A Prediction-based Curriculum Analysis using the Modified Artificial Bee Colony Algorithm

Reir Erlinda E. Cutad[1], Bobby D. Gerardo[2]

Graduate Programs, Technological Institute of the Philippines, Quezon City, Philippines[1]
West Visayas State University, Iloilo City, Philippines[2]

Abstract—**Due to the vast amount of students' information and the need of quick retrieval, establishing databases is one of the top lists of the IT infrastructure in learning institutions. However, most of these institutions do not utilize them for knowledge discovery which can aid in informed decision-making, investigation of teaching and learning outcomes, and development of prediction models in particular. Prediction models have been utilized in almost all areas and improving the accuracy of the model is sought- after this study. Thus, the study presents a Scoutless Rule-driven binary Artificial Bee Colony (SRABC) as a searching strategy to enhance the accuracy of the prediction model for curriculum analysis. Experimental verification revealed that SRABC paired with K-Nearest Neighbor (KNN) increases the prediction accuracy from 94.14% to 97.59% than paired with Support Vector Machine (SVM) and Logistic Regression (LR). SRABC is efficient in selecting 14 out of 60 variables through majority voting scheme using the data of the BSIT students of Davao Del Norte State College (DNSC), Davao del Norte, Philippines.**

Keywords—Binary artificial bee colony; rule-driven mechanism; prediction model; curriculum analysis

I. INTRODUCTION

One of the important factors for nation's sustainable economy which aimed in the improvement of proficiency of general public is education [1]. Knowledge discovery from learning institution's databases through different techniques in data mining can aid in informed decision-making [2],[3] and when paired with advances of machine learning algorithms, it can build prediction models with enhanced accuracy[4],[5]. However, the application of these techniques in learning institutions still need more interest [6].

The goal of higher learning institutions is to have higher percentage of on-time graduates particularly for colleges and universities that are government-funded because when students enrol for another semesters, this will put in more expenditure in the government budget. Hence, to attain such goal, curriculum analysis must be join in the process of addressing the problems. In curriculum analysis, courses that students have difficulty to pass can be determine, can aid in the inquiry if graduates' skills are relevant to job market at least at the minimum and a tool for verification if observance to standards from the governing bodies is followed.

According to studies of [7]-[9], the most influential factors in prediction model in the field of education is academic data. Also in [10], it is stipulated that academic data, the course performance in particular, is utilized in constructing model for curriculum analysis; thus, this study.

This study will discover the most relevant courses in relation to on-time graduation. The output will empower the learning institution academic heads to determine the most common skills among the on-time graduates which can be utilize to verify if it matches the job skills requirements of the nation and locality in particular.

Discovering most relevant courses can be characterized as a variable selection problem and is solved through optimization. Variable selection is a technique of finding the highly most important subset of variables having the highest accuracy and minimum error [11], [12]. However, traditional variable selection methods suffered from cost of computational time if implemented in high dimension problems. This is where swarm-based intelligences came in as one of the solutions [13].

Swarm-based algorithms, which form a sub-field of artificial intelligence, has gained particular importance because of workability and simplicity in implementation [14], [15]. Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC) and Ant Colony Optimization (ACO) were some of the state-of-the-art algorithms [16], [17]. Among these, ABC has a simple structure, few parameters for tuning [18], [19], adaptable and robust [20] which become its advantage compared with other swarm-based algorithms.

Nevertheless, as stated in the literature, same with other algorithms, ABC has a widely-known performance problem which is poor exploitation ability [21]-[23]. Hence, modification is necessary before its application to actual case scenario, specifically if the variables dimension is high like in this case. Also, based on readings, there is no recent verified performance of ABC and its variants which was utilized in academic data.

To realize the goal of this study, a novel Scoutless Rule-driven binary Artificial Bee Colony (SRABC) is presented as tool to enhance the accuracy of prediction model through selecting the most relevant variables. This is done through the use of simple reverse mechanism, replacement of operators from arithmetic to mathematical set, fourths-uniform crossover, and on-mutation strategies. Being a scoutless ABC is inspired from the study of [24]. Before the implementation of SRABC to actual case scenario, its performance was verified and compared to recent ABC's in the literature such as Novel binary ABC algorithm (NBABC) [25] and hybrid

ABC and gradient boosting decision tree algorithm (ABCoDT) [26] using some of their utilized datasets from UCI, and the results used for comparison was obtained from their respective studies.

The dataset utilized in the study has 60 variables representing the student's course grades from School Years 2015- 2016 to 2018-2019 of the BSIT curriculum, College entrance exam and High School's general average or general average from previous school attended for transferees. These information were extracted from the Enrolment Information System of Davao Del Norte State College (DNSC), Davao del Norte, Philippines. The framework of the study is made up of SRABC paired with three machine learning algorithms such as KNN, SVM and LR to discover the best model among them. Further, majority voting scheme is applied in the final selection process for the most relevant variables to address non-deterministic behavior of the algorithm.

II. LITERATURE REVIEW

A. Prediction-Based Models Applied in Learning Institutions

In this section, the most recent and varied studies based on prediction which employed in learning institutions are presented.

In the study of [27], a hybrid GA-KNN was employed to build a prediction model using records of student responses to reduce the variables in the evaluation instrument of faculty instructional performance in State Colleges and Universities in Caraga Region, Philippines. The study achieved an accuracy of 95.53% and 66.66% data reduction. The study of [28] was implemented to find the most relevant activities for physics course at the Universidad Loyola Andaluc´ua, in Seville, Spain, it utilized different classification algorithms such as Naive Bayes, C5.0, SVM and Fast Correlation Based Filter (FCBF), Scatter Search (SS) and Genetic Algorithm (GA) as searching strategy for variable selection. Two datasets containing evaluation based on daily work (D-5) and evaluation focused on knowledge (D-7) were created. Based on experimental results, in D-5, the SS obtained the lowest number of subset with an average accuracy of 89.43% for all classifier algorithms, while in D-7, FCBF obtained the lowest subset with an average of 76.43%. An adaptive support vector machine framework is called as RF-CSCA-SVM was implemented to predict college students' entrepreneurial intention and was built through a combination of random forest, support vector machine, and sine cosine algorithm (SCA), a new swarm- based intelligence algorithm. SCA is modified by introducing a chaotic local search. The performance of the modified SCA (CSCA) has been verified through comparison to other swarm-based algorithms and the results showed that it outperformed the others by having an average prediction accuracy of 83.5% [29].

B. Original ABC Algorithm

The Artificial Bee Colony (ABC) algorithm, proposed by Karaboga and Basturk [30], is an optimization algorithm that reached the global optima in an iterative manner. The original algorithm's approach is composed of initialization phase, repeat, employed bees phase, onlooker bees phase, scout bees phase, memorize the best solution achieved so far, until the maximum cycle number[31]. The employed bees have in their mind the position of a particular food source which can be a possible solution, then, the onlooker bees select the food source with high quality through the dance of the employed bees. The scout bees were employed bees that exhausted a particular food source, it will generate new food source in a random manner once activated[32]. At the initialization phase the food source is generated using the equation.

$$x_{ij} = x_j^{min} + rand(0,1) (x_j^{max} - x_j^{min}) \qquad (1)$$

Where $i=1,...N$, the size of the food source, $j=1,...D$, the dimension of the food source, x_{ij} is the jth dimension of the ith food source, and x_j^{min} and x_j^{max} are the lower and upper bound of jth dimension.

At the employed bees phase, after sending them to the assigned solution in her memory, they will find new candidate solution using the equation.

$$v_{ij} = x_{ij} + rand(-1,1)(x_{ij} - x_{kj}) \qquad (2)$$

After a candidate and a current solution is evaluated and compared, a greedy selection is applied, the better solution will be place in the memory, and reset the abandonment counter, otherwise, is incremented by 1. Evaluation of the quality of the food source is through their probability (P_i) and fitness (fit_i) values using the equations (3) and (4).

$$fit_i = \begin{cases} |f_i|, & fi < 0 \\ 1/f_i, & fi \geq 0 \end{cases} \qquad (3)$$

$$P_i = \frac{fit_i}{\sum_{n=1}^{N} fit_n} \qquad (4)$$

At the onlooker bees phase, to exploit the found better solution, sources with higher probability have chances to be selected using a selection method such as the roulette wheel and undergoes same procedure as the employed bees. Lastly, scout bees phase will be activated once the abandonment counter is higher than the predetermined limit value and utilized again the equation (1) to generate new food source[33].

C. Binary Artificial Bee Colony

To implement ABC for binary optimization, equations (1) and (2) should be modified. The equation (1), which represents the food source is converted into binary vector through the Bernoulli process, whereas, arithmetic operators in equation (2), was modified into logical, bitwise or genetic operators.

1) Multiple strategies: MDisABC, a modified binary ABC, introduced an effective mechanism of information sharing, instead of selecting one neighborhood in the generation of a new solution, a selection mechanism to create three neighborhood mutant solution was introduced and embed a recombination operator for selection mechanism among the solutions [34].

2) Logical and bitwise operators: BitABC replaced the arithmetic operator of equation (2) into combined bitwise operations of XOR, AND, and OR [35], while binABC[36] utilized the XOR operation in the generation of solution.

3) Mutation and crossover: GB-ABC, a binary ABC which utilized genetic operators such as two-point crossover to the current and neighborhood food source to produce a child food source and swap operator for grandchildren food source. This is also utilized to select the best food source[37]. Recently, PrBABC, a probabilistic binary artificial bee colony introduced a strategy which composed of partial-mapped crossover operator and bitwise AND operator. The random generation of two points as the start and endpoint was implemented before the two operators were applied, and after each iteration, a probability equation was used to determine if which operator will be selected [38].

III. SCOUTLESS RULE-DRIVEN BINARY ARTIFICIAL BEE COLONY

To address the exploitation ability problem of artificial bee colony, new strategies are applied in the exploration and exploitation equations of the algorithm. These strategies will improve the exploitation ability of the ABC and also addresses the computational cost specifically for high dimension problems. In this study, the crossover and mutation are modified along with the introduction of reverse mechanism in the initialization phase and the mathematical set operators as replacement to bitwise operators. Fig. 1 shows the diagram of the modified ABC algorithm.

A. Initialization Phase

1) Migration: To generate food sources, SRABC utilize the equation (5). This equation is not new because most of the algorithm for binary optimization used the concept with distinction on threshold value. Each food sources is a fixed-size array depending upon the number of variables in the dataset and each arrays is composed of 1's and 0's since the algorithm is implemented for variable selection problem.

$$x_{iD} = \begin{cases} 1, & n \geq 0.5 \\ 0, & n < 0.5 \end{cases} \tag{5}$$

The migration occurs when the food sources were reversed in which the nth position in the array will become the new 1st position. Algorithm 1 and Fig. 2 display the concept.
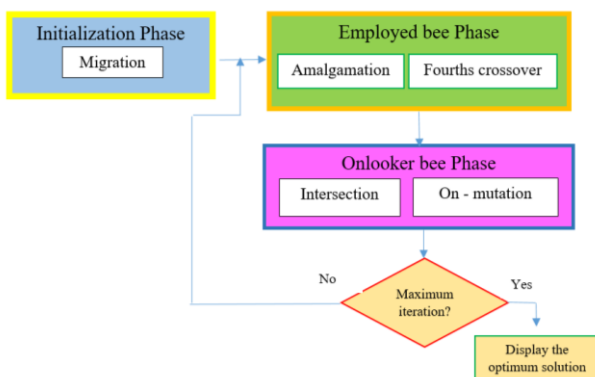

Fig. 1.   Diagram of SRABC.

Algorithm 1: Migration
1. Utilize the equation (5) to check the generated value and if the value is greater than the threshold, append 1, otherwise, append 0, repeat until the fixed-size array is produced.
2. Reverse each array as shown in figure 2.
3. Execute same process for all food sources.

B. Employed Bees Phase

Multiple strategies were implemented in this phase which aimed to reinforce the strong exploration ability of ABC. The random and greedy selections were still observed in the process of finding the neighborhood and better food source.

Algorithm 2: Employed bees phase
1. For each food source, a new neighborhood food source is selected through a random selection procedure;
2. For each current and selected neighborhood food sources;
   2.1 create new solution through the application of amalgamation which takes in all the position with value of 1's.
   2.2 produce another two new solutions through the application of fourths-crossover, if one of the solutions has an empty array consider the current food source;
3. Apply greedy selection to the three created solutions;
4. The found better solution will be the new current food source.

1) Amalgamation: Instead of implementing the widely utilized OR bitwise operator in binary optimization, the strategy implements a union mathematical set operator to produce a new solution. The strategy has the same concept with OR bitwise, but computationally not costly because it will not compare each position in the array through iterative manner.

2) Fourths-crossover: To produce two new solutions, the two food sources are split into 25% for the first-fourths and 75% for the three-fourths. The first solution is composed of the first-fourths of the current food source and the last three-fourths of the neighborhood food source. While, the other solution is composed of the last three-fourths of the current food source and the first-fourths of the neighborhood food source. The diagram is shown in Fig. 3.
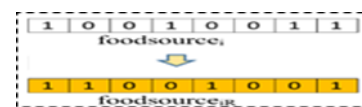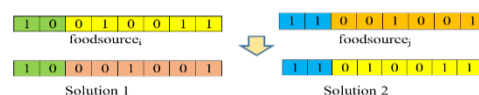

Fig. 2.   Reverse Method.

Fig. 3.   Fourths Crossover Method.

C. Onlooker Bees Phase

An intensive multiple exploitation strategies are implemented in this phase to improve the found better solution. The concept of higher fitness will have a higher probability to be chosen are still observed in this algorithm.

1) Intersection: A replacement of the widely utilized AND bitwise operator for binary optimization. The strategy implements the intersection mathematical set operator between the current and neighborhood food sources. The new solution is produced through selection of positions with the same value of 1, if the same position has different values or same value but is 0, a zero value is append in that particular position. Fig. 4 illustrates the concept.

2) On-mutation: The on-mutation strategy will change the value from 0 to 1 of the position whose location is the same in the generated random number. However, if the position's value is 1, no changes will be made. The number of times that random number will be generated is equivalent to the 20% of the dataset's variables. This strategy is both implemented in the current and neighborhood food sources to exploit more the found better food sources.
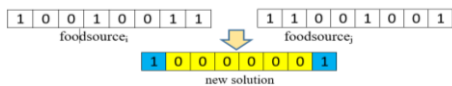

Fig. 4. Intersection Method.

IV. EXPERIMENTS

A. Benchmark Dataset and Settings

The effectiveness of SRABC is verified through the comparison of prediction accuracy and number of selected variables using three datasets from the UCI machine learning repository [39] as shown in Table I.

Each datasets is split into 70% for training and 30% for testing sets respectively. To predict the accuracy of the selected variables, 10-folds cross-validation and KNN classifier having K=5 were used, these parameters were same with those algorithms that SRABC was compared.

The equation (6) is the fitness function formula for multi-objective variable selection problem where α is set to 90% to give importance to the accuracy. SF defines the number of selected variables for each solution and TF is the total number of variables in the dataset.

$$\text{Fitness} = \alpha * \text{accuracy} + (1 - \alpha)\left(1 - \frac{SF}{TF}\right) \tag{6}$$

The program is implemented in Python 3.6.4 and experiments are conducted on a computer with Intel Core i5-6200U CPU at 2.30 GHz and 4G of installed memory using 64 bit Windows 10 operating system.

TABLE. I. PROPERTIES OF THE DATASETS

Dataset	Instances	Features	Classes
Wine	178	13	3
Ionosphere	351	34	2
Arcene	900	10,000	2

B. Comparative Study

NBABC and ABCoDT are recent state-of-art ABC which has experiments conducted for variable selection problem using publicly available datasets. Table II is the description of the performance of the algorithms.

Table III illustrates the comparative results of the experiments, the == in the said table signifies that the experiment for such dataset is not conducted in their study. Based on the experiments conducted, SRABC exhibit a superior performance in terms of prediction accuracy and number of selected variables, though for wine, ABCoDT performance is better in terms of number of selected variables but the accuracy was deteriorating, while Table IV shows the efficiency of SRABC in selecting the minimum number of variables without compromising the prediction accuracy, particularly in the very high dimension dataset.

TABLE. II. DESCRIPTION OF THE STATE-OF-THE-ART BINARY ABC VARIANTS

Algorithm	Description
NBABC	The algorithm was compared to other five binary ABC variants and tested in eight datasets with 166 features as the highest. Based on experiments conducted, the algorithm is superior to the other five variants in six out of eight datasets in which wine is one among the two datasets that the algorithm is not superior.
ABCoDT	The algorithm was tested in eight datasets with 10,000 features as highest. It was compared to six state-of-the-art methods and results showed that it was effective in reducing the number of features. However, there were three datasets that the algorithm is not superior and among them is the arcene dataset.

TABLE. III. COMPARATIVE RESULTS OF ACCURACY AND NUMBER OF SELECTED VARIABLES

Dataset	SRABC	NBABC	ABCoDT
Wine Testing accuracy Selected variables	95.00 3.00	66.80 7.20	90.91 1.00
Ionosphere Testing accuracy Selected variables	93.00 3.00	84.20 3.55	==
Arcene Testing accuracy Selected variables	85.00 353	==	80.20 2,103

TABLE. IV. THE REDUCTION RATE OF SRABC FROM THE ORIGINAL VARIABLES

Dataset	Original variables	# of selected variables	Reduction rate
Wine	13	3.0	76.9 %
Ionosphere	34	3.0	91.17%
Arcene	10,000	353	96.47%

Aside from the better performance of SRABC in terms of its ability to reach a very good accuracy and minimal number of selected variables which is called the global optima, it also exhibit a good balance between exploration and exploitation to avoid trapping in the local optima. It is shown in Fig. 5 that during fitness evaluations the algorithm selected varied number of variables which supports the strong exploration ability of ABCs and make use of the found better solution for exploitation towards convergence. While, Fig. 6 shows that convergence rate is enhanced when reaching the global optima thru a smaller number of iterations.

C. Curriculum Analysis-the Case Study

The conceptual framework of the study is shown in Fig. 7. The process includes extraction of data from the College Information System database and students' manual archive. The data were Pre-processed before the application of SRABC with majority voting scheme to find the most relevant variables with increase prediction accuracy. SRABC is paired with three machine learning algorithms to attain a best model.
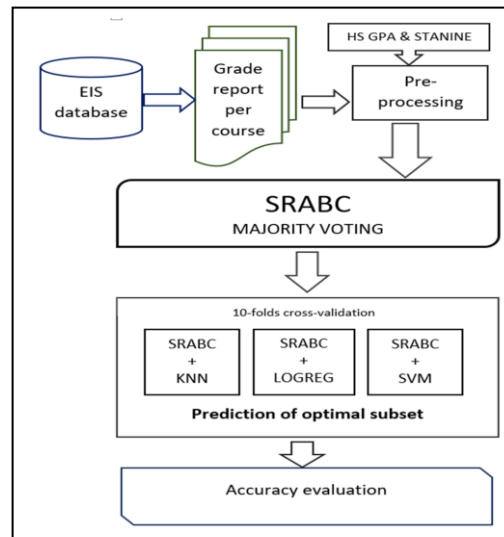

Fig. 5.   Illustration of Exploration-Exploitation Ability.
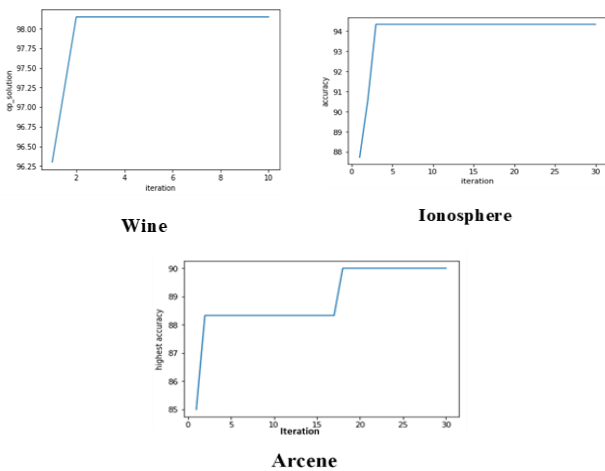
Fig. 6.   Highest Prediction Accuracy Per Iteration.

Fig. 7.   Diagram of the Prediction-based Model for Curriculum Analysis.

1) Data collection: The study utilized the high school general average or general average from previous school attended for transferees, stanine and course grades from School Year 2015- 2016 to School Year 2018-2019 of the Information Technology students of Davao Del Norte State College (DNSC). The Grade reports per semester were extracted from the Enrolment Information System while others were manually taken from the students' archive. A total of 343 student records were obtained including those who dropped or transferred to other learning institutions.

2) Data preparation: The collected grade reports per semester and other data were converted into a dataset having column name of their corresponding code number as shown in Fig. 8. Following such, is the transformation of values for suitability in building the model. The class variable with a value of 1 means that the students graduated on-time and 0 for not on-time. The transformed data were processed to remove records of students who stop enrolling in the BSIT program which resulted into a dataset of 216 records, comprised of students who graduated on-time, graduated during summer and soon to graduate within the School year 2019-2020.
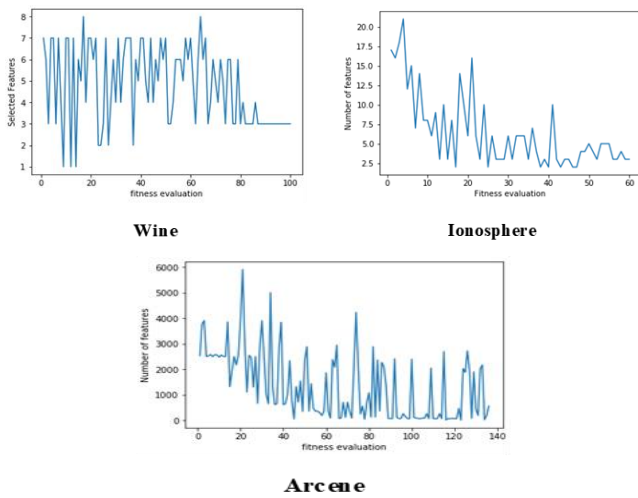

Fig. 8.   Courses, GPA and Stanine with their Corresponding Code Number.

3) Data analysis approach: In the literature, many modifications for swarm-based algorithms have integrated schemes of directions during the search process because this algorithms are stochastic in nature. But still, they behave with different outputs in different runs, called non-deterministic behavior, specifically when implemented in very high dimensions problems. To address such, a majority voting is applied to the outputs of SRABC to develop a model that behaves in a deterministic manner in the process of finding the optimal subset of variables. The optimal subset is composed of variables getting the most number of votes in relation to the average number of variables from the 30 runs. This technique is integrated to ensure that experimental results is precise in all conducted experiments. Also, the said technique fosters stability of the model, as stated in [40] that stability in variable selection technique is similarly significant to the accuracy of the prediction model. The study of [41] also a utilized majority voting scheme.

Experiments were conducted for different types of machine learning algorithms such as KNN, SVM and LR using 10-folds cross-validation to come up a model having best prediction accuracy and lowest root mean squared error. Some statistical tools such as Precision, Recall and F-measure were also utilized to evaluate the model.

4) Majority voting results: Each expirements conducted is composed of 30 runs; in all runs, Capstone Project is always present as expected, because it is the thesis course of the curriculum. The rest of the courses that comprised the optimal subset were IT3, IT7, IT8, IT18, ITE ELECT4, PE1, NSTP2, HUM2, ENG1, PHY1, HUM1, ENG2, and SS3. The new subset has a total of 14 courses and having a data reduction rate of 76.6%. The tally of scores is shown in Fig. 9.

5) Prediction evaluation: When using the KNN algorithm, tuning of the algorithm is conducted through iteration of k values from 1 up to 10 to find the value of k that has the best prediction accuracy as shown in Fig. 10, based on the result, K=1 has the highest prediction accuracy.

Also, an experiment to verify the prediction accuracy of the machine learning algorithms using all variables was conducted to be used in the comparison of accuracy. Table V shows the prediction accuracy using the three machine learning algorithms.
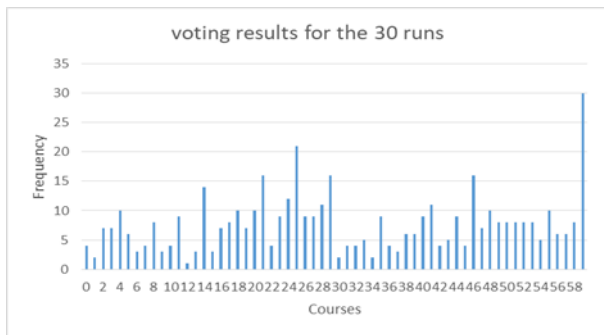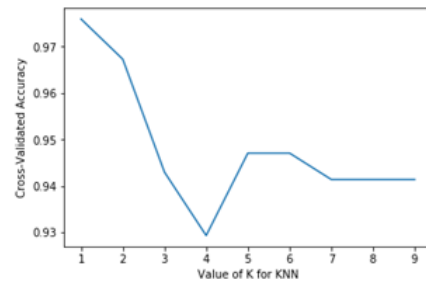

Fig. 9.   Majority Voting Results.

Fig. 10. Accuracy Per K Value.

TABLE. V.   PREDICTION ACCURACY USING ALL VARIABLES

Algorithms	Prediction accuracy
KNN	94.14%
SVM	94.79%
LR	91.98%

TABLE. VI.   EVALUATION MEASURES OF SRABC PAIRED WITH KNN, SVM AND LR

Model	Accuracy	RMSE	Precision	Recall	F-Measure
SRABC + KNN	**97.59%**	**0.1**	**.97**	**.98**	**.9736**
SRABC + SVM	92.38%	0.24	.93	.91	.9161
SRABC + LR	91.54%	0.26	.90	.94	.9114

Table VI shows the evaluation measures of the prediction-based model of SRABC paired with KNN, SVM and LR in the selected 14 courses as optimal subset.

Based on the results, the prediction accuracy of the model has increase from 94.14% to 97.59% and reduced the dimensionality from 60 variables to an optimal subset of 14 variables.

V.   CONCLUSION

The goal to develop a prediction-based model for curriculum analysis is achieved with the integration of SRABC as the searching strategy paired with KNN. The SRABC discovered the most relevant variables and exhibit an increased prediction accuracy through the aid of majority voting scheme which addresses the non-deterministic behavior of the algorithm, specifically for high dimensions problem. The superior performance of SRABC is due to the reverse mechanism in the initialization of food sources, new exploration and exploitation strategies in the employed and onlooker bees phase which composed of amalgamation, fourths-crossover, intersection and on-mutation.

Hence, it is recommended that the prediction-based model for curriculum analysis will be utilize to the remaining program offerings of Davao del Norte State College and further be implemented to higher learning institutions in the region.

For future works, it is recommended to verify the performance of SRABC in various actual case scenario as prediction-based model and further study can be investigated to implement SRABC to decision trees.

REFERENCES

[1] H. Sun, S. Wei-feng, Y. Geng, X. Yang, and B. K. Edziah, "How does natural resource dependence affect public education spending?," Environ. Sci. Pollut. Res., vol. 26, no. 4, pp. 3666–3674, 2019.

[2] A. Peña-ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," Expert Syst. with Appl., vol. 14, no. 4, pp. 1432–1462, 2014.

[3] I. A. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," Heliyon, vol. 5, no. 2, p. e01250, 2019.

[4] H. Ji, K. Park, J. Jo, and H. Lim, "Mining students activities from a computer supported collaborative learning system based on peer to peer network," Peer-to-Peer Netw. Appl., vol. 9, no. 3, pp. 465-476., 2016.

[5] S. Fedushko and T. Ustyianovych, "Predicting Pupil's Successfulness Factors Using Machine Learning Algorithms and Mathematical Modelling Methods," in International Conference on Computer Science, Engineering and Education Applications, 2019, pp. 625–636.

[6] H. Aldowah, H. Al-samarraie, and W. M. Fauzy, "Educational Data Mining and Learning Analytics for 21st century higher education: A Review and Synthesis," Telemat. Informatics, 2019.

[7] A. A. Saa, M. Al-Emran, and K. Shaalan, "Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques," Technol. Knowl. Learn., pp. 1–32, 2019.

[8] S. Morsy and G. Karypis, "A Study on Curriculum Planning and Its Relationship with Graduation GPA and Time To Degree," in Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 2019, pp. 26–35.

[9] N. M. Suhaimi, S. Abdul-Rahman, S. Mutalib, N. Hamimah, A. Hamid, and A. M. Malik, "Review on Predicting Students' Graduation Time Using Machine Learning Algorithms," Int. J. Mod. Educ. Comput. Sci., vol. 11, no. 7, pp. 1–13, 2019.

[10] S. A. Priyambada, E. Mahendrawathi, and B. N. Yahya, "Curriculum Assessment of Higher Educational Institution Using Aggregate Profile Clustering," in Procedia Computer Science, 2018, vol. 124, pp. 264–273.

[11] S. Yu and J. Principe, "Simple Stopping Criteria for Information Theoretic Feature Selection," entropy, vol. 21, no. 1, 2019.

[12] M. Mafarja, I. Jaber, S. Ahmed, and T. Thaher, "Whale Optimisation Algorithm for high-dimensional small-instance feature selection," Int. J. Parallel, Emergent Distrib. Syst., pp. 1–17, 2019.

[13] N. Nayar, S. Ahuja, and S. Jain, "Swarm Intelligence for Feature Selection: A Review of Literature and Reflection on Future Challenges," Adv. Data Inf. Sci., pp. 211–221, 2019.

[14] J. C. Bansal, A. Gopal, and A. K. Nagar, "Analysing Convergence, Consistency, and Trajectory of Artificial Bee Colony Algorithm," IEEE Access, vol. 6, pp. 73593–73602, 2018.

[15] A. Nayyar, V. Puri, and G. Suseendran, "Artificial bee Colony optimization population-based meta-heuristic swarm intelligence technique.," Data Manag. Anal. Innov. Springer, Singapore, pp. 513–525, 2019.

[16] E. Figueiredo, M. Macedo, H. Valadares, C. J. Santana, A. Gokhale, and C. J. A. Bastos-filho, "Swarm intelligence for clustering — A systematic review with new perspectives on data mining," Eng. Appl. Artif. Intell., vol. 82, no. May, pp. 313–329, 2019.

[17] L. Cui et al., "A smart artificial bee colony algorithm with distance-fitness-based neighbor search and its application," Futur. Gener. Comput. Syst., no. 89, pp. 478–493, 2018.

[18] S. Aslan, H. Badem, and D. Karaboga, "Improved quick artificial bee colony (iqABC) algorithm for global optimization," Soft Comput., 2019.

[19] K. Liu, C. Wang, and S. Liu, "Artificial Bee Colony Algorithm Combined with Previous Successful Search Experience," IEEE Access, vol. 7, pp. 34318–34332, 2019.

[20] T. Saw and P. H. Myint, "Swarm Intelligence Based Feature Selection for High Dimensional Classification: A Literature Survey," Int. J. Comput., vol. 33, No 1, pp. 69–83, 2019.

[21] S. K. Agarwal and S. Yadav, "A Comprehensive Survey on Artificial Bee Colony Algorithm as a Frontier in Swarm Intelligence," Ambient Commun. Comput. Syst., pp. 125–134, 2019.

[22] W. Xiang, Y. Li, R. He, X. Meng, and M. An, "An improved artificial bee colony algorithm with fitness-based information," IEEE Access, vol. 7, pp. 41052–41065, 2019.

[23] H. Gao, Y. Shi, C. Pun, and S. Kwong, "An Improved Artificial Bee Colony Algorithm with its Application," IEEE Trans. Ind. Informatics, vol. 15, no. 4, pp. 1853–1865, 2018.

[24] K. Hussain, M. Najib, M. Salleh, S. Cheng, Y. Shi, and R. Naseem, "Artificial bee colony algorithm: A component-wise analysis using diversity measurement," J. King Saud Univ. - Comput. Inf. Sci., 2018.

[25] C. Santana Jr, M. Macedo, H. Siqueira, A. Gokhale, and C. J. Bastos-filho, "A novel binary artificial bee colony algorithm," Futur. Gener. Comput. Syst., vol. 98, pp. 180–196, 2019.

[26] H. Rao, X. Shi, A. Kouassi, and J. Feng, "Feature selection based on artificial bee colony and gradient boosting decision tree," Appl. Soft Comput. J., vol. 74, pp. 634–642, 2019.

[27] A. J. P. Delima, A. M. Sison, and R. P. Medina, "Variable Reduction-based Prediction through Modified Genetic Algorithm," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 5, 2019.

[28] M. Garcia-Torres, D. Becerra-alonso, F. A. Gomez-Vela, F. Divina, I.L. Cobo, and Martinez-Alvarez, F., "Analysis of Student Achievement Scores: A Machine Learning Approach," In International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems(CISIS 2019) and 10th International Conference on EUropean Transactional Education(ICEUTE 2019), 2019, pp. 275–284.

[29] J. Tu, A. Lin, H. Chen, Y. Li, and C. Li, "Predict the Entrepreneurial Intention of Fresh Graduate Students Based on an Adaptive Support Vector Machine Framework," Math. Probl. Eng., 2019.

[30] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," J. Glob. Optim., pp. 459–471, 2007.

[31] D. Karaboga and B. Gorkemli, "A quick artificial bee colony (qABC) algorithm and its performance on optimization problems," Appl. Soft Comput., vol. 23, pp. 227–238, 2014.

[32] J. Yang et al., "An adaptive encoding learning for artificial bee colony algorithms," J. Comput. Sci., 2018.

[33] H. Habbi, Y. Boudouaoui, D. Karaboga, and C. Ozturk, "Self-generated fuzzy systems design using artificial bee colony optimization," Inf. Sci. (Ny)., vol. 295, pp. 145–159, 2015.

[34] E. Hancer, B. Xue, D. Karaboga, and M. Zhang, "A binary ABC algorithm based on advanced similarity scheme for feature selection," Appl. Soft Comput. J., vol. 36, pp. 334–348, 2015.

[35] D. Jia, X. Duan, and M. Khurram, "Binary Artificial Bee Colony optimization using bitwise operation," Comput. Ind. Eng., vol. 76, pp. 360–365, 2014.

[36] M. S. Kiran and M. Gunduz, "XOR-based artificial bee colony algorithm for binary optimization," Turkish J. Electr. Eng. Comput. Sci., pp. 2307–2328, 2013.

[37] C. Ozturk, E. Hancer, and D. Karaboga, "A novel binary artificial bee colony algorithm based on genetic operators," Inf. Sci. (Ny)., vol. 297, pp. 154–170, 2015.

[38] Z. B. Ozger, B. Bolat, and B. Diri, "A Probabilistic Multi-Objective Artificial Bee Colony Algorithm for Gene Selection," J. Univers. Comput. Sci., vol. 25, no. 4, pp. 418–443, 2019.

[39] A. Asuncion and D. Newman, "UCI machine learning repository," 2007.

[40] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," J. King Saud Univ. - Comput. Inf. Sci., 2019.

[41] S. S. Bhowmick, D. Bhattacharjee, and L. Rato, "Identification of tissue specific tumor biomarker using different optimization algorithms," Genes Genomics, vol. 41, no. 4, pp. 431–443, 2018.