OPEN ACCESS
International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

An Experimental Comparison of Hybrid Modified Genetic Algorithm-based Prediction Models

Allemar Jhone P. Delima

Abstract*: The quest for an optimal prediction model is still a hot topic in the field of data mining and machine learning. An optimal model is achieved when the algorithm used posses the highest performance rating based on the evaluation matrix the researchers sought to satisfy. Through this study, a hybrid modified genetic algorithm-based prediction was modeled along with the selected data mining algorithms namely the K-Nearest Neighbor, Naive Bayes, C4.5, and Rule Base algorithms such as DT, JRip, OneR, and PART. The crossover operator of the genetic algorithm was also modified to optimize the minimization process of the variables before prediction. The simulation results showed that the MGA-KNN outperformed the MGA-NB, MGA-C4.5 and MGA-RB with DT, JRip, OneR and PART algorithms with the prediction accuracy of 94%, 86%, 89%, 85%, 92%, 75%, and 92%, respectively.

***Index Terms*: Hybrid prediction model, Modified genetic algorithm, IBAX operator, Prediction accuracy enhancement**

I. INTRODUCTION

Data Mining (DM), otherwise known as Knowledge Discovery in Databases (KDD), is one of the quickest developing fields due to the tremendous need for added value from large-scale databases. It is defined as the extraction of information from expansive databases to discover essential and valuable data [1]. According to [2], the objective of a data mining is either to create a descriptive model or a predictive model. A descriptive model exhibits the data in a succinct form which is basically a summary of the data points, discover patterns in the data and link the connections between attributes represented by the data. Some of the tasks under the descriptive model include association rules, clustering, summarizations, and sequence discovery. Meanwhile, the predictive model works by predicting future values of the data, which utilizes known results found from various previous datasets. The predictive data mining model includes classification, prediction, regression, and analysis of time series. Prediction [3] is one of the renowned data mining approach that is commonly used in educational data mining (EDM) [4]-[6], crime mining [7], [8], business and finance [9], [10], health [11], [12], and more.

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
Allemar Jhone P. Delima*, College of Engineering and Information Technology, Surigao State College of Technology, Surigao City, Philippines.

The literature in forecasting and prediction is extensive. Various models were developed and utilized in response to the problems the researchers sought to answer. According to [3], there are two general categories for forecasting and prediction namely the classical and modern methods. Classical methods include econometrics-based approaches, statistical inferences, and traditional mathematical programming while the modern method employs soft computing algorithms and artificial intelligence.

For example, the study of [13] proposed a forecasting approach that combines the strengths of the neural network and multivariate time series models. In the proposed approach, forecasting the exchange rate of UK, USA, and Japan was done first by time series, and then GRNN was used to correct the forecasting errors. On the other hand, [14] examined the forecasting accuracy of the exchange rate in Brazil using different approaches. They employed intelligent systems like multilayer perceptron and radial basis function neural networks and the Takagi–Sugeno fuzzy system versus the traditional methods of forecasting such as autoregressive moving average (ARMA) and ARMA-generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) linear models. It was found out that the intelligent-based methods provided more accurate results than the traditional ones.

Recently, [15] developed a prediction model for OTOP's products using K-Nearest Neighbor (KNN) in a 5, 10, 15, 20, and 25 k-fold cross-validations and K-NN with k value assigned with 3, 5, and 7. The model with 5 folds cross-validation and K-NN with k=3 yields the best prediction with 87.73% accuracy. Moreover, the authors suggested to compare the reliability of the results using Naïve Bayes, C4.5 and Rule base algorithms in order to search for the optimal model for prediction, hence, this study. The proposed prediction models incorporated the modified genetic algorithm (MGA) with its new crossover mating scheme to the NB, C4.5, and RB algorithms. The MGA-KNN, MGA-NB, MGA-C4.5, and MGA-RB prediction models were compared and evaluated to search for the optimal model for prediction.

For real encoding problems using the arithmetic function, the average crossover (AX) [16] of the genetic algorithm was modified in this study.

II. RELATED LITERATURE

A. Hybrid Prediction Models

With the advent of hybridization of various algorithms, prediction models have become more effective and efficient in performing the job. For instance, the imperialist competitive algorithm (ICA), which is a type of evolutionary algorithm (EA), was used to optimize the artificial neural network (ANN) in the study of [17]. The result showed that prediction using ICA-ANN model outperformed the ANN model when used alone. In the study of [18], a model comprising the genetic algorithm and SVM for vaccine design was conducted. An average prediction with 93.50% accuracy on IEDB dataset was achieved while the accuracy of 95.125% was generated using Wang benchmark dataset. Meanwhile, linear and nonlinear models using ARIMA and ANN respectively were used in the study of [19]. The adaptation of hybrid methodology combining ARIMA and Deep Neural Network (DNN), which is an ANN model with multiple hidden layers, was considered as the optimal model for predicting roll motion compared to the non-hybrid models. It was found out that DNN-ARIMA hybrid model showed improved forecast accuracy and was identified to be very effective. Further, a drought prediction model was proposed by [20] using a de-noised empirical mode decomposition (EMD) and a deep belief network (DBN). The proposed method was applied to predict different time scale drought indices across the Colorado River basin using a standardized streamflow index (SSI) as the drought index. The results obtained using the proposed method was compared with standard methods such as multilayer perceptron (MLP) and support vector regression (SVR). The proposed hybrid model showed improvement in prediction accuracy, especially for multi-step-ahead predictions. In the study of [21], they predicted the monthly solar radiation by hybridizing adaptive neuro-fuzzy inference system (ANFIS) with PSO, GA, and differential evolution (DE) algorithm using widely available meteorological data. After the simulation process, it was observed that the hybrid ANFIS-PSO model showed a better performance than the standalone ANFIS model, ANFIS-GA and ANFIS-DE algorithms. Meanwhile, [22] proposed a method for cancer mining using the hybrid particle swarm optimization (PSO) and genetic algorithm for gene selection and a fuzzy support vector machine algorithm to be used in prediction. A good result was depicted for the proposed prediction model with an accuracy of 100% using leukemia dataset, 96.67% for colon cancer dataset and 98% for breast cancer dataset. Further, [23] used three popular data mining algorithms namely Naïve Bayes, Radial Basis Function (RBF) Network, and J48 to develop a prediction model for breast cancer survivability. The paper of [3] integrated fuzzy C-Means (FCM), data envelopment analysis (DEA), and artificial neural network (ANN) in online prediction performance of companies in stock exchange. The study of [24] combined support vector regression (SVR) and hybrid neural network (HNN) to obtain earthquake prediction. In the study of [25], the tuned J48 was used along with Naïve Bayes, Bayes Net, Multilayer Perceptron, Support Vector Machine (SVM), REPTree and Random Forest in the analysis of performance prediction in EDM. Also, [26] developed a hybrid Support Vector Machine (SVM) combined with Firefly Algorithm (FFA) techniques (SVM-FFA) for prediction of field capacity and permanent wilting point of soil. The result showed that the combined SVM-FFA model has increased prediction accuracy way better than SVM and ANN models alone. Previously, [27] used decision table (DT)-based predictive models for breast cancer survivability and employed the under-sampling C5 technique and bagging algorithm to deal with the imbalanced problem, improving the predictive performance on breast cancer. In the study of [28], feature selection algorithms such as Information Gain Ratio (IGR) attribute evaluation, Correlation-based Feature Selection (CFS), Symmetrical Uncertainty (SU) and Particle Swarm Optimization (PSO) Algorithms were used to improve the performance accuracy of MLP, Simple Logistic, Rotation Forest, Random Forest and C4.5 prediction algorithms.

Further, [29] used a hybrid GA-BP neural network in predicting long-term skid resistance of epoxy asphalt mixture. The GA-BP model produced a great accuracy result when tested using the training set, validation set, and test set. Lastly, the study of [30] applied the genetic algorithm, Levenberg-Marquardt (LM) algorithm and backpropagation neural network in fault prediction of drying furnace equipment. The hybrid GA-LM-BP model showed an increased prediction accuracy compared to both BP neural network and GA-BP neural network.

B. Genetic Algorithm

The Genetic Algorithm (GA), among the many evolutionary algorithms anchored on the biological concept, is deliberately one of the commonly used methods for data reduction and feature selection. GA is one of the competent data reduction, feature selection, and global optimization algorithms used for those problems. It works by controlling some individuals by setting optimal operators on its selection, crossover, and mutation functions [31].

C. K-Nearest Neighbor Algorithm

The K-Nearest Neighbor (KNN) algorithm, introduced by Fix and Hodges, is another known data mining algorithm that adopts instance-based learning for prediction. This non-parametric classifier does not produce assumptions on the input data distribution making it renowned and used in various applications [32], [33]. The KNN is executed by assigning k values of the nearest neighbor of an instance and perform the calculation on the Euclidian distance. The K neighboring attributes that have the lowest Euclidian distance is chosen.

D. Naive Bayes (NB) Algorithm

The NB classifier is a probabilistic classifier based on Bayes' theorem. It has a stiff conditional independence assumption since attributes are fully independent [34].

The simplicity of its structure and robustness to noise and irrelevant attributes are some of the notable advantages of the Naive Bayes algorithm [35]. The formula below shows the Naive Bayes based on Bayes theorem:

$$P(A|B_i) = \frac{P(A)P(B_i|A)}{P(B_i)} \quad (1)$$

NB sets data into appropriate categories based on the highest posterior probability $P(A|B_i)$. Classification of data will commence once the posterior probability data A from condthe ition X_i $\{P(A|B_i)\}$ is smaller than posterior probability data A on $\{P(A|B_j)\}$ [36].

E. C4.5 Algorithm

According to [37], the C4.5 algorithm is a descendant to ID3 model which was created by J. Ross Quinlan and is also grounded on Hunt's algorithm. According to Salzberg, (1994), C4.5 is the most famous decision tree algorithm in machine learning. This claim was asserted to be true up to date by [39]. C4.5 uses both categorical and continuous attributes in building a decision tree (DT). Splitting the attribute values into two according to the identified threshold where all values greater than the threshold is considered as one child hence, the other shows how robust C4.5 in handling continuous attributes. This algorithm is also known to handle missing attribute values. In building the DT, C4.5 uses Gain Ratio as attribute selection measure by removing the biases of information gain if many outcome values of an attribute are identified. To perform, compute the gain ratio of each attribute first. Attributes whose gain ratio is at maximum will be identified as the root node of the tree. The algorithm uses a pessimistic pruning approach in removing unnecessary branches in the decision tree to increase the classification accuracy.

F. Rule Base (RB) Algorithm

Rule base algorithms work by discovering knowledge in the form of rules. The RB algorithms are most reasonable for investigating data with a blend of numerical and qualitative attributes. Some of the RB algorithms found in the study of [40] are the following:

Decision Table

Decision tables (DTs) uses a tabular representation for describing and analyzing situations. The decision, i.e., an action is taken depending upon the number of conditions and their inter-relationships [41].

JRip

Optimized version algorithm proposed by William W. Cohen. This algorithm will try to add every possible rule until it becomes accurate. It Optimizes rule set using discretion length [42].

One Rule (OneR)

It is the most straightforward learning algorithm for discrete attributes [43].

PART

Combines divide and conquer strategy. Incomplete C45 tree is built in each step and rule is build using the best leaf [44].

III. METHODOLOGY

A. Datasets

In this study, a total of 597 records of student-respondents in the evaluation of the faculty instructional performance from the four State Universities and Colleges (SUC) in Caraga Region, Philippines were used as the dataset. The thirty variables that represent the instructional performance of the faculty were reduced to obtain a maximized accuracy in the prediction. 70% of the data were used as the training set, and the remaining 30% was used for testing.

B. Modified Genetic Algorithm for Variable Reduction

In this study, the crossover which is one of the most important operators of GA that deals with the mating scheme, is modified. The genetic algorithm with the integration of the novel mating scheme called Inversed Bi-segmented Average Crossover (IBAX) is used in this study where its process is depicted in Fig. 1. This novel crossover is a modified version of the traditional average crossover of GA that is commonly used for real encoding problems using the arithmetic function. For the IBAX operator to be realized, the following steps must be executed:

Step 1: Take the parents from the selection pool.

Step 2: Count the number of genes found in the chromosomes. Identify if the dataset is in odd or even numbers.

Step 3: Segment the chromosomes (x and y) by dividing the total number of genes in the chromosomes into two and make sure that both first and second segments must contain an equal number of genes in an even count.

Step 4: On the first segment, create offspring Z for each gene by inversely pairing the first gene from chromosome X to the last gene on chromosome Y. Repeat until the last gene of the chromosome X and the first gene of the chromosome Y have inversely mated and have produced an offspring using the formula:

$$z = [x + y]/\,2 \quad (1)$$

Step 5: Execute the same process on the second segment until genes from all segments have produced offspring. . In the case of odd datasets, the last genes of the chromosomes will not be combined in the second segment and will automatically be mated with each other to produce offspring.

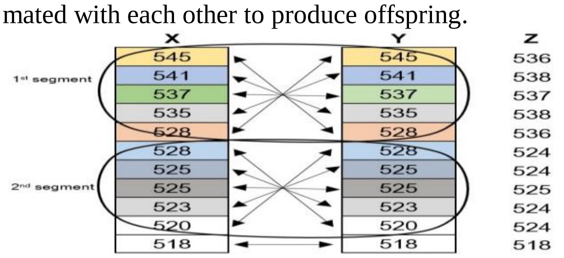

Fig. 1. Inversed Bi-Segmented Average Crossover with rank-based selection function

C. The Proposed Hybrid Prediction Models

This study proposed a hybrid MGA-KNN, MGA-NB, MGA-C4.5, and MGA-RB prediction models in the search for an optimal model in predicting instructional performance using the obtained datasets.

The genetic algorithm with a new crossover mating scheme is integrated prior to prediction using KNN, NB, C4.5, and RB algorithms. The variables used were minimized using the genetic algorithm with the proposed IBAX operator having the rank-based as the selection function. The variables left after the ten generations were used in the prediction algorithms. For simulation purposes, the Waikato Environment for Knowledge Analysis (WEKA) version 3.8.2 was instrumental in the prediction.

IV. RESULTS AND DISCUSSION

With the integration of the modified genetic algorithm, the prediction model using KNN, NB, C4.5 and DT algorithms has showed promising results in the prediction. The simulation result showed that among the proposed models, the optimal model for prediction is the MGA-KNN followed by the JRip which is under the umbrella of the MGA-RB with the corresponding prediction accuracy of 94.41% and 92.73%, respectively. The comparison of the results with the prediction accuracy, Precision, Recall, and F-Measures are depicted in Table 1.

Table 1. Indexed hybrid prediction results

Prediction Models		Accuracy %	Precision	Recall	F- Measure
mGA-KNN		94.4134%	0.944	0.944	0.943
mGA-NB		86.5922%	0.877	0.866	0.869
mGA-C4.5		89.3855%	0.893	0.894	0.894
mGA-RB	DT	85.4749%	0.864	0.855	0.857
	JRip	92.7374%	0.929	0.927	0.928
	OneR	75.9777%	0.773	0.760	0.765
	PART	92.1788%	0.923	0.922	0.920

Fig. 2 and 3 showed the graphical representation of the proposed hybrid prediction models with their corresponding percentage accuracy, precision, recall, and f-measure values.

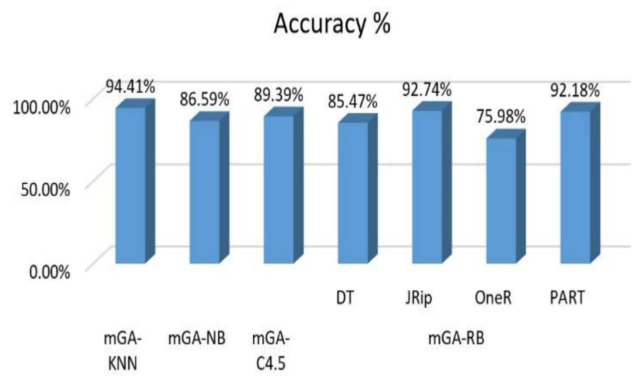

Fig. 2. Prediction model accuracy

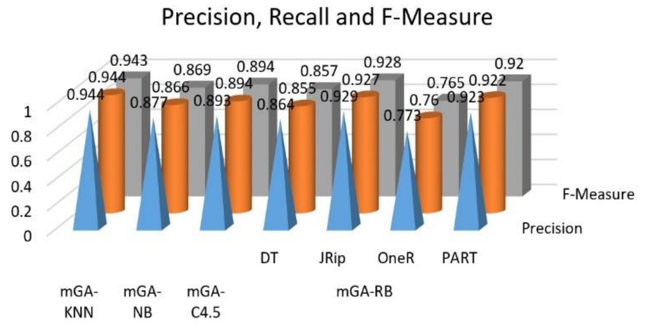

Fig. 3. Precision, Recall, and F-Measure of Prediction Models

V. CONCLUSION

Through the study, a novel approach for prediction was introduced and added to the body of knowledge. The proposed modified GA-based prediction models showed promising results, but the MGA-KNN was identified to be the optimal model for predicting the effectiveness of the instructional performance of faculty in the four SUCs in the Caraga Region, Philippines. The MGA-KNN prediction model obtained a 94.41% prediction accuracy having a k value of 1. This discovery showed parallelism on the study conducted by [15] in the quest to know the performance of KNN as prediction model, especially when compared to Naive Bayes, C4.5, and rule based algorithms. Among the four rule-based algorithms used in this study, the JRip obtained the highest accuracy and outperformed the other three RB algorithms with the prediction accuracy of 92.73%.

REFERENCES

1. S. Wahyuni, K. S. S, and M. Iswan, "The Implementation of Decision Tree Algorithm C4.5 Using Rapidminer in Analyzing Dropout Students," in *4th International Conference on Technical and Vocation Education and Training*, 2017.
2. K. S. Deepashri and A. Kamath, "Survey on Techniques of Data Mining and its Applications," *Int. J. Emerg. Res. Manag. &Technology*, 2017.
3. M. J. Rezaee, M. Jozmaleki, and M. Valipour, "Integrating dynamic fuzzy C-means , data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange," *Physica A*, vol. 489, pp. 78–93, 2018.
4. M. Zaffar, M. Ahmed, K. S. Savita, and S. Sajjad, "A Study of Feature Selection Algorithms for Predicting Students Academic Performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 541–549, 2018.
5. R. Ahuja, A. Jha, R. Maurya, and R. Srivastava, *Analysis of Educational Data Mining*, vol. 741. Springer Singapore, 2019.
6. U. O. Cagas, A. J. P. Delima, and T. L. Toledo, "PreFIC : Predictability of Faculty Instructional Performance through Hybrid Prediction Model," *Int. J. Innov. Technol. Explor. Eng.*, vol. 8, no. 7, pp. 22–25, 2019.
7. S. Prabakaran and S. Mitra, "Survey of Analysis of Crime Detection Techniques Using Data Mining and Machine Learning," *J. Phys. Conf. Ser.*, vol. 1000, no. 1, pp. 1–10, 2018.
8. P. Vrushali, M. Trupti, G. Pratiksha, and G. Arti, "Crime Rate Prediction using KNN," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 6, no. 1, pp. 124–127, 2018.
9. V. Ravi Jain, M. Gupta, and R. Mohan Singh, "Analysis and Prediction of Individual Stock Prices of Financial Sector Companies in NIFTY50," *Int. J. Inf. Eng. Electron. Bus.*, vol. 10, no. 2, pp. 33–41, 2018.
10. P. Carmona, F. Climent, and A. Momparler, "Predicting failure in the U.S. banking sector: An extreme gradient boosting approach," *Int. Rev. Econ. Financ.*, pp. 1–54, 2018.

11. A. Kiruthika, P. Deepika, S. Sasikala, and S. Saranya, "Predicting Ailment of Thyroid Using Classification and Recital Indicators," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 3, no. 3, pp. 1481–1485, 2018.
12. K. Lakshmi, D. I. Ahmed, and G. Siva Kumar, "A Smart Clinical Decision Support System to Predict diabetes Disease Using Classification Techniques," *2018 Ijsrset*, vol. 4, no. 1, 2018.
13. A. Chen and M. T. Leung, "Regression neural network for error correction in foreign exchange forecasting and trading," *Comput. Oper. Res.*, vol. 31, pp. 1049–1068, 2004.
14. A. A. P. Santos, N. C. A. da Costa, and L. dos S. Coelho, "Computational intelligence approaches and linear models in case studies of forecasting exchange rates," *Expert Syst. Appl.*, vol. 33, pp. 816–823, 2007.
15. J. Tarapitakwong, B. Chartrungruang, and N. Tantranont, "A Classification Model for Predicting Standard Levels of OTOP ' s Wood Handicraft Products by Using the K-Nearest Neighbor," *Int. J. Comput. Internet Manag.*, vol. 25, no. 2, pp. 135–141, 2017.
16. G. Pavai and T. V Geetha, "A Survey on Crossover Operators," *ACM Comput. Surv.*, vol. 49, no. 4, pp. 1–43, 2016.
17. D. J. Armaghan, M. Hasanipanah, and E. T. Mohamad, "A combination of the ICA - ANN model to predict air - overpressure resulting from blasting," *Eng. Comput.*, 2015.
18. B. A. Moghram, E. Nabil, and A. Badr, "Ab-Initio Conformational Epitope Structure Prediction Using Genetic Algorithm and SVM for Vaccine Design," *Comput. Methods Programs Biomed.*, 2017.
19. N. Suhermi, Suhartono, D. D. Prastyo, and B. Ali, "Roll motion prediction using a hybrid deep learning and ARIMA model," *Procedia Comput. Sci.*, vol. 144, pp. 251–258, 2018.
20. N. A. Agana and A. Homaifar, "EMD-Based Predictive Deep Belief Network for Time Series Prediction : An Application to Drought Forecasting," *Hydrology*, vol. 5, no. 18, 2018.
21. L. M. Halabi, S. Mekhilef, and M. Hossain, "Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation," *Appl. Energy*, vol. 213, no. November 2017, pp. 247–261, 2018.
22. N. Y. Moteghaed, K. Maghooli, and M. Garshasbi, "Improving Classification of Cancer and Mining Biomarkers from Gene Expression Profiles Using Hybrid Optimization Algorithms and Fuzzy Support Vector Machine Improving Classi fi cation of Cancer and Mining Biomarkers from Gene Expression Pro fi les Using H," *J. Med. Signals Sensors*, 2018.
23. [V. Chaurasia, S. Pal, and B. B. Tiwari, "Prediction of benign and malignant breast cancer using data mining techniques," *J. Algorithms Comput. Technol.*, vol. 12, no. 2, pp. 119–126, 2018.
24. K. M. Asim, A. Idris, T. Iqbal, and F. M. Alvarez, "Earthquake prediction model using support vector regressor and hybrid neural networks," *PLoS One*, pp. 1–22, 2018.
25. M. Anoopkumar and A. M. J. Z. Rahman, "Model of Tuned J48 Classification and Analysis of Performance Prediction in Educational Data Mining," *Int. J. Appl. Eng. Res.*, vol. 13, no. 20, pp. 14717–14727, 2018.
26. M. A. Ghorbani, S. Shamshirband, D. Z. Haghi, A. Azani, H. Bonakdari, and I. Ebtehaj, "Application of firefly algorithm-based support vector machines for prediction of field capacity and permanent wilting point," *Soil Tillage Res.*, vol. 172, no. September 2016, pp. 32–38, 2017.
27. L. Ya-qin, W. Cheng, and Z. Lu, "Decision tree based predictive models for breast cancer survivability on imbalanced data," in *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, 2009, pp. 1–4.
28. H. Almayan and W. Al Mayyan, "Improving accuracy of students' final grade prediction model using PSO," in *Proceedings of the 6th International Conference on Information Communication and Management, ICICM 2016*, 2016, pp. 35–39.
29. D. Zheng, Z. Qian, Y. Liu, and C. Liu, "Prediction and sensitivity analysis of long-term skid resistance of epoxy asphalt mixture based on GA-BP neural network," *Constr. Build. Mater.*, vol. 158, pp. 614–623, 2018.
30. W. Zhou, D. Liu, and T. Hong, "Application of GA-LM-BP Neural Network in Fault Prediction of Drying Furnace Equipment," in *MATEC Web of Conferences*, 2018, vol. 232, pp. 1–5.
31. A. B. Hassanat, V. B. S. Prasath, M. A. Abbadi, S. A. Abu-Qdari, and H. Faris, "An improved Genetic Algorithm with a new initialization mechanism based on Regression techniques," *Inf.*, vol. 9, no. 7, 2018.
32. A. Tharwat, H. Mahdi, M. Elhoseny, and A. E. Hassanien, "Recognizing human activity in mobile crowdsensing environment using optimized k-NN algorithm," *Expert Syst. Appl.*, vol. 107, pp. 32–44, 2018.
33. M. Huang, R. Lin, S. Huang, and T. Xing, "A novel approach for precipitation forecast via improved K-nearest neighbor algorithm," *Adv. Eng. Informatics*, vol. 33, pp. 89–95, 2017.
34. D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli, and I. O. Ellis, "A 'non-parametric' version of the naive Bayes classifier," *Knowledge-Based Syst.*, vol. 24, no. 6, pp. 775–784, 2011.
35. W. Chen, X. Yan, Z. Zhao, H. Hong, D. T. Bui, and B. Pradhan, "Spatial prediction of landslide susceptibility using data mining-based kernel logistic regression, naive Bayes and RBFNetwork models for the Long County area (China)," *Bull. Eng. Geol. Environ.*, pp. 1–20, 2018.
36. D. Stiawan, S. Sandra, E. Alzahrani, and R. Budiarto, "Comparative analysis of K-Means method and Naïve Bayes method for brute force attack visualization," *2017 2nd Int. Conf. Anti-Cyber Crimes, ICACC 2017*, pp. 177–182, 2017.
37. S. K. Yadav and S. Pal, "Data Mining : A Prediction for Performance Improvement of Engineering Students using Classification," *World Comput. Sci. Inf. Technol. J. WCSIT*, vol. 2, no. 2, pp. 51–56, 2012.
38. S. L. Salzberg, "Book Review: C4.5: by J. Ross Quinlan. Inc., 1993.," *Mach. Learn.*, vol. 16, pp. 235–240, 1994.
39. R. Benkercha and S. Moulahoum, "Fault detection and diagnosis based on C4 . 5 decision tree algorithm for grid connected PV system," *Sol. Energy*, vol. 173, no. April, pp. 610–634, 2018.
40. K. Elekar, M. M. Waghmare, and A. Priyadarshi, "Use of rule base data mining algorithm for Intrusion Detection," in *2015 International Conference on Pervasive Computing (ICPC)*, 2015.
41. R. Kohavi, "The Power of Decision Tables," in *8th European Conference on Machine Learning*, 1995, pp. 174–189.
42. W. Cohen, "Fast Effective Rule Induction," in *12th International Conference on Machine Learning*, 1995, pp. 115–123.
43. R. C. Holte, "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets," *Mach. Learn.*, vol. 11, no. 1, pp. 63–90, 1993.
44. E. Frank and H. Ian, "Generating Accurate Rule Sets Without Global Optimization," in *15th International Conference on Machine Learning*, 1998, pp. 144–151.

AUTHORS PROFILE

Allemar Jhone P. Delima was born at Surigao City, province of Surigao del Norte, Philippines on March 24, 1994. He completed master in information technology degree at Surigao State College of Technology, Surigao City, Philippines, last 2016. With his quest in improving his professional career, he took doctoral degree and is now a student of Doctor in information technology at Technological Institute of the Philippines – Quezon City, Philippines. He is currently connected in Surigao State College of Technology as a Faculty in the College of Engineering and Information Technology. His research interest is in data mining and data analytics.