the first part of our algorithm, the relevant features are obtained. In the second part algorithm, the redundant features in the relevant feature set are removed. For this function, the effective non-negative matrix factorization based clustering method is used. The acquired relevant features are allowed to the NMF based clustering to remove the redundant features. The main goal is to factorize the input X into non-negative  $m \times k$  matrix A and  $k \times n$  matrix  $B^T$ . After these two non-negative matrices are normalized and the matrix of A is transposed for obtaining  $B^T$  and achieved the clustering of features. By using clusters, the exemplary feature is collected and it is known as final selected features without irrelevant features as well as redundant features. Finally, the desired features are obtained which can increase the classification performance.

#### IV. EXPERIMENTAL RESULT

In this section, the classification performance with proposed NMFC classification and without feature selection method results are compared in terms of classification accuracy, True Positive rate (TP rate), and True Negative rate (TN rate). From this experimental result, we can say that the proposed method has high efficiency than the other techniques.

### A. Database Description

For our experimentation, we are considering the student’s database which contains 297 data instance that are collected from the different colleges. In our database 40 attributes are available which includes student’s name, course, age, gender, and nature of college such as engineering/medical/arts, college type such as government/self I financed, location feature, family type such as joint family/nuclear family, family factors such as occupation & educational qualification of relatives, economic aspects, college aspects, social aspects and spending time in television, mobile, computer and etc., personal factors, academic factors. Here, location features are referred as the area in which student’s home, school and college located such as rural, urban, semi-urban region. College aspects are one of the traits that gives the data about whether the student refer lecturer notes which is given by lecturers or books, training system such as lecturer method or black board, number of students in class, whether the college permitted mobile phones or not, and so on. Social aspects are such that direction of relatives for studies, number of friends and educational performance of friends.

In our experimentation, we are evaluating the student’s performance such that good or poor in the institute in accordance with the features presented in the information. Data samples are given to the feature selection method to collect the effective features. Then these collected features are given to classifiers to evaluate the performance. For our experimentation, we are using two classifiers such as Prism and J48. We are providing 150 data samples as training data (with class label) to the classifier for learning process and remaining data samples are taken as test data (without class label) which are given to the classifier for finding the class label. Eventually, the output variable or attribute or class is to be chosen which gives the educational status or student’s performance that has two consequences either PASS (student who pass the course) or FAIL (student who needs to repeat the course).

### B. True Positive rate

True Positive rate (TP rate) also called as sensitivity or recall, is the proportion of true positives which are identified as positive and is computed by Equation 13. Based on this experiment, if the result class label from the prediction is PASS and the true class label is also PASS, it is called as TP rate.

$$TP = \frac{\text{True Positive}}{\text{True positive} + \text{False Negative}} \quad (13)$$

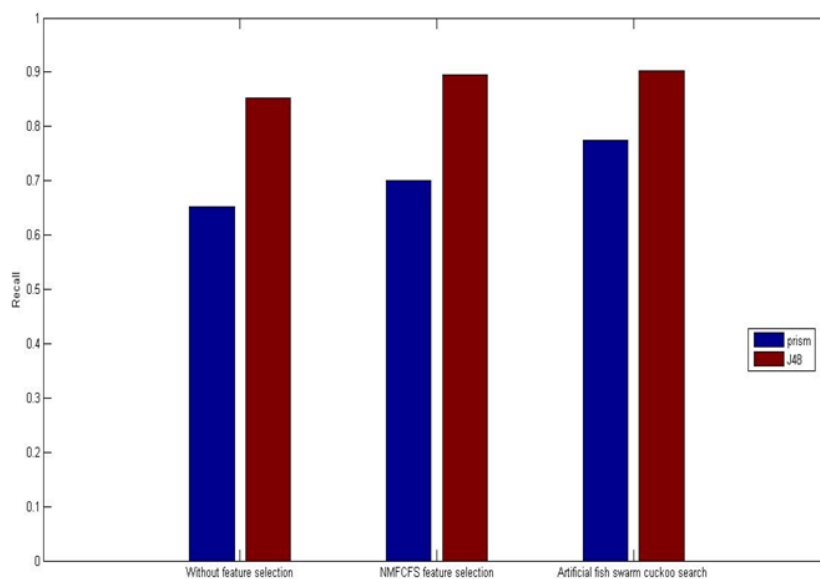


Fig. 1. TP rate comparison

Fig. 1 demonstrates that comparison of the TP parameter between the classification without feature selection and the classification with NMFC which uses the existing method symmetric uncertainty as well as our proposed method artificial fish swarm-cuckoo search optimization approach. In the graph, X-axis denotes classification methods such as Prism, J48 and Y-axis denotes TP rate. From this graph, we conclude that the classification with out proposed NMFC uses artificial fish swarm-cuckoo search optimization has more efficient in TP performance relatively.

### C. True Negative rate

True Negative rate (TN rate) also called as specificity, is proportion of true negatives which are identified as negative and is computed by Equation 14. Based on this experiment, if the result class label from the prediction is FAIL and the true class label is also FAIL, it is called as TN rate.

$$TP = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (14)$$

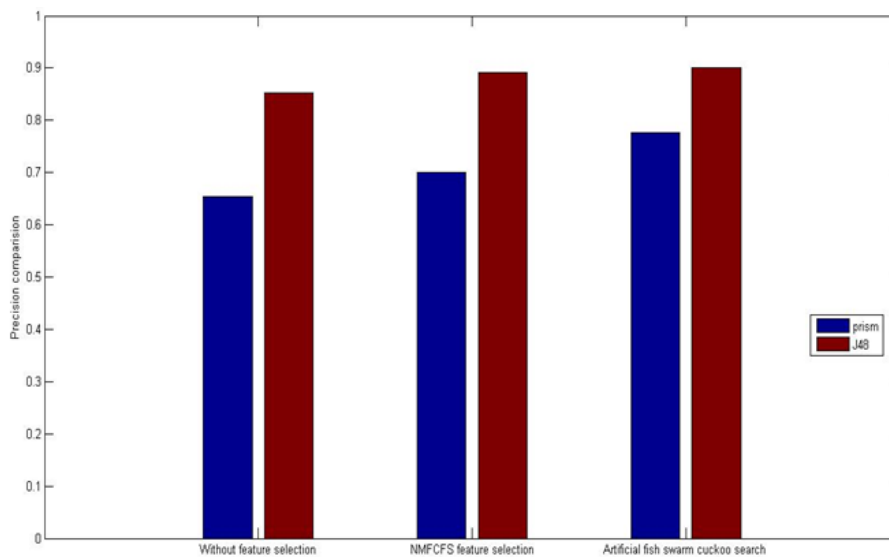


Fig. 2. TN rate comparison

Fig. 2 demonstrates that comparison of the TN parameter between the classification without feature selection and the classification with NMFCS which uses the existing method symmetric uncertainty as well as our proposed method artificial fish swarm-cuckoo search optimization approach. TN rate is calculated by using the formula. In the graph, X-axis denotes classification methods such as Prism, J48 and Y-axis denotes TN rate. From this comparison graph, we conclude that the classification with out proposed NMFCS uses artificial fish swarm-cuckoo search optimization has more efficient in TN performance relatively.

*D. Accuracy rate*

Accuracy or classification accuracy is referred as the ability of the classifier and is computed by Equation 15. In this equation, TP rate and TN rate are given in the above sections 4.2 and 4.3 respectively. False Positive rate (FP rate) is the proportion of true negatives which are identified as positive. Based on this experiment, if the result class label from prediction is PASS and the true class label is FAIL, it is called as FP rate. Likewise, False Negative rate (FN rate) is the proportion of true positives which are identified as negative. Based on this experiment, if the result class label from prediction is FAIL and the true class label is PASS, it is called as FN rate.

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (15)$$

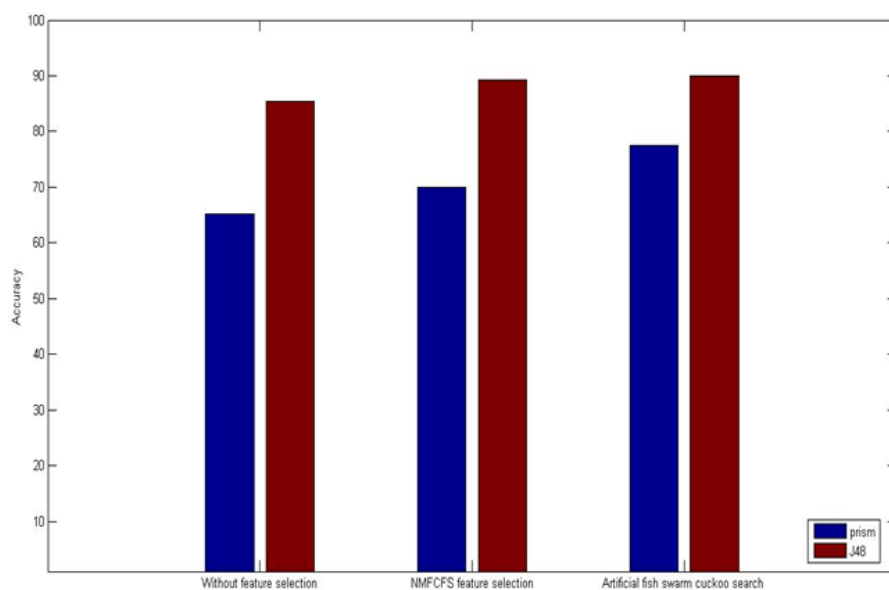


Fig. 3. Accuracy comparison

Fig. 3 shows that comparison of the accuracy parameter between the classification without feature selection and the classification with NMFC which uses the existing method symmetric uncertainty as well as our proposed method artificial fish swarm-cuckoo search optimization approach. Accuracy of classification is measured by using the formula. In the graph, X-axis denotes classification methods such as Prism, J48 and Y-axis denotes accuracy rate. From this accuracy comparison graph, we conclude that the classification with our proposed NMFC uses artificial fish swarm-cuckoo search optimization has more efficient in accuracy performance relatively.

## V. CONCLUSION

Various strategies can be developed and implemented which enables the educational institutions for transforming the wealth of information into the wealth of predictability, stability and profits. Our proposed system is establishing a new feature selection method to classify the performance of the students in the educational institutions. This method has various advantages such as removal of irrelevant features and also removal of redundant features which are presented in the irrelevant features. This will increase the result of classification accuracy. In our experimentation, we are using the classification technique such as Prism and J48 to classify the student's performance of different colleges. Experimental result shows that accuracy level, TP rate and TN rate of our proposed method has higher accuracy than the classifier without feature selection process. In addition, this new novel technique can reduce the time complexity and memory complexity of the classification process. Thus, we can improve the performance of the student's failure and dropout prediction by using our proposed technique.

## REFERENCES

- [1] Gautam, R., & Pahuja, D. A Review on Educational Data Mining.
- [2] Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. arXiv preprint arXiv:1201.3417.
- [3] Delibašić, B., Vukićević, M., Jovanović, M., & Suknović, M. (2013). White-Box or Black-Box Decision Tree Algorithms: Which to Use in Education?. *IEEE Transactions on Education*, 56(3), 287-291.
- [4] Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, 13(1), 61-72.
- [5] Nikam, S. S. (2015). A Comparative Study of Classification Techniques in Data Mining Algorithms. *Oriental Journal of Computer Science & Technology*, 8(1), 13-19.
- [6] El Gamal, A. F. (2013). An Educational Data Mining Model for Predicting Student Performance in Programming Course. *International Journal of Computer Applications*, 70(17).
- [7] Singh, B., Kushwaha, N., & Vyas, O. P. (2014). A Feature Subset Selection Technique for High Dimensional Data Using Symmetric Uncertainty. *Journal of Data Analysis and Information Processing*, 2(04), 95.
- [8] Singh, M., Singh, J., & Rawal, A. (2014, December). Feature Extraction Model to Identify At-Risk Level of Students in Academia. In *Information Technology (ICIT), 2014 International Conference on* (pp. 221-227). IEEE.
- [9] Mustafa, M. N., Chowdhury, L., & Kamal, M. S. (2012, May). Students dropout prediction for intelligent system from tertiary level in developing country. In *Informatics, Electronics & Vision (ICIEV), 2012 International Conference on* (pp. 113-118). IEEE.
- [10] Rachburee, N., & Punlumjeak, W. (2015, October). A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining. In *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 420-424). IEEE.
- [11] Kaur, P., Singh, M., & Josan, G. S. (2015). Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. *Procedia Computer Science*, 57, 500-508.
- [12] Shahiri, A. M., & Husain, W. (2015). A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, 414-422.
- [13] Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39(10), 9468-9476.
- [14] Emary, E., Zawbaa, H. M., & Hassanien, A. E. (2016). Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172, 371-381.