ISSN 2278-3091
Volume 8, No.6, November – December 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse122862019.pdf
https://doi.org/10.30534/ijatcse/2019/122862019

# Prediction-Based Model for Student Dropouts using Modified Mutated Firefly Algorithm

**Ariel O. Gamao**[1], **Bobby D. Gerardo**[2]
[1]Graduate Programs, Technological Institute of the Philippines, Quezon City, Philippines, ariel.gamao@dnsc.edu.ph
[2]West Visayas State University, Ilo-ilo, Philippines, bobby.gerardo@gmail.com

**ABSTRACT**

Academic database is considered as the heart and soul of every higher education institutions. This database contains a vast amount of useful information that is useful for analysis. Algorithms for machine learning play a significant role in mining academic databases and have been proven to be effective when applied in the academic field. Prediction models are made using relevant classification algorithms for dropout analysis. The success of the prediction model depends on the performance of the feature selection algorithm used for dimensionality reduction. The study utilized the Modified Mutated Firefly Algorithm (MMFA) as a dimensionality reduction strategy to enhance the accuracy of the prediction model for dropout analysis. The results of the simulation revealed that the Decision Tree (DT) classifier outperformed the Naïve Bayesian using the three UCI datasets. After the test of benchmark datasets, a students' cumulative dataset was used to come up with a predictive model for dropout analysis of Davao del Norte State College, Davao del Norte, Philippines. The results of the experiment confirmed that the MMFA+DT obtained an accuracy rate of 95.82%, while MMFA+NB only has 92.85% using 10-fold cross-validation.

**Key words :** Dropout analysis, firefly algorithm, mutation process, prediction model, stochastic approach.

## 1. INTRODUCTION

Knowledge mining from large databases, especially from the field of academics, plays a significant role in all aspects of human life [1]. From a global perspective, it is an incontestable fact that the nation's progress is likely dependent on the education of its citizens [2]-[3]. Democratizing access to Higher Education (HE) has increased the diversity of students, and this new situation requires a deeper understanding of the students' paths leading to them dropping out or completing their courses [4]. The student's performance prediction is an important research topic because it can help prevent students from dropping out before final exams and identify students that need additional assistance[5]-[6].

Machine learning algorithms have been proven to be effective when applied in the academic field [7]. The optimization problem has been a hard task for many researchers to find the best local searching method. This problem also leads to a branch of knowledge which is the evolutionary computing. The methods were greatly influenced by nature. A few decades ago, many methods were developed, for instance, Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), Firefly Algorithm (FFA) or Artificial Immune System (AIS) that have been used to solve difficult optimization problems [8]-[9]. Classification is a data mining technique used to predict group membership for data instances in different types of problems [10]-[11]. Some classification algorithms may perform quite well in general, but these may be easily outperformed by other algorithms in terms of performance when dimensionality reduction technique or feature selection is not correctly performed [12].

The success of the classification process depends on the quality of datasets. Features may contain unreliable data, which may lead the classification process to produce undesirable results; thus, a feature selection approach is considered a solution for this kind of problem [13]. Also, the selection of appropriate highlights assumes a fundamental job in the selection process [14]. Feature selection (FS) is an essential machine learning technique for classification applications to achieve an optimal subset of input features [15]. The accuracy of prediction is a primary challenge in training a model [16]. Feature selection techniques do not alter the original features of the variables, but merely selects a subset of them [17]. Also, feature selection is a crucial task in applying machine learning in various fields. Also, the increase of data dimensionality poses a significant challenge to many existing feature selection methods concerning effectiveness and efficiency [7].

A novel Modified Mutated Firefly Algorithm (MMFA) is presented as a tool to enhance the accuracy of the prediction model by selecting the best features. The MMFA is tested with the use of UCI datasets for benchmarking purposes. The results of the benchmarking experiment were compared with the other related studies. After performing the benchmarking experiment, another dataset was used that was obtained from the Guidance and Testing Office of the Davao del Norte State College, Davao del Norte, Philippines. The dataset contains the students' cumulative record, covering SY2016-2017 to SY2018-19. The study used the MMFA integrated with classification algorithms to explore the best model as means in the analysis of students' dropout to be used by academic administrators in crafting policies as intervention in minimizing dropout rates.

## 2. LITERATURE REVIEW

### 2.1 Feature Selection

The basic idea of all selection methods is to select features that maximize the accuracy of prediction or classification [19]. The data preprocessing method is an essential part of any regression and classification problems. A classification or regression problem involves a high time complexity and can have poor performance when a large number of variables or features are used. However, it has high performance for a minimum size and the most practical features [20].

Feature selection is one of the essential techniques widely employed for reducing dimensionality. It aims to choose a small subset of the relevant features from the available ones according to specific relevance evaluation criteria, which usually leads to a better learning performance, lower model error, and better model generalization [21]. Datasets with higher dimensionality may lead to noise, redundant and irrelevant features, which may cause overfitting of models and can increase error rates in the optimization process. A dimensionality reduction technique can be applied in the preprocessing phase to clean up a noisy, redundant, and irrelevant data [22]. Due to existing high throughput technologies, and their recent advancements are resulting in high dimensional data due to which feature selection is being treated as handy and mandatory in such datasets [23].

### 2.2 Firefly Algorithm-based Feature Selection

The firefly algorithm is a nature-inspired meta-heuristic approach for global optimization, and it is based on the natural behavior of the flashing characteristics of a swarm of fireflies [24]. The flashing characteristics can be summarized by the following three essential rules [25], firstly, all fireflies are attracted to other fireflies regardless of their sex. Secondly, the attractiveness of the fireflies is proportional to their brightness, which means for any couple of flashing fireflies, the firefly with less brightness moves towards the brighter one. Lastly, the brightness or features of a firefly is calculated using the objective functions of the underlying optimization model.

A mutated FA algorithm is based on monitoring the movement of fireflies by using different probability for each firefly and then perform mutation on each firefly according to its probability. Simulations are performed to show the performance of the proposed algorithm with standard firefly algorithm, based on ten standard benchmark functions. The results reveal that the proposed algorithm improves the convergence speed, accurateness, and prevents premature convergence [26]. The modified FFA algorithm adaptively balances the exploration and exploitation to find the optimal solution quickly. The FFA can quickly search the feature space for optimal or near-optimal feature subset minimizing a given fitness function [27]. The number of attacks in recent times has tremendously increased due to the increase in Internet activities. This security issue has made the Intrusion Detection Systems (IDS) a significant channel for information security. The issue of classification time is significantly reduced in the IDS through feature selection [28].

Recently, there are available a vast amount of data in the field of medicine that helps the physicians in diagnosing diseases. Data mining techniques can be applied to medical data to extract knowledge so that disease prediction becomes accurate and more straightforward. Cardiotocogram (CTG) data is analyzed using Support Vector Machine (SVM) for predicting fetal risk. Opposition based firefly algorithm (OBFA) is used to extract the relevant features that maximized the classification performance of SVM [29]. Nature Inspired Algorithms are a famous meta-heuristic search algorithm used in solving combinatorial optimization problems [30].

The artificial fireflies are designed to represent the feature subset, and they move in a hyperdimensional space to obtain the best features. The features are extracted using a Discrete Cosine Transform (DCT) and Haar wavelets based Discrete Wavelet Transform (DWT). The algorithm is validated using benchmark face databases, namely ORL and Yale while outperforming various existing techniques [31]. Classification-based feature selection using firefly algorithm and fuzzy entropy, to increase the accuracy and performance of glistening's detection with the basic k-Nearest Classifier using all features and compared the result with feature selection method using standard firefly algorithm by evaluating average accuracy, precision, and F-measure, against ophthalmologist's hand-drawn ground-truth [32].

The feature selection is a crucial step in a pattern recognition system. The main objective of this selection is to reduce the features number by eliminating irrelevant and redundant attributes. Besides, it also maintains or improves the classifier performance using a neural network algorithm. Nevertheless, a new stochastic search strategy inspired by the clonal selection theory in an artificial immune system is proposed for feature subset selection. We have used the firefly and clonal selection algorithms to select the most relevant features in a dataset. In our proposed strategy, feature selection is an optimization algorithm that searches optimum with a reduced number of features in the feature space with reasonable accuracy rates [33].

Several variable selection algorithms in multivariate calibration can be accelerated using Graphics Processing Units (GPU). Among these algorithms, the Firefly Algorithm (FA) is a recently proposed metaheuristic that may be used for variable selection. This paper presents a GPU-based FA (FA-MLR) with multi-objective formulation for variable selection in multivariate calibration problems and compares it with some traditional sequential algorithms in the literature [34].

Motor Imagery (MI) electroencephalography (EEG) is widely studied for its non-invasiveness, easy availability, portability, and high temporal resolution. As for MI EEG signal processing, the high dimensions of features represent a research challenge. It is necessary to eliminate redundant features, which not only create an additional overhead of managing the space complexity but also might include outliers, thereby reducing classification accuracy. The firefly algorithm (FA) can adaptively select the best subset of features and improve classification accuracy [35]. An essential process relevant to prediction analysis using feature selection have been introduced with favorable accuracy rates applying bio-inspired search algorithms that produce optimal attribute set [36].

### 2.3 Feature Selection-based Classification Algorithms

The success of applying machine learning methods to real-world problems depends on many factors. One such factor is the quality of available data. The more the collected data contain irrelevant or redundant information, or contain noisy and unreliable information, the more difficult for any machine learning algorithm to discover or obtain acceptable and practicable results. Feature selection refers to the process of identifying and removing redundant and irrelevant features in the dataset. Regardless of whether a learner attempts to select features itself, or ignores the issue, feature selection before learning has obvious merits [37].

Several data mining techniques become the leading techniques that can be applied in the medical field; the following are Decision tree algorithm, K-Nearest Neighbor algorithm, on a large dataset from the "Hepatitis dataset‟ (derived from the UCI Machine Learning Repository) that comprises of 20 attributes (including class) and 155 instances. Also, it is investigated on the importance of feature selection and applied three feature selection algorithms, namely Fisher filtering, Relief filtering, Step Disc, and classified the dataset using 15 most common classifiers [38]-[39].

Searching for meaningful patterns from big datasets has become a challenging task. So, data miners, on the other hand, try to adopt innovative methods to solve problems by applying feature selection for dimensionality reduction [40]. Due to the massive influence of prediction models in different sectors of society, various researchers have employed hybrid algorithms with improved accuracy rates of the prediction model. Genetic Algorithms (GAs) sufficiently enhances the performance of another prediction model by integrating the GA with a novel Inversed Bi-segmented Average Crossover (IBAX) operator paired with rank-based selection function to the KNN algorithm [41].

Feature selection was applied in the cell apoptosis/survival dataset to achieve a good result by dividing it into three main categories, namely: wrapper method (WM), filtering method (FM), and Embedded Method (EM). After applying the feature selection (FS) algorithm, seven different marker proteins were obtained[42].

A novel feature selection method based on a decision tree for price forecasting is proposed in this work. The method uses a genetic algorithm along with a decision tree classifier to obtain the minimum number of features giving an optimum forecast accuracy. The usefulness of the approach is established through the performance test of the forecaster using the feature selected by this approach. It is found out that the forecast with the reduced features consistently outperformed a more extensive feature set [43].

A decision tree is a well-established model that offers not only excellent predictive performance but also provides a rich feature important set. While practitioners often employ variable importance methods that rely on this impurity-based information, these methods remain poorly characterized from a theoretical perspective [44].

Feature selection improves the accuracy of Naive Bayes for five data sets to maintain or degrade accuracy for six data sets. The accuracy of Naive Bayes with feature selections improves on mushroom data set (from 95.8% to maximum 97.7%), vote (from 90.1% to maximum 91.7%), credit (from 77.7% to maximum 84.9%), audiology (from 73.5% to maximum 73.9%) and M2 (from 66.4% to maximum 67.1%). Maximum improvement is 7.2% [45].

One of the most successful and widely used methods of classification algorithms is the Naive Bayesian learning (NB). The NB represents each class with a probabilistic summary and finds the most likely class for each example, and it is asked to classify [46].

Attribute selection results can be of value in determining one type of data packet traffic. In obtaining the results (data), this study uses several stages of Data Capture, Feature Extraction, and Feature Selection, but in this study only focuses on the process of feature selection using the gain and Entropy Information and Naïve Bayes algorithm. The testing process by dividing raw data into parts is 70 percent for Training data and 30 percent for testing data [47].

## 3. METHODOLOGY

### 3.1 MMFA for Optimized Feature Selection

Fireflies inspire a Modified Mutated Firefly Algorithm (MMFA) in sending information by producing light intensity, and this is also called a swarm intelligence optimization algorithm grounded on the population search approach [48]. The MMFA applies the same principle of the MFA. Thus, every firefly is attracted to other fireflies regardless of sex. Attractiveness is directly proportional to the light intensity of firefly, where a firefly that emits less light may relocate closer to a firefly that has brighter light intensity.

### 3.1 Initialization Phase

The initialization phase determines the attributes of the dataset in terms of the number of instances, features, and classes. Objective functions such as Ackley, Rosenbrock, and Sphere were needed to establish randomly generated real numbers serving as the light intensity and position of the fireflies. The objective functions calculate the light intensity of fireflies. With the attractiveness and distance, MMFA uses the attractiveness and light intensity. The light intensity of a firefly is considered as $I(x) \propto f(x)$ where I is the intensity of a firefly, and x is a particular location. Attractiveness β changes when distance $r_{ij}$ also changes from firefly i to firefly j [49]. Light intensity $I(r)$ applies the inverse square law of light intensity variations and absorption coefficient γ [50]. Thus, [51] defined attractiveness β of firefly in (1):

$$\beta = \beta_0 \exp(-\gamma r^2) \qquad (1)$$

Where $\beta_0$ is the attractiveness at distance r = 0, and γ is the light absorption coefficient. The distance r of any two fireflies is calculated using the Euclidean distance from a lesser intensity firefly $i_{th}$ towards brighter firefly $j_{th}$ [25] as defined in (2):

$$r = \sqrt{(i_2 - i_2)^2 + (j_2 - j_1)^2} \qquad (2)$$

The movement a firefly with lesser light intensity from position *i* to position *j* is the attraction towards the firefly with greater light intensity, as shown in (3):

$$x_i = x_i + \beta_0 e(-\gamma r^2_{ij})(x_j - x_i) + \alpha\varepsilon_{i,} \qquad (3)$$

The initial positions of n fireflies are uniformly distributed in the search space whenever the number of fireflies is greater than the number of local optima [52]. The fundamental advantage of this attraction phenomenon subsequently allows the exploration of fireflies for search space, and update their position [53]. After the generation of fireflies, each firefly with its corresponding light intensity undergoes a discretization process to mark the candidate firefly with either 1 or 0 [54] and [55]. The binary values are essential as these are the basis for fireflies to qualify in the selection phase [56] as shown in (4):

$$F(li) = \begin{cases} 1, li \geq 0.5 \\ 0, li \leq 0.5 \end{cases} \qquad (4)$$

### 3.2 Selection of Firefly using Light Intensity Values

To establish the size of qualified fireflies, the MMFA counts all that have 1 value to serve as the size of the features of the dataset. In selecting qualified features, the algorithm utilizes random permutation to generate unique random numbers based on the size of features of the dataset to be used as indices of the vectors. This strategy was used to make sure the random numbers generated in the stochastic approach are not repeated [57]-[58].

The number generated is used in searching for fireflies or features in the dataset. Once a number is generated, it is removed from the vector in order not to generate the same number in the future. The removed random numbers form part in the vector of fireflies, which means the fireflies in the vector in uniformly unique numbers serving as indices of the vector, as shown in Figure 1.

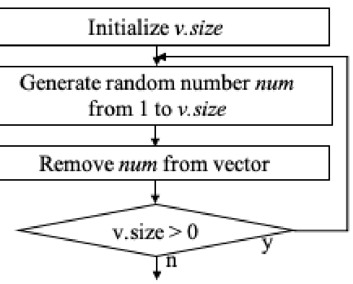


**Figure 1:** Random Permutation

Randomization is used to augment the exploitation capabilities in the search space of relevant features. As shown in Figure 2, for example, based on the discretization process, there are 10 qualified fireflies. The said number of fireflies serves as the size of a vector (vector1), which is subject to random permutation. Then a random number is generated from within the range of the size of the vector, suppose rand = 6, whereby 6 elements from the vector1 are randomly picked to form a new vector (vector2) using random permutation as shown in Figure 2. The new values in vector2 serves as the indices subject for stochastic searching.

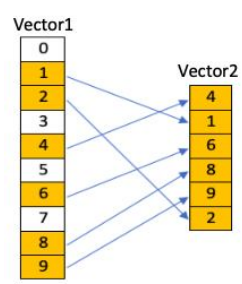


**Figure 2:** Random Permutation

### 3.2 Mutation Process of Fireflies

During the attraction process, mutation may happen for the lesser light intensity by improving its features from a brighter firefly. Mutation means that several features to be added depending on the result of the mutation probability calculation. Only the lesser intensity firefly always has a significant chance of mutation. The mutation operator is used to improve features of selected fireflies with a probability $p_m$ (mutation probability), leading to additional firefly features to help the search process escape from local optimal traps. Mutation probability $p_m$ allows every firefly to undergo its necessary mutation phenomenon; $p_m$ affects the entire performance of the algorithm. Usual values of $p_m$ are also adapted from GA, i.e., 0.001 to 0.05. The mutation probability (MP) is calculated in (5):

$$MP = f_{new} - f_{old}, \tag{5}$$

Where $f_{new}$ is the fitness value of the newly generated firefly and $f_{old}$ the fitness value of the original firefly, for an iteration that undergoes $n_m$ mutation operations, the average mutation progress value AMP is denoted in equation (6): The mutation rate updates its value during the iteration process using the AMP values.

$$AMP = \frac{1}{n_m} \sum MP \tag{5}$$

### 3.3 Stochastic Approach

The MMFA employs a stochastic approach for balanced searching capability in the search space such as Las Vegas and Monte Carlo Algorithms. However, this study only focuses on the Las Vegas algorithm, where it outperforms the Monte Carlo Algorithm in terms of performance to explore the optimum result regardless of time constraint [59], as shown in Algorithm 1.

**Algorithm 1**: Las Vegas Algorithm

```
Las_Vegas_Search()
{
   while(i <= noOfFireflies){
      randomly selects from n elements;
      if(n is found)
         return n;
      i = i + 1;
   }
}
```

### 3.4 Student Cumulative Record Prediction Analysis Model

The study used a model in analyzing students' cumulative data using who most likely to drop from college (DNSC), as shown in Figure 2. The dataset was fed into the MMFA for dimensionality reduction process to fit in classification algorithms using Naïve Bayesian and Decision Tree. The success of this model plays a vital role in the decision-making of academic administrators where drop-outs can be minimized if not prevented by establishing remediation activities for students most likely to drop.

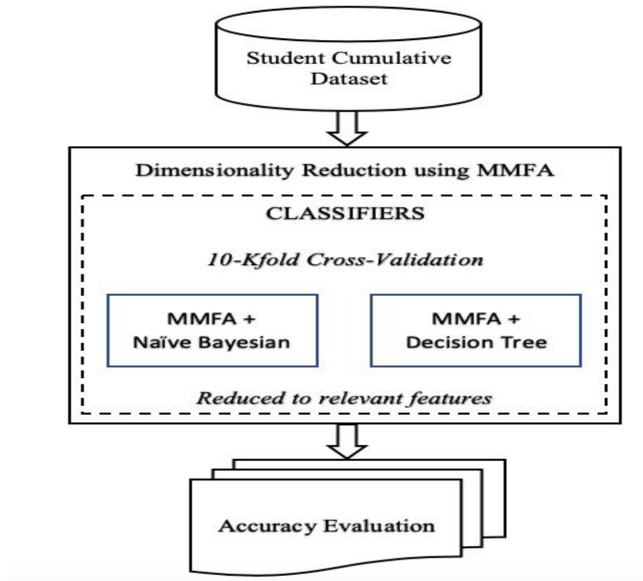


**Figure 2:** Student Cumulative Prediction Analysis Model

### 3.5 Using Benchmark Datasets

In testing the accuracy of MMFA, the study utilized three public datasets, namely: Glass Dataset, Parkinson Dataset 2, and Breast Tissue downloaded from the UCI repository [60],[33] (see Table 1).

**Table 1:** Attributes of Benchmark Databases

| Dataset | Attributes | | |
|---|---|---|---|
| | No. of Features | No. of Classes | No. of Instances |
| Glass Dataset | 10 | 6 | 214 |
| Breast Tissue | 9 | 6 | 106 |
| Parkinson 2 | 23 | 2 | 195 |

All the datasets were used and tested through the MMFA for dimensionality reduction with the integration of classifier algorithms to achieve the optimum prediction accuracy. The datasets were split into training and testing sets using a 10-fold cross-validation for both Naïve Bayesian (NB) and Decision Tree (DT) classifiers. The DT has consistently outperformed the NB for the three datasets, as shown in Figures 3-5.

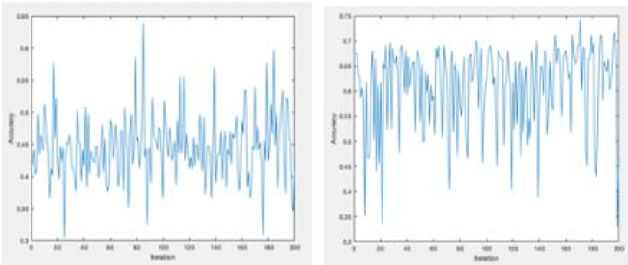


**Figure 3**: Accuracy Test Using Glass Dataset

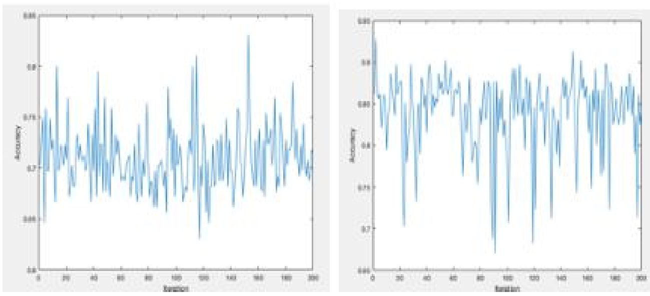


**Figure 4**: Accuracy Test Using Parkinson Dataset

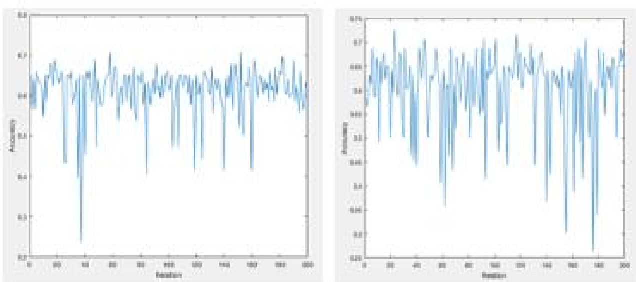


**Figure 5:** Accuracy Test Using Breast Tissue Dataset

In the comparison of accuracy results, the study benchmarked the existing Immune Firefly Algorithm (IFA) as the basis for the performance of the MMFA when applied for feature selection with the integration of relevant classifier algorithms. The comparative results of the accuracy show that both MMFA+NB and MMFA+DT outperformed the existing IFA. The result of the performance of IFA is just a benchmark result taken from the other relative research for comparative purposes only [33]. However, the MMFA+DT shows a better performance in terms of prediction accuracy (see Table 2).

**Table 2:** Attributes of Student Cumulative Dataset

| ALGORITHM | | DATASETS | | |
|---|---|---|---|---|
| | | **GLASS DATASET** | **PARKINSON 2** | **BREAST TISSUE** |
| **IFA** | **No. of Features** | 8 | 2 | 5 |
| | **Accuracy** | 60.32 | 83.14 | 68.49 |
| **MMFA+ NB** | **No. of Features** | 3 | 4 | 6 |
| | **Accuracy** | 62.83 | 83.01 | 70.76 |
| **MMFA+ DT** | **No. of Features** | 7 | 7 | 6 |
| | **Accuracy** | 73.30 | 92.82 | 72.64 |

### 3.6 Data Collection

The study utilized a cumulative record of freshmen college students of the Guidance and Testing Office of the Davao del Norte State College containing the personal and family data, socioeconomic status, educational data, record of outstanding performance, and tests taken, such as OLSAT and SATT. The said dataset is an official record of the Guidance Office of the college, where students were required to fill in the cumulative student form during the enrollment period, and this contains 45 (see Table 3).

**Table 3:** Attributes of Students' Cumulative Dataset

| Dataset | Attributes | | |
|---|---|---|---|
| | No. of Features | No. of Classes | No. of Instances |
| Student Cumulative Dataset | 45 | 2 | 1862 |

In the experiment, the study utilized two classification algorithms, namely: Naïve Bayesian and Decision Tree, using a 10-fold cross-validation strategy to arrive on the best model with the best prediction accuracy. The said model can be used in the analysis of student dropouts using the students' cumulative record for the academic administrations to come-up with necessary remediation activities to avoid if not minimize dropouts for students in the whole duration of students' journeys in the college [61], [62], [63], [64].

Two prediction models were drawn for student cumulative record analysis. Both models underwent dimensionality reduction to come up with the best models concerning each classifier algorithm. It is observed that Decision Tree has a higher accuracy of 95.82% compared to the Naïve Bayesian with only 92.85%, whereby still consistently performing better than the Naïve Bayesian and still holds even with the benchmark datasets. With its best model, the Naïve Bayesian was reduced to 27% of its features, while the Decision Tree also reduced to 23% of relevant features necessary for the predictive model (see Table 4).

**Table 4:** Prediction Accuracy Result of Students' Cumulative Dataset

| **Datasets** | **MMFA + NB** | | **MMFA + DT** | |
|---|---|---|---|---|
| | **No. of features** | **Accuracy** | **No. of features** | **Accuracy** |
| Students' Cumulative Record | 12 | 92.85 | 10 | 95.82 |

During the dimensionality reduction process, models were created for features. There were 12 relevant features for the MMFA+NB and 10 relevant features for the MMFA+DT. The variables in the table represent variables in the students' cumulative record of the college where academic administrators may consider the said factors that can affect student dropouts (see Table 5).

**Table 5:** Dimensionality Reduction Result

| **MMFA + Naïve Bayesian** | **MMFA + Decision Tree** |
|---|---|
| Var9, Var13, Var20, Var5, Var24, Var12, Var18, Var17, Var10, Var8, Var28, Var2 | Var14, Var18, Var17, Var7, Var10, Var16, Var12, Var15, Var5, Var8 |

## 4. CONCLUSION

With the experiments conducted in the study, the goal of developing a predictive model for the analysis of dropouts using the students' cumulative record was achieved through the integration of the MMFA with the relevant classification algorithms such as NB and DT. The MMFA was able to explore the optimized solution or model through the search space with the use of firefly behavior using the mutation process.

Thus, the accuracy achieved through the study can be used by academic administrators of the Davao del Norte State College in crafting academic policies that enhance students' performance for both curricular and extra-curricular activities to minimize dropouts. Future researchers may consider integrating the MMFA with other relevant classification algorithm to explore its potential use for whatever it serves best.

## REFERENCES

[1] S. Chokkadi, M. S. Sannidhan, K. B. Sudeepa, and A. Bhandary, "A study on various state of the art of the art face recognition system using deep learning techniques," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 4, pp. 1590–1600, 2019. https://doi.org/10.30534/ijatcse/2019/84842019

[2] L. A. Choudhary AI, "Economic Effects of Student Dropouts: A Comparative Study," *J. Glob. Econ.*, vol. 03, no. 02, pp. 2–5, 2015. https://doi.org/10.4172/2375-4389.1000137

[3] I. S. Makki and F. Alqurashi, "An adaptive model for knowledge mining in databases 'EMO_MINE' for tweets emotions classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 7, no. 3, pp. 52–60, 2018. https://doi.org/10.30534/ijatcse/2018/04732018

[4] J. R. Casanova, A. Cervero, J. C. Núñez, L. S. Almeida, and A. Bernardo, "Factors that determine the persistence and dropout of university students," *Psicothema*, vol. 30, no. 4, pp. 408–414, 2018.

[5] M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381–407, 2019. https://doi.org/10.1007/s10462-018-9620-8

[6] A. K. P. Moore, "DROPPED OUT: FACTORS THAT CAUSE STUDENTS TO LEAVE BEFORE GRADUATION," *ABA J.*, vol. 102, no. 4, pp. 24–25, 2017.

[7] M. Kumar and A. J. Singh, "Performance analysis of students using machine learning & data mining approach," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 3, pp. 75–79, 2019.

[8] V. Osuna-Enciso, E. Cuevas, and H. Sossa, "A comparison of nature inspired algorithms for multi-threshold image segmentation," *Expert Syst. Appl.*, vol. 40, no. 4, pp. 1213–1219, 2013. https://doi.org/10.1016/j.eswa.2012.08.017

[9] G. K. Orman and V. Labatut, "Digital Information and Communication Technology and Its Applications," *Digit. Inf. Commun. Technol. Its Appl. (part II)*, vol. 167, no. June 2011, pp. 265–279, 2011.

[10] E. R. S.Neelamegam, "An Overview of Classification Algorithm in Data mining," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 4, no. 12, pp. 255–257, 2015.

[11] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Inf.*, vol. 10, no. 4, pp. 1–68, 2019. https://doi.org/10.3390/info10040150

[12] J. Gama and P. Brazdil, "Characterization of Classi cation Algorithms 2 Normalizing Meta-Datasets," no. October 2013, pp. 1–12, 2000.

[13] E. M. Mashhour, E. M. F. El Houby, and K. T. Wassif, "Feature Selection Approach based on Firefly Algorithm and," vol. 8, no. 4, pp. 2338–2350, 2018.

[14] M. I. Mohmand, A. Bhaumik, M. Humayun, and Q. Shah, "Science Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse31852019.pdf The Performance and Classifications of Audio-Visual Speech Recognition by," vol. 8, no. 5, 2019. https://doi.org/10.30534/ijatcse/2019/31852019

[15] L. Zhang, L. Shan, and J. Wang, "Optimal feature selection using distance-based discrete firefly algorithm with mutual information criterion," *Neural Comput. Appl.*, vol. 28, no. 9, pp. 2795–2808, 2017.

[16] R. Moazenzadeh, B. Mohammadi, S. Shamshirband, and K. W. Chau, "Coupling a firefly algorithm with support vector regression to predict evaporation in northern iran," *Eng. Appl. Comput. Fluid Mech.*, vol. 12, no. 1, pp. 584–597, 2018.

https://doi.org/10.1080/19942060.2018.1482476
[17] S. Dash and B. Patra, "Feature selection algorithms for classification and clustering in bioinformatics," *Artif. Intell. Concepts, Methodol. Tools, Appl.*, vol. 3, no. January, pp. 2071–2091, 2016.
[18] A. A. Yahya, "Feature Selection for High Dimensional Data : An Evolutionary Filter Approach Feature Selection for High Dimensional Data : An Evolutionary Filter Approach," no. December, 2014.
[19] I. B. Yashkov, "Feature selection using decision trees in the problem of JSM classification," *Autom. Doc. Math. Linguist.*, vol. 48, no. 1, pp. 6–11, 2014.
[20] Y. Masoudi-Sobhanzadeh, H. Motieghader, and A. Masoudi-Nejad, "FeatureSelect: A software for feature selection based on machine learning approaches," *BMC Bioinformatics*, vol. 20, no. 1, 2019.
[21] Y. AKHIAT, M. CHAHHOU, and A. ZINEDINE, "Ensemble Feature Selection Algorithm," *Int. J. Intell. Syst. Appl.*, vol. 11, no. 1, pp. 24–31, 2019.
[22] B. Venkatesh and J. Anuradha, "A review of Feature Selection and its methods," *Cybern. Inf. Technol.*, vol. 19, no. 1, pp. 3–26, 2019.
[23] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2019. https://doi.org/10.1016/j.jksuci.2019.06.012
[24] X. S. Yang, "Firefly algorithm, Levy flights and global optimization," *Res. Dev. Intell. Syst. XXVI Inc. Appl. Innov. Intell. Syst. XVII*, pp. 209–218, 2010.
[25] B. Alizadeh and S. Bakhteh, "A modified firefly algorithm for general inverse p-median location problems under different distance norms," *Opsearch*, vol. 54, no. 3, pp. 618–636, 2017.
[26] S. Arora, S. Singh, S. Singh, and B. Sharma, "Mutated Firefly Algorithm," *2014 Int. Conf. Parallel, Distrib. Grid Comput. Mutated*, pp. 33–38, 2014.
[27] E. Emary, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien, and B. Parv, "Firefly optimization algorithm for feature selection," *ACM Int. Conf. Proceeding Ser.*, vol. 02-04-Sept, 2015.
[28] R. F. Najeeb and B. N. Dhannoon, "A feature selection approach using binary Firefly Algorithm for network intrusion detection system," *ARPN J. Eng. Appl. Sci.*, vol. 13, no. 6, pp. 2347–2352, 2018.
[29] S. V and M. D, "Opposition Based Firefly Algorithm Optimized Feature Subset Selection Approach for Fetal Risk Anticipation," *Mach. Learn. Appl. An Int. J.*, vol. 3, no. 2, pp. 55–64, 2016.
[30] T. S. Durga and V. Yasaswini, "An Enhancement for the optimization of feature selection to perform classification Using Meta Heuristic Algorithms," pp. 64–70, 2016.
[31] V. Agarwal and S. Bhanot, "Firefly inspired feature selection for face recognition," *2015 8th Int. Conf. Contemp. Comput. IC3 2015*, pp. 257–262, 2015.
[32] P. Jitpakdee, B. Uyyanonvara, and C. Hull, "Feature Selection Using Fuzzy-based Firefly Algorithm for Glistenings Detection on Intraocular Lenses," vol. 30, no. 2, pp. 103–115, 2019.
[33] B. K. Kihel and S. Chouraqui, "Firefly optimization using artificial immune system for feature subset selection," *Int. J. Intell. Eng. Syst.*, vol. 12, no. 4, pp. 337–347, 2019. https://doi.org/10.22266/ijies2019.0831.31
[34] L. C. M. De Paula, A. S. Soares, T. W. De Lima, A. C. B. Delbem, C. J. Coelho, and A. R. G. Filho, "A GPU-based implementation of the firefly algorithm for variable selection in multivariate calibration problems," *PLoS One*, vol. 9, no. 12, pp. 1–22, 2014.
[35] A. Liu, K. Chen, Q. Liu, Q. Ai, Y. Xie, and A. Chen, "Feature selection for motor imagery EEG classification based on firefly algorithm and learning automata," *Sensors (Switzerland)*, vol. 17, no. 11, 2017.
[36] M. A. Basir, Y. Yusof, and M. S. Hussin, "Optimization of attribute selection model using bio-inspired algorithms," *J. Inf. Commun. Technol.*, vol. 18, no. 1, pp. 35–55, 2019.
[37] G. Guo, D. Neagu, and M. T. D. Cronin, "Using kNN model for automatic feature selection," *Lect. Notes Comput. Sci.*, vol. 3686, no. PART I, pp. 410–419, 2005.
[38] A. Soofi and A. Awan, "Classification Techniques in Machine Learning: Applications and Issues," *J. Basic Appl. Sci.*, vol. 13, no. August, pp. 459–465, 2017.
[39] A. . Nancy.P, Sudha.V, "Analysis of feature selection and classification algorithms on hepatitis data," *Int. J. Adv. Res. Comput. Eng. Technol.*, vol. 6, no. 1, pp. 19–23, 2017.
[40] M. Naseriparsa, "Improving Performance of a Group of Classification Algorithms Using Resampling and Feature Selection," *World Comput. Sci. Inf. Technol. J.*, vol. 3, no. 4, pp. 70–76, 2013.
[41] 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, pp. 356–363, 2019. https://doi.org/10.14569/IJACSA.2019.0100544
[42] S. Jain and A. O. Salau, "An image feature selection approach for dimensionality reduction based on kNN and SVM for AkT proteins," *Cogent Eng.*, vol. 6, no. 1, pp. 1–14, 2019.
[43] A. K. Srivastava, D. Singh, A. S. Pandey, and T. Maini, "A novel feature selection and short-term price forecasting based on a decision tree (J48) model," *Energies*, vol. 12, no. 19, pp. 1–17, 2019.
[44] S. J. Kazemitabar, A. A. Amini, A. Bloniarz, and A. Talwalkar, "Variable importance using decision trees," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 426–435, 2017.
[45] J. Novakovic, "The Impact of Feature Selection on the Accuracy of Naive Bayes Classifier," *18th Telecommun. forum TELFOR*, vol. 2, pp. 1113–1116, 2010.
[46] C. Ratanamahatana and D. Gunopulos, "Feature selection for the naive bayesian classifier using decision trees," *Appl. Artif. Intell.*, vol. 17, no. 5–6, pp. 475–487, 2003.
[47] A. Fali Oklilas, Tasmi, S. Desy Siswanti, M. Afrina, and H. Setiawan, "Attribute selection using information gain and naïve bayes for traffic classification," *J. Phys. Conf. Ser.*, vol. 1196, no. 1, 2019.
[48] C. F. Wang and W. X. Song, *A novel firefly algorithm based on gender difference and its convergence*, vol. 80.

Elsevier B.V., 2019.
[49] T. Apostolopoulos and A. Vlachos, “Application of the Firefly Algorithm for Solving the Economic Emissions Load Dispatch Problem,” *Int. J. Comb.*, vol. 2011, pp. 1–23, 2010.
[50] L. Tighzert, C. Fonlupt, and B. Mendil, “A set of new compact firefly algorithms,” *Swarm Evol. Comput.*, vol. 40, pp. 92–115, 2018.
[51] Y. Zhang, X. fang Song, and D. wei Gong, “A return-cost-based binary firefly algorithm for feature selection,” *Inf. Sci. (Ny).*, vol. 418–419, pp. 561–574, 2017.
[52] Z. W. K. J. H. Geem, “A New Heuristic Optimization Algorithm: Harmony Search,” *Optimization*, pp. 35–54, 2015.
[53] M. A. Bramer and R. Ellis, “Firefly Algorithm, Levy Flights and Global Optimization,” *Res. Dev. Intell. Syst. XXVI Inc. Appl. Innov. Intell. Syst. XVII*, no. June, 2010.
[54] J. Zhang, Q. Wang, Y. Li, D. Li, and Y. Hao, “A method for chinese text classification based on three-dimensional vector space model,” *Proc. - 2012 Int. Conf. Comput. Sci. Serv. Syst. CSSS 2012*, pp. 1324–1327, 2012. https://doi.org/10.1109/CSSS.2012.334
[55] S. L. Marie-sainte and N. Alalyani, “Firefly Algorithm based Feature Selection for Arabic Text Classification,” *J. King Saud Univ. - Comput. Inf. Sci.*, 2018.
[56] R. E. E. Cutad and B. D. Gerardo, “A Prediction-based Curriculum Analysis using the Modified Artificial Bee Colony Algorithm,” vol. 10, no. 10, pp. 117–123, 2019.
[57] J. Gao and D. Wang, “Permutation Generation: Two New Permutation Algorithms,” p. 7, 2003.
[58] A. Bacher, O. Bodini, H. K. Hwang, and T. H. Tsai, “Generating random permutations by coin tossing: Classical algorithms, new analysis, and modern implementation,” *ACM Trans. Algorithms*, vol. 13, no. 2, 2017.
[59] H. H. Hoos and T. Stützle, “Evaluating Las Vegas Algorithms: Pitfalls and Remedies,” *Proc. Fourteenth Conf. Uncertain. Artif. Intell.*, pp. 238–245, 1998.
[60] A. Kedia, M. Narsaria, S. Goswami, and J. Taparia, “Empirical Study to Evaluate the Performance of Classification Algorithms on Healthcare Datasets,” vol. 5, no. 1, pp. 1–11, 2017.
[61] S. Geiser and M. V. Santelices, “Validity of high-school grades in predicting student success beyond the freshman year: High school record vs. standardized tests as indicators of four-year college outcomes,” *CSHE Res. Occas. Pap. Ser.*, p. 35, 2007.
[62] B. A. Friedman and R. G. Mandel, “Motivation predictors of college student academic performance and retention,” *J. Coll. Student Retent. Res. Theory Pract.*, vol. 13, no. 1, pp. 1–15, 2011.
[63] F. J. da Costa, M. de S. Bispo, and R. de C. de F. Pereira, “Dropout and retention of undergraduate students in management: a study at a Brazilian Federal University,” *RAUSP Manag. J.*, vol. 53, no. 1, pp. 74–85, 2018. https://doi.org/10.1016/j.rauspm.2017.12.007
[64] E. M. Sosu and P. Pheunpha, “Trajectory of University Dropout: Investigating the Cumulative Effect of Academic Vulnerability and Proximity to Family Support,” *Front. Educ.*, vol. 4, no. February, pp. 1–10, 2019. https://doi.org/10.3389/feduc.2019.00006