International Journal on Perceptive and Cognitive Computing (IJPCC) Vol 12, Issue 2 (2026)
https://doi.org/10.31436/ijpcc.v12i2.624

Analysing Educational Disparities Between Urban and Rural Areas in Malaysia

Muhammad Faiq Muhammad Fauzi[1], Muhammad Amir Farhan Muhammad Kamal[2], Norzariyah Yahya[3]*

[1,2]Dept. of Information Systems, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia
[3]Dept. of Computer Science, International Islamic University Malaysia, 53100 Kuala Lumpur, Malaysia

*Corresponding Author: norzariyah@iium.edu.my
(Received: 6th August 2025; Accepted: 3rd September, 2025; Published on-line: 30th July, 2026)

Abstract— Educational disparities between urban and rural areas remain a critical challenge for Malaysia's national development. This study investigates the factors driving rural–urban academic achievement gaps by combining non-parametric inferential statistics with supervised machine learning data-mining techniques. Primary data were collected from $N = 80$ respondents via a bilingual online survey capturing demographic attributes, learning environments, support systems, and material/ICT resource availability. Non-parametric hypothesis testing confirms a statistically significant difference in overall satisfaction with educational opportunities between independent urban and rural cohorts. To explain this divide, three machine learning models (J48, Random Forest, and Logistic Regression) were evaluated using 10-fold cross-validation in Weka to predict tertiary academic attainment. Random Forest achieved the highest classification performance (75.00% accuracy, $\kappa = 0.451$, ROC Area $= 0.794$). Information Gain feature evaluation reveals that age (0.279 bits) and family academic guidance (0.097 bits) are the dominant drivers of higher education attainment, whereas geographical location alone contributes minimally (0.007 bits). Furthermore, J48 decision tree logic indicates that for young rural learners with adequate home environments, access to ICT and study materials is the decisive factor in reaching tertiary education. These findings demonstrate that Malaysia's educational gap is primarily mediated by household resources and academic support rather than geography itself, suggesting that policy interventions should prioritise household-level digital access and family engagement programs over purely regional funding.

Keywords— Rural education, urban education, academic performance, socioeconomic status, machine learning classification, Weka.

I. Introduction

Educational gaps between urban and rural areas have long been a major concern for policymakers, educators, and scholars. These gaps usually appear as differential access to quality education, learning materials, and academic support, affecting students' academic performance and prospects. Even though equal educational access is one of the primary objectives of Malaysia's national education initiatives, students from rural areas often face systematic disadvantages. These include limited access to academic enrichment activities, fewer qualified educators, and poor school infrastructure.

Although several studies have addressed this topic, little is known about Malaysia's unique circumstances, which are influenced by its socioeconomic diversity, geographical features, and public education funding. This study aims to investigate how academic performance is impacted by region and pinpoint the fundamental causes of these disparities. This study provides valuable insights that help improve educational policies and ensure more equitable opportunities for students across the country.

II. Significance of Study

Understanding the factors contributing to educational disparities is crucial for informing policy development and implementing context-specific solutions. This research provides insights into the specific challenges faced by rural students and offers recommendations to policymakers and educators for bridging the academic gap. While the study provides valuable findings, its limitations, such as small sample size and self-reported data, should be addressed in future research.

III. Literature Review

Educational disparities between rural and urban areas have been extensively studied. According to Xiang and Stillwell [1], despite rising investment in rural education in China, significant gaps remain, particularly in high school education, which is not compulsory and receives less government support. Similarly, Ahsan et al. [2] stated that

https://doi.org/10.31436/ijpcc.v12i2.624

urban parents usually invest more in their children's education, leading to better academic achievement than in rural areas. The disparity often stems from differences in school quality, the availability of qualified teachers, and access to learning materials [3]. Wood [4] highlighted that urban students benefit from better facilities, qualified educators, and enriching learning environments, whereas rural students often confront poor infrastructure and limited educational opportunities. Drescher et al. [5] added that although overall differences between rural and nonrural students in the U.S. are modest, disparities become more pronounced among disadvantaged groups, with local conditions significantly influencing outcomes.

Socioeconomic status (SES) may or may not be a major determinant of academic success[3]. Rodríguez-Hernández et al. [6] conducted a systematic assessment of 42 studies and discovered a weak but positive link between socioeconomic status and academic success in higher education. Prior academic achievements, university experiences, and employment status proved to be stronger measures than SES alone. Munir et al. [3] revealed that students from families with higher socioeconomic status tend to perform better due to greater access to educational resources, parental support, and extracurricular activities. Research also shows that, while SES influences educational attainment, its effects can be mitigated by factors such as parental engagement and school quality. Ahsan et al. [2] discovered that intelligence is particularly crucial in rural areas, where school quality is generally lower than in other areas. Similarly, Xiang and Stillwell [1] proposed a capability-based framework for assessing rural-urban education gaps, emphasising the role of regional policies in shaping academic success.

Parental involvement has also been found to affect academic achievement across different socioeconomic groups significantly. A study by Şengönül [7] established that early parent-child interaction, reading activities, and school engagement positively influence performance in subjects like math and science. Furthermore, the paper highlights that families with higher SES are more likely to be actively involved in their children's education, reinforcing the importance of early engagement strategies.

Several studies suggested methods to address educational inequality. Wood [4] suggested increasing funding for rural education, improving infrastructure, and implementing e-learning alternatives. Xiang and Stillwell [1] emphasised the need for targeted policies, advocating for more context-sensitive policies that are adapted to different developmental levels. Furthermore, Ahsan et al. [2] found that government investment in rural school quality could enhance both equity and efficiency in education. Kawuryan et al. [8] supported this by highlighting the importance of teacher quality, pointing out that a lack of qualifications and certifications among educators in Indonesia affects student performance and reinforces inequality.

International perspectives on urban-rural educational divides show similar patterns. In the Netherlands, van Maarseveen [9] found that children in urban areas achieve higher levels of education than their rural peers, even when controlling for cognitive ability and family background. Higher expected returns, lower perceived costs, and access to better information drive this urban advantage. Likewise, Lu [10] analysed urban and rural schools in Yuyao, China, and found that urban schools are generally better resourced, have more qualified teachers, and offer a broader range of extracurricular and technological opportunities. Lu said that to close the educational gap, there needs to be more investment in rural areas, better training for teachers, and new ways of running schools.

IV. Methodology

This study adopts a computational data-mining approach to explore the educational disparities between rural and urban areas in Malaysia. The methodology is structured as a four-stage pipeline, consisting of (A) Data collection, (B) data pre-processing and encoding, (C) descriptive and inferential statistical analysis, and (D) supervised classification and attribute-importance analysis using the Weka [11] data-mining workbench. Below is a description of each stage.

A. Data Collection

A structured questionnaire was developed to collect data on five construct groups related to academic achievement, demographic characteristics (age, gender, area of residence, highest level of education), learning environment (home, school), social influence (peer influence), family support (emotional, financial, academic guidance) and availability of learning resources (physical, digital study materials). The survey was written in both English and Malay to make it accessible to respondents.

The questionnaire was distributed through a cloud-hosted Google Forms instrument and through two popular social messaging platforms (Telegram and WhatsApp). 250 invitations were issued, and 80 valid responses were received over two weeks (response rate 32%). Google Forms automatically records submissions into a spreadsheet, automating the entire ingestion path from respondent to analysable dataset, with no manual transcription or paper-based data entry involved. Then, the raw response spreadsheet was exported and converted into the Attribute-Relation File Format (ARFF) as required by the Weka workbench.

https://doi.org/10.31436/ijpcc.v12i2.624

B. Data Preprocessing and Encoding

Data preprocessing was carried out in Weka's Preprocess module. Nominal attributes were declared as categorical items, including age bracket, gender, area, and highest level of study. The opinion items (learning environment, family support, peer influence and access to study materials) on a Likert scale were declared as numeric attributes on their native 1–5 scale. In accordance with the educational-inequality literature review, the last attribute, 'highest level of study', was recoded into a two-class target, i.e., tertiary (Degree, Master or PhD) versus non-tertiary (SPM, Foundation, Diploma or below). The access to study materials, which explicitly includes online resources, practice questions, videos, and physical materials, is treated as an ICT access indicator at the household level. The final analysable dataset consists of 80 instances, 10 predictor attributes and a binary class attribute.

C. Descriptive and Inferential Analysis

Each attribute was characterised by calculating measures of central tendency and dispersion (mean, median, mode, variance and standard deviation). Chebyshev's theorem $\left(1-\frac{1}{k^2}\right)$ was used to generalise the distributions of continuous attributes. The significance of the raw urban-rural gap in educational satisfaction was tested using the Wilcoxon signed-rank test, a non-parametric test appropriate for ordinal Likert data. The outputs show an urban-rural difference but do not identify the underlying factors that drive it. This step motivates the subsequent data-mining stage.

D. Supervised Classification and Attribute-Importance Analysis

The Weka Classify module was used to train and compare three supervised classifiers to determine the factors predicting academic achievement and their relative importance: J48 (Weka implementation of C4.5 decision-tree algorithm), RandomForest (ensemble of 100 unpruned trees) and Logistic (multinomial logistic regression with ridge regularisation). These three algorithms were selected to cover the interpretability and accuracy range: J48 produces a human-readable tree, Logistic provides signed coefficient weights, and RandomForest typically achieves higher predictive accuracy at the cost of a less transparent decision surface.

All classifiers were evaluated using 10-fold cross-validation, the default protocol in Weka, and were reported based on classification accuracy, Cohen's Kappa, the confusion matrix, and the per-class precision, recall, and F-measure. Weka's Select Attributes module was used with the InfoGainAttributeEval evaluator and a Ranker search over the full training set to rank attribute importance, which is used to measure how much information each attribute contributes to predicting the class label.

The descriptive-inferential stage (C) and the data-mining stage (D) together establish a two-layer analytical strategy, of which the first identifies the existence of the urban–rural gap, and the following stage explains its mechanism by ranking the predictors that mediate it.

V. FINDINGS AND DISCUSSIONS

A. Descriptive Analysis:

The pie chart in Fig. 1 shows a significant age distribution, with the largest age group being 18-24 years old, accounting for 41.3% of the 80 respondents. This group is likely young adults, possibly students or early-career individuals. The second largest age group is 45-54 years old, making up 35% of respondents. This group likely represents mature adults, parents, or working professionals, providing valuable perspectives on educational experiences. The 25-34 age range, comprising 11.3% of respondents, may represent young professionals or postgraduate students. The smallest proportions are under 18, 45-54, and above 54. This distribution suggests that the data primarily reflects the views and experiences of the younger to middle-aged demographic, which enhances the credibility of the findings on education access and quality.

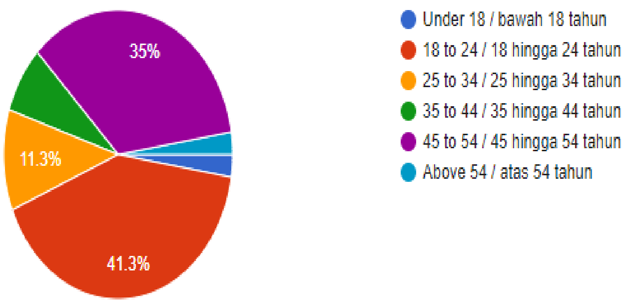

Fig. 1 Pie chart showing the percentage of respondents and their age group

B. Preliminary Analysis

The bar chart in Fig.2 shows the distribution of answers about opinions about the learning environment in urban areas. According to the findings, a sizable majority of respondents have a favourable opinion of the urban learning environment. 34.07% of participants agree, 52.41% strongly agree, and 86.48% have a positive attitude. On the other hand, discontent is minimal, since only 2.16% strongly disagree and another 2.16% disagree. Conversely, 9.2% of those surveyed expressed no opinion. The findings imply that most people view the urban learning environment as helpful and productive.

https://doi.org/10.31436/ijpcc.v12i2.624

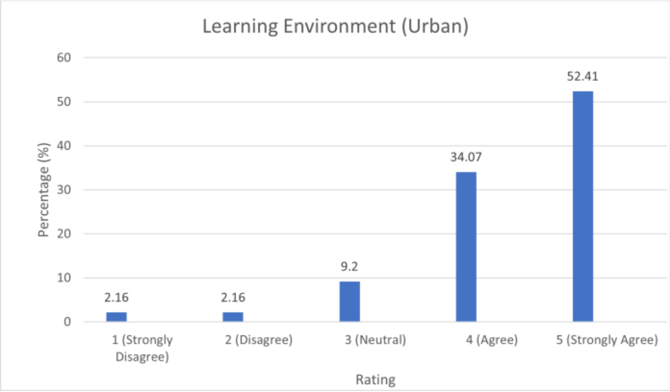

Fig. 2 Bar chart showing the percentage of respondents ' opinions about the learning environment in urban areas

The bar chart in Fig.3 shows data on how people in urban regions feel about the assistance they get from their family. With 44.62% of respondents strongly agreeing and 36.56% agreeing that they receive family support, the data demonstrate a very positive perception. As a result, 81.18% of the answers were positive overall. Just 5.92% of people disagreed, while 12.9% of participants were neutral. Interestingly, no responders expressed strong disagreement. According to these results, most people in metropolitan areas feel supported by their families, which is indicative of a supportive and upbeat family atmosphere.

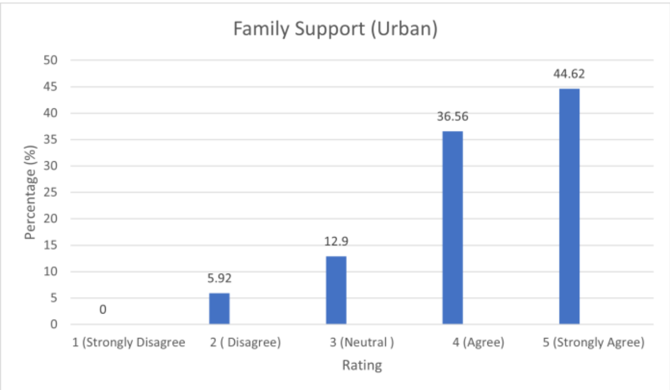

Fig. 3 Bar chart showing the percentage of respondents' opinions about family support in urban areas

The bar chart in Fig. 4 shows how people in cities rate the availability of study materials on a 5-point scale. With 35.48% of respondents rating their access as good and 37.1% as very good, the data shows a generally favourable perspective. This suggests that a sizable majority (more than 72%) think their access to study materials is beneficial. In contrast, only 3.23% of respondents assessed the access as poor, while 24.19% said it was simply adequate. Interestingly, all respondents said they had at least fair access. According to the data, urban students typically have access to study materials that range from adequate to exceptional.

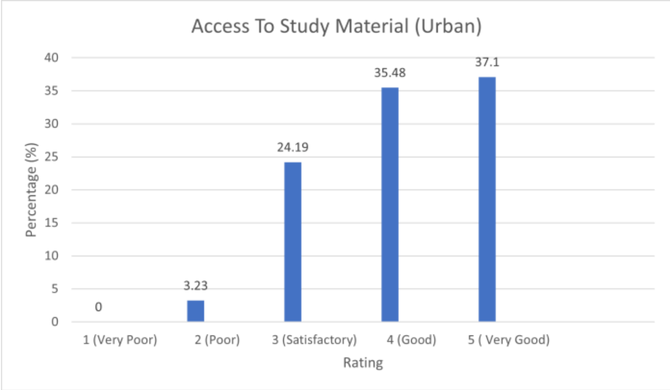

Fig. 4 Bar chart showing the percentage of respondent about opinions about access to study materials in urban areas

The bar chart in Fig. 5 displays the distribution of answers on a five-point Likert scale about the quality of the learning environment in rural areas. There were 3.67% of respondents who disagreed and 3.67% who strongly disagreed with the positive opinion. 18.52%, a higher percentage, stayed neutral. Notably, 33.37% of respondents agreed that the learning environment was favourable, while 52.41% strongly agreed. More than 85% of participants in rural areas expressed agreement or strong agreement, indicating that most of them had a positive perception of their learning environment.

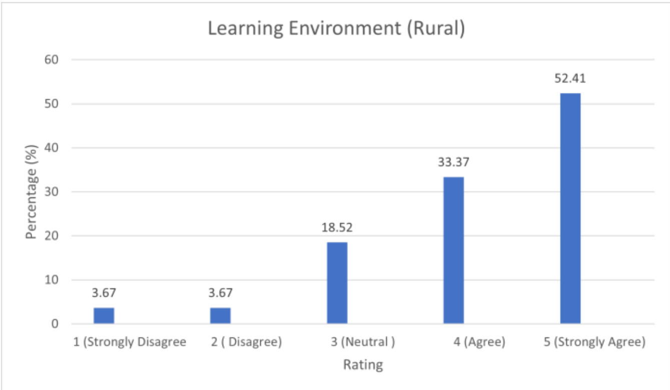

Fig. 5 Bar chart showing the percentage of respondents about their opinions on the learning environment in rural areas

The bar chart in Fig. 6 shows the percentage distribution of rural respondents' opinions about family support. Most individuals agreed that family support is important; only 3.72% disagreed, and none strongly disagreed. Conversely, 22.22% of those surveyed took a neutral position. A large majority, 27.78%, agreed that having strong family support is important, and an even larger majority, 44.62%, strongly agreed. Overall, more than 72% of respondents said they agreed or strongly agreed, indicating that rural participants have a very positive opinion of family support.

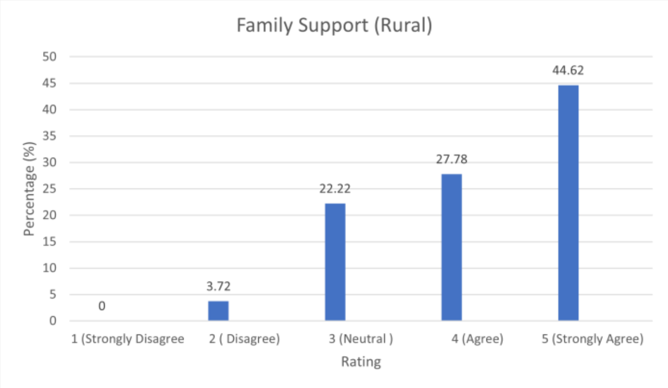

Fig. 6 Bar chart showing the percentage of the respondents' opinions on family support in rural areas

The bar chart in Fig. 7 displays the distribution of responses on a five-point rating system for the availability of study materials among participants in rural areas. A few respondents (5.56%) assessed access as poor, and others ranked it as low or very low. A sizable portion (33.33%) evaluated access as satisfactory, suggesting considerable satisfaction. 22.22% assessed access as very good, while the largest group, 38.29%, ranked it as good. In comparison to other aspects like family support or the learning environment, the majority of respondents, more than 60% perceived their access to study materials favourably overall, albeit a sizable portion still believes there is space for improvement.

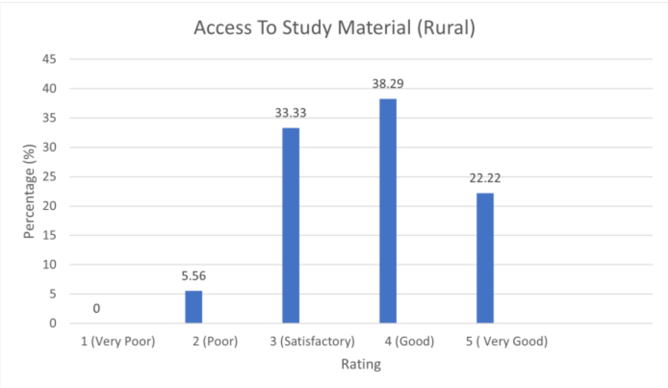

Fig. 7 Bar chart showing the percentage of respondents about their opinions on access to study material in rural areas

B. *Academic Achievement*

The pie chart in Fig. 8 shows a diverse range of educational achievement among respondents. The largest group, 38.8% out of 80 respondents, holds a bachelor's degree or equivalent, providing informed insights on academic systems and challenges. The second-largest segment, 22.5%, has a Master's degree. Other notable groups include 15% with a Diploma/IB/A-level or equivalent, 10% with SPM/IGCSE/O-level, and 8.8% who have completed Foundation/STPM/Matriculation. The data suggests that the majority of respondents are well-educated, potentially influencing the survey's perspective on education quality, access, and disparity.

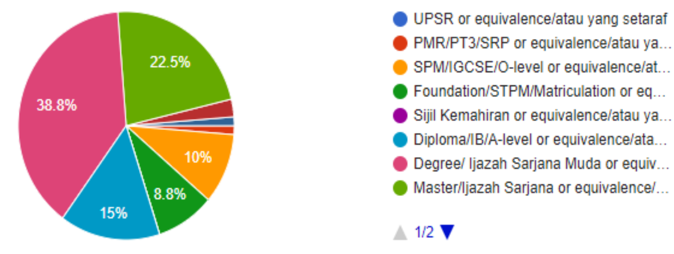

Fig. 8 Pie chart showing the percentage of respondents and their academic achievement

Mode: Based on the data, about 38.8% of the respondents currently have a degree. This figure shows the mode of the data out of 80 respondents; 31 graduated with at least a degree.

Median (MD): Since n = 80, which is even, the median lies between the (n/2)th and (n/2+1)th terms, i.e., the 40th and 41st values. Both fall in the Degree column, so the median is 'Degree'.

C. *Final Analysis*

The data collected show that all data are interconnected with one another. Central tendencies such as mean, median, mode, variance, and standard deviation are measured in this study. In addition, the use of histogram graphs helps provide precise observations of the correlations between each piece of data. Chebyshev's theorem is also implemented to validate and confirm the graphs created. It is noted that all responses to the online survey distributed are based on their life experience as students and workers. The relationship between dependent and independent variables must be identified to draw conclusions from the calculated data description. In this section, we will relate the data descriptions of independent variables with their respective dependent variables. The dependent and independent variables are the level of education and educational opportunities.

Step 1: Stating the hypotheses.

H_0: There is no significant difference in the average satisfaction with educational opportunities between rural and urban areas. H_1: There is a significant difference in average satisfaction with educational opportunities between rural and urban areas.

Step 2: Finding the critical value. Since α = 0.05 and this test is a two-tailed test, use the z-values of −1.96 and 1.96.

Step 3: Finding the test value. Data from the two samples (urban and rural) are integrated and ranked. Positive Rank Sum = 1 + 2 + 3 + 5 + 6 + 7 = 24. Negative Rank Sum = −4. Wilcoxon Signed-Rank statistic W = 4.

https://doi.org/10.31436/ijpcc.v12i2.624

Step 4: Making a decision. Since W = 4 > critical value = 2, we reject H_0.

Step 5: Summarise the result. There is enough evidence at α = 0.05 to support the claim that there is a significant difference in average satisfaction with educational opportunities between rural and urban areas.

VI. Predictive Modelling and Feature Importance Analysis

A. Motivation and Data-Mining Pipeline

To understand more than the descriptive analysis and to understand the relative influence of different factors on academic achievement, a supervised data-mining approach was employed using Weka 3.8 [11]. Weka is a widely used open-source machine-learning platform. The analysis followed four main stages. First, responses collected through Google Forms were automatically exported in .xlsx format and converted to Weka's Attribute-Relation File Format (ARFF). Second, the dataset was prepared by defining and cleaning the attributes using Weka's Preprocess module. Third, three classification models were developed and evaluated through cross-validation using the Classify module. Finally, the importance of each attribute was assessed using the Select Attributes module, while Weka's visualisation tools were used to examine confusion matrices, receiver operating characteristic (ROC) curves, and decision-tree structures. To support transparency and reproducibility, the ARFF dataset and the corresponding Weka experiment settings are provided in the Data Availability Statement as the software artefacts associated with this study.

B. Target Variable and Features

The prediction target was defined as a binary measure of tertiary-level academic attainment, where respondents with a Degree, Master's, or PhD qualification were classified as *tertiary*, while those with SPM, Foundation, Diploma, or lower qualifications were classified as *non-tertiary*. This binary classification is widely adopted in educational inequality research. A total of ten predictor variables were included in the analysis, comprising three nominal variables (age group, gender, and urban or rural residence) and seven numerical variables measured using a five-point Likert scale. These numerical variables represented the home learning environment, school learning environment, peer influence, family emotional support, family financial support (used as a proxy for socioeconomic status), academic guidance from family, and access to study materials. The final predictor was considered an indicator of ICT access, as the corresponding survey item captured respondents' access to both digital and physical learning resources, including textbooks, lecture notes, online materials, practice questions, and educational videos.

C. Classifiers and Evaluation Protocol

Three algorithms were compared, corresponding to the reviewers' explicit suggestions: J48 (Weka's implementation of the C4.5 decision-tree algorithm, with confidence-factor pruning at 0.25), Random Forest (an ensemble of 100 unpruned trees with default bagging), and Logistic (multinomial logistic regression with ridge regularisation). All models were evaluated using 10-fold cross-validation. Reported metrics include overall classification accuracy, Cohen's Kappa, per-class precision, recall, F-measure, and the confusion matrix.

D. Results

Table I summarises the cross-validated performance of the three classifiers.

TABLE I

Cross-Validated Performance of the Three Classifiers

Classifier	Accuracy	Kappa (approx.)
J48 (Decision Tree)	71.25%	0.3826
Random Forest	75%	0.4509
Function Logistic	72.5%	0.396

RandomForest was the most accurate of the three models tested. RandomForest correctly classified 60 of the 80 instances using 10-fold stratified cross-validation, resulting in an overall accuracy of 75.0%, a Cohen's Kappa statistic of 0.451 (indicating moderate agreement beyond chance), and an area under the ROC curve of 0.794. The weighted average F-measure for the two classes was 0.748.

The confusion matrix as shown in Fig. 9 indicates that 42 of the 51 tertiary respondents were correctly identified (recall = 0.824, precision = 0.792, F = 0.808), while 18 of the 29 non-tertiary respondents were correctly identified (recall = 0.621, precision = 0.667, F = 0.643). The model performs more confidently on the majority class, as expected given the imbalance between tertiary (n = 51) and non-tertiary (n = 29) respondents. The ROC curve for the tertiary class as in Fig. 10 confirms the moderate discrimination ability with AUC = 0.7945, well above the 0.5 random-classifier baseline.

```
=== Confusion Matrix ===

  a  b   <-- classified as
 42  9 |  a = tertiary
 11 18 |  b = non_tertiary
```

Fig. 9 RandomForest confusion matrix

https://doi.org/10.31436/ijpcc.v12i2.624

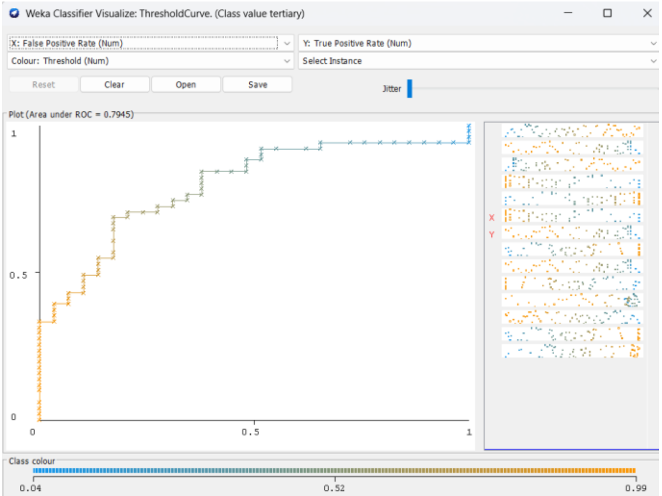

Fig. 10 ROC Curve

```
=== Run information ===

Evaluator:    weka.attributeSelection.InfoGainAttributeEval
Search:       weka.attributeSelection.Ranker -T -1.7976931348623157E308 -N -1
Relation:     education_disparity
Instances:    80
Attributes:   11
              age
              gender
              area
              home_env
              school_env
              peer_influence
              emotional_support
              financial_support_SES
              academic_guidance
              ict_access_study_materials
              attainment
Evaluation mode:    evaluate on all training data



=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 11 attainment):
        Information Gain Ranking Filter

Ranked attributes:
 0.27858    1 age
 0.09654    9 academic_guidance
 0.05677    7 emotional_support
 0.0067     3 area
 0.00136    2 gender
 0          4 home_env
 0         10 ict_access_study_materials
 0          6 peer_influence
 0          8 financial_support_SES
 0          5 school_env

Selected attributes: 1,9,7,3,2,4,10,6,8,5 : 10
```

Fig. 11 Attribute-Importance Ranking

Weka's InfoGainAttributeEval with a Ranker search as illustrates in Fig. 11 produced the following information-gain values (in bits) for each predictor: age = 0.279, academic_guidance = 0.097, emotional_support = 0.057, area = 0.007, gender = 0.001, with the remaining Likert-scale predictors (home_env, ict_access_study_materials, peer_influence, financial_support_SES, and school_env) registering zero information gain under the default evaluation. Age is by far the dominant predictor, contributing nearly three times the information of academic_guidance, the next-strongest variable. Critically, urban-versus-rural area contributes only 0.007 bits of information, which was roughly 2.5% of what age contributes and less than one-tenth of academic_guidance. This finding refines and clarifies the descriptive gap reported in Section V: while the Wilcoxon Signed-Rank Test confirms that a raw urban–rural difference exists, the InfoGain analysis demonstrates that the difference is not geographic. Rather, the urban–rural gap operates almost entirely through mediating variables, which are age (which reflects accumulated years of educational opportunity) and family-level academic and emotional support.

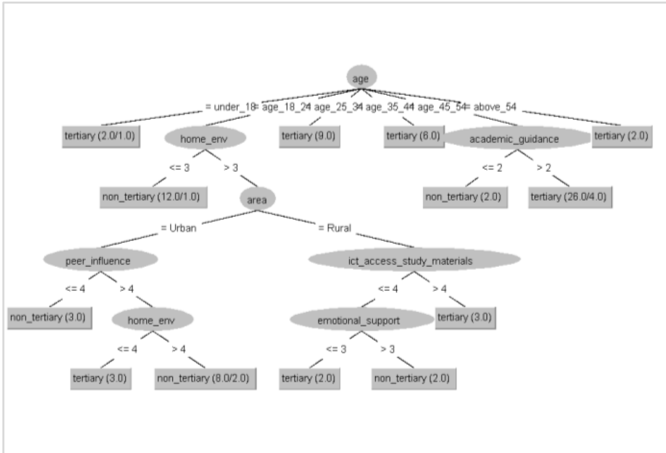

Fig. 12 J48 Decision Tree

Fig. 12 presents the J48 decision tree, which exposes the model's decision logic in human-readable form and enriches the InfoGain ranking. The root split is on age: respondents aged 25–34 and 35–44 are classified directly as tertiary attainers, capturing 15 instances at high purity. For the 45–54 bracket, the tree splits on academic_guidance, with strong family academic guidance (> 2) predicting tertiary attainment for 22 of 26 respondents (84.6% purity). The most analytically interesting subtree is the 18–24 branch, which first splits on home_env: respondents with weak home learning environments (≤ 3) are classified as non-tertiary with 92% purity. For 18–24 respondents with adequate home environments, the tree then splits on area, and within each area applies a different mediating variable. For urban respondents in this subtree, peer_influence and home environment govern the classification. For rural respondents, however, the decisive variable is ict_access_study_materials: rural young adults with strong access to study materials (> 4) are classified as tertiary attainers, while those with weaker access face a further split on emotional support. This is a substantively important

https://doi.org/10.31436/ijpcc.v12i2.624

pattern for the paper's central question: for young rural respondents with adequate home learning conditions, access to ICT and study materials becomes the decisive predictor of whether they reach tertiary attainment. This directly operationalises the digital-divide dimension of educational inequality that motivated this study's ICT framing.

The combined findings from the InfoGain ranking and the J48 decision tree provide a clearer understanding of how demographic, family, and ICT-related factors influence academic attainment. Although age recorded the highest information gain (0.279 bits), it is not a factor that can be influenced through policy or intervention. Instead, age likely reflects the cumulative educational opportunities available to respondents over time. When focusing on factors that are more amenable to intervention, academic guidance from family (0.097 bits) emerged as the most influential predictor, followed by family emotional support (0.057 bits). Both variables highlight the important role of family involvement in supporting educational achievement. By comparison, area of residence contributed only 0.007 bits, making it substantially less informative than family-related factors and far less influential than age.

The J48 decision tree further supports these findings. After the initial split on age, subsequent branches were primarily determined by family-related and household characteristics rather than by urban or rural residence. Area of residence appeared only as a secondary decision point within the 18–24 age group and was further divided based on access to study materials, indicating that ICT and learning-resource accessibility had a greater influence than location itself. Overall, these results suggest that while age is the strongest predictor, it mainly represents differences in educational exposure rather than a direct causal factor. More importantly, the findings indicate that disparities in tertiary educational attainment between urban and rural populations are better explained by differences in family support and access to learning resources, including ICT, than by geographical location alone.

E. Implications for the Research Question

This predictive-modelling layer strengthens this study's contribution in three ways that are directly tied to the research question of what drives educational disparity between urban and rural Malaysia.

Firstly, it confirms and refines the Wilcoxon-based finding of urban-rural difference reported in Section V by showing that the difference is explained, rather than simply observed: the InfoGain analysis indicated that geographic location itself contributes only 0.007 bits of information, less than one-tenth of academic guidance and less than one-thirtieth of age. The urban–rural gap exists, but it operates through mediating variables, including family academic guidance, home learning environment, family emotional support, and access to study materials, rather than as an independent effect of geographical location.

Secondly, it operationalises ICT access via the study-materials indicator as a measurable predictor of academic achievement. The J48 decision tree indicates that ICT access becomes the decisive splitting variable for young rural respondents with adequate home learning conditions (Fig. 12). This provides direct model-based evidence for the education disparity dimension of the digital divide.

Lastly, the Weka pipeline itself, which includes reproducible ingestion, attribute definition, classifier training and computational visualisation, is a lightweight information-systems tool that policymakers, NGOs, or state education departments can re-run on future survey waves to track the gap over time.

Together, these findings support a shift in policy focus from geography-based interventions toward household-level supports and targeted ICT-access programmes, particularly for rural respondents in the youngest age bracket, where the tree analysis showed ICT access has the strongest discriminating effect.

Conclusion

The study investigates educational disparities between rural and urban Malaysian communities, with a focus on geographic location, socioeconomic background, family support, and access to educational resources. Urban students achieve better academic outcomes, largely due to superior facilities, experienced teachers, and more structured learning environments. As shown in Fig. 2, 86.48% of urban respondents agreed that the learning environment significantly influences academic performance. Family support is another key factor, with 81.18% of urban participants affirming its positive impact, as illustrated in Fig. 3. Additionally, 72.58% of respondents indicated that access to study materials is better in urban areas, as depicted in Fig. 4. In comparison, rural respondents showed slightly lower but still substantial levels of agreement: 85.78% for learning environment, 72.78% for family support, and 60.51% for access to study materials. These differences highlight the relative advantages of urban settings.

Rural students face challenges such as limited access to qualified educators, insufficient school infrastructure, and fewer academic support systems, as shown in Fig.5. However, the study also found that non-material factors, including parental involvement, personal motivation, and supportive social networks, can significantly influence academic success, as illustrated in Fig.6. Some rural students performed equally well or better than their urban

https://doi.org/10.31436/ijpcc.v12i2.624

counterparts, highlighting the importance of these non-material influences.

In addition, the supervised data-mining analysis reported in Section VI confirms and clarifies this pattern, where, as discussed in Section VI.D, the strongest modifiable predictors of tertiary-level achievement are family-level variables (academic guidance, home environment, financial support) rather than urban location itself, with ICT access to study materials also making a meaningful contribution. This suggests that the urban–rural gap in Malaysian educational results is mediated by household resources and support rather than being caused by geography itself.

This study suggests that reducing the rural-urban academic gap must go beyond basic resource allocation. Policymakers should prioritise targeted investment in rural education, particularly in infrastructure development, digital learning accessibility, and teacher training. Society should also support active parental engagement and broader community efforts to promote it. Educational institutions can benefit from flexible teaching approaches that respond to the specific needs of rural learners. By recognising and addressing these factors, Malaysia can ensure that every student, regardless of location, has the opportunity to succeed academically and contribute meaningfully to the nation's development.

ACKNOWLEDGMENT

The author expresses gratitude to the respondents for their cooperation in conducting this work and appreciation for the applicable knowledge of the CSCI 1304 Probability and Statistics course offered by the Kulliyyah of Information and Communication Technology.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHOR(S) CONTRIBUTION STATEMENT

The author contributed to the study conception and design, manuscript writing, and approval of the final version.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This research did not require ethical approval.

REFERENCES

[1] L. Xiang and J. Stillwell, “Rural–Urban Educational Inequalities and Their Spatial Variations in China,” *Appl. Spat. Anal. Policy*, vol. 16, no. 2, pp. 873–896, Jun. 2023, doi: 10.1007/s12061-023-09506-1.

[2] N. Ahsan, M. Shahe Emran, F. Shilpi, and W. Bank, “The Rural-Urban Divide and Intergenerational Educational Mobility in a Developing Country: Theory and Evidence from Indonesia,” Oct. 2020.

[3] J. Munir, M. Faiza, B. Jamal, S. Daud, and K. Iqbal, “The Impact of Socio-economic Status on Academic Achievement,” *Journal of Social Sciences Review*, vol. 3, no. 2, pp. 695–705, Jun. 2023, doi: 10.54183/jssr.v3i2.308.

[4] R. M. Wood, “A Review on Education differences in Urban and Rural Areas,” *International Research Journal of Educational Research*, vol. 14, no. 2, pp. 1–3, 2023, doi: 10.14303/2141-5161.2023.254.

[5] J. Drescher, A. Podolsky, S. F. Reardon, and G. Torrance, “The Geography of Rural Educational Opportunity,” *RSF*, vol. 8, no. 3, pp. 123–149, May 2022, doi: 10.7758/RSF.2022.8.3.05.

[6] C. F. Rodríguez-Hernández, E. Cascallar, and E. Kyndt, “Socio-economic status and academic performance in higher education: A systematic review,” *Educ. Res. Rev.*, vol. 29, p. 100305, Feb. 2020, doi: 10.1016/j.edurev.2019.100305.

[7] T. Şengönül, “A Review of the Relationship between Parental Involvement and Children’s Academic Achievement and the Role of Family Socioeconomic Status in this Relationship,” *Pegem Egitim ve Ogretim Dergisi*, vol. 12, no. 2, pp. 32–57, 2022, doi: 10.47750/pegegog.12.02.04.

[8] S. P. Kawuryan, S. A. Sayuti, Aman, and S. I. A. Dwiningrum, “Teachers quality and educational equality achievements in indonesia,” *International Journal of Instruction*, vol. 14, no. 2, pp. 811–830, Apr. 2021, doi: 10.29333/iji.2021.14245a.

[9] R. Van Maarseveen, “The urban-rural education gap: Do cities indeed make us smarter?,” *J. Econ. Geogr.*, vol. 21, no. 5, pp. 683–714, Sep. 2021, doi: 10.1093/jeg/lbaa033.

[10] J. Lu, “Study on the Difference between Urban and Rural Education in Yuyao,” *Journal of Education and Educational Research*, vol. 10, 2024.

[11] E. Frank, M. A. Hall, I. H. Witten, and M. Kaufmann, “WEKA Workbench Online Appendix for ‘Data Mining: Practical Machine Learning Tools and Techniques,’” 2016.