Genetic programming in machine learning based on the evaluation of house affordability classification

Suraya Masrom[1], Norhayati Baharun[2], Nor Faezah Mohamad Razi[2], Abdullah Sani Abd Rahman[3], Nor Hazlina Mohammad[2], Nor Aslily Sarkam[2]

[1]Computer Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Perak Branch, Perak, Malaysia
[2]Mathematical Sciences Studies, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA Perak Branch, Perak, Malaysia
[3]Faculty of Sciences and Information Technology, Universiti Teknologi PETRONAS, Perak, Malaysia

Article Info

Article history:

Received Sep 22, 2023
Revised Feb 24, 2024
Accepted Mar 31, 2024

Keywords:

Crossover rate
Genetic programming
House affordability
Machine learning
Mutation rate
Population size

ABSTRACT

One of the big challenges in machine learning is difficulty of achieving high accuracy in a short completion time. A more difficulties appeared when the algorithm needs to be used for solving real dataset from the survey-based data collection. Imbalance dataset, insufficient strength of correlations, and outliers are common problems in real dataset. To accelerate the modelling processes, automated machine learning based on meta-heuristics optimization such as genetic programming (GP) has started to emerge and is gaining popularity. However, identifying the best hyper-parameters of the meta-heuristics' algorithm is the critical issue. This paper demonstrates the evaluation of GP hyper-parameters in modeling machine learning on house affordability dataset. The important hyper-parameters of GP are population size (PS), that has been observed with different setting in this research. The machine learning with GP was used to predict house affordability among employers with transport expenditure and job mobility as some of the attributes. The results from testing that run on hold-out samples show that GP machine learning can reach to 70% accuracy with split ratio 0.2 and GP PS 30. This research contributes to the advancement of automated machine learning techniques, offering potential for faster and more accurate real survey-based datasets.

Corresponding Author:

Norhayati Baharun
Mathematical Sciences Studies, College of Computing, Informatics and Mathematics
Universiti Teknologi MARA Perak Branch
35400 Tapah Road, Perak, Malaysia
Email: norha603@uitm.edu.my

1. INTRODUCTION

Machine learning has been so prevalent in various domains of real-world problems due to the evolution of industrial 4.0. The ever-increasing realm of machine learning has helped industries, businesses, government agencies, public, and private people in making fast decision for simple and complex problems. In medical [1], [2], education [3]–[6], agriculture [7], finance and economy [8], [9], building and property [10]–[12], as well as in engineering [13], the utilization of machine learning is highly substantial. As a result, critical demands are needed to simplify the implementation complexity of machine learning to be used by inexpert or inexperienced data scientists from various research fields. To introduce rapid tools for the novice machine learning users is highly significant.

Genetic programming (GP) [14] is a popular meta-heuristic [15] algorithm that can be used to automate some important tasks in machine learning. Most research on automated machine learning that used GP aim to automate features selection task. GP based feature selection was introduced in [16] for optimizing the selection of machine learning parameters when tested on the particular problems. Focus on high-dimensional data, Ma and Gao [14] presented multi-features construction based on multi-tree GP representation to resolve single-features GP representation. Multi-features construction provides multi-layers of features mimics to deep learning neural network to be utilized by machine learning. Concerning class imbalance issues that can create bias prediction in a classification model, Devarriya *et al.* [17] used entropy gain-based fitness function in GP that potentially deployed for machine learning classification problems. Santoso *et al.* [18] demonstrate the applicability of GP in solving binary classification problems, highlighting its effectiveness in comparison to traditional machine learning techniques. Regardless of the various applications of GP in machine learning, all related studies emphasize the advantageous impact GP has contributed to the tested domains.

To demonstrate the practicality of GP based machine learning, this research used a set of real datasets on cases of housing affordability among employees in Perak, Malaysia. Housing affordability has been a decisive issue for more than a decade because of the continuous growth of property prices. Housing prices in Malaysia for example are becoming severely unaffordable either in urban or rural areas. Significant growing work on housing affordability has been concerned on housing costs and household income. Recently, a new indicator of housing affordability which includes transportation expenditure has been widely discussed [19]–[21]. Housing and transportation are the two biggest expenses for most households. However, research by Dewita *et al.* [19] suggested that transportation costs when integrated with housing affordability measurements reveal a different pattern of affordability. Research to investigate how the combination of housing and transportation costs can impact housing affordability measurements in machine learning approach is very rare in the literature. The structural of housing is the main concern to be analyzed with machine learning house affordability models [22]. Based on the machine learning house affordability models, the projection of housing affordability can be strategized for the benefit of vulnerable tenants. Admitting that transport and job mobility are two critical problems in Malaysia, the development of effective and intelligent prediction model is highly significant. In the previous research on conventional machine learning for the same dataset of housing affordability, low accuracy and high classification error have been faced that enforced the researchers to extend the implementation with the GP machine learning to be presented in this paper.

To address the gap, this presents a more robust GP based machine learning is introduced. By using tree-based pipeline optimization tool (TPOT) Python [23], [24], automation of machine learning not just on the features selection but to the extent of optimizing the best machine learning algorithms with the suitable hyper-parameters setting. However, findings of research on the GP hyper-parameters in TPOT are not yet widely accessible in the literature, hence need more discovery.

Therefore, this paper introduces automated machine learning from Python libraries that only need a very short programming to be replicated by researchers. The contributions of this paper are two-folds. First, it presents simple and rapid implementation of machine learning. For inexpert data scientists, the easy framework can be directly replicated for their data domain. For the expert, the rapid framework can accelerate the process of machine learning modelling. The second contribution provides insight into the hyper-parameters impact on the GP based machine learning. As meta-heuristics, GP search optimization involves parameters that will influence the search direction to converge at the most optimal. Research to study the effects of GP hyper-parameters keep in progressing and deserves more intention due to the wider benefits of GP in many domains. As evolutionary based algorithm, the GP performances usually depending on the three important parameters namely population size (PS), mutation rate and crossover rate as well as the dataset used for the evaluation. As this is a new attempt for GP on the house affordability cases, focusing on the default setting of mutation and crossover rates but observing different PS will be a useful contribution for future research extension.

2. METHOD

2.1. The dataset

The dataset for evaluating the GP machine learning was the existing collected data that consists of 300 records of employees in Perak, Malaysia. The features can be divided into three groups; demography, transport expenditure, and employment status as given in Table 1. The DV is house affordability by means of the ability of the person to provide something for buying a new house not considering repayment of the existing house loans. It is denoted as true or false to indicate afford or not afford respectively.

Table 1. The house affordability classification model feature

Features group	Attribute
Demography	Age, gender, marital status, family size, lifestyle
Transport expenditure	Vehicle ownership, vehicle usage, public transit
Employment status	Position, company status, distance, job mobility

2.2. GP in machine learning

Figure 1 presents the implementation method of GP machine learning executed with Python Google Colab platform. Every experiment method in Figure 1 was repeated with different split ratio (0.2, 0.3,0.4) and each split ratio is repeated with 4 PS (10,20,30,40) and 3 validations from the 3 GP generations. Thus, the total experimental run is 12 and the total validations from GP is 3×12=36. The average of the tree validation was recorded in each of the 12 runs.

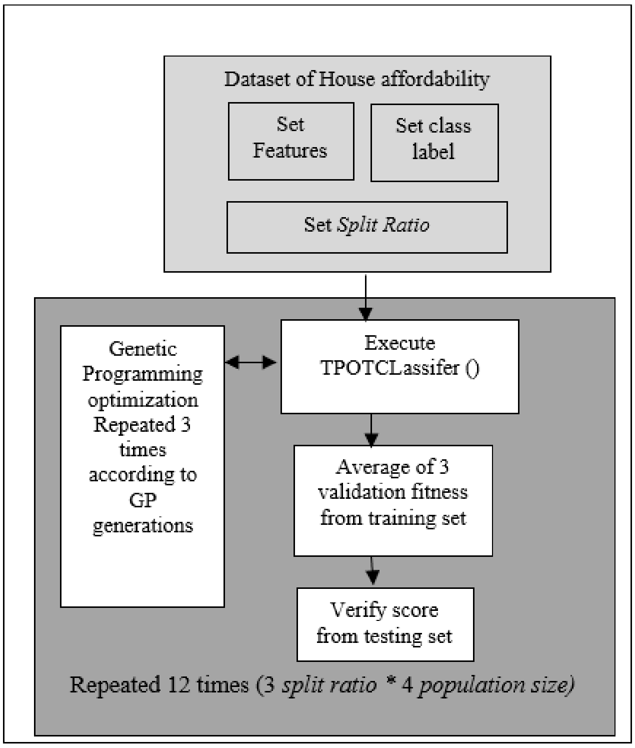

Figure 1. The experimental method

The total row of the house dataset is 300, which is divided into training and testing set as depicted in Figure 2. If the split ratio is 0.3, 210 out of 300 were deployed for training set, and the rest of 90 data was left for testing. For validation, TPOT used cross-validation to divide the training dataset into training and validation sets according to the number of k-folds. This research used 5 k-folds, which can be illustrated in Figure 2.

2.3. Performance metric

The validation and testing accuracy as well as area under curve (AUC) [25] were observed to access the performances of the machine learning with GP. AUC is more reliable than accuracy in the assessment of the classification model because AUC can depict the trade-offs between the model sensitivity and specificity. For this paper, only the best split size setting is measured with AUC to be discussed. The precision or true positive rate is the number of correct predictions for house affordability class from the total of real house affordability cases. On the other hand, recall or false positive rate is the number of incorrect house affordability classification from the total of house unaffordability cases. AUC can be observed from receiver operating characteristic (ROC) plot that can depict the curve to map the relationship between the precision and recall.

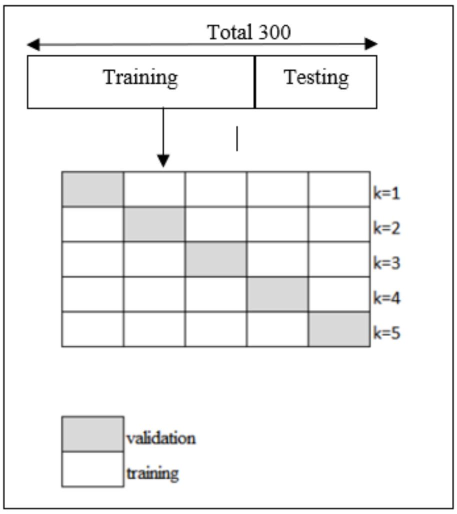

Figure 2. Training and testing dataset

3. RESULTS AND DISCUSSION

Figure 3 presents the output from a single run of GP machine learning with TPOT library in Python codes. The codes are very short to be replicated by researchers. The same codes were executed 12 times according to the 3 split ratios and 4 PSs. Split ratio 80:20 for training and testing can be set with test_size parameter as in Figure 3.

```
X_train = train_df.drop(["DV"], axis=1).values
Y_train = train_df["DV"]
Features = X_train
Class = Y_train

Feature_Train, Feature_Test, Class_Train, Class_Test  = train_test_split(X_train, Y_train, test_size=0.2)
tpot2 = TPOTClassifier(generations=3, population_size=20, mutation_rate=0.9,crossover_rate=0.1,verbosity=2, cv=5)
tpot2.fit(Feature_Train, Class_Train)
print(tpot2.score(Feature_Test, Class_Test))

Generation 1 - Current best internal CV score: 0.5407407407407409

Generation 2 - Current best internal CV score: 0.5518518518518518

Generation 3 - Current best internal CV score: 0.562962962962963

Best pipeline: GradientBoostingClassifier(input_matrix, learning_rate=1.0, max_depth=10, max_features=0.70000000000
0.7
/usr/local/lib/python3.8/dist-packages/sklearn/metrics/_scorer.py:794: FutureWarning: sklearn.metrics.SCORERS is de
  warnings.warn(
```

Figure 3. The Python codes with optimal setting and output

Furthermore, Figure 4 presents the comparison of testing accuracy between four PSs at different split ratios (0.10, 0.20,0.30,0.40,0.50). The testing accuracy score is the proportion of correct predictions made by the model on an unseen (hold-out) dataset that has never been exposed during the training process. When comparing the different PS in Figure 4, population size 20 (PS20) outperforms the others, including PS30, which indicates that larger PS does not significantly lead to better accuracy. This could be due to the fact that PS20 strikes a balance between exploration and exploitation in the genetic algorithm machine learning optimization process. Too large a population may lead to excessive computation time without significant gains in accuracy.

Regardless to split ratios, the graph in Figure 4 shows that as the split ratio decreases from 0.50 to 0.20, there is a general increase in accuracy. This suggests that the models benefit from a larger training dataset, which a lower split ratio provides, as it gives the model more data to learn from. It is interesting to note that when the split ratio is less than 0.2, the increase in accuracy becomes less pronounced. This could indicate that the improvement in accuracy becomes less pronounced when the split ratio drops below 0.2.

This shows that the models have reached a point where they have enough data to capture the underlying pattern and that additional data from the training set does not significantly change their performance. This finding could suggest that optimizing the split ratio around 0.2 may be a good stating point for GP machine learning optimization. Table 2 presents the accuracy results from 12 classification models across different PS and split ratios.

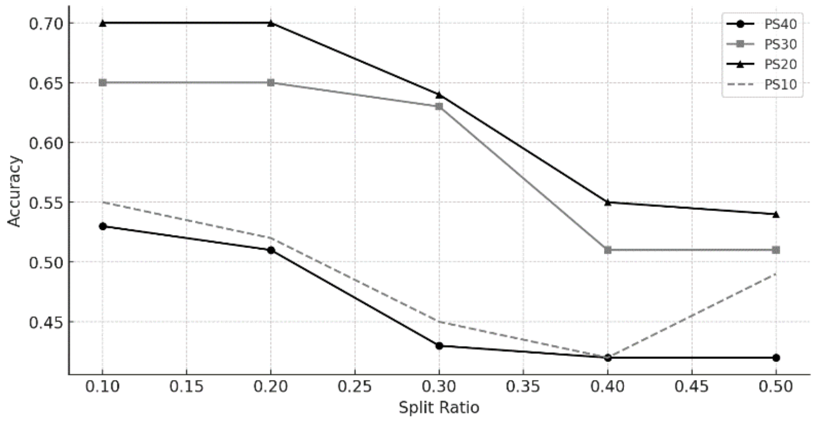

Figure 4. Comparison of testing accuracy from different PS

Table 2. Accuracy results from the 12 classification models

Population size	Validation of training set	Testing set
Split ratio (80:20)/*test_size 0.1*		
40	0.48	0.53
30	0.48	0.65
20	0.68	0.70
10	0.51	0.55
Split ratio (70:30)/test_size 0.2		
10	0.47	0.51
20	0.48	0.65
30	0.66	0.70
40	0.58	0.52
Split ratio (60:40)/test_size 0.3		
10	0.39	0.43
20	0.59	0.63
30	0.59	0.64
40	0.41	0.45
Split ratio (60:40)/test_size 0.3		
10	0.41	0.42
20	0.48	0.51
30	0.53	0.55
40	0.41	0.42

The data shows how accuracy varies with changes in the PS and the proportion of data allocated to training in comparison to testing. There is some level of consistency, particularly in the relationship between PS and model performance. There is a clear pattern in both training and testing results that the PS of 20 and 30 show higher performances. Although the consistency of these results varies with different split ratios, the superiority in terms of performance oscillates between PS 20 and 30 only, regardless of the specific split ratio used.

Furthermore, Figure 5 confirms that the PS of 20 and 30 not only offer robust performance but also the highest AUC values, emphasizing their efficacy in model prediction. As PS was set to 10, the AUC was 0.58, signaling a relatively low predictive performance. However, increasing the PS to 20 caused a significant improvement, with the AUC rising to 0.93, suggesting a substantial enhancement in predictive accuracy. Further increasing the PS to 30 continued to produce positive results, as the AUC value improved to 0.95, indicating a high level of predictive capability. Surprisingly, when the PS was increased to 40, the AUC dropped to 0.56, suggesting a decrease in predictive performance compared to the previous setting. These findings highlight the importance of choosing an appropriate PS in GP to achieve optimal predictive accuracy in modeling machine learning tasks.

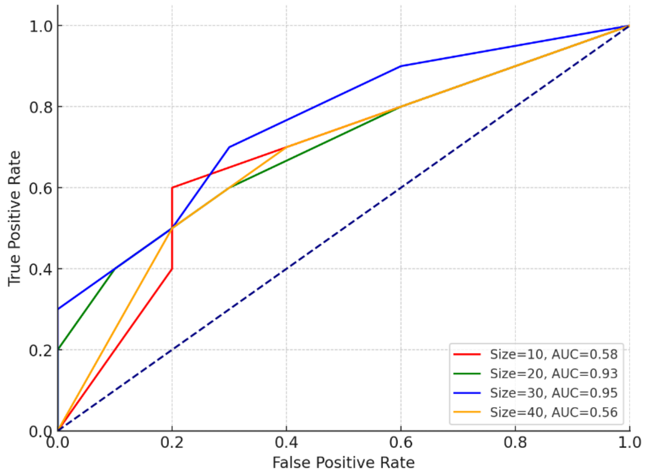

Figure 5. AUC-ROC at split ratio 0.2

4. CONCLUSION

This paper provides in-depth insight into the potential of machine learning with GP to be evaluated on the house affordability dataset. House affordability is one of the critical issues in many countries that contributed to major impact on socioeconomic and societal well-being. Research on identifying the effective and efficient approach for housing affordability is useful for the government and the property industry to strategize future planning. Due to problems of real datasets that hinder machine learning to generate high accuracy results, using GP is beneficial. The evaluation of hyper-parameters in this study highlights the critical challenge of identifying the optimal settings for GP in machine learning, particularly in the context of modeling house affordability, where a 70% accuracy was achieved with a PS of 30. This research contributes to the advancement of automated machine learning techniques using meta-heuristics optimization, offering the potential for faster and more accurate preliminary analysis of real survey-based datasets, such as those related to house affordability.

ACKNOWLEDGEMENTS

The authors would like to thank Universiti Teknologi MARA for the full support of this study.

REFERENCES

[1] A. Alamoudi, "Acceptance of artificial intelligence (AI) and machine learning (ML) among radiologists in Saudi Arabia," *International Journal of Advanced and Applied Sciences*, vol. 9, no. 1, pp. 154–157, Jan. 2022, doi: 10.21833/ijaas.2022.01.018.

[2] M. M. Ahsan and Z. Siddique, "Machine learning-based heart disease diagnosis: A systematic literature review," *Artificial Intelligence in Medicine*, vol. 128, p. 102289, Jun. 2022, doi: 10.1016/j.artmed.2022.102289.

[3] M. Arashpour *et al.*, "Predicting individual learning performance using machine-learning hybridized with the teaching-learning-based optimization," *Computer Applications in Engineering Education*, vol. 31, no. 1, pp. 83–99, Jan. 2023, doi: 10.1002/cae.22572.

[4] H. S. Alenezi and M. H. Faisal, "Utilizing crowdsourcing and machine learning in education: Literature review," *Education and Information Technologies*, vol. 25, no. 4, pp. 2971–2986, Jul. 2020, doi: 10.1007/s10639-020-10102-w.

[5] O. Chamorro-Atalaya *et al.*, "Supervised learning using support vector machine applied to sentiment analysis of teacher performance satisfaction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 516–524, Oct. 2022, doi: 10.11591/ijeecs.v28.i1.pp516-524.

[6] O. Chamorro-Atalaya *et al.*, "Text mining and sentiment analysis of teacher performance satisfaction in the virtual learning environment," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 525–534, 2022, doi: 10.11591/ijeecs.v28.i1.pp525-535.

[7] A. Nandy and P. K. Singh, "Application of fuzzy DEA and machine learning algorithms in efficiency estimation of paddy producers of rural Eastern India," *Benchmarking: An International Journal*, vol. 28, no. 1, pp. 229–248, Jan. 2021, doi: 10.1108/BIJ-01-2020-0012.

[8] A. Priyadarshini, S. Mishra, D. P. Mishra, S. R. Salkuti, and R. Mohanty, "Fraudulent credit card transaction detection using soft computing techniques," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1634–1642, Sep. 2021, doi: 10.11591/ijeecs.v23.i3.pp1634-1642.

[9] P. H. D. A. Samad, S. Mutalib, and S. Abdul-Rahman, "Analytics of stock market prices based on machine learning algorithms," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 2, pp. 1050–1058, Nov. 2019, doi: 10.11591/ijeecs.v16.i2.pp1050-1058.

[10] S. Jamil, T. Mohd, S. Masrom, and N. A. Rahim, "Machine learning price prediction on green building prices," in *2020 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, Jul. 2020, pp. 1–6, doi: 10.1109/ISIEA49364.2020.9188114.

[11] A. Baldominos, I. Blanco, A. Moreno, R. Iturrarte, Ó. Bernárdez, and C. Afonso, "Identifying real estate opportunities using machine learning," *Applied Sciences*, vol. 8, no. 11, p. 2321, Nov. 2018, doi: 10.3390/app8112321.

[12] N. Shinde and K. Gawande, "Valuation of house prices using predictive techniques," *International Journal of Advances in Electronics and Computer Science*, vol. 5, no. 6, pp. 34–40, 2018.

[13] M. B. Slimene and A. Aljaloud, "Neural network estimation of a photovoltaic system based on the MPPT controller," *International Journal of Advanced and Applied Sciences*, vol. 7, no. 2, pp. 85–90, Feb. 2020, doi: 10.21833/ijaas.2020.02.012.

[14] J. Ma and X. Gao, "Designing genetic programming classifiers with feature selection and feature construction," *Applied Soft Computing*, vol. 97, p. 106826, Dec. 2020, doi: 10.1016/j.asoc.2020.106826.

[15] S. K. Joshi and J. C. Bansal, "Parameter tuning for meta-heuristics," *Knowledge-Based Systems*, vol. 189, p. 105094, Feb. 2020, doi: 10.1016/j.knosys.2019.105094.

[16] J. Ma and X. Gao, "A filter-based feature construction and feature selection approach for classification using Genetic Programming," *Knowledge-Based Systems*, vol. 196, p. 105806, May 2020, doi: 10.1016/j.knosys.2020.105806.

[17] D. Devarriya, C. Gulati, V. Mansharamani, A. Sakalle, and A. Bhardwaj, "Unbalanced breast cancer data classification using novel fitness functions in genetic programming," *Expert Systems with Applications*, vol. 140, p. 112866, Feb. 2020, doi: 10.1016/j.eswa.2019.112866.

[18] L. W. Santoso, B. Singh, S. S. Rajest, R. Regin, and K. H. Kadhim, "A genetic programming approach to binary classification problem," *EAI Endorsed Transactions on Energy Web*, vol. 8, no. 31, p. e11, Jul. 2018, doi: 10.4108/eai.13-7-2018.165523.

[19] Y. Dewita, B. T. H. Yen, and M. Burke, "The effect of transport cost on housing affordability: Experiences from the Bandung Metropolitan Area, Indonesia," *Land Use Policy*, vol. 79, pp. 507–519, Dec. 2018, doi: 10.1016/j.landusepol.2018.08.043.

[20] M. A. Hassan, H. Hamdan, J. Abdullah, and Y. A. Abdullah, "Housing and transport expenditure: An assessment of location Housing affordability," *Planning Malaysia*, vol. 16, Sep. 2018, doi: 10.21837/pm.v16i6.465.

[21] M. E. A. Haffner and K. Hulse, "A fresh look at contemporary perspectives on urban housing affordability," *International Journal of Urban Sciences*, vol. 25, no. sup1, pp. 59–79, Jan. 2021, doi: 10.1080/12265934.2019.1687320.

[22] T. Ye *et al.*, "Using machine learning to help vulnerable tenants in New York City," in *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*, Jul. 2019, pp. 248–258, doi: 10.1145/3314344.3332484.

[23] R. S. Olson and J. H. Moore, "TPOT: A tree-based pipeline optimization tool for automating machine learning," in *Automated Machine Learning: Methods, Systems, Challenges*, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds. Cham: Springer International Publishing, 2019, pp. 151–160, doi: 10.1007/978-3-030-05318-5_8.

[24] T. Zhang, J. Zhang, G. Peng, and H. Wang, "Automated machine learning for steel production: A case study of TPOT for material mechanical property prediction," in *2022 IEEE International Conference on e-Business Engineering (ICEBE)*, Oct. 2022, pp. 94–99, doi: 10.1109/ICEBE55470.2022.00025.

[25] C. Halimu, A. Kasem, and S. H. S. Newaz, "Empirical comparison of area under ROC curve (AUC) and Mathew correlation coefficient (MCC) for evaluating machine learning algorithms on imbalanced datasets for binary classification," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, Jan. 2019, pp. 1–6, doi: 10.1145/3310986.3311023.

BIOGRAPHIES OF AUTHORS

Ts. Dr. Suraya Masrom received her Ph.D. in Information Technology and Quantitative Science from UiTM in 2015. She started her career in the information technology industry as an Associate Network Engineer at Ramgate Systems Sdn. Bhd (a subsidiary of DRB-HICOM) in June 1996 after receiving her bachelor's degree in computer science from Universiti Teknologi Malaysia (UTM) in Mac 1996. She started her career as a lecturer at UTM after receiving her master's degree in computer science from Universiti Putra Malaysia in 2001. She transferred to the Universiti Teknologi MARA (UiTM), Seri Iskandar, Perak, Malaysia, in 2004. She is an active researcher in the meta-heuristics search approach, machine learning, and educational technology. She can be contacted at email: suray078@uitm.edu.my.

Dr. Norhayati Baharun is an Associate Professor of Statistics, Universiti Teknologi MARA Perak Branch, Tapah Campus. She received her PhD in Statistics Education from the University of Wollongong Australia in 2012. Her career started as an academic from January 2000 to date at the Universiti Teknologi MARA that specialized in statistics. Other academic qualifications include both Master Degree and Bachelor Degree in Statistics from Universiti Sains Malaysia and Diploma in Statistics from Institute Teknologi MARA. Among her recent academic achievements include twelve on-going and completed research grants (local and international), four completed supervision of postgraduate studies, fifteen indexed journal publications, two academic and policy books, twenty-six refereed conference proceedings and book chapter publications, a recipient of 2013 UiTM Academic Award on Teaching, and fourteen innovation projects with two registered Intellectual Property Rights by RIBU, UiTM. She is also a certified Professional Technologist (Ts.) (Information & Computing Technology) of the Malaysia Board of Technologist (MBOT), a Fellow Member of the Royal Statistical Society (RSS), London, United Kingdom, a Professional Member of Association for Computing Machinery (ACM), New York, USA, and a Certified Neuro Linguistic Program (NLP) Coach of the Malaysia Neuro Linguistic Program Academy. Her research interests continue with current postgraduate students under her supervision in the area of decision science now expanding to a machine learning application. She can be contacted at email: norha603@uitm.edu.my.

Nor Faezah Mohamad Razi is a Senior lecturer in University Teknologi MARA (UiTM). She is the Resource Person for Fundamentals of Research. She received her BSc in Statistics and MSc in Applied Statistics from UiTM in 2009 and 2012 respectively. Currently she pursues her PhD study in Decision Science area. Her research interest in structural equation modeling (SEM), data envelopment analysis, and educational technology. She can be contacted at email: norfa122@uitm.edu.my.

Ts. Abdullah Sani Abd Rahman obtained his first degree in Informatique majoring in Industrial Systems from the University of La Rochelle, France in 1995. He received a master's degree from Universiti Putra Malaysia in Computer Science, with specialization in Distributed Computing. Currently, he is a lecturer at the Universiti Teknologi PETRONAS, Malaysia and a member of the Institute of Autonomous System at the same university. His research interests are cybersecurity, data analytics and machine learning. He is also a registered Professional Technologist. He can be contacted at email: sani.arahman@utp.edu.my.

Nor Hazlina Mohammad is a lecturer from School of Mathematical Sciences, College of Computing, Informatics and Mathematics, Universiti Teknologi MARA (UiTM), Perak Branch, Malaysia. She received her BSc in Statistics and MSc in Applied Statistics from UiTM in 2010 and 2013 respectively. In view of her past experience and contribution, she was teaching statistics more than 10 years and engaged with several researches of social sciences. She can be contacted at email: norha869@uitm.edu.my.

Dr. Nor Aslily Sarkam is a senior lecturer at the University Teknologi MARA (UiTM). She is the Coordinator Postgraduate of Colege of Computing, Informatics and Mathematics at Universiti Teknologi MARA (UiTM) Perak Branch. She received her Ph.D. in Management from the Tohoku University Japan in 2018. She started her career as a lecturer at UiTM Perak Branch in 2000 that specialized in Statistics. Other academic qualifications include both Master Degree and Bachelor's Degree in Statistics from Universiti Teknologi MARA. Among her recent achievements include on-going and complete research grants (leader and members), 17 indexed journal publications. Her research interest is in structural equation modeling (SEM) and educational technology. She can be contacted at email: noraslilysarkam@uitm.edu.my.