International Journal of Engineering and Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 14 Issue 2 February 2025 || PP 56-61

Predicting Academic Performance and Major Suitability Using Deep Learning: A Case Study

Minh-Phuong Han[1]
[1]*Information System Department, ThuongMai University, Vietnam*

ABSTRACT: *Choosing an appropriate academic major is a pivotal decision for students, impacting both their educational experiences and career trajectories. This study explores the use of deep learning techniques to predict students' aca-demic performance and major suitability at Greenwich Vietnam. A Deep Neural Network (DNN) model was developed using a dataset comprising students' demographic details, entrance exam scores, and academic records. The model achieved a training accuracy of 93% and a test accuracy of 73%, highlighting its predictive capability. The findings demonstrate that aca-demic scores, gender, and regional factors significantly influence students' success. While the study provides valuable insights into optimizing aca-demic advising, limitations such as data scope and binary classification of GPA necessitate future enhancements. Expanding the dataset, integrating qualitative factors, and refining predictive models are suggested to improve forecasting accuracy. This research contributes to the advancement of data-driven decision-making in higher education institutions, promoting better academic and career outcomes for students.*

KEYWORDS - *academic performance prediction, deep learning, neural network, student success, major selection, educational data mining*

Date of Submission: 14-02-2025 Date of Acceptance: 28-02-2025

I. INTRODUCTION

Selecting an academic major is one of the most critical decisions in a student's life, influencing not only their learning experience but also shaping their future career. When students choose a major that aligns with their interests and abilities, they are more likely to be engaged in their studies, thereby enhancing academic performance. Conversely, selecting an unsuitable major can lead to difficulties in knowledge acquisition, poor academic outcomes, and an increased risk of dropping out [1], [2]. A well-informed choice can also open up numerous career opportunities and personal growth prospects [3].

In today's competitive education landscape, higher education institutions must not only offer a diverse range of programs but also support students in selecting a major that aligns with their abilities and aspirations. Universities implement various initiatives to assist students in career orientation [4], including organizing seminars and discussions with industry experts, providing comprehensive guides on academic pro-grams, career prospects, and entry requirements, and facilitating orientation programs where prospective students can interact with faculty members and current students.

However, these activities primarily focus on marketing and recruitment, aiming to disseminate information about academic programs, admission requirements, infra-structure, and job prospects to attract students to the institution. Career counseling centers, staffed with expert advisors, play a crucial role in guiding students; however, their efforts are often centered on admissions rather than personalized academic counseling based on students' abilities. Therefore, it is essential to enhance advisory ser-vices by incorporating reliable predictive data to support students in making well-informed decisions about their major selection [5], [6]. Students should not only choose their major based on personal interest but also consider their aptitude and potential for success.

II. RELATED WORKS

Early research on major selection prediction primarily focused on evaluating the compatibility between applicants' profiles and academic criteria. For instance, Zubaedah et al. (2021) calculated students' likelihood of enrolling in engineering programs based on their scores in Mathematics, Physics, Chemistry, Biology, Language, and Technical Skills [7]. The Analytic Hierarchy Process (AHP) was applied to compare and prioritize subject relevance for each major, helping students make informed decisions. Similarly, Bautista et al. (2016) determined students' enrollment potential based on key factors such as gender and academic performance in Algebra, Integral Calculus, Differential Calculus, Physics I, and Physics II [8].

Other studies, such as Iyer and Variawa (2019), analyzed first-year students' potential inclination towards specific majors by comparing their profiles with those of senior students, thereby suggesting suitable academic pathways [9]. While these approaches have proven effective, relying solely on academic compatibility does not guarantee optimal major selection. A more comprehensive evaluation should consider students' academic performance and career outcomes post-graduation, thus facilitating the development of recommendation models that integrate both personal preferences and alumni success metrics [10], [11].

Previous studies predominantly relied on academic data collected over an extended period. However, academic performance alone may not be a sufficient predictor of students' long-term success, as it overlooks other influential factors. Research by Latifah et al. (2019) sought to address this limitation by incorporating personal at-tributes alongside academic performance to improve the accuracy of predictions [12]. Similarly, studies by Al-Shalabi (2019) and Stein et al. (2020) introduced recommendation models that accounted for skills, interests, and prior experiences [13], [14]. However, some studies, such as Al-Shalabi's, focused only on students with high GPAs (between 3.0 and 4.0), which limited the generalizability of findings.

Predictive models have also leveraged standardized test scores such as ACT, SAT, and NCEE to estimate students' academic trajectories and design personalized learning plans. Xiao and Yi (2021) developed an artificial intelligence (AI) model that achieved a 95% accuracy rate in predicting students' academic paths based on standardized test scores [15]. Mengash (2020) implemented a decision-support system using HSGA, SAAT, and GAT scores to predict students' future academic performance [16]. While effective, these models primarily relied on academic indicators, failing to account for non-academic factors that may influence students' success as mentioned in several researches [17], [18], [19].

In addition to machine learning approaches, expert rule-based systems have been employed to assist with major selection. Al Ahmar (2012) introduced a rule-based expert system designed to help high school students choose suitable university majors [20]. However, such systems often incorporate subjective biases and lack com-prehensive assessments of students' skills. Similarly, fuzzy logic techniques, as explored by Salaki et al. (2015), have been used to assist vocational students in selecting their fields of study [21]. However, both studies failed to integrate real-world student success data, which could improve model accuracy and applicability.

As predictive modeling in education evolves, recent studies have explored more advanced machine learning techniques, including artificial neural networks (ANN) and gradient boosting algorithms. Dhar and Jodder (2020) developed a recommendation system using LightGBM and CatBoost, which provided highly accurate predictions for students in science, humanities, commerce, and engineering programs [22]. Additionally, ANN-based models have demonstrated superior performance in predicting students' academic success [23], [24], [25].

Given these advancements, this study aims to build upon prior research by developing a predictive model that incorporates both academic and demographic factors to guide students in selecting their most suitable major. By integrating a deep learning approach, this research seeks to improve the accuracy and relevance of major selection recommendations, ultimately enhancing student retention and success rates.

III. RESEARCH METHODOLOGIES

To effectively predict student success or failure, this study employs deep neural networks to analyze academic performance data. By leveraging advanced machine learning techniques, the model provides data-driven insights, assisting prospective students in selecting the most suitable major for their capabilities and strengths.

3.1 Data Collection and Pre-processing

This study utilizes admission and academic performance data from Greenwich Vietnam for the academic years 2020 and 2021. The dataset comprises 797 student records, including demographic details, entrance exam scores, and academic performance during their studies. The use of real-world institutional data enhances the applicability of this research and provides valuable insights into factors influencing student success.

The dataset includes the following features:

- Demographic Information: Province of residence (P), gender (G)
- Academic Performance: Scores in Mathematics (M), Literature (L), and English (E); scaled from 0 to 10.
- Admission Information: Chosen major (e.g., Information Technology (IT), Business Administration (BA), Graphic Design (GD)).
- Academic Outcomes: Course grades, semester GPA, pass/fail status, and year of study.

To ensure robustness, the dataset was randomly split into training (70%) and test (30%) subsets. This practice, commonly used in machine learning, prevents model overfitting and ensures reliable evaluation.

The target variable (dependent variable) in this study is students' academic performance, measured by their Grade Point Average (GPA). For simplicity, the GPA was categorized into:

- 0 if GPA < 6.5 (below average).
- 1 if GPA ≥ 6.5 (above average).

By transforming GPA into a binary classification task, the model aims to predict whether students will achieve above-average academic performance in their chosen major.

3.2 DEEP NEURAL NETWORK MODEL DESIGN

To analyze the relationship between student attributes and academic success, a Deep Neural Network model (DNN) was implemented with the following architecture:

- Input Layer: 6 nodes (corresponding to 6 input features: province, gender, three subject scores, and chosen major).
- Hidden Layers: 3 layers with 25 neurons each.
- Output Layer: 1 node (binary classification of academic performance).

The ReLU (Rectified Linear Unit) activation function was used in hidden layers to introduce non-linearity and improve learning capacity. A Sigmoid activation function was applied to the output layer for binary classification.

The Adamax optimizer was selected due to its adaptability in handling varying gradients during training. A learning rate of 0.005 was used to ensure stable convergence.

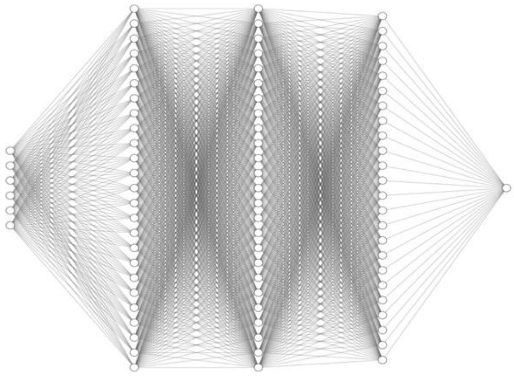

Fig.1 deep neural network model design

3.3 MODEL EVALUATION METRICS

The performance of the predictive model was evaluated using:

- Accuracy: Measures the proportion of correctly classified instances.
- Mean Squared Error (MSE): Assesses the difference between predicted and actual outcomes.
- Overfitting Analysis: Compares training and test accuracy to determine whether the model generalizes well to unseen data.

By leveraging deep learning techniques, this study seeks to enhance predictive ac-curacy and support evidence-based academic advising, ultimately helping students make informed decisions about their major selection.

The performance of the predictive model was evaluated using:

- Accuracy: Measures the proportion of correctly classified instances.
- Mean Squared Error (MSE): Assesses the difference between predicted and actual outcomes.
- Overfitting Analysis: Compares training and test accuracy to determine whether the model generalizes well to unseen data.

By leveraging deep learning techniques, this study seeks to enhance predictive accuracy and support evidence-based academic advising, ultimately helping students make informed decisions about their major selection.

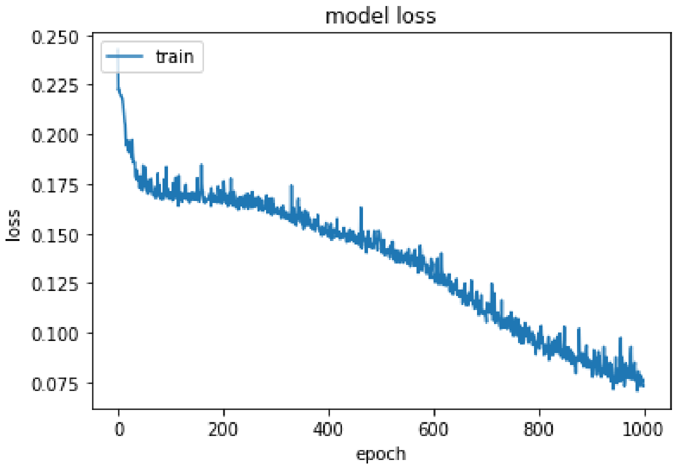

Fig.2 loss function value across epochs

IV. EXPERIMENT RESULTS

The evaluation of the predictive model revealed the following performance metrics:

- Training Accuracy:93%
- Test Accuracy:73%

These results indicate that the model performs well on the training data but experiences a slight drop in accuracy on the test set, suggesting potential overfitting. Although the performance gap is not severe, further optimization techniques such as regularization or hyperparameter tuning could enhance the model's generalizability.

Convergence Analysis: The Mean Squared Error (MSE) was monitored over 1000 epochs, showing stable convergence, which suggests that the model has successfully learned from the data without major oscillations. However, adding validation monitoring during training could provide further insights into optimal stopping points, reducing overfitting risks.

Predictive Capability: The model effectively predicts students' academic success based on their admission profiles. For example, when given the input:

[10, 0, 8.4, 4, 6.6, 0]

The model predicts the highest probability of success in Information Technology (IT) with a 77% likelihood of achieving a GPA above 6.5. In contrast, the probability of success in Business Administration (BA) and Graphic Design (GD) drops to 5% and 38%, respectively.

Table 1: Example of experimental results for the major selection prediction model

Input features						**Prediction of Learning Ability**			**Conclusion**
P	G	M	L	E	CM	IT (0)	BA (1)	GD (2)	
10	0	8,4	4,0	6,6	2	0,77	0,05	0,38	Right choice
0	0	7,8	5,75	6,4	1	0,49	0,18	0	Wrong choice

These insights suggest that the model can not only forecast students' academic performance but also recommend suitable majors based on their entrance scores and demographic information.

V. DISCUSSIONS

This predictive model offers valuable insights into the relationship between student potential and academic outcomes. The results confirm that factors such as academic scores, gender, and region of study significantly influence student performance. Specially, students with higher scores in core subjects tend to achieve better academic results, validating these scores as not just predictors but also reflections of actual learning capacity.

The model's ability to predict student success probabilities within different majors highlights its potential for personalized academic advising. However, the study acknowledges limitations, including the limited timeframe of the data, the simplification of academic performance, and the exclusion of certain non-academic factors.

Future improvements could involve expanding the dataset, incorporating more de-tailed academic performance metrics, integrating additional features (e.g., extracurricular activities, psychological assessments) and developing dynamic models that adapt to evolving educational trends. Combining quantitative analysis with qualitative research (e.g., interviews, surveys) could further enrich the understanding of factors influencing student success, leading to more holistic and personalized recommendations.

VI. CONCLUSIONS

This study demonstrates the potential of deep learning in predicting student academic performance and providing valuable insights for academic advising and major selection. The DNN model successfully predicts the probability of student success based on admission data, highlighting the influence of academic scores, gender, and study region on academic outcomes. While the model shows promising results, future research should address the limitations and explore potential improvements to enhance its accuracy, generalizability, and interpretability. By incorporating more comprehensive data and advanced techniques, we can develop even more effective tools to support students in their academic journeys and empower them to make informed decisions about their future.

REFERENCES

[1] K. Shaukat, I. Nawaz, S. Aslam, S. Zaheer, and U. Shaukat, "Student's performance in the context of data mining", International Multi-Topic Conference (INMIC), 2016: IEEE, pp. 1-8.
[2] M. Hussain, W. Zhu, W. Zhang, and S. M. R. Abidi, "Student Engagement Predictions in an e- Learning System and Their Impact on Student Course Assessment Scores", Computational intelligence and neuroscience, vol. 2018, no. 1, p. 6347186, 2018.
[3] A. Baskota and Y.-K. Ng, "A graduate school recommendation system using the multi-class support vector machine and KNN approaches", IEEE International Conference on Information Reuse and Integration (IRI), 2018: IEEE, pp. 277-284.
[4] S. A. El-Qulity, A. W. Mohamed, A. O. Bafail, and R. Abdelaal, "A multistage procedure for optimal distribution of preparatory-year students to faculties and departments: A mixed integer nonlinear goal programming model with enhanced differential evolution algorithm", Journal *of Computational and Theoretical Nanoscience,* vol. 13, no. 11, pp. 7847-7863, 2016.
[5] S. Fong, Y.-W. Si, and R. P. Biuk-Aghai, "Applying a hybrid model of neural network and decision tree classifier for predicting university admission", International Conference on Information, Communications and Signal Processing (ICICS), 2009: IEEE, pp. 1-5.
[6] S. Alghamdi, N. Alzhrani, and H. Algethami, "Fuzzy-Based Recommendation System for University Major Selection", in IJCCI, 2019, pp. 317-324.
[7] R. Zubaedah, M. Lintang, and N. P. Putra, "Decision support system for departemen selection for prospective students using the naïve bayes method and analytical hierarchy process model at faculty of engineering Universitas Musamus", IOP Conference Series: Materials Science and Engineering, 2021, vol. 1125, no. 1: IOP Publishing, p. 012030.
[8] R. M. Bautista, M. Dumlao, M. A. Ballera, and V.-Q. City, "Recommendation system for engineering students' specialization selection using predictive modeling", The Third International Conference on Computer Science, Computer Engineering, and Social Media (CSCESM2016), 2016, p. 34.
[9] S. Iyer and C. Variawa, "Using machine learning as a tool to help guide undeclared/undecided first-year engineering students towards a discipline", Proceedings of the Canadian Engineering Education Association (CEEA), 2019.
[10] Y. Meng and M. Fu, CMRS: Towards Intelligent Recommendation for Choosing College Majors, Proceedings of the 4th International Conference on Advances in Image Processing, Chengdu, China, 2021. [Online]. Available: https://doi.org/10.1145/3441250.3441272.
[11] V. Powar, S. Girase, D. Mukhopadhyay, A. Jadhav, S. Khude, and S. Mandlik, Analysing recommendation of colleges for students using data mining techniques, International Conference on Advances in Computing, Communication and Control (ICAC3), 2017: IEEE, pp. 1-5.
[12] S. N. Latifah, R. Andreswari, and M. A. Hasibuan, "Prediction analysis of Student specialization suitability using artificial neural network algorithm, International Conference on Sustainable Engineering and Creative Computing (ICSECC), 2019: IEEE, pp. 355-359.
[13] L. Al-Shalabi, A Data Mining Model for Students' Choice of College Major Based on Rough Set Theory, J. Comput. Sci, vol. 15, pp. 1150-1160, 2019.
[14] S. A. Stein, G. M. Weiss, Y. Chen, and D. D. Leeds, A College Major Recommendation System, *Proceedings of the 14th ACM Conference on Recommender Systems*, 2020, pp. 640-644.
[15] M. Xiao and H. Yi, Building an efficient artificial intelligence model for personalized training in colleges and universities, Computer Applications in Engineering Education, vol. 29, no. 2, pp. 350-358, 2021.
[16] H. A. Mengash, Using data mining techniques to predict student performance to support decision making in university admission systems, IEEE Access, vol. 8, pp. 55462-55470, 2020.
[17] K. Sethi, V. Jaiswal, and M. D. Ansari, Machine learning based support system for students to select stream (subject), Recent advances in computer science and communications (Formerly: Recent patents on computer science), vol. 13, no. 3, pp. 336-344, 2020.
[18] Y. Wei et al., An improved grey wolf optimization strategy enhanced SVM and its application in predicting the second major, Mathematical Problems in Engineering, vol. 2017, no. 1, p. 9316713, 2017.
[19] N. Kamal, F. Sarker, and K. A. Mamun, A comparative study of machine learning approaches for recommending university faculty, 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), 2020: IEEE, pp. 1-6.
[20] M. A. Al Ahmar, A prototype rule-based expert system with an object-oriented database for university undergraduate major selection, Proceeding of International Journal of Applied Information Systems (IJAIS), pp. 1-5, 2012.
[21] R. J. Salaki, C. R. Kawet, R. Manoppo, and F. Tumimomor, Decision support systems major selection vocational high school in using fuzzy logic android-based, Int. Conf. Electr. Eng. Informatics, Its Educ, 2015.
[22] J. Dhar and A. K. Jodder, An Effective Recommendation System to Forecast the Best Educational Program Using Machine Learning Classification Algorithms, Ingénierie des Systèmes d Inf., vol. 25, no. 5, pp. 559-568, 2020.

[23] G. Abosamra and A. Faloudah, Machine learning based marks prediction to support recommendation of optimum specialization and study track, *International Journal of Computer Applications,* vol. 975, p. 8887, 2019.

[24] M. Ezz and A. Elshenawy, Adaptive recommendation system using machine learning algorithms for predicting student's best academic program, *Education and Information Technologies,* vol. 25, no. 4, pp. 2733-2746, 2020.

[25] T. M. Alam, M. Mushtaq, K. Shaukat, I. A. Hameed, M. Umer Sarwar, and S. Luo, A novel method for performance measurement of public educational institutions using machine learning models, *Applied Sciences,* vol. 11, no. 19, p. 9296, 2021.