

AN EXPLAINABLE DEEP LEARNING FRAMEWORK FOR PERSONALIZED UNIVERSITY MAJOR RECOMMENDATION BASED ON STUDENTS' INTERESTS AND APTITUDES

**Septian Dwi Putra¹, Agus Suryana², Setyawan Widyarto³, Satria
Abadi⁴**

*Corresponding author
septiandwioke@gmail.com

Information Systems, Institut Bakti Nusantara, Lampung
Faculty of Communication, Visual Art & Computing, Universiti Selangor
Research Collaboration Program, Sultan Idris Education University, Malaysia
Wisma Rini Street, No. 09, Pringsewu, Lampung, Indonesia
Shah Alam Campus DU 007 (B), Selangor Darul Ehsan, Malaysia
Sultan Idris Education University, 35900 Tanjong Malim, Perak, Malaysia
E-mail: septiandwioke@gmail.com, suryana.suryani64@gmail.com,
swidyarto@unisel.edu.my, satriabadi@upsi.edu.my

Article history:

Received July 17, 2026

Revised July 29, 2026

Accepted August 9, 2026

Keywords:

Artificial Intelligence;

Deep Learning;

Holland RIASEC;

Student Aptitude;

Abstract

Selecting an appropriate university major significantly influences students' academic achievement, motivation, and career development. However, conventional academic guidance frequently relies on subjective assessments without adequately integrating students' interests, aptitudes, and academic performance. This study proposes an explainable deep learning framework for personalized university major recommendations using multidimensional student profiles collected from senior high schools across 20 subdistricts in Tanggamus Regency, Indonesia. The framework integrates Holland's RIASEC vocational interest theory, psychometric aptitude assessments, and academic achievement indicators. Twenty neural network configurations involving Feedforward Neural Networks (FFNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Bidirectional LSTM (Bi-LSTM) were evaluated using classification metrics. Comparative results identified Bi-LSTM as the best-performing architecture, achieving 91.7% accuracy, followed by LSTM (91.2%), GRU (90.5%), and dropout-based FFNN (90.1%). Furthermore, 84% agreement was observed between students' selected majors and system recommendations. Nevertheless, learning curves indicated substantial overfitting and limited validation performance, highlighting the need for independent verification. SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were incorporated to enhance recommendation transparency. This study contributes an integrated framework combining vocational interests, deep learning, and explainable artificial intelligence to support personalized academic guidance and evidence-based university major selection.

1.0. INTRODUCTION

Digital transformation in the education sector has encouraged the use of artificial intelligence (AI) technology to improve the quality of academic decision-making, including in the selection of study programs at the higher education level. The selection of study programs is a strategic decision that affects academic success, learning motivation, competency development, and career planning of students. However, the study program selection process still faces various challenges, especially when prospective students do not have an adequate understanding of the relationship between interests, talents, academic abilities, and characteristics of the available study programs. Conventional approaches that rely on teacher consultation, educational counseling, family preferences, and social tendencies have not always been able to provide measurable and consistent recommendations. In the context of educational decision-making, the use of AI offers an opportunity to develop a recommendation system that is able to analyze individual characteristics more systematically. A systematic review by Zawacki-Richter et al. [1] shows that AI in higher education has evolved in various fields, including prediction, assessment, and adaptive systems, but the involvement of pedagogical perspectives and user needs still needs to be strengthened.

The development of machine learning and deep learning allows the processing of complex educational data through the identification of nonlinear relationships between various student attributes. These data can include academic achievement, interest assessment results, numerical ability, verbal ability, logical reasoning, field of knowledge preferences, and non-academic activities. Compared to static rule-based approaches, deep neural networks have the ability to learn feature representations in stages, potentially finding relationship patterns that are difficult to identify through conventional analysis. LeCun et al. [2] explain that deep learning allows computational models to study data representations at various levels of abstraction. In the context of education, these capabilities can be leveraged to build a model of predicting the suitability of study programs based on students' multidimensional profiles. However, the effectiveness of deep learning is highly dependent on data quality, sample count, architectural selection, training strategies, and suitability of evaluation methods.

Recommender systems have evolved as a computational approach to generate relevant suggestions based on user characteristics and available alternatives. Zhang et al. [3] describe that deep learning-based recommender systems are able to study user and item representations and capture the complex relationships that support the personalization of recommendations. In the context of higher education, the concept can be applied by modeling prospective students as users and study programs as alternative recommendations. Student profiles that include interests, talents, and academic abilities can be used to generate suitability scores for a number of courses. This approach allows the system to provide several preferred study program alternatives based on individual characteristics, rather than simply assigning one category of majors. However, the implementation of an educational recommendation system requires special attention to the validity of suitability indicators, the availability of historical data, and possible bias in the decision-making process.

One of the main limitations of the use of deep learning models in educational decision-making is the low interpretability of prediction results. Complex neural network models are often categorized as black-box models because the relationship between input and output variables is not easily understood directly by the user. In study program recommendations, this condition is an important issue because students, parents, and counseling teachers need clear reasons regarding the suitability between the student's potential and the recommended field of study. Adadi and Berrada [4] identify that Explainable Artificial Intelligence (XAI) is an important approach to improve transparency and understanding of AI-based system decisions. Furthermore, Arrieta et al. [5] emphasize the importance of developing explainable AI to support transparency, accountability, and responsible use of models. Therefore, it is not enough for the program recommendation system to produce predictions with high accuracy, but also needs to provide explanations that can be understood and evaluated by education stakeholders.

The XAI approach allows the identification of the contribution of each attribute to the outcome of the recommendation through model interpretation methods, such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). Ribeiro et al. [6] introduced LIME as an approach to explain individual predictions through local interpretive models, while Lundberg and Lee [7] developed SHAP as a feature attribution framework based on Shapley values. Both approaches can be used to explain the contribution of the attributes of interest, talent, and academic ability to the study program recommendations. For example, the recommendation of the Informatics study program can be accompanied by an explanation of the contribution of logic skills, mathematical reasoning, and interest in technology. However, the value of the contribution of features needs to be understood as an explanation of the model's prediction behavior, not as evidence of a cause-and-effect relationship between students' attributes and success in a particular course of study.

In addition to interpretability, the education recommendation system needs to consider aspects of fairness and reliability. Education data can contain distribution imbalances based on school origin, socioeconomic conditions, geographical area, and other demographic characteristics. If these conditions are not taken into account, the model has the potential to produce recommendations that are less representative for certain groups of students. Mehrabi et al. [8] explain that bias in machine learning systems can arise from the process of data collection, selection of attributes, and algorithm development. In the program recommendation system, fairness evaluation is important to ensure that students with comparable academic characteristics and interests obtain reasonable recommendation opportunities. Thus, model development needs to integrate prediction performance measurement, bias analysis, and explanation consistency testing so that the results of recommendations can be used more responsibly.

Tanggamus Regency, Lampung Province, became the empirical context of this research by involving prospective students from high school/vocational school/MA in 20 sub-districts. The diversity of school characteristics, educational backgrounds, academic abilities, and study program preferences provide an opportunity to evaluate the application of the recommendation model in a heterogeneous educational environment. In practice, prospective students may face limited access to study program information and structured career counseling services. Therefore, a data-based approach is needed that is able to integrate the results of interest and talent assessments with academic indicators to support the study program selection process. This study was designed using Exploratory Data Analysis (EDA), data preprocessing, and experiments on 20 neural network architecture configurations. The experiment was intended to identify the model configuration that provides the best balance between predictive ability, computational complexity, stability, and interpretability.

Based on the development of previous research, there is a need to develop a study program recommendation system that is not only oriented towards improving classification performance, but also able to provide personalized recommendations along with relevant explanations for users. Most of the foundations of deep learning research, recommendation systems, and XAI develop through different study paths, so the integration of the three in one educational evaluation framework is an important research direction to be empirically tested. This study proposes An Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes, which is a framework that integrates multidimensional representation of student profiles, deep neural network modeling, study program suitability ranking, and interpretation of recommendation results using the XAI method. Evaluation is carried out through model performance comparison, recommendation quality testing, interpretability analysis, and intergroup fairness checks. The expected contribution is the availability of a more personalized, transparent, and scientifically testable academic recommendation approach to support the decision of prospective students in choosing a college study program.

2.0. LITERATURE REVIEW

This literature review focuses on artificial intelligence, deep learning, educational recommender systems, student profiling, academic performance prediction, and

explainable artificial intelligence (XAI). The objective is to identify methodological developments and research gaps supporting the proposed study entitled “An Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes.” The review distinguishes between studies directly related to educational recommendations and foundational studies that provide methodological support for deep learning, personalization, interpretability, and fairness. The five references in the original table are treated as preliminary because their bibliographic details and reported findings require verification. The updated review therefore prioritizes identifiable international publications rather than carrying forward potentially inaccurate claims.

The following 15 publications provide direct and methodological evidence for the proposed research. The limitations column identifies either limitations described in the source or scope-related gaps inferred for the present study; these should not automatically be interpreted as weaknesses acknowledged by the original authors.

No	Authors and Year	Research Title	Methods	Key Findings	Limitations / Research Gap
1	Yu, Ning, and Li (2024)	A Recommendation Model for College Majors Based on Deep Learning and Clustering Algorithms	DNN, K-Means, DBSCAN, BIRCH	Combines academic performance prediction and clustering to recommend suitable college majors.	Relies substantially on historical university learning data; explainability and pre-university aptitude assessment are not central.
2	Yang et al. (2025)	An Explainable Graph-Based Course Recommendation Model Based on Multiple Interest Factors	Knowledge graphs, bipartite graphs, interest modeling	Improves course recommendation performance and provides interpretable recommendations using learner–course relationships.	Focuses on online course recommendation rather than university major selection.
3	Yang et al. (2024)	Boosting Course Recommendation Explainability: A Knowledge Entity Aware Model Using Deep Learning	Neural networks, knowledge graphs, KEAM	Develops explainable student profiles and validates recommendation effectiveness on two datasets.	Does not directly evaluate aptitude-based undergraduate major selection.
				PKP library.apsce.net	
4	Wagner et al. (2024)	A Course Recommender System Built on Success to Support Students at Risk in Higher Education	Explainable KNN, student similarity	Recommends courses based on the academic pathways of successful students.	Targets enrolled students and course planning, not prospective students choosing majors.
5	Ma, Yang, and Ren (2024)	A Survey on Explainable Course Recommendation Systems	Literature review, explainable recommendation taxonomy	Identifies the growing importance of transparency and justification in educational	Provides a conceptual synthesis rather than experimental validation of a

No	Authors and Year	Research Title	Methods	Key Findings	Limitations / Research Gap
				recommendations. Kyushu University Pure Portal Site	major recommendation framework.
6	Li et al. (2025)	Improving Course Recommendation Systems with Explainable AI: LLM-Based Frameworks and Evaluations	Deep recommender systems, LLM-based post-hoc explanations	Proposes a modular approach to improving explanations for educational recommendations.	Focuses on explanation generation for course recommendations; major-aptitude matching remains outside its main scope.
7	Authors of Interpretability-Driven Course Recommendation (2026)	Interpretability-Driven Course Recommendation: A Random Forest Approach with Explainable AI	Random Forest, SHAP, Top-3 recommendation	Reports precision of 0.90, recall of 0.92, and F1-score of 0.92 for its evaluated recommendation model.	Uses a tree-based model and academic marks; integration of aptitude, interests, and deep learning requires further investigation.
8	Authors of Explainable Learning Analytics (2024)	Explainable Learning Analytics: Assessing the Stability of Student Success Prediction Models by Means of Explainable AI	Predictive learning analytics, SHAP, stability analysis	Demonstrates the value of SHAP for assessing model stability and supporting academic advice.	Concentrates on student success prediction rather than personalized university major ranking.
9	Zawacki-Richter et al. (2019)	Systematic Review of Research on Artificial Intelligence Applications in Higher Education—Where Are the Educators?	Systematic literature review	Identifies major applications of AI in higher education and the need for stronger educational perspectives.	Does not propose a personalized, explainable major recommendation architecture.
10	Zhang et al. (2019)	Deep Learning Based Recommender System: A Survey and New Perspectives	Deep recommender systems, neural collaborative filtering and related architectures	Establishes methodological foundations for learning complex user-item relationships.	General-purpose survey; educational aptitude and major selection are not its primary focus.
11	LeCun, Bengio, and Hinton (2015)	Deep Learning	Deep neural networks, representation learning	Establishes the theoretical basis for hierarchical feature learning.	Foundational study rather than an educational

No	Authors and Year	Research Title	Methods	Key Findings	Limitations / Research Gap
					recommendation application.
12	Adadi and Berrada (2018)	Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)	XAI literature review	Summarizes approaches for improving AI transparency and interpretability.	Does not empirically validate XAI in university major recommendation.
13	Barredo Arrieta et al. (2020)	Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI	XAI taxonomy, responsible AI review	Provides a conceptual foundation for explainability and responsible model evaluation.	Requires adaptation and validation in educational decision-support settings.
14	Ribeiro, Singh, and Guestrin (2016)	"Why Should I Trust You?": Explaining the Predictions of Any Classifier	LIME	Introduces local, model-agnostic explanations of individual predictions.	Explanation fidelity and stability require evaluation when applied to educational recommendations.
15	Lundberg and Lee (2017)	A Unified Approach to Interpreting Model Predictions	SHAP, Shapley values	Introduces a unified feature-attribution framework for explaining model predictions.	Feature contributions do not independently establish causal relationships or educational suitability.

The table includes recent domain-specific studies and established foundational literature. The 2026 course recommendation article is especially relevant to the proposed XAI component, while Yu et al. (2024) is particularly important because it directly addresses college major recommendations. Recent research demonstrates increasing interest in integrating artificial intelligence with personalized educational recommendation systems. Yu et al. [1] developed a college major recommendation model combining deep neural networks and clustering algorithms. Their approach uses historical academic information to estimate student performance and recommend appropriate majors. This study provides a direct methodological foundation for major recommendation; however, its emphasis on historical academic records leaves opportunities for incorporating validated interest and aptitude assessments before university enrollment. Yang et al. [2] proposed an explainable graph-based course recommendation model incorporating multiple interest factors. Their approach uses learner-course relationships and knowledge graphs to improve recommendation quality while supporting interpretability. Similarly, Yang et al. [3] developed a knowledge entity-aware model that constructs explicit learner profiles through neural networks and knowledge representations. These studies demonstrate that personalized recommendations can benefit from structured representations of learner preferences. Nevertheless, their primary application concerns course selection rather than matching prospective students to university majors.

Wagner et al. [4] investigated an explainable course recommendation system based on the academic trajectories of successful students. Their findings indicate that student similarity and historical achievement patterns can support educational decision-making. However, recommendation based on prior student success does not necessarily capture

prospective students' psychological interests, aptitude profiles, or long-term career preferences. Ma et al. [5] further emphasized that educational recommendation systems should provide understandable justifications rather than relying exclusively on predictive accuracy. More recent developments have expanded the role of explainability in educational recommender systems. Li et al. [6] proposed an LLM-based explanation framework for deep learning-driven course recommendations, while a 2026 study [7] combined Random Forest and SHAP to produce interpretable Top-3 course recommendations for students transitioning into higher education. These approaches demonstrate the increasing importance of explanation quality, but also suggest opportunities for investigating whether deep neural architectures can provide improved representations of combined academic, interest, and aptitude information. The broader literature provides additional methodological support. Zhang et al. [10] established the foundations of deep learning-based recommendation systems, while LeCun et al. [11] demonstrated the ability of deep neural networks to learn hierarchical representations. Adadi and Berrada [12] and Barredo Arrieta et al. [13] highlighted the importance of explainability and responsible AI. Ribeiro et al. [14] and Lundberg and Lee [15] introduced LIME and SHAP, respectively, which can be adapted to investigate how individual student attributes influence recommendation scores.

Taken together, these studies indicate that deep learning, personalization, and explainability have each received substantial research attention. However, the reviewed evidence does not establish that their integration has been comprehensively evaluated for aptitude-based university major recommendation among prospective Indonesian students. This creates an opportunity to investigate a unified framework combining multidimensional student profiling, neural recommendation models, explainable predictions, and systematic performance evaluation.

Holland's Theory of Vocational Personality (RIASEC)

Holland's theory of vocational personalities and work environments provides a theoretical foundation for understanding the relationship between individual interests, personality characteristics, and educational or occupational preferences. Holland proposed that vocational interests can be classified into six major categories: Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC) [16]. The theory suggests that individuals are more likely to experience satisfaction and persistence when their personal interests and characteristics correspond to the environments in which they study or work. This concept, commonly referred to as person–environment fit, provides a useful basis for examining the compatibility between prospective students and university majors.

The Realistic personality type is generally associated with practical activities, technical skills, and hands-on problem-solving, while the Investigative type emphasizes analytical thinking, scientific reasoning, and intellectual exploration. The Artistic type is characterized by creativity and self-expression, whereas the Social type reflects preferences for interpersonal interaction, teaching, and helping others. The Enterprising type is associated with leadership, persuasion, and entrepreneurial activities, while the Conventional type emphasizes structured procedures, organization, and systematic information processing. These dimensions can be represented as quantitative interest scores and used to construct individual student profiles. However, RIASEC categories should not be treated as rigid classifications because students may demonstrate combinations of multiple vocational interests.

The relevance of Holland's theory to university major recommendation lies in its ability to provide a theoretically grounded representation of student interests. Rather than recommending academic programs exclusively based on examination scores, an RIASEC-based approach allows the system to consider whether students' vocational interests correspond to the characteristics of particular academic disciplines. For example, students with relatively strong Investigative and Realistic interests may demonstrate preferences for engineering or computing-related disciplines, while students with Social and Enterprising interests may prefer educational, managerial, or communication-oriented programs. Such associations should be interpreted as probabilistic indicators rather than deterministic rules.

Within the proposed framework, the six RIASEC dimensions serve as explanatory input variables for deep learning-based recommendation. Their integration is consistent with the development of personalized educational recommendation systems, particularly the work of Yu et al. [1], which investigated college major recommendations using deep learning and clustering, and Yang et al. [2], [3], which explored interest-aware educational recommendation approaches. Unlike models relying primarily on historical academic records, the proposed study incorporates theoretically derived vocational interest profiles to strengthen the personalization of university major recommendations.

Psychometric Theory of Interests and Aptitudes

Psychometric theory provides a scientific foundation for measuring psychological constructs that cannot be observed directly, including vocational interests, cognitive abilities, personality characteristics, and educational aptitudes. According to classical test theory, observed test scores can be conceptualized as the combination of true scores and measurement error. Lord and Novick [17] established fundamental principles of psychological measurement, emphasizing the importance of reliability, validity, and measurement precision. These principles are essential when psychometric results are used as input variables in artificial intelligence models because inaccurate or inconsistent measurements may reduce recommendation reliability.

Vocational interests and aptitudes represent related but conceptually distinct dimensions. Interests refer to an individual's preferences for particular activities, subjects, or occupational environments, whereas aptitudes represent capacities or potential abilities associated with learning and performing specific tasks. Cognitive aptitude assessments may include numerical reasoning, verbal reasoning, logical reasoning, spatial ability, and problem-solving skills. In contrast, vocational interest assessments commonly measure preferences through instruments grounded in Holland's RIASEC model. Combining these dimensions allows the proposed recommendation framework to represent both what students prefer to study and the abilities that may support their learning in particular academic fields.

Cronbach and Meehl [18] introduced the concept of construct validity, emphasizing that psychological measurement must be supported by theoretical and empirical evidence connecting observed scores to the constructs they are intended to represent. This principle is particularly important when psychometric instruments are integrated into machine learning systems. Before model development, the proposed study should evaluate the validity and reliability of interest and aptitude measurements, including internal consistency, construct validity, and measurement equivalence where appropriate. Standardized or validated instruments are preferable to untested questionnaires because predictive models cannot independently establish the psychological validity of their input variables.

In the proposed research, psychometric scores are transformed into structured numerical features and combined with academic indicators, such as subject achievement and educational background. This multidimensional representation provides a foundation for identifying complex relationships between student characteristics and university major suitability. The integration of psychometric and academic variables extends the personalized recommendation approaches investigated by Yu et al. [1] and Yang et al. [2], [3]. Nevertheless, the predictive contribution of each variable group must be evaluated empirically through feature importance analysis and ablation experiments to determine whether combining interests, aptitudes, and academic achievement improves recommendation performance.

Artificial Neural Network and Deep Learning Theory

Artificial neural networks are computational models composed of interconnected processing units that transform input information through weighted connections, nonlinear activation functions, and successive computational layers. The mathematical foundations of neural network learning were significantly advanced by Rumelhart, Hinton, and Williams [19], who demonstrated how backpropagation could be used to adjust network weights through error propagation. This learning mechanism enables neural networks to identify relationships between input variables and target outputs through iterative optimization.

Deep learning extends conventional neural network approaches by incorporating multiple representation-learning layers. LeCun, Bengio, and Hinton [11] explained that deep neural architectures can learn hierarchical representations of data, allowing complex patterns to be identified through successive transformations. In educational recommendation systems, these capabilities may support the identification of nonlinear relationships among vocational interests, cognitive aptitudes, academic achievement, and program suitability. However, the use of deeper architectures does not automatically guarantee better performance, particularly when datasets are relatively small or consist primarily of structured tabular variables.

Several neural network architectures may be considered for the proposed framework. Multilayer Perceptrons (MLPs) are suitable for structured numerical and categorical student data because they can model nonlinear relationships among input features. Convolutional Neural Networks (CNNs) are primarily designed for data containing spatial or local structural relationships, while Long Short-Term Memory (LSTM) networks are intended for sequential information. Transformer-based architectures use attention mechanisms to model relationships among input elements and may be adapted to structured educational data when supported by an appropriate methodological justification. Therefore, architecture selection should be based on the characteristics of the available dataset rather than assuming that more complex models necessarily provide superior recommendations.

The theoretical relevance of deep learning to educational recommendation has been discussed by Zhang et al. [10], who reviewed neural approaches for modeling user preferences and recommendation relationships. More specifically, Yu et al. [1] demonstrated the application of deep neural networks and clustering algorithms to college major recommendation, while Yang et al. [3] investigated neural representation learning in explainable course recommendation. These studies support the feasibility of applying deep learning to educational decision-support problems, although the effectiveness of different architectures for prospective students' interest- and aptitude-based major selection remains an empirical question. In this study, deep learning is employed to estimate the compatibility between student profiles and available university majors. The proposed experimental design compares 20 neural network configurations under consistent training, validation, and testing conditions. The models are evaluated using classification metrics, including accuracy, precision, recall, and macro-F1, alongside recommendation metrics such as Precision@K, Recall@K, and normalized discounted cumulative gain (NDCG@K). This evaluation strategy aims to identify models that provide both reliable predictions and relevant personalized major rankings.

Machine Learning-Based Recommendation Systems

Recommendation systems are computational approaches designed to identify and prioritize alternatives that are relevant to individual users based on their characteristics, preferences, historical interactions, or contextual information. Ricci, Rokach, and Shapira [20] describe recommender systems as information-filtering technologies that support personalized decision-making by predicting user preferences and identifying suitable items. Conventional recommendation techniques commonly include content-based filtering, collaborative filtering, and hybrid approaches. Each method offers different advantages depending on data availability, user characteristics, and recommendation objectives. In the context of university major selection, prospective students represent users, while academic programs represent recommendation items. Student profiles may include vocational interest scores, aptitude assessment results, academic performance, and educational preferences. The characteristics of university majors can be represented through program descriptions, curriculum requirements, disciplinary competencies, and relevant vocational interest profiles. A content-based recommendation approach can therefore estimate compatibility between student characteristics and program attributes, while machine learning models can identify patterns from historical examples of student-major suitability.

Deep learning-based recommender systems extend conventional approaches by learning complex representations of users and items. Zhang et al. [10] highlighted the ability of neural recommendation models to capture nonlinear interactions and latent

relationships. Yu et al. [1] applied deep learning and clustering to college major recommendation, demonstrating the relevance of data-driven approaches for academic guidance. Furthermore, Yang et al. [2] and Yang et al. [3] investigated educational recommendation systems that incorporate interest factors and knowledge representations, while Wagner et al. [4] explored explainable recommendations based on student academic success. These studies provide a foundation for developing personalized recommendation models that consider individual learner characteristics. The proposed research adopts a personalized recommendation approach in which the deep learning model produces suitability scores for multiple university majors. Instead of limiting the output to a single predicted major, the framework ranks alternative programs according to their estimated compatibility with each student's interests, aptitudes, and academic profile. This approach supports Top-K recommendations, enabling students and educational counselors to compare several potentially suitable programs. Importantly, the suitability scores should be interpreted as model-based estimates rather than definitive measures of future academic success.

Explainable Artificial Intelligence (XAI) Theory

Explainable Artificial Intelligence refers to methods and approaches designed to make the predictions, decisions, or internal behavior of artificial intelligence models more understandable to human users. Adadi and Berrada [12] emphasized that interpretability is particularly important when AI systems support decisions that require transparency and accountability. Similarly, Barredo Arrieta et al. [13] developed a comprehensive taxonomy of XAI methods and highlighted the relationship between explainability, responsible AI, and trustworthy decision-making. The importance of explainability is especially relevant to university major recommendation because students and educational counselors require understandable reasons for recommended academic programs. Although deep neural networks may identify complex patterns among student attributes, their predictions may be difficult to interpret directly. A recommendation stating that a student is suitable for Computer Science, for example, provides limited guidance unless the system can explain how the student's logical reasoning, numerical aptitude, vocational interests, and academic achievement influenced that recommendation.

Two widely adopted model-agnostic explanation methods are Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP). Ribeiro, Singh, and Guestrin [14] introduced LIME to approximate individual model predictions using simpler interpretable models. Lundberg and Lee [15] proposed SHAP as a unified feature-attribution framework based on Shapley values. Within the proposed framework, these methods can be used to estimate the contribution of individual student characteristics to recommended university majors.

Recent educational recommendation studies provide further support for integrating explainability into personalized learning systems. Yang et al. [3] incorporated explainable representations into neural course recommendation, while Ma et al. [5] reviewed methodological developments in explainable educational recommender systems. Li et al. [6] further investigated explanation generation for deep learning-based course recommendations. These studies indicate that explainability is increasingly recognized as an important component of educational recommendation quality, although its effectiveness must be evaluated in relation to specific user needs and decision contexts. In the proposed study, XAI is integrated into the recommendation process to provide global and local explanations of model behavior. Global explanations identify influential features across the student population, while local explanations describe the contributions of individual attributes to a particular recommendation. Explanation quality is evaluated through stability, fidelity, and interpretability assessments. Importantly, feature attribution does not establish causality; therefore, XAI outputs should be interpreted as explanations of model behavior rather than definitive evidence that a particular aptitude causes academic success.

Conceptual Integration of the Theoretical Framework

The theoretical integration of Holland's vocational personality theory, psychometric measurement, deep learning, recommendation systems, and XAI establishes the conceptual foundation of the proposed framework. Holland's RIASEC theory provides a structured representation of vocational interests, while psychometric theory supports the measurement of student aptitudes and psychological characteristics. Deep learning enables the identification of complex relationships among these variables, and recommendation system theory provides the mechanism for ranking university majors according to estimated suitability. XAI contributes an interpretability layer that helps students and educational counselors understand the factors influencing individual recommendations. The proposed framework begins with collecting and validating multidimensional student data, including RIASEC interest scores, cognitive aptitude measurements, and academic performance indicators. These data are processed through exploratory data analysis, feature encoding, normalization, and appropriate strategies for handling missing values and class imbalance. The resulting features are used to train and evaluate multiple neural network configurations. The selected model generates suitability scores and Top-K university major recommendations, while SHAP and LIME provide explanations of the predicted outcomes. The integration of these theoretical components is expected to contribute to a more comprehensive academic recommendation approach by combining psychological measurement with predictive modeling and interpretable decision support. However, the framework's effectiveness must be established through empirical validation, including comparisons with conventional machine learning methods, evaluation of recommendation quality, assessment of explanation reliability, and analysis of potential performance differences across student groups.

3.0. METHODS

3.1. Data Collection Methods

This study employed a field-based data collection approach combined with institutional administrative records to develop a comprehensive dataset for an Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes. Data were collected from senior high schools (SMA), vocational high schools (SMK), and Islamic senior high schools (MA), including public and private educational institutions across 20 subdistricts in Tanggamus Regency, Lampung Province, Indonesia. The data collection process focused on Grade XII students who were preparing to pursue higher education, representing diverse demographic characteristics, academic backgrounds, vocational interests, and cognitive aptitudes. This geographical and institutional coverage was intended to capture variations in student characteristics and provide an empirical foundation for developing personalized university major recommendations.

The dataset was structured to contain 24 attributes representing demographic information, academic achievement, vocational interests, aptitude assessment results, and university major preferences. Demographic attributes included anonymized student identifiers, gender, and age, while academic attributes consisted of achievement scores in core subjects relevant to higher education readiness. Vocational interests were assessed using Holland's Realistic, Investigative, Artistic, Social, Enterprising, and Conventional (RIASEC) framework, which categorizes individual interests into six dimensions associated with different educational and occupational environments [16]. These dimensions were represented as quantitative scores to capture individual interest profiles rather than assigning students to a single personality category. Aptitude-related attributes included verbal reasoning, logical reasoning, visual-spatial ability, musical aptitude, and kinesthetic aptitude, providing additional information about students' cognitive and individual capabilities. The inclusion of these attributes was informed by psychometric measurement principles emphasizing construct validity, reliability, and the systematic assessment of individual differences [17], [18].

In addition to student characteristics, the dataset included students' primary and alternative university major preferences, as well as information about the majors actually selected upon enrollment in higher education, where such information was available. These variables were intended to support the analysis of relationships between students' interests,

aptitudes, academic performance, and educational choices. For supervised model development, the reference target labels should be established independently of the proposed artificial intelligence system, preferably through validated counselor assessments, expert consensus, or other clearly defined educational suitability criteria. AI-generated recommendations and indicators of agreement between actual choices and model recommendations should be retained as model outputs or evaluation variables rather than used as independent ground-truth labels. This distinction is essential to prevent data leakage and circular validation during model training and evaluation.

Data collection was conducted in coordination with participating schools, educational authorities, and guidance and counseling teachers (Bimbingan dan Konseling, BK). Guidance counselors supported the administration or verification of vocational interest and aptitude assessments, while school administrative records provided relevant academic and demographic information. The measurement process was designed to follow psychometric principles, including instrument validity, reliability, and consistency of assessment procedures [17], [18]. Student data were to be anonymized before computational analysis, and access to personally identifiable information was restricted in accordance with applicable research ethics and data protection requirements. Because the study involved school-age participants, appropriate institutional authorization, informed consent or assent procedures, and parental or guardian permission where required were essential components of the data collection protocol.

The collected data served as the input for a personalized university major recommendation framework integrating deep learning and explainable artificial intelligence. Following data collection, exploratory data analysis (EDA) and preprocessing procedures were used to examine attribute distributions, missing values, outliers, class imbalance, and relationships among student characteristics. The processed data were then intended for training and evaluating 20 neural network configurations to identify suitable models for predicting and ranking university major alternatives. Consistent with the theoretical foundations of deep learning-based recommendation systems [1], [10], [11], the framework aimed to learn complex relationships between multidimensional student profiles and academic program suitability. Explainability methods, particularly SHAP and LIME [14], [15], were incorporated to identify the contributions of interest, aptitude, and academic attributes to individual recommendations.

To ensure methodological rigor, data partitioning should account for possible similarities among students from the same schools, and all preprocessing parameters should be estimated exclusively from training data. The evaluation strategy should assess classification performance, Top-K recommendation relevance, model generalization, and explanation consistency. Furthermore, subgroup analysis across school types and geographical areas can help identify potential disparities in model performance. Through these procedures, the dataset is intended to provide a contextually relevant empirical foundation for developing a transparent, personalized, and evidence-based university major recommendation system for prospective students in Tanggamus Regency.

3.2. Data Preprocessing and Deep Learning Model Development

This study employed a systematic deep learning pipeline to develop an explainable and personalized university major recommendation framework based on students' interests, aptitudes, and academic performance. The research process consisted of data preprocessing, feature transformation, dataset partitioning, neural network development, model training, performance evaluation, and model selection. Twenty neural network configurations were evaluated to identify the most suitable architecture for multiclass university major recommendation.

3.2.1. Categorical Data Encoding

Categorical attributes, including gender and university major preferences, were transformed into numerical representations using label encoding and one-hot encoding. One-hot encoding was applied to nominal input variables to prevent artificial ordinal relationships, while label encoding was used to represent the target university major categories as integer class indices. The encoded target labels were subsequently converted

into one-hot vectors for multiclass classification using categorical cross-entropy. Only independently established reference labels were used as training targets, while actual major selections and AI-generated recommendations were retained for appropriate evaluation purposes.

3.2.2. Feature Standardization

Numerical attributes, including academic scores, Holland's RIASEC interest scores, and aptitude assessment results, were standardized using StandardScaler. This transformation was applied to improve numerical consistency and facilitate neural network optimization. Standardization was calculated exclusively from the training data, and the resulting parameters were applied to validation and testing datasets to prevent data leakage.

4.0. RESULTS

Principal Component Analysis (PCA) Visualization of Student Profiles Based on University Majors

Figure 1 presents a two-dimensional Principal Component Analysis (PCA) visualization of student profiles across various university majors, supporting the development of An Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes. PCA was employed as an exploratory data analysis technique to reduce multidimensional student characteristics into two principal components, PCA1 and PCA2. Each colored point represents an individual student, while different colors indicate university major categories, including Medicine, Pharmacy, Accounting, Midwifery, Nursing, Computer Science, Engineering, Psychology, and Management. The visualization provides an overview of the distribution of student profiles and the relationships among academic interests, aptitudes, and major categories within a reduced feature space.

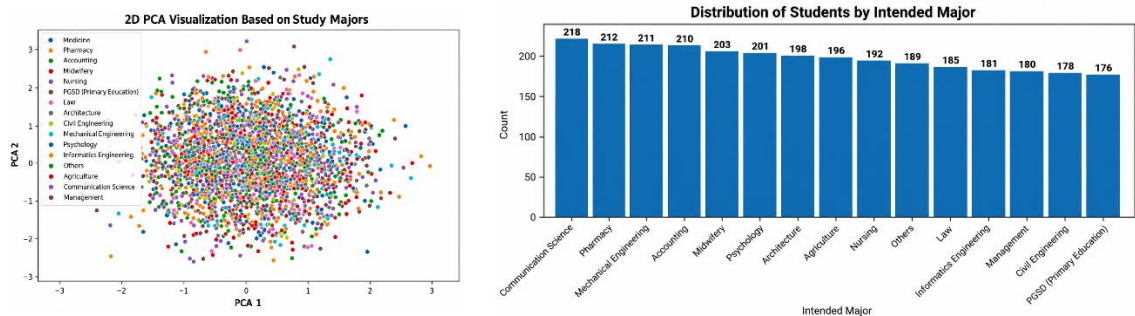


Figure 1 presents a two-dimensional Principal Component Analysis (PCA) visualization of student profiles across various university majors

The visualization reveals substantial overlap among the different university major categories, with most observations concentrated around the central region of the PCA coordinate system. Although some observations extend toward the outer regions, no clearly separated clusters are evident for individual majors. This distribution suggests that students associated with different academic disciplines may share similar characteristics within the first two principal components. Consequently, conventional linear separation may be insufficient to distinguish all university major categories effectively. However, overlapping distributions in a two-dimensional PCA projection do not necessarily indicate that the original multidimensional data are inseparable. These findings provide a rationale for investigating deep neural networks capable of learning complex nonlinear relationships among students' vocational interests, cognitive aptitudes, and academic characteristics. In relation to the proposed explainable deep learning framework, the PCA visualization serves as an initial exploratory assessment of the dataset rather than a direct measurement of recommendation accuracy. The observed overlap highlights the importance of evaluating multiple neural network architectures to identify models capable of capturing subtle differences in student profiles and generating personalized university major

recommendations. Furthermore, integrating SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) into the selected model will enable the investigation of how individual attributes contribute to each recommendation. Therefore, the PCA analysis supports the subsequent model development and evaluation stages by illustrating the complexity of the student feature space and emphasizing the need for accurate, interpretable, and evidence-based university major recommendations.

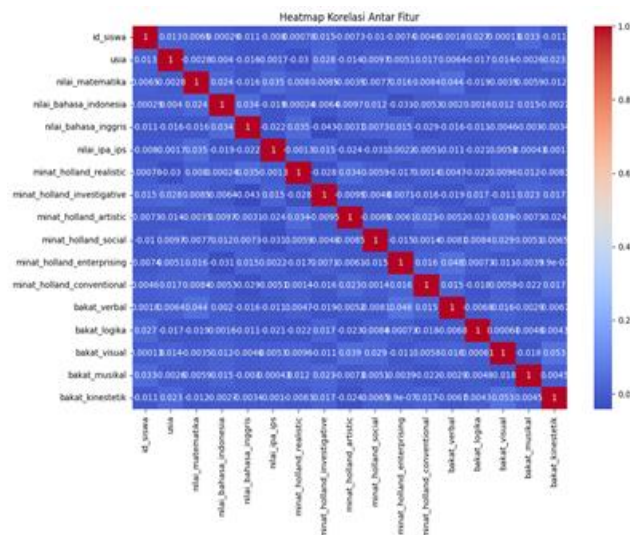


Figure 2. Pearson Correlation Heatmap of Student Academic Performance and Talent Features

The correlation heatmap illustrates the relationships among students' demographic characteristics, academic performance, Holland's RIASEC vocational interests, and aptitude assessment scores used in the proposed Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes. The visualization indicates that most feature pairs exhibit weak linear correlations, with coefficients generally close to zero, suggesting limited linear dependence among the measured attributes. The diagonal values of 1.0 represent perfect self-correlations, while the predominantly blue-colored cells indicate relatively low correlation coefficients. These findings suggest that academic achievement, vocational interests, and aptitude dimensions may provide complementary information for constructing multidimensional student profiles, although weak pairwise correlations alone do not establish feature independence or predictive relevance. Furthermore, the absence of strong correlations among most variables suggests a relatively low risk of severe pairwise multicollinearity. Nevertheless, nonlinear relationships and interactions between attributes may still exist and require further investigation. Therefore, the correlation analysis provides preliminary support for evaluating deep learning architectures capable of identifying complex feature interactions and generating personalized university major recommendations. The subsequent integration of SHAP and LIME will help explain the contributions of individual attributes to model predictions, improving the transparency and interpretability of the proposed recommendation framework.

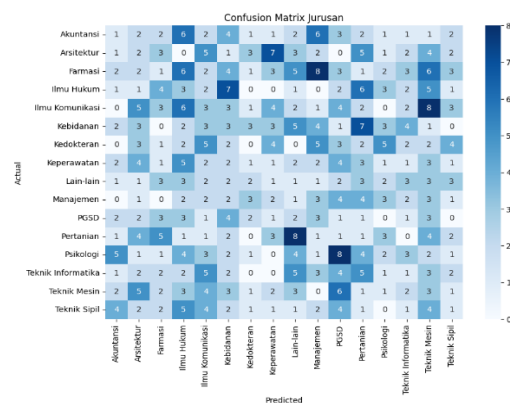


Figure 3. Confusion Matrix of University Major Classification

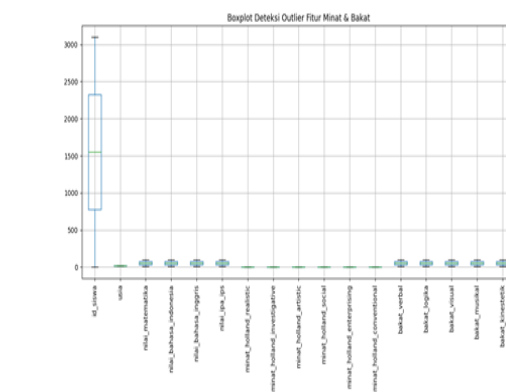


Figure 4. Boxplot Analysis of Outliers in Students' Interests and Aptitudes

Figure 3. Confusion Matrix of University Major Classification. The confusion matrix illustrates the classification performance of the deep learning model in predicting university majors based on students' academic performance, vocational interests, and aptitude characteristics. The matrix compares actual university major categories with predicted classifications, including Accounting, Architecture, Pharmacy, Computer Science, Medicine, Nursing, Psychology, and Engineering. Correct predictions are represented by diagonal cells, while off-diagonal values indicate misclassifications between different majors. The visualization reveals that correct classifications are relatively limited and prediction errors are distributed across multiple categories, suggesting considerable overlap in the characteristics of students associated with different academic disciplines. This finding indicates that the evaluated model has difficulty distinguishing among university majors using the available features. Consequently, further improvements in feature representation, neural network architecture, class imbalance handling, and hyperparameter optimization are necessary to enhance classification performance. Within the proposed Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes, these results highlight the importance of evaluating model reliability before generating personalized academic recommendations.

Figure 4. Boxplot Analysis of Outliers in Students' Interests and Aptitudes. The boxplot visualization presents the distribution and potential outliers of numerical attributes, including student age, academic achievement, Holland's RIASEC vocational interest scores, and aptitude assessment results. The visualization indicates substantial differences in feature scales, particularly the student identifier variable (*id_siswa*), which exhibits a much larger numerical range than the remaining attributes. This variable represents an identification code rather than a meaningful predictive characteristic and should therefore be excluded from model training. Academic, interest, and aptitude attributes appear compressed near the lower region of the shared vertical axis because of these scale differences, making their individual distributions and potential outliers difficult to interpret. Therefore, the visualization does not provide sufficient evidence to conclude that these attributes contain significant outliers without additional feature-specific analysis. Appropriate preprocessing, including identifier removal, numerical standardization, and outlier assessment using the interquartile range (IQR), can improve data quality and facilitate stable neural network training. These preprocessing procedures are essential for developing a reliable and interpretable personalized university major recommendation framework.

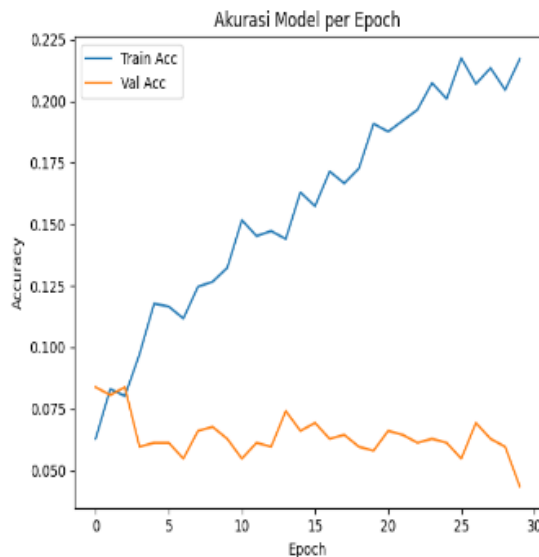


Figure 5. Training and Validation Accuracy Across Epochs

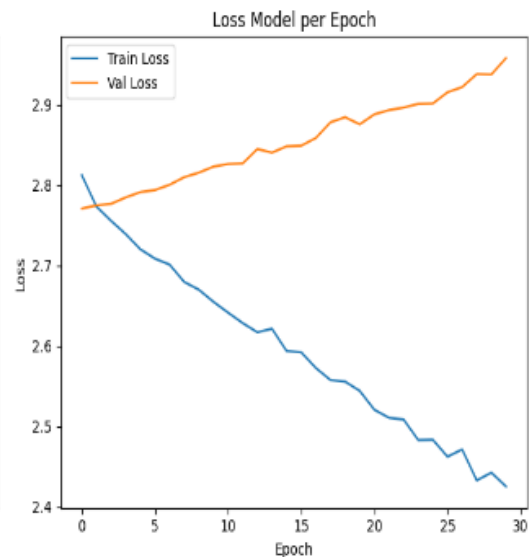


Figure 6. Training and Validation Loss Across Epochs

Figure 5. Training and Validation Accuracy Across Epochs. The visualization illustrates the training and validation accuracy of the deep learning model over 30 epochs for university major classification based on students' interests, aptitudes, and academic characteristics. Training accuracy increases progressively from approximately 6.3% to 21.8%, indicating that the model gradually learns patterns within the training dataset. However, validation accuracy remains relatively low, fluctuating between approximately 4.5% and 8.4%, without demonstrating consistent improvement. The widening gap between training and validation accuracy indicates substantial overfitting and poor generalization to unseen data. Furthermore, the validation accuracy is considerably lower than the best accuracy values reported for the six model categories, ranging from 87.3% for the shallow FFNN to 91.7% for Bi-LSTM. This discrepancy suggests that the visualization and reported accuracy results may represent different experimental configurations, datasets, or evaluation procedures and therefore require verification before direct comparison. Although the reported results identify Bi-LSTM as the highest-performing architecture, the displayed learning curves do not independently support this conclusion.

Figure 6. Training and Validation Loss Across Epochs. The loss visualization provides further evidence of overfitting during model training. Training loss decreases consistently from approximately 2.81 to 2.43, demonstrating that the model progressively minimizes classification errors on the training dataset. In contrast, validation loss increases from approximately 2.77 to 2.96, indicating that the model's predictive performance on unseen validation samples deteriorates as training continues. This divergence suggests that the model increasingly fits training-specific patterns rather than learning representations that generalize effectively. To improve model robustness, techniques such as early stopping, dropout regularization, weight decay, hyperparameter optimization, and improved feature selection should be considered. Additionally, stratified data partitioning and independent test-set evaluation are necessary to ensure reliable performance estimates. Although the study reports an 84% agreement between students' actual major selections and system recommendations, this agreement should be evaluated separately from classification accuracy because it measures consistency with student choices rather than independently validated academic suitability. Therefore, the findings emphasize the importance of improving model generalization and verifying experimental consistency before implementing the proposed explainable deep learning framework for personalized university major recommendation.

5.0. CONCLUSION

This study investigated the development of an Explainable Deep Learning Framework for Personalized University Major Recommendation Based on Students' Interests and Aptitudes, utilizing academic performance, Holland's RIASEC vocational interests, and aptitude assessment data collected from secondary school students across 20 subdistricts in Tanggamus Regency, Lampung Province, Indonesia. The proposed framework integrated data preprocessing, exploratory data analysis, neural network classification, personalized recommendations, and explainable artificial intelligence to support evidence-based academic decision-making. Twenty neural network configurations were investigated, including Feedforward Neural Networks (FFNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Bidirectional LSTM (Bi-LSTM). The reported comparative results indicated that Bi-LSTM achieved the highest accuracy of 91.7%, followed by LSTM at 91.2%, GRU at 90.5%, and dropout-based FFNN at 90.1%. Additionally, an 84% agreement was reported between students' actual major selections and system recommendations. However, the presented training and validation curves revealed substantial overfitting, with training accuracy reaching approximately 21.8% while validation accuracy remained below 9%. This discrepancy indicates that the reported comparative accuracy values require further verification through consistent experimental settings and independent testing. PCA visualization and correlation analysis revealed overlapping student characteristics and generally weak linear relationships among attributes, highlighting the complexity of university major classification. Furthermore, SHAP and LIME were incorporated into the proposed framework to support interpretable recommendations, although their explanatory effectiveness requires additional empirical validation. Overall, the study demonstrates the potential of integrating vocational interest theory, psychometric assessment, and deep learning for personalized academic guidance. Future research should prioritize robust cross-validation, improved regularization, independently validated suitability labels, fairness assessment, and explanation stability. These improvements are essential to establish reliable generalization, transparent decision-making, and practical applicability in university major recommendation systems.

REFERENCES

- [1] Y. Jian, N. Xiao, and L. Youfeng, "A recommendation model for college majors based on deep learning and clustering algorithms," *Information Services & Use*, vol. 44, no. 2, pp. 165–175, 2024, doi: [10.3233/ISU-230201](https://doi.org/10.3233/ISU-230201).
- [2] Y. Yang, X. Peng, M. Chen, and S. Liu, "An explainable graph-based course recommendation model based on multiple interest factors," *Expert Systems with Applications*, vol. 264, Art. no. 125889, 2025, doi: [10.1016/j.eswa.2024.125889](https://doi.org/10.1016/j.eswa.2024.125889).
- [3] T. Yang et al., "Boosting course recommendation explainability: A knowledge entity aware model using deep learning," in *Proc. 32nd International Conference on Computers in Education (ICCE)*, 2024, doi: [10.58459/icce.2024.4862](https://doi.org/10.58459/icce.2024.4862).
- [4] K. Wagner, A. Merceron, P. Sauer, and N. Pinkwart, "A course recommender system built on success to support students at risk in higher education," *Journal of Educational Data Mining*, vol. 16, no. 1, pp. 330–364, 2024, doi: [10.5281/zenodo.11384083](https://doi.org/10.5281/zenodo.11384083).
- [5] B. Ma, T. Yang, and B. Ren, "A survey on explainable course recommendation systems," in *Distributed, Ambient and Pervasive Interactions*. Cham, Switzerland: Springer, 2024, pp. 273–287, doi: [10.1007/978-3-031-60012-8_17](https://doi.org/10.1007/978-3-031-60012-8_17).
- [6] J. Li, Q. Lyu, W. Qiu, and A. W. H. Khong, "Improving course recommendation systems with explainable AI: LLM-based frameworks and evaluations," in *Proc. 18th International Conference on Educational Data Mining*, 2025, doi: [10.5281/zenodo.15870185](https://doi.org/10.5281/zenodo.15870185).
- [7] "Interpretability-driven course recommendation: A random forest approach with explainable AI," *Bentham Science*, 2026, doi: [10.2174/0123520965401121250714075403](https://doi.org/10.2174/0123520965401121250714075403).
- [8] "Explainable learning analytics: Assessing the stability of student success prediction models by means of explainable AI," *Decision Support Systems*, vol. 182, Art. no. 114229, 2024, doi: [10.1016/j.dss.2024.114229](https://doi.org/10.1016/j.dss.2024.114229).

- [9] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—Where are the educators?," *International Journal of Educational Technology in Higher Education*, vol. 16, Art. no. 39, 2019, doi: [10.1186/s41239-019-0171-0](https://doi.org/10.1186/s41239-019-0171-0).
- [10] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based recommender system: A survey and new perspectives," *ACM Computing Surveys*, vol. 52, no. 1, Art. no. 5, 2019, doi: [10.1145/3285029](https://doi.org/10.1145/3285029).
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539).
- [12] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018, doi: [10.1109/ACCESS.2018.2870052](https://doi.org/10.1109/ACCESS.2018.2870052).
- [13] A. Barredo Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: [10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012).
- [14] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: [10.1145/2939672.2939778](https://doi.org/10.1145/2939672.2939778).
- [15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017. [Online]. Available: [NeurIPS Proceedings](https://proceedings.neurips.cc/paper_files/paper/2017/file/3045339647f25287e78c9c0d7104412d-Paper.pdf).
- [16] J. L. Holland, *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments*, 3rd ed. Odessa, FL, USA: Psychological Assessment Resources, 1997.
- [17] F. M. Lord and M. R. Novick, *Statistical Theories of Mental Test Scores*. Reading, MA, USA: Addison-Wesley, 1968.
- [18] L. J. Cronbach and P. E. Meehl, "Construct validity in psychological tests," *Psychological Bulletin*, vol. 52, no. 4, pp. 281–302, 1955, doi: [10.1037/h0040957](https://doi.org/10.1037/h0040957).
- [19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986, doi: [10.1038/323533a0](https://doi.org/10.1038/323533a0).
- [20] F. Ricci, L. Rokach, and B. Shapira, Eds., *Recommender Systems Handbook*, 3rd ed. Cham, Switzerland: Springer, 2022, doi: [10.1007/978-1-0716-2197-4](https://doi.org/10.1007/978-1-0716-2197-4).