

GENDER PATTERNS FROM STUDENT ADMISSION DATA: DATA MINING FOR RELEVANT GENDER AND DEVELOPMENT INITIATIVE

REX P. FLEJOLES¹ and HELEN N. PERLAS²

^{1, 2} Assistant Professor, Iloilo Science and Technology University Miagao Campus, Igtuba, Miagao, Iloilo, Philippines. Email: ¹rexflejoles@gmail.com, ²helen.perlas@isatu.edu.ph
ORCID: ¹<https://orcid.org/0000-0002-8650-6867>, ²<https://orcid.org/0000-0001-9151-6771>

Abstract

Data mining functionalities have been employed to discover patterns from big data. Taking into consideration such capability, this study was conducted to employ such technique to gain insights from the university admission data. Aside from aggregation, this investigation utilized the Decision Tree and the Random Forest algorithms to predict the degree program and specialization preference. It was revealed that more of the applicants were female. From among the models generated, female preference differs basically based on their municipal residence, while male preference is attributed to their Senior High School, average grade, and municipal residence.

Keywords: University Admission, Gender Pattern, Data Mining, Characterization

INTRODUCTION

Recognizing the benefits of data mining, it has been employed in research endeavor to discover interesting patterns through its different methods or functionalities, generally categorized as descriptive or predictive (Chaudhary, 2015). Characterization is among the descriptive task, while classification is one of the predictive functionalities (Han & Kamber, 2006). The use of the Decision Tree (Hamsa et al., 2016, 2016) and the Random Forest (Ali et al., 2023; Orte et al., 2023) for classification are evident in many studies.

This data mining approach is useful where big data are available. Such big data may be obtained from various sources, including the student admission. By using appropriate tools, certain insights may be generated to help the school, such as taking initiatives for the Gender and Development (GAD) program.

Iloilo Science and Technology University Miagao Campus (ISATU-MC) fosters equality, recognizing the basic human rights of every person, regardless of gender. Every incoming college students may choose any of the degree program offerings of the school. Presently, it offers 10 degree programs: Bachelor of Technical Vocational Teacher Education (BTVTED), Bachelor of Technology and Livelihood Education (BTLED), Bachelor of Science in Hospitality Management (BSHM), Bachelor of Secondary Education (BSED), Bachelor of Science in Tourism Management (BSTM), Bachelor of Science in Information Technology (BSIT), Bachelor of Science in Information Systems (BSIS), Bachelor of Science in Entrepreneurship (BSEntrep), Bachelor of Industrial Technology (BIT), and Bachelor of Elementary Education (BEED).

There are specialization in some degree programs. The BIT has nine specializations, namely, Welding and Fabrication Technology (WAFT), Food Technology (FoodTech), Fashion and Apparel Technology (FAT), Electronic Technology (Electronics), Heating Ventilating Air

Conditioning-Refrigeration Technology (HVACRT), Electrical Technology (Electrical), Cosmetology, Automotive Technology (Automotive), and Architectural Drafting Technology (ADT). For the BSED, the five specializations are Social Studies (SocStud), Mathematics (Math), Filipino, English, and Science. The BTLED has two specializations: Industrial Arts (IA) and Home Economics (HE). Six specializations are offered in BVTED, namely, Electrical Technology (Electrical), Drafting Technology (Drafting), Food and Service Management (FSM), Electronic Technology (Electronics), Garments, Fashion and Design (GFD), and Automotive Technology (Automotive). Despite accommodating students in any degree program, variation in the distribution of male and female students was observed. Hence, this study was initiated to explore the student admission data to determine gender patterns. Consequently, a relevant GAD activity may be recommended.

METHOD

This research employed the data mining as a knowledge discovery from data process, which is composed of four stages: (1) cleaning and integration, (2) selection and transformation, (3) data mining, and (4) pattern evaluation and presentation. Repetition in certain activities may be expected depending on the needed functionality (Han & Kamber, 2006). The data mining functionalities applied in this study were both descriptive and predictive using the RapidMiner Studio. The dataset was obtained from the Office of the Campus Registrar and Admission, ISATU-MC, Iloilo, Philippines. It contained selected data of student applicants for the academic year 2021-2022 admission. In compliance with the Data Privacy Act, only relevant data for this investigation were collected, which include: gender, municipal residence, type of Senior High School (SHS) attended, SHS average grade, and the first preferred degree program to enroll. The dataset contained 2,215 records. During the data cleaning and integration, records with missing data were eliminated, which resulted to 2,187 instances. In the selection and transformation stage, the preferred degree program field was separated into two to come up with another field distinguishing the specialization. The degree program value was copied for the specialization value of the degree programs without specialization. In the data mining stage, aggregation was employed for descriptive while classification for predictive. In the evaluation and presentation stage, results were assessed and various visualizations were considered.

For classification, two target attributes were explored, namely, the degree program with 10 classes and the specialization with 28 classes. The splitting was done taking into consideration three different ratios (training-testing): 70-30, 75-25, and 80-20. While two classification algorithms were used: the Decision Tree and the Random Forest. With the Random Forest, three different numbers of trees were explored: 100, 200, and 300. This resulted to a total of 24 models investigated.

RESULTS AND DISCUSSION

Out of the 2,187 applicants, 1,387 were female while 800 of them were male as illustrated in Figure 1. This shows that more female students had interest to enroll in ISATU-MC than the males. Although the results was not the actual enrollment, the case was different in Northwest Syria, where more male students enrolled in the two universities being studied (Almelhem et al., 2022).

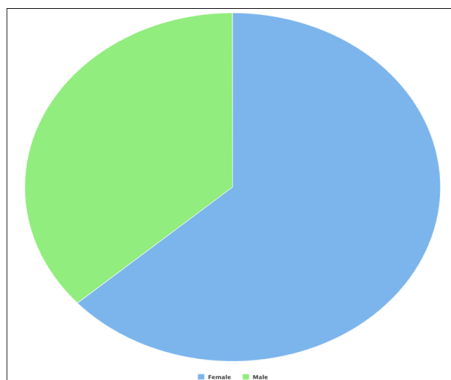


Figure 1: Distribution of applicants according to gender

In terms of the municipal residence, applicants came from different municipalities of Iloilo Province and its nearby provinces as shown in Figure 2. However, most of them were from the seven municipalities of the First District of Iloilo, namely, Miagao, Tigbauan, San Joaquin, Guimbal, Igararas, Oton, and Tubungan. This may be attributed to the fact that ISATU-MC is situated in Miagao, First District of Iloilo. Taking into consideration the associated costs for going to school, students and their parents would greatly consider accessibility in their school preference.

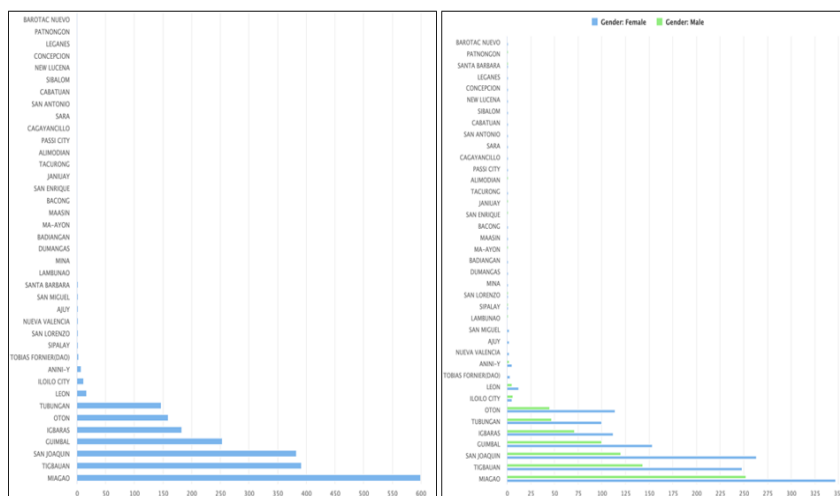


Figure 2: Municipal residence of the applicants

Regarding the type of SHS attended by the applicants 2,095 were from public schools while 92 of them were from private schools as depicted in the Figure 3. Moreover, the figure shows that 1,335 female applicants were from public schools while 52 of them were from private schools, and 760 of the male applicants were from the public schools while 40 of them were from the private schools. This means that most of the applicants, regardless of gender were from public schools. The result is opposite with the number of undergraduate dental students in Taiwan (Cheng et al., 2023). For the result of this study, it justifies the fact that many public schools offer the SHS curriculum. Moreover, this may construe that parents took advantage of the benefits of sending their children in public schools, particularly in terms of costs since public schools do not collect school fees.

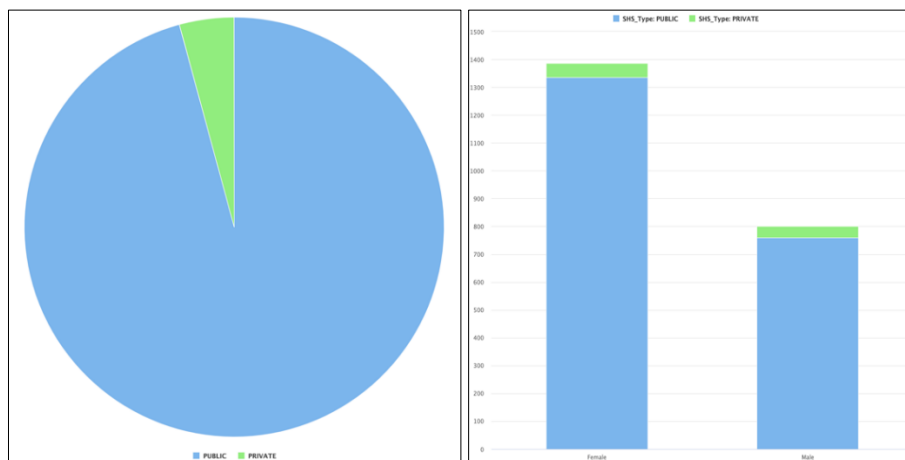


Figure 3: Applicants' type of last school attended

With regards to the average grade of the applicants, the minimum (Q0), lower quartile (Q1), median (Q2), upper quartile (Q3), and maximum (Q4) values were 74, 85.5, 89, 91, and 98.75, respectively. As to gender, the Q0, Q1, Q2, Q3, and Q4 average grades of the female applicants were 74, 87, 90, 92, and 98.22, respectively. While for the male applicants, the corresponding values were 75, 84, 86, 89.0375, and 98.75. Although the highest grade was observed from the male, the median grade implies that female students obtained higher average grade. This observation is different from the quasi-experimental involving students of the University of Barcelona, in which male students outperformed the females (Montolio & Taberner, 2021). The corresponding illustrations are presented in Figure 4.

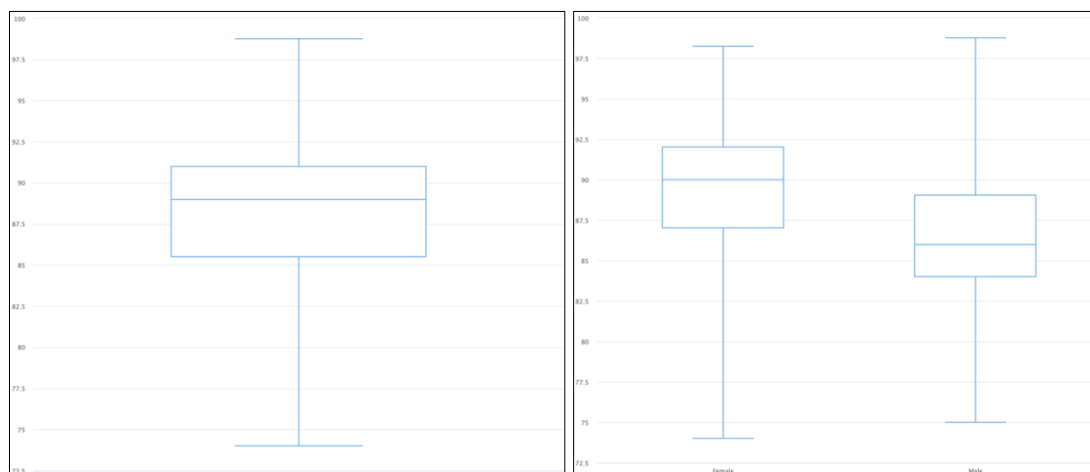


Figure 4: Average grade of the applicants

The preferred degree program of the student applicants were as follows: 161 for BTVTED, 34 for BTLED, 444 for BSHM, 425 for BSED, 135 for BSTM, 312, for BSIT, 17 for BSIS, 56 for BSEntrep, 459 for BIT, and 144 for BEED. When classified according to gender, the preference of the females were as follows: 60 for BTVTED, 28 for BTLED, 345 for BSHM, 354 for BSED, 116 for BSTM, 172 for BSIT, 9 for BSIS, 44 for BSEntrep, 129 for BIT, and 130 for BEED; while of the males were as follows: 101 for BTVTED, 6 for BTLED, 99 for BSHM, 71 for BSED, 19 for BSTM, 140 for BSIT, 8 for BSIS, 12 for BSEntrep, 130 for BIT, and 14

for BEED. Although each program had a number of male and female applicants, it is observed that there were more females who preferred BTLED, BSHM, BSED, BSTM, BSIT, BSIS, BSEntrep, and BEED; while there were more males who preferred BTVTLED, and BIT. The corresponding visualizations are shown in Figure 5.

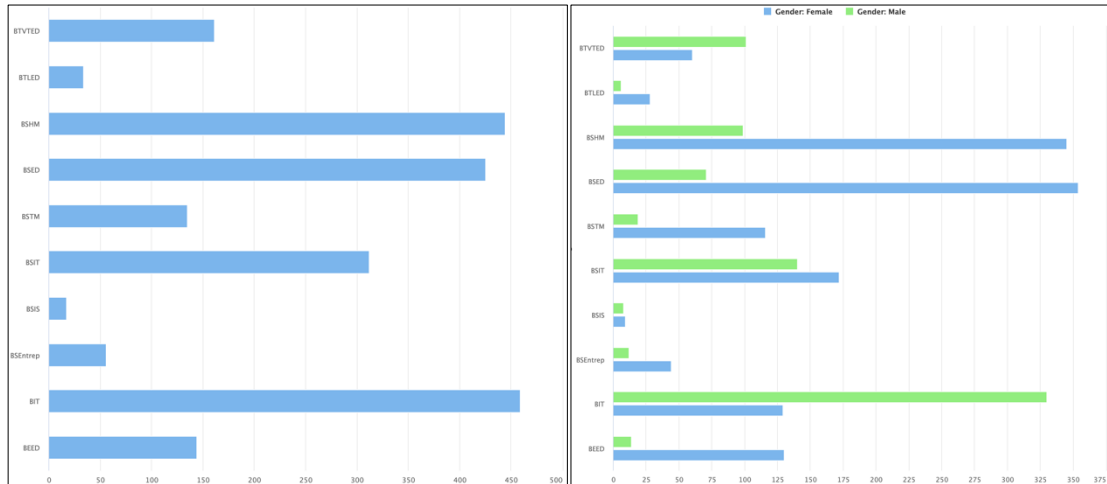


Figure 5: Preferred degree program of the applicants

In terms of BIT specializations, the preference of the applicants were as follows: 58 for WAFT, 131 for FoodTech, 10 for FAT, 23 for Electronics, 12 for HVACRT, 25 for Electrical, 10 for Cosmetology, 128 for Automotive, and 32 for ADT. When grouped according to gender, the preference of the females were as follows: 5 for WAFT, 89 for FoodTech, 9 for FAT, 2 for Electronics, 2 for Electrical, 8 for Cosmetology, 1 for Automotive, and 13 for ADT; while of the males were as follows: 53 for WAFT, 42 for FoodTech, 1 for FAT, 21 for Electronics, 12 for HVACRT, 53 for Electrical, 2 for Cosmetology, 127 for Automotive, and 19 for ADT. Results showed that there were more females who preferred FoodTech, FAT, and Cosmetology; while there were more males who preferred WAFT, Electronics, Electrical, Automotive, and ADT. Moreover, there was no female who preferred the HVACRT. Figure 6 presents the corresponding illustrations.

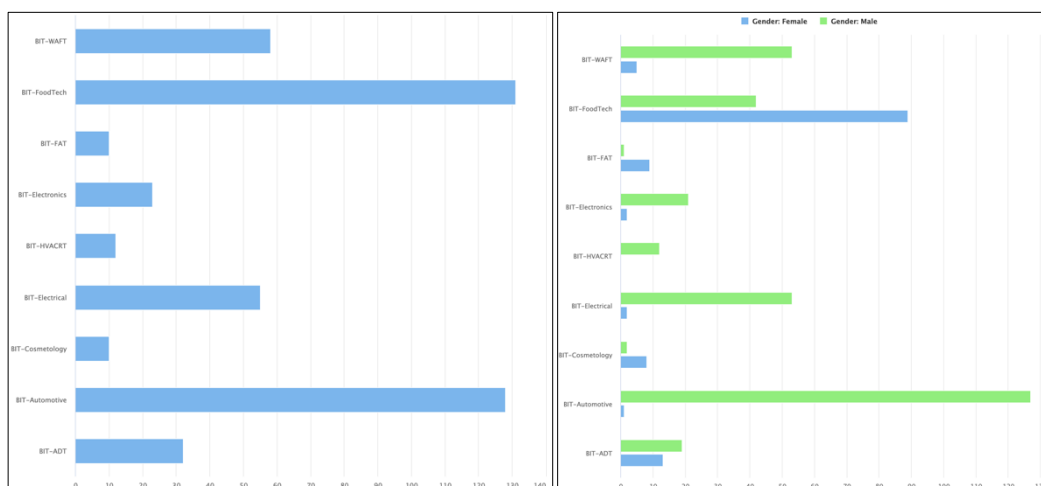


Figure 6: Distribution of students who preferred the BIT degree program

With regards to the BSED specializations, the preference of the applicants were as follows: 71 for SocStud, 52 for Math, 101 for Filipino, 147 for English, and 54 for Science. When classified according to gender, the preference of females were as follows: 52 for SocStud, 40 for Math, 90 for Filipino, 129 for English, and 43 for Science; while for males were as follows: 19 for SocStud, 12 for Math, 11 for Filipino, 18 for English, and 11 for Science. Results revealed that there were more females who preferred any of the specializations. Figure 7 shows the corresponding visualizations.

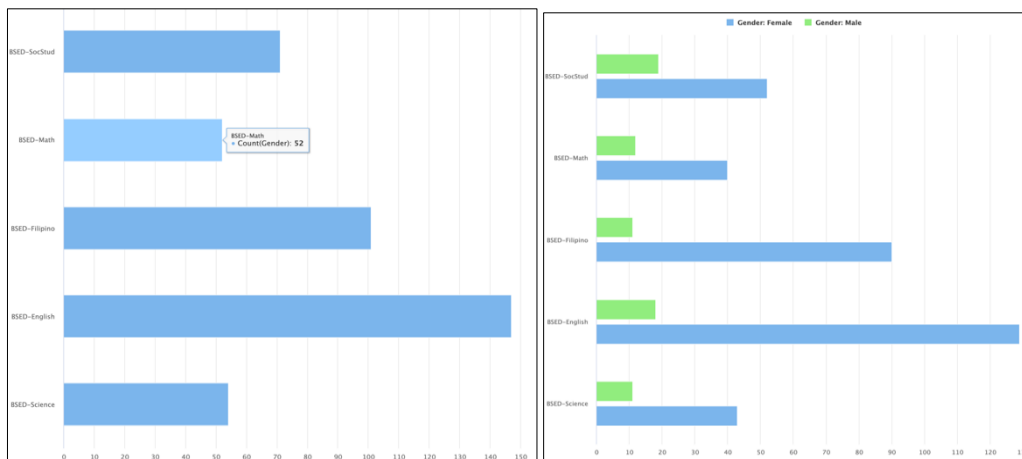


Figure 7: Distribution of students who preferred the BSED degree program

For BTLED specializations, 5 preferred IA and 29 preferred HE. When grouped according to gender, 3 females preferred IA and 25 preferred HE; while 2 males preferred IA and 4 males preferred HE. Results showed that there were more females who preferred any of the specializations. The corresponding illustrations are presented in Figure 8.

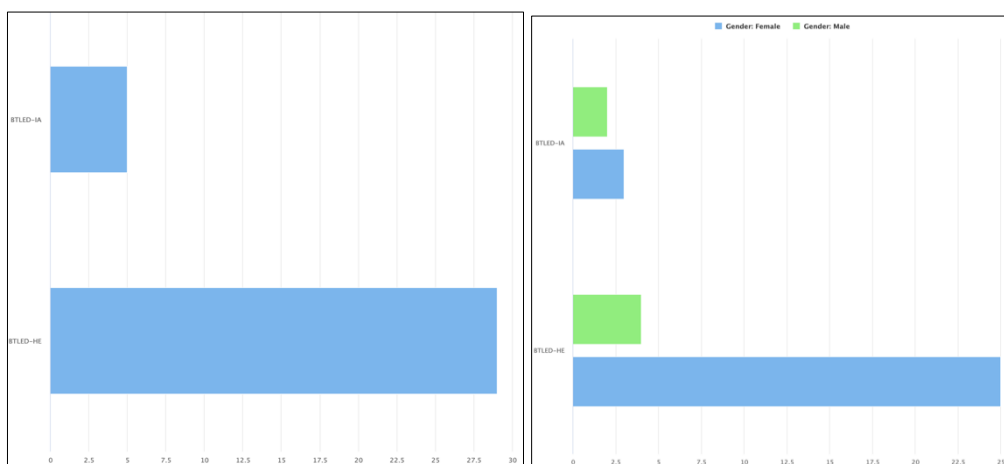


Figure 8: Distribution of students who preferred the BTLED

With the BTVTED specializations, the preferences were as follows: 19 for Electrical, 7 for Drafting, 47 for FSM, 14 for Electronics, 23 for GFD, and 51 for Automotive. When classified according to gender, the preference of females were as follows: 3 for Electrical, 2 for Drafting, 33 for FSM, 20 for GFD, and 2 for Automotive; while of the males were 16 for Electrical, 5 for Drafting, 14 for FSM, 14 for Electronics, 3 for GFD, and 49 for Automotive. Results

revealed that there were more females who preferred FSM and GFD; while there were more males who preferred Electrical, Drafting, and Automotive. Moreover, there was no female who preferred Electronics. The corresponding visualizations are presented in Figure 9.

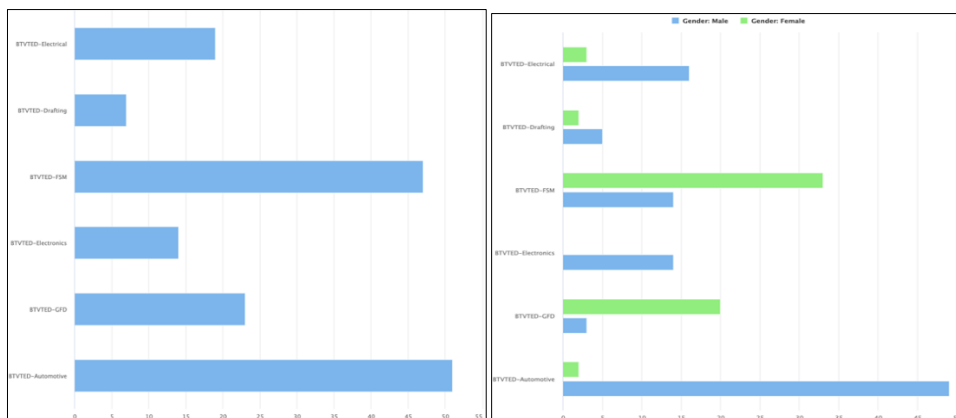


Figure 9: Distribution of students who preferred the BTVTED

Taking into consideration the differences in the degree program preferences of the student applicants, a classification was explored using the Decision Tree and Random Forest algorithms. Since the latter may use a number of trees, three were considered: 100, 200, and 300 trees. Three ratios for the training and testing data were set, namely, 70-30, 75-25, and 80, 20. While for the target attribute, those parameters were tried to classify the specialization and the degree. Table 1 presents the different models with their respective parameters and accuracy results.

Table 1: Classification Models and Their Respective Accuracy Results

Algorithm	Target Attribute	Training : Testing	Accuracy (%)
Decision Tree	Specialization	70 : 30	22.02
Decision Tree	Specialization	75 : 25	21.06
Decision Tree	Specialization	80 : 20	21.79
Decision Tree	Degree	70 : 30	30.99
Decision Tree	Degree	75 : 25	30.40
Decision Tree	Degree	80 : 20	32.27
Random Forest (100 trees)	Specialization	70 : 30	21.41
Random Forest (100 trees)	Specialization	75 : 25	22.34
Random Forest (100 trees)	Specialization	80 : 20	22.94
Random Forest (200 trees)	Specialization	70 : 30	22.17
Random Forest (200 trees)	Specialization	75 : 25	23.26
Random Forest (200 trees)	Specialization	80 : 20	22.94
Random Forest (300 trees)	Specialization	70 : 30	22.32
Random Forest (300 trees)	Specialization	75 : 25	23.08
Random Forest (300 trees)	Specialization	80 : 20	22.71
Random Forest (100 trees)	Degree	70 : 30	33.89
Random Forest (100 trees)	Degree	75 : 25	33.88
Random Forest (100 trees)	Degree	80 : 20	35.47
Random Forest (200 trees)	Degree	70 : 30	34.02
Random Forest (200 trees)	Degree	75 : 25	33.88
Random Forest (200 trees)	Degree	80 : 20	36.38
Random Forest (300 trees)	Degree	70 : 30	33.89
Random Forest (300 trees)	Degree	75 : 25	34.25
Random Forest (300 trees)	Degree	80 : 20	36.16

From the 24 models, the model generated from employing the Random Forest with 200 trees, at 80-20 ratio, which classifies the preferred degree program obtained the highest accuracy (36.38%). The results were not really promising. However, looking at the first model that chose gender as the first attribute at the root, it reveals certain patterns. The aforesaid model reveals the following about female applicants: (a) females from Ajuy, Cabatuan, Cagayancillo, Concepcion, New Lucena, and Sibalom would prefer BSED; (b) females from Badiangan and Maasin would prefer BSIT; (c) females from Mina would prefer BIT; (d) females from San Lorenzo would prefer BTVTLED; and (e) females from Tobias Fornier (Dao) would prefer BSHM. It should be noted that Cagayancillo, Sibalom, San Lorenzo, and Tobias Fornier are municipalities outside of Iloilo.

For male applicants, the same model reveals the following: (a) males from private school, who reside in Igaras would prefer BSIT, (b) males from private school, who in Oton would prefer BEED, (c) males from private school, who reside in San Joaquin would prefer BSTM, (d) males from public school, who obtained an average grade of greater than 94.50 and reside in Igaras would take either BSHM or BSTM, (e) males from public school, who obtained an average grade of greater than 94.50 and reside in Miagao would take either BIT or BSED, (f) males from public school, who obtained an average grade of greater than 94.50 and reside in San Enrique or Tubungan would prefer BSIT, and (g) males from public school, who obtained an average grade of greater than 94.50 and reside in Tigbauan would take BIT.

CONCLUSIONS AND RECOMMENDATIONS

Most of the students who applied for admission at ISATU-MC were female. This may imply that the degree program offerings of the university cannot attract more male students or that the population of males within the district is lower than of females. Another contributing factor perhaps is the presence of a private school in the same district, which offers maritime programs that are attractive to males. Hence, a possible GAD initiative should focus on attracting male students. Taking into consideration the existing program offerings of the University, the difference between the female and male applicants in BSIT and BSIS was minimal. However, the programs with gender disparity dominated by females include the BTLED, BSHM, BSED, BSTM, BSEntrep, and BEED; while which dominated by males include BTVTLED and BIT. Hence, a possible GAD initiative, may take into consideration revisiting the career fair activities prepared for the SHS students. Maybe, there is a need to showcase the skills gained from the different programs, emphasizing the contributions of both males and females. Moreover, a presentation in which successful male and female graduates in every program may be produced. The produced models showed low accuracy. Caution must be taken in its implementation. It is recommended to conduct similar study with the integration of data from other academic years. Nonetheless, relevant information may be considered as a stepping stone in preparing GAD-related activity.

Declaration of Interest

The authors declare no potential conflicts of interest with respect to this research, authorship, and publication.

References

1. Ali, Md. R., Nipu, S. Md. A., & Khan, S. A. (2023). A decision support system for classifying supplier selection criteria using machine learning and random forest approach. *Decision Analytics Journal*, 7, 100238. <https://doi.org/10.1016/j.dajour.2023.100238>
2. Almelhem, S., Almshor, E., Alabdullah, S., Kadan, B., Alzoabi, M., & Jhar, A. (2022). Factors affecting gender balance in higher education in northwest Syria: Challenges and potential actions. *International Journal of Educational Research Open*, 3, 100164. <https://doi.org/10.1016/j.ijedro.2022.100164>
3. Chaudhary, D. P. (2015). Data Mining System, Functionalities and Applications: A Radical Review. *International Journal of Innovations in Engineering and Technology*, 5(2).
4. Cheng, F.-C., Wang, Y.-L., Wang, L.-H., Yu-Fong Chang, J., Liu, S.-Y., & Chiang, C.-P. (2023). Overview of dental education system and gender distribution of undergraduate dental students in Taiwan. *Journal of Dental Sciences*, 18(1), 420–427. <https://doi.org/10.1016/j.jds.2022.11.011>
5. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student Academic Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm. *Procedia Technology*, 25, 326–332. <https://doi.org/10.1016/j.protcy.2016.08.114>
6. Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). Morgan Kaufmann Publishers.
7. Montolio, D., & Taberner, P. A. (2021). Gender differences under test pressure and their impact on academic performance: A quasi-experimental design. *Journal of Economic Behavior & Organization*, 191, 1065–1090. <https://doi.org/10.1016/j.jebo.2021.09.021>
8. Orte, F., Mira, J., Sánchez, M. J., & Solana, P. (2023). A random forest-based model for crypto asset forecasts in futures markets with out-of-sample prediction. *Research in International Business and Finance*, 64, 101829. <https://doi.org/10.1016/j.ribaf.2022.101829>