Educenter : Jurnal Ilmiah Pendidikan
Vol 4 No 2 2025
ISSN: 2827-8542 (Print) ISSN: 2827-7988 (Electronic)
Open Access: https://jurnal.arkainstitute.co.id/index.php/educenter/index

Implementation of deep learning in the recommendation system for high school admissions in Semarang City

Alfa Hikma Zaidan[1*], Novita Kurnia Ningrum[2]
[1,2] Universitas Dian Nuswantoro, Indonesia
[1]111202214234@mhs.dinus.ac.id, [2]novita.kn@dsn.dinus.ac.id

Article Info

Article history:
Received September 8th 2025
Revised September 25th 2025
Accepted October 15th 2025

Keyword:
Deep Learning; Multi-Layer Perceptron;
Recommendation System;
Euclidean Distance

ABSTRACT

The New Student Admission System for senior high schools in Semarang still faces challenges in achieving fair and proportional selection. The dominant zoning policy often ignores students' academic potential; therefore, a more comprehensive recommendation system is needed. This study proposes the development of a deep learning-based school recommendation system using a Multi-Layer Perceptron (MLP) architecture with a backpropagation algorithm. The dataset consists of 16 public senior high schools in Semarang, with the main variables including exam scores, age, school capacity, and distance from student residence calculated using the Euclidean distance method. The data is divided into a training set and a test set, with normalization applied to all numeric features. The training results show high accuracy. The system is able to generate school recommendation rankings that are visualized in tabular formats and interactive maps. Experimental results indicate that distance and school capacity contribute significantly to determining preference scores. Therefore, this study confirms that the deep learning approach is more adaptive than the rule-based linear method and can be an alternative solution to support a fairer and more transparent Student Admissions policy. For further research, it is recommended to develop the system by adding more diverse variables, real-time data integration, and implementing a more complex deep learning architecture to optimize the quality of recommendations.

INTRODUCTION

The High School New Student Admissions System in Semarang City still faces challenges in ensuring a fair, transparent, and proportional selection process. Based on Minister of Education and Culture Regulation (Permendikbud) No. 1 of 2021, the New Student Admissions mechanism in Indonesia is required to consider three main aspects: distance from residence (zoning), academic achievement, and non-academic achievements. However, implementation in the field shows that zoning criteria still dominate, potentially overlooking students' academic achievements and talents.

This situation gives rise to various problems, one of which is unfairness for students from sub-districts without public high schools, such as Candisari, Gajahmungkur, and Tugu. Although students from these areas have good academic achievements, they are often unable to compete due to limited access to zoning. A report from the Semarang City Education Office reinforced these findings, followed by various media reports highlighting public complaints regarding the zoning policy, which is considered unfair and even encourages dishonest practices in the admissions process (Liu et al., 2021; Sattin-Bajaj & Jennings, 2020).

Efforts to address the New Student Admissions selection process have been made through the implementation of Decision Support Systems (DSS). Erlina & Okfalisa (2023) applied the TOPSIS method in an online New Student Admissions system to recommend study programs that better align with students' talents and interests. Their research findings indicate that a weighting-based approach can help improve the objectivity of student placement.

Another study by Wong et al., (2024) combined Profile Matching and Simple Additive Weighting (SAW) methods to recommend student programs under the K-13 curriculum. This approach successfully increased recommendation accuracy to 79.2%, demonstrating the significant potential of multi-criteria methods in New Student Admissions. Furthermore, research by Wongvilaisakul et al., (2023) applied the Weighted Product method in a junior high school New Student Admissions recommendation system, resulting in a system that was more adaptable to varying student selection criteria.

A similar approach was also employed by Kumar & Banerji (2025), who used the Analytic Hierarchy Process (AHP) method to support school selection decisions. The research results show that AHP can help balance achievement and distance factors, although it is limited in capturing complex relationships between variables. Meanwhile, research by Ajol et al., (2020) utilized Fuzzy Logic methods to assess student eligibility for New Student Admissions, resulting in a flexible system but requiring complex parameter settings.

Although various previous studies have made important contributions, most approaches are still based on linear rules. This limitation makes it difficult for models to capture non-linear relationships between variables such as distance, academic grades, and school capacity. These factors often interact complexly and cannot be simply represented using traditional weighting methods. Therefore, a more adaptive approach capable of modeling data patterns in depth is needed.

This research offers novelty by integrating Deep Learning methods, specifically Multi-Layer Perceptron (MLP), into the New Student Admissions recommendation system for high schools in Semarang City. Unlike previous studies that rely on linear methods such as TOPSIS, SAW, or AHP, the deep learning-based approach is able to capture non-linear relationships between variables. Furthermore, this research utilizes spatial data (geographic coordinates) to calculate student-to-school distances using the Euclidean method, enabling the system to provide more adaptive, personalized school recommendations tailored to each student's unique profile.

The primary objective of this research is to develop a deep learning-based New Student Admissions recommendation system that integrates zoning, academic achievement, and school capacity. This system is expected to help students select schools that best suit their profiles, while also supporting the Education Office in creating a fairer, more transparent, and more efficient selection process. Academically, this research expands the application of deep learning in education, which has previously been dominated by Multi-Criteria Decision Making (MCDM) methods. Thus, this research not only provides practical contributions to improving the New Student Admissions system but also theoretical contributions to the development of artificial intelligence-based recommendation models.

RESEARCH METHODS

1. Data Collection Method

The research data was obtained from several official and validated sources:

a. School coordinates were collected via Google Maps.

b. Student capacity data for public senior high schools in Semarang City was downloaded from the Open Data Central Java portal.

c. The National School Identification Number (NPSN) and school addresses were accessed through the official Kemendikbudristek School Data Portal.

d. Data validation was conducted through interviews with the Semarang City Education Office, which directed the researchers to use the official Kemendikbudristek portal.

This data collection method aligns with quantitative research practices based on secondary data in the field of educational information systems (Cheong et al., 2023).

2. Data Processing Method

After data collection, the following preprocessing stages were performed:

a. Data Cleaning: removing duplicates, correcting formatting errors, and handling missing values (Sharifnia et al., 2025).

b. Data Normalization: all numerical variables (academic scores, distance, capacity) were normalized using Min-Max Scaling (Ali, 2022).

$$x^{!} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Explanation:

x = original value (raw data).

x_{min} = the smallest value of all data in the variable.

x_{max} = the largest value of all data in the variable.

x′ = the normalized value (new data on a scale of 0–1).

c. Spatial Distance Calculation: student-to-school distances were calculated using the Euclidean Distance method (Soukhov et al., 2025).

d. Categorical Data Encoding: non-numeric data was converted into numeric form using label encoding (Alamuri et al., 2024).

e. Data Splitting: the dataset was divided into 80% training data and 20% testing data (Vrigazova, 2021).

These steps follow standard practices in educational data mining research (Papadogiannis et al., 2024).

3. Research Algorithm

This study applies the Multi-Layer Perceptron (MLP), a type of feedforward artificial neural network widely used in modern recommendation systems (Bao et al., 2022).

Main System Workflow:

a. Geocoding student addresses → converting addresses into geographic coordinates using an API.

b. Calculating distances to schools → using the Distance Matrix API, with Euclidean Distance as a fallback.

$$Distance\ (km) = \sqrt{(lat_1 - lat_2)^2 + (lng_1 - lng_2)^2} \times 111$$

Explanation:

lat_1, lat_2 = latitude of points 1 and 2

lng_1, lng_2 = longitude of points 1 and 2

c. Dataset preparation → features: latitude, longitude, exam scores, age, distance; target: preference score.

d. Building the MLP model → Architecture:

1) Input layer → Dense(64, ReLU) → Dropout(0.2) →Dense(32, ReLU) → Dropout(0.2) →Dense(16, ReLU) →Output layer (Sigmoid activation).
2) Loss Function: Mean Squared Error (MSE)
3) Optimizer: Adam

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - {y_i}^2)^2$$

Explanation:

yi = actual value (true data) at point i.

yi^ = predicted or estimated value at point i.

n = number of data (samples).

e. Recommendation prediction → the model ranks schools based on the highest preference scores.

f. Result visualization → top 3 school routes are displayed using the Static Map API.

4. Rationale for Algorithm Selection

a. Euclidean Distance was chosen for its simplicity and efficiency in spatial calculations (Behrens et al., 2018).

b. MLP was selected for its ability to model non-linear relationships among variables, offering better performance compared to linear methods such as SAW or TOPSIS (Iswanto et al., 2021; Wątróbski et al., 2024).

The combination of both produces a recommendation system that is more accurate, adaptive, and spatially aware (Jumiarti & Suwarno, 2023).

RESULTS AND DISCUSSION

1. Model Training Results

The model training process was carried out using the Deep Learning Multi-Layer Perceptron (MLP) algorithm with data from public senior high school students in Semarang City. The dataset was divided into two subsets: training data (80%) and testing data (20%). The main model parameters were as follows:

a. Epochs: 100

b. Batch Size: 32

c. Optimizer: Adam

d. Hidden Layer Activation Function: ReLU

e. Output Layer Activation Function: Softmax

f. Loss Function: Mean Squared Error (MSE)

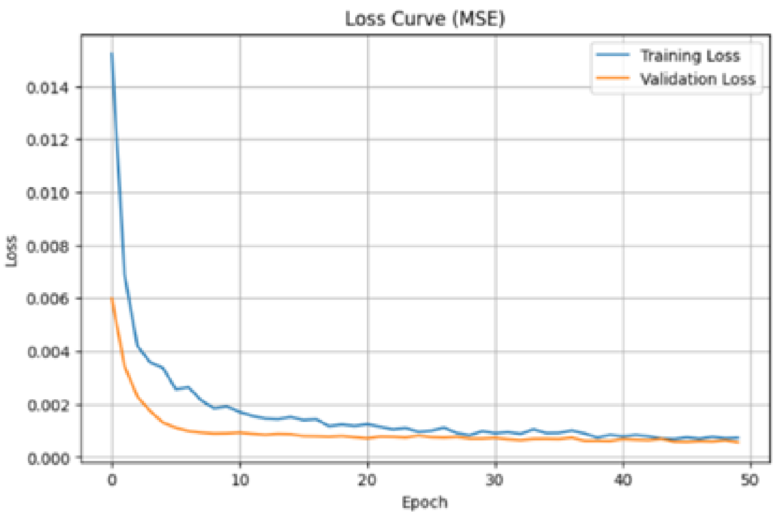

Figure 1. Training and validation loss curve

The graph above shows the loss curve (MSE) during the model training process up to 50 epochs. The X-axis represents the number of epochs or training iterations, while the Y-axis shows the loss value calculated using the Mean Squared Error (MSE) formula. The blue line depicts the training loss obtained from the training data, while the orange line shows the validation loss obtained from the validation data. At the beginning of training, the loss value was relatively high for both the training and validation data, indicating that the model still made many errors. However, as the number of epochs increased, both loss values decreased sharply until they finally reached a stable point near zero. This indicates that the model is able to learn well from the training data while generalizing patterns in the validation data. There are no signs of overfitting because the training loss and validation loss curves decrease in the same direction and are both stable at low values. Thus, this graph indicates that the trained model has good performance and is able to produce accurate predictions.

The loss curve indicates that the error on both training and validation data decreased significantly as the number of epochs increased. At the beginning of training, the loss value was relatively high, but after several epochs it gradually decreased and stabilized at the convergence point. This demonstrates that the model successfully learned without showing signs of overfitting.

2. Model Evaluation

The model's performance was evaluated using Mean Squared Error (MSE) and Mean Absolute Error (MAE) as metrics.

a. Test Loss (MSE): 0.0006

b. Test MAE: 0.0182

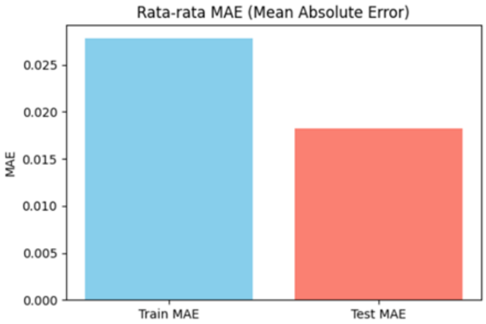

Figure 2. Mean MAE diagram

The graph shows that the MAE value for the test data is lower than the MAE value for the training data. This means that the model actually produces more accurate predictions on the test data than on the training data. This situation is quite rare, but it can occur if the test data is "simpler" or easier to predict than the training data, or because of more complex variations in the training data. In general, the MAE values for both datasets are equally low (below 0.03), indicating a relatively small average prediction error.

This graph shows that the model has good predictive performance, indicated by low MAE values for both the training and test data. In fact, the prediction results for the test data are more accurate than those for the training data, demonstrating that the model does not experience overfitting and is capable of good generalization.

The very small MAE value shows that the average prediction error of the model compared to actual data is minimal, indicating that the model has strong predictive performance in generating school recommendations.

3. Prediction Results

To measure the alignment between predicted and actual values, a comparison was conducted using a scatter plot.

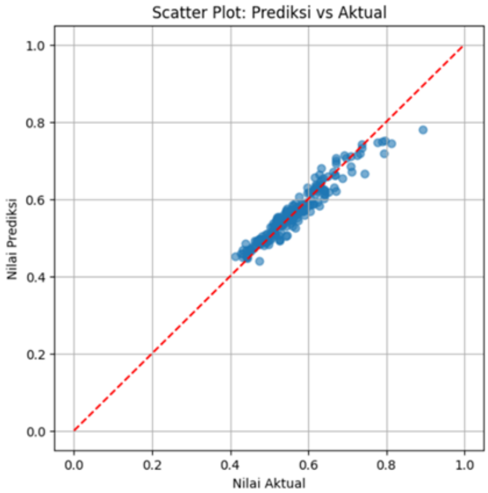

Figure 3. Scatter plot of predicted vs. actual values

The scatterplot above shows the comparison between the actual values (X-axis) and predicted values (Y-axis) from the model. Each blue dot represents a pair of actual and predicted values in the data. The dashed red line depicts the ideal line where the predicted values exactly match the actual values (y=x). The graph shows that most of the points are densely distributed around the red line, indicating that the model's predictions are very close to the actual values. The relatively narrow distribution of points following the diagonal line indicates that the model has good accuracy and relatively small prediction errors. Some points deviate slightly from the line, but overall, this pattern shows consistency between the actual and predicted values. Thus, this scatterplot indicates that the model is capable of producing fairly accurate and reliable predictions.

The scatter plot shows that most prediction points closely follow the diagonal line (y = x). This indicates that the model's predictions are highly accurate and closely match the actual values.

4. Real-World Testing

To evaluate system performance in real-world conditions, a test was conducted using the following student data:

a. Student Address: Jl. Bintoro II No.9, RT.04/RW.07, Pandean Lamper, Gayamsari District, Semarang City
b. Exam Score: 76
c. Age: 16 years
d. Student Coordinates: (-6.9920648, 110.4419308)

The system executed the following processes:

a. Geocoding → converting the address into coordinates
b. Distance Calculation → using Euclidean/Haversine method to compute distances between the student and all public high schools in Semarang City
c. Preference Scoring → combining academic scores, distance, and school capacity
d. School Ranking → ordering schools from the highest to the lowest preference score

5. School Recommendation Results

Table 1. Recommended Schools

Ranking	School Name	Distance (km)	Duration (minutes)	Capacity Preference	Score
1	SMA Negeri 11	2.44	8.1	504	0.6019
2	SMA Negeri 1	4.06	11.3	408	0.5511

Ranking	School Name	Distance (km)	Duration (minutes)	Capacity Preference	Score
3	SMA Negeri 2	4.06	11.3	432	0.5511
4	SMA Negeri 5	5.32	14.4	432	0.4985
5	SMA Negeri 3	5.46	15.4	408	0.4978
6	SMA Negeri 4	5.60	15.9	360	0.4923
7	SMA Negeri 9	6.20	17.1	360	0.4766
8	SMA Negeri 7	6.85	18.9	360	0.4614
9	SMA Negeri 6	7.10	19.4	384	0.4597
10	SMA Negeri 12	7.50	20.2	360	0.4478
11	SMA Negeri 10	8.00	21.3	360	0.4386
12	SMA Negeri 81	8.40	22.4	360	0.4312
13	SMA Negeri 13	9.10	23.8	360	0.4195
14	SMA Negeri 14	9.50	25.1	360	0.4120
15	SMA Negeri 15	10.20	26.8	360	0.4033
16	SMA Negeri 16	10.80	28.0	360	0.3957

Interpretation: SMA Negeri 11 ranked first because it has the shortest distance (2.44 km), fastest travel time (8.1 minutes), and the largest capacity (504 students). These factors gave it a higher preference score compared to other schools.

6. Visualization of Recommendations on Map

In addition to tabular format, the system also presents recommendations on an interactive map. The student's location is marked with a ⚫ symbol, the top recommended school is marked with a 🔵 symbol, and other schools are represented with different markers. An additional feature provides direct Google Maps links, allowing users to view real-time routes.

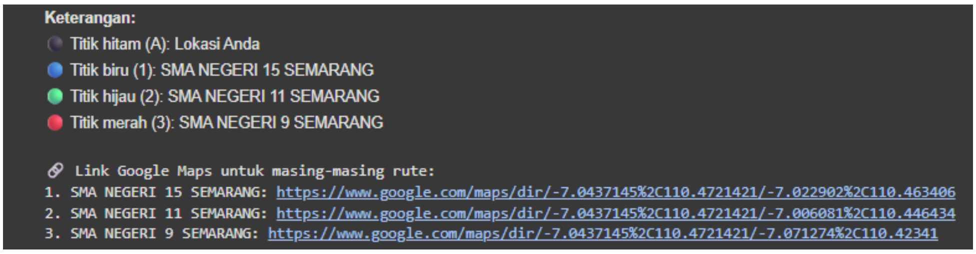

Figure 4. Description

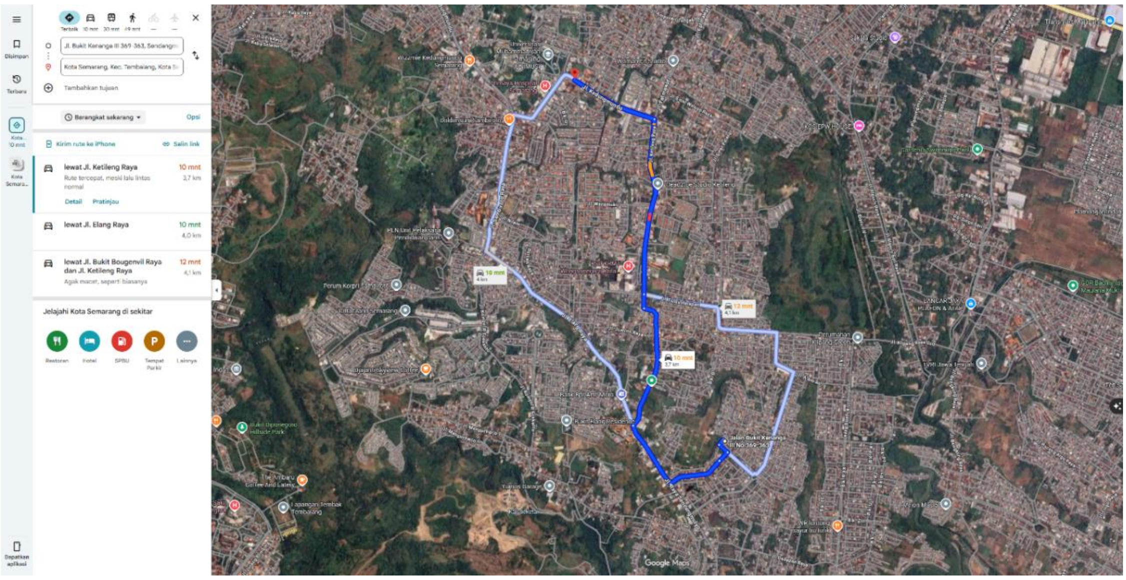

Figure 5. Location of State Senior High School 15 Semarang based on Google Maps

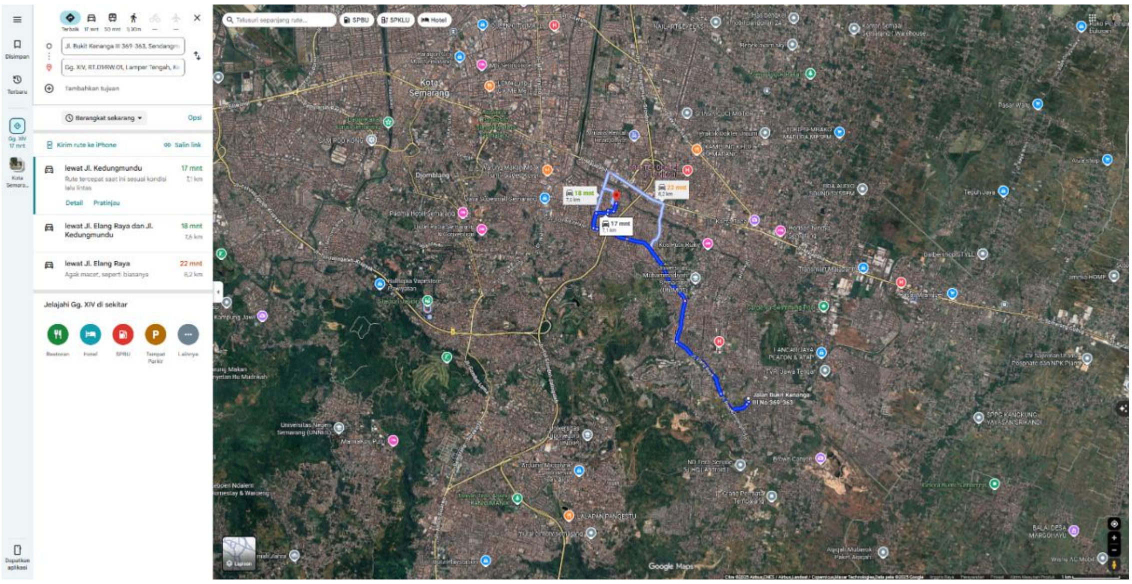

Figure 6. Location of State Senior High School 11 Semarang based on Google Maps

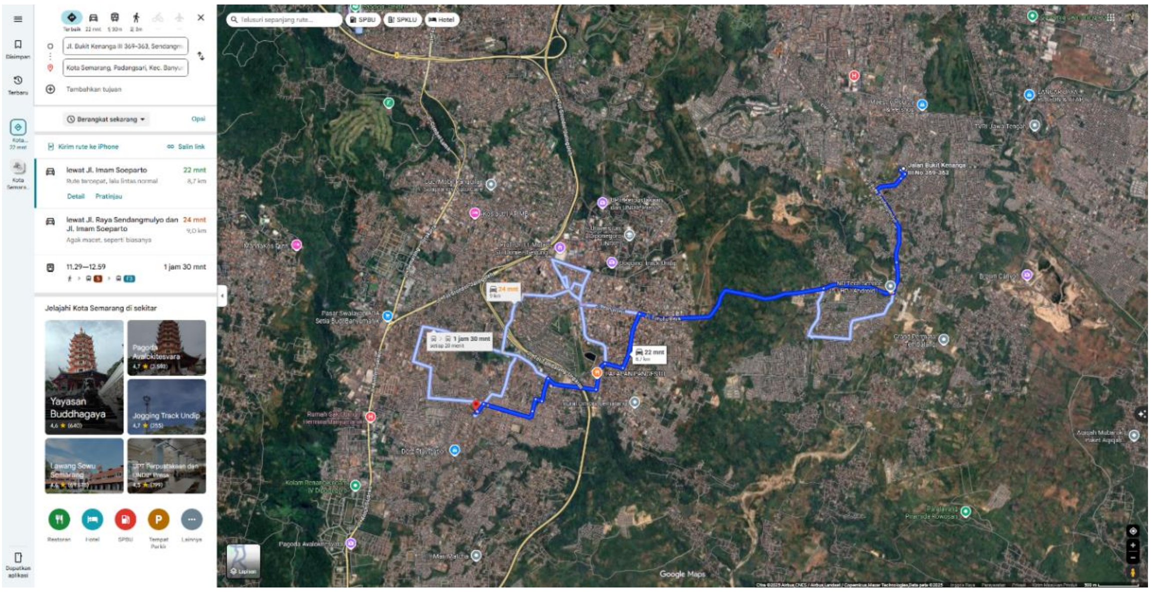

Figure 5. Location of State Senior High School 9 Semarang based on Google Maps

7. Discussion

Based on experimental and testing results, this study demonstrates that the Multi-Layer Perceptron (MLP)-based Deep Learning model is capable of providing school recommendations with high accuracy. This is demonstrated by the low Mean Squared Error (MSE) and Mean Absolute Error (MAE) values, proving that the system can predict consistently with minimal error. Furthermore, the analysis also reveals that distance and school capacity are dominant factors in determining recommendations, in line with the applicable zoning policy mechanism. The integration of academic aspects with spatial factors in this model enables the creation of a recommendation system that is both fair and practical for both students and parents, particularly with the visualization features in the form of graphs, tables, and maps that facilitate interpretation of the results.

These research findings align with several previous studies. Tiwari & Roy (2025) demonstrated that the TOPSIS method is effective in providing New Student Admissions recommendations based on zoning and achievement criteria, despite being limited to a linear model. Kurniawan et al., (2020) also found that the combination of Profile Matching and SAW methods increased the accuracy of the recommendation system by up to 79.2%, underscoring the importance of integrating various variables in the New Student Admissions system. Meanwhile, Xhafaj et al., (2022) used the AHP method to demonstrate that distance and achievement remain key factors in school selection, although this approach is less able to capture non-linear relationships between variables. Therefore, this study strengthens previous findings while also providing new developments through the use of Deep Learning.

The main contribution of this study lies in the application of a deep learning model capable of capturing the non-linear relationship between distance, academic achievement, and school capacity. This approach differs from previous studies, which primarily used linear weighting-based methods. Furthermore, this study provides a practical contribution by producing a more adaptive, personalized, and user-friendly recommendation system, which can assist the Education Office and schools in improving the transparency and fairness of the Student Admissions process in Semarang City.

For future research, it is recommended that the system be expanded to include additional variables such as socio-economic conditions, affirmative action pathways, and students' non-academic achievements. Integrating real-time data from the official New Student Admissions system could also increase the relevance of the recommendation results to the field. Furthermore, model testing can be extended to other regions with different characteristics to test the system's scalability and generalizability. With further development, this recommendation system has the potential to become a comprehensive solution in supporting equal access to secondary education in Indonesia.

CONCLUSION

This research demonstrates that a Multi-Layer Perceptron (MLP)-based deep learning model is capable of integrating various important factors in the New Student Admissions process for high schools in Semarang City. By considering variables such as exam scores, student age, school capacity, and distance from home, the developed recommendation system offers a more comprehensive approach than conventional methods that tend to prioritize zoning or academic achievement alone. The application of deep learning allows for the exploration of non-linear relationships between variables, resulting in predictions that are more adaptive and representative of real-world conditions. These findings demonstrate the potential of artificial intelligence technology to support fairer, more transparent, and data-driven education policies.

While the results are promising, this study still has limitations, particularly the limited dataset coverage of schools in Semarang City and the lack of inclusion of non-academic factors such as student interests, extracurricular achievements, and socioeconomic conditions. Therefore, future research is recommended to expand the data coverage to the regional and national levels and integrate more diverse variables to enhance the accuracy and inclusiveness of the recommendation system. In addition, exploration of other model architectures such as ensemble learning or hybrid methods has the potential to provide new contributions in developing educational recommendation systems that are smarter and more responsive to the dynamics of community needs.

REFERENCES

Ajol, T. A., Gran, S. S., Kanyan, A., & Lajim, S. F. (2020). An Enhanced Systematic Student Performance Evaluation Based on Fuzzy Logic Approach for Selection of Best Student Award. *Asian Journal of University Education*, *16*(4), 10–20. https://eric.ed.gov/?id=EJ1288013

Alamuri, M., Surampudi, B., & Negi, A. (2024). Multi Dimensional Deep Encoding for Categorical Feature Space. *Proceedings of the 2024 13th International Conference on Computing and Pattern Recognition*, 120–128. https://doi.org/10.1145/3704323.3704378

Ali, P. J. M. (2022). Investigating the Impact of min-max data normalization on the regression performance of K-nearest neighbor with different similarity measurements. *ARO-The Scientific Journal of Koya University*, *10*(1), 85–91. https://doi.org/10.14500/aro.10955

Bao, S., Wang, T., Zhou, L., Dai, G., Sun, G., & Shen, J. (2022). Two-Layer Matrix Factorization and Multi-Layer Perceptron for Online Service Recommendation. *Applied Sciences*, *12*(15), 7369. https://doi.org/10.3390/app12157369

Behrens, T., Schmidt, K., Viscarra Rossel, R. A., Gries, P., Scholten, T., & MacMillan, R. A. (2018). Spatial modelling with Euclidean distance fields and machine learning. *European Journal of Soil Science*, *69*(5), 757–770. https://doi.org/10.1111/ejss.12687

Cheong, H., Lyons, A., Houghton, R., & Majumdar, A. (2023). Secondary qualitative research methodology using online data within the context of social sciences. *International Journal of Qualitative Methods*, *22*, 16094069231180160. https://doi.org/10.1177/16094069231180160

Erlina, M., & Okfalisa, O. (2023). The Study Program Selection System: Integrated Analytical Hierarchy Process (AHP) and Technique For Others Preference by Similarity to Ideal Solution (TOPSIS) Approach. *IT Journal Research and Development*, *8*(1), 32–47. https://doi.org/10.25299/itjrd.2023.13562

Iswanto, M. E., Siregar, M. U., 'Uyun, S., & Nuruzzaman, M. T. (2021). Recommender systems for specializing new students in the K-13 curriculum using the profile matching, SAW, and a combination of both. *Jurnal Teknologi Dan Sistem Komputer*, *9*(2), 96–105. https://doi.org/10.14710/jtsiskom.2021.13902

Jumiarti, E., & Suwarno, J. (2023). Penerapan Metode TOPSIS Untuk Merekomendasikan Pemilihan Jurusan Dalam Sistem Penerimaan Peserta Didik Baru (PPDB) Online Di SMK YAPPIKA Legok. *Sainstech Jurnal Penelitian Dan Pengkajian Sains Dan Teknologi*, *33*(2), 66–71. https://doi.org/10.37277/stch.v33i2.1588

Kumar, U., & Banerji, H. (2025). Estimating weightage and prioritizing major domains of inclusive education assessment framework using analytical hierarchy process. *International Journal of System Assurance Engineering and Management*, 1–12. https://link.springer.com/article/10.1007/s13198-025-02869-3

Kurniawan, H., Swondo, A. P., Sari, E. P., Ummi, K., & Tanjung, M. R. (2020). Analysis and Comparative Between Profile Matching and SAW Method in Decision Support. *2020 8th International Conference on Cyber and IT Service Management (CITSM)*, 1–7. https://ieeexplore.ieee.org/abstract/document/9268857/

Liu, L., Zhang, W., & Han, C. (2021). A survey for the application of blockchain technology in the media. *Peer-to-Peer Networking and Applications*, *14*(5), 3143–3165. https://link.springer.com/article/10.1007/s12083-021-01168-5

Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational data mining: A foundational overview. *Encyclopedia*, *4*(4), 1644–1664. https://doi.org/10.3390/encyclopedia4040108

Sattin-Bajaj, C., & Jennings, J. L. (2020). School counselors' assessment of the legitimacy of high school choice policy. *Educational Policy*, *34*(1), 21–42. https://doi.org/10.1177/0895904819881774

Sharifnia, A. M., Kpormegbey, D. E., Thapa, D. K., & Cleary, M. (2025). A Primer of Data Cleaning in Quantitative Research: Handling Missing Values and Outliers. *Journal of Advanced Nursing*. https://doi.org/10.1111/jan.16908

Soukhov, A., Higgins, C. D., Paez, A., & Mohamed, M. (2025). Ten Years of School Closures and Consolidations in Hamilton, Canada and the Impact on Multimodal Accessibility. *Networks and Spatial Economics*, 1–30. https://link.springer.com/article/10.1007/s11067-025-09677-z

Tiwari, D., & Roy, U. K. (2025). Preferences for student rental housing in Dehradun, India: Using TOPSIS technique. *Proceedings of the International Conference of Contemporary Affairs in Architecture and Urbanism-ICCAUA*, *8*(1), 1719–1731. http://journal.iccaua.com/jiccaua/article/view/725

Vrigazova, B. (2021). The proportion for splitting data into training and test set for the bootstrap in classification problems. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, *12*(1), 228–242. https://ojs.srce.hr/index.php/bsr/article/view/17347

Wątróbski, J., Bączkiewicz, A., Król, R., & Rudawska, I. (2024). A Neural Network Regression Model Supported by Multi-Criteria Methods for Ranking Prediction in Sustainable Development Assessment. In *Advances in Information Systems Development: Information Systems Development, Organizational Aspects, and Societal Trends* (pp. 1–21). Springer. https://link.springer.com/chapter/10.1007/978-3-031-57189-3_1

Wong, J. S., Lee, C., & Beck, N. (2024). The effects of aftercare/resettlement services on crime and violence in children and youth: A systematic review. *Campbell Systematic Reviews*, *20*(2), e1404. https://doi.org/10.1002/cl2.1404

Wongvilaisakul, W., Netinant, P., & Rukhiran, M. (2023). Dynamic multi-criteria decision making of graduate admission recommender system: AHP and fuzzy AHP approaches. *Sustainability*, *15*(12), 9758. https://doi.org/10.3390/su15129758

Xhafaj, E., Qendraj, D. H., Xhafaj, A., & Gjikaj, N. (2022). A hybrid integration of PLS-SEM, AHP, and FAHP methods to evaluate the factors that influence the use of an LMS. *International Journal of Decision Support System Technology (IJDSST)*, *14*(1), 1–17. https://www.igi-global.com/article/a-hybrid-integration-of-pls-sem-ahp-and-fahp-methods-to-evaluate-the-factors-that-influence-the-use-of-an-lms/286697