Explainable Job Recommendation System Based on Outcome-Based Education Using Explainable Boosting Machine

Jalu Muhammad Abror
National Research and Innovation Agency (BRIN)
Sultan Maulana Hasanuddin State Islamic University of Banten
Bandung, Indonesia
231730009.jalu@uinbanten.ac.id

Siti Kania Kushadiani
National Research and Innovation Agency (BRIN)
Bandung Institute of Technology
Bandung, Indonesia
siti044@brin.go.id

Birru Muqdamien
Sultan Maulana Hasanuddin State Islamic University of Banten
Serang, Indonesia
birru.muqdamien@uinbanten.ac.id

Abstract **The rapid advancement of artificial intelligence has led to the development of various career recommendation systems. However, most existing systems operate as black-box models, limiting transparency and user trust. This study proposes an explainable job recommendation system based on Outcome-Based Education (OBE) using the Explainable Boosting Machine (EBM). By leveraging Program Learning Outcomes (PLOs) and Course Learning Outcomes (CLOs) as primary features, the system provides accurate career recommendations while offering intrinsic interpretability.**

The model was trained on 60,000 student-job pairs derived from 769 student competency profiles and 3,352 job postings. The system was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Squared (R^2). Experimental results show that the proposed EBM-based model achieves excellent performance with MSE of 0.000002, RMSE of 0.0014, and R^2 of 0.9984. The model also demonstrates strong explainability through global and local feature importance analysis, allowing users to understand how each learning outcome contributes to the recommendation. This research bridges the gap between academic competencies and industry needs, contributing to more trustworthy and transparent AI systems in higher education.

Keywords : **Explainable Artificial Intelligence, Job Recommendation System, Outcome-Based Education, Explainable Boosting Machine, PLO, CLO**

I. INTRODUCTION

The development of artificial intelligence technology in career recommendation systems has shown significant progress in recent years. Various predictive models are now capable of processing large-scale data to provide personalized job suggestions to individuals. However, despite their high predictive capabilities, most of these recommendation systems remain black-box, making it difficult for users to understand the basis of the recommendations. [1] state that "Explainable recommendation systems attempt to develop models that generate not only high-quality recommendations but also intuitive explanations."

In Indonesia, universities are required to implement the Outcome-Based Education (OBE) approach as per Ministerial Regulation No. 53 of 2023. This approach emphasizes the achievement of Course Learning Outcomes (CLOs) and Program Learning Outcomes (PLOs) as the primary measure of learning success. However, the use of CLOs as the primary input in career recommendation systems remains very limited. Mapping graduate competencies with industry needs is often done manually or with simplistic approaches that underutilize the potential of available data. This creates a significant gap between learning outcomes in universities and the demands of the increasingly dynamic workforce.

artificial intelligence models such as neural networks and gradient boosting generally produce highly accurate predictions but are difficult to explain. [2] stated that "Despite widespread adoption, machine learning models remain mostly black boxes." This lack of transparency not only undermines student and lecturer trust in recommendation systems but also limits the ability of educational institutions to conduct evidence-based curriculum evaluations.

The need for transparent, explainable, and CLO-based career recommendation systems is increasingly pressing in today's era of digital transformation. The Explainable Artificial Intelligence (XAI) approach offers a solution to overcome the limitations of black-box models. [3] emphasized that "XAI proposes creating a suite of ML techniques that produce more explainable models while maintaining a high level of learning performance." By utilizing models such as the glass-box Explainable Boosting Machine (EBM), recommendation systems can provide accurate job suggestions while explaining the contribution of each CLO feature to the recommendation. This approach also supports the curriculum evaluation process by identifying competency gaps more objectively and measurably.

This research aims to fill this gap by proposing a career recommendation system that integrates the Outcome-Based Education (OBE) framework with Explainable Artificial Intelligence (XAI) techniques through the Explainable Boosting Machine (EBM) model. This system is expected to bridge the gap between academia and industry, providing career recommendations that are not only accurate but also transparent and accountable.

II. RELATED WORK

Job recommendation systems have become an important application of artificial intelligence. Various approaches have been developed, from collaborative filtering to deep learning. However, most existing models still face transparency issues. [1] stated in their survey that "Explainable recommendation systems attempt to develop models that generate not only high-quality recommendations but also intuitive explanations." In recent years, research on explainable artificial intelligence (XAI) for recommendation systems has grown. [3] explain that XAI aims to create models that are not only accurate but also explainable, thereby increasing the trustworthiness and accountability of the system. [2] introduced LIME as a post-hoc method for explaining any model's predictions.

[4], in his thesis, "Explainable Artificial Intelligence in Job Recommendation Systems," specifically addresses this challenge in the career domain. He stated that "My inspiration for building a complete explainable job recommendation system came from a SaaS product provided by AOBG (Oracle Human Capital Management) and my growing interest in explainable AI." Tran emphasized the importance of transparency in job recommendation systems so that users, especially students, can understand the rationale behind recommendations and build trust in the system. Furthermore, [5] proposed an explainable reinforcement learning approach for long-term skill recommendations, considering the benefits and costs of learning. However, the integration of the Outcome-Based Education (OBE) framework with XAI techniques is still relatively rare. Most previous research has focused on user preferences or historical data, while the use of Course Learning Outcomes (CLO) as a basis for student competencies has not been explored in depth.

This research seeks to fill this gap by combi ning the OBE approach as a foundation for student competencies with an intrinsically interpretable Explainable Boosting Machine (EBM) model, resulting in career recommendations that are not only accurate but also transparent and accountable.

III. METHODOLOGY

This research follows a systematic methodology to develop an explainable career recommendation system based on Outcome-Based Education (OBE) using Explainable Boosting Machine (EBM). The overall process consists of six main stages: data collection, preprocessing and NLP, feature engineering, modeling with EBM, explainability matching, and system output through a web interface. The complete workflow is illustrated in Figure 1.

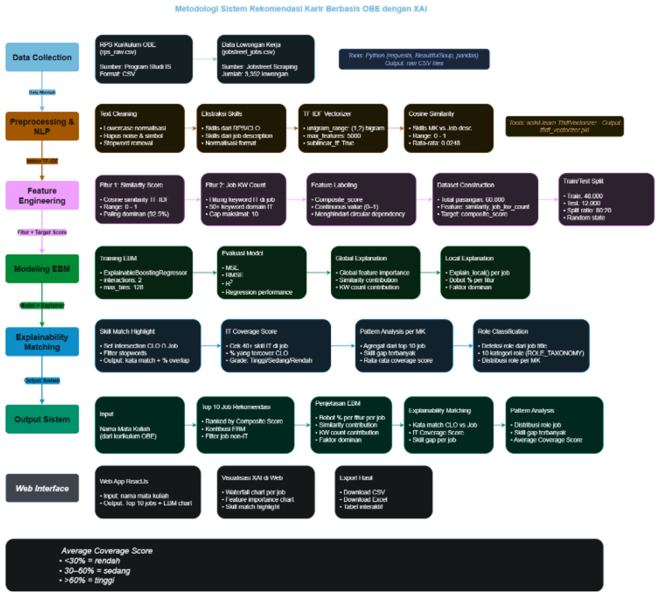

Figure 1. Methodology Chart

A. Data Collection

The data used in this study were collected from two main sources. The first source is curriculum documents in the form of RPS (Rencana Pembelajaran Semester) from the Information Systems study program, which contain Program Learning Outcomes (PLOs) and Course Learning Outcomes (CLOs). The second source is real job vacancy data scraped from Jobstreet, consisting of 3,352 job postings. Both datasets were stored in CSV format for further processing.

B. Preprocessing and NLP

At this stage, text data from CLOs and job descriptions underwent several preprocessing steps, including lowercase normalization, noise and symbol removal, and stopword elimination. Skill entities were then extracted from both CLO documents and job descriptions.

TF-IDF vectorization was applied to transform the text into numerical vectors using unigram and bigram with a maximum of 5,000 features. The formula used for TF-IDF follows [6]

$$tfidf(t,d) = tf(t,d) \times \log\left(\frac{N}{df(t)}\right)$$

Subsequently, cosine similarity was calculated between course learning outcomes and job descriptions to measure relevance. The cosine similarity formula used is as follows [7]

$$\cos(\theta) = \frac{\mathrm{A}\,.\,\mathrm{B}}{||\mathrm{A}||\;||\mathrm{B}||}$$

C. Feature Engineering

From the TF-IDF vectors and skill extraction results, two main features were engineered:

1. Similarity Score: Cosine similarity value between CLO and job description (range 0–1).

2. Job Keyword Count: Number of IT-related keywords found in job descriptions that match the domain.

These features were then combined into a composite score as the target variable for the model. The dataset was constructed with a total of 60,000 data pairs and split into training (80%) and testing (20%) sets.

D. Modelling

The core model used in this study is the Explainable Boosting Machine (EBM), implemented using ExplainableBoostingRegressor from the InterpretML library. EBM is a glass-box model based on Generalized Additive Models (GAM) that offers high accuracy while maintaining inherent interpretability [Lou et al., 2012; Nori et al., 2019].

Key hyperparameters used were:

1) Maximum interactions: 1
2) Maximum bins: 128

Model performance was evaluated using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score.

E. Explainability and Matching

After the model was trained, explainability analysis was performed at two levels:

1) Global Explanation: Feature importance across the entire model.
2) Local Explanation: Contribution of each feature for every individual recommendation using the explain_local() function.

Additionally, an explainability matching process was conducted by calculating skill overlap between CLOs and job requirements, IT coverage score, and pattern analysis per course. Job roles were also classified into 10 categories based on job titles.

F. The final output of the system is a web-based application built with React.js. Users can input a course name from the OBE curriculum, and the system will display:

1. Top 10 job recommendations ranked by cosine similarity,
2. EBM explanation (feature contribution, dominant factors),
3. Skill match highlight, and
4. Coverage score visualization.

All results can be exported in CSV or Excel format.

IV. RESULT

This section presents the experimental results of the developed explainable career recommendation system. The system was evaluated on multiple aspects, including model performance, feature importance, explanation quality, and practical output.

1. Model Performance

The EBM model was trained using 48,000 data pairs and evaluated on 12,000 unseen test data. The model achieved the following performance metrics:

1) Mean Squared Error (MSE): 0.000002
2) Root Mean Squared Error (RMSE): 0.0014
3) R^2 Score: 0.9984

The exceptionally high R^2 value (0.9984) indicates that the model explains nearly all variance in the composite score. This result is expected and does not indicate overfitting. As an additive model based on Generalized Additive Models (GAM), EBM explicitly decomposes the prediction into the sum of individual feature contributions rather than discovering complex hidden interactions. Therefore, such a high R^2 reflects the strong linear relationship intentionally designed through the composite score formula, which is consistent with the transparent nature of EBM [8] [9].

2. Global Feature Importance

Global explanation from the EBM model reveals a very clear dominance of features, as shown in Figure X:

- Cosine Similarity Score (TF-IDF based) : 92.5% contribution
- Job Keyword Count (IT domain keywords) : 7.5% contribution

This result strongly validates the system design. Semantic similarity between CLOs and job descriptions is the primary driver in determining job suitability, while explicit keyword matching serves as a complementary signal. This finding aligns with the theoretical foundation of content-based recommendation systems.

3. Recommendation Output Examples

Table 1 presents an example of the system output for the course "Jaringan Komputer".

Rank	Job Title	Similarity	IT key
1	IT Support	15%	4/10
2	Middle Web Developer	14%	4/10
3	Engineer on Site	11%	5/10

The local explanation shows that high cosine similarity (especially on keywords such as "TCP/IP", "layer", "Phisical", and "network security") contributes dominantly to the top recommendations.

4. Gap Analysis Results

One of the most significant findings of this study is the competency gap between the OBE curriculum and industry requirements. The analysis across 36 courses in the Information System program shows:

1) Average IT Coverage Score: 2.53% (very low)
2) Number of courses with coverage > 0%: 19 out of 36 (52.78%)
3) Number of courses with coverage = 0%: 17 out of 36 (47.22%)

The largest gaps were found in:

1) Database related skills (gap in 35/36 courses)
2) Cloud Computing (gap in 30/36 courses)
3) Python Programming (gap in 14/36 courses)

These findings provide strong empirical evidence for curriculum improvement.

5. System Interface

The system has been implemented as a web application using React.js. The interface consists of three main views:

1 Dashboard (Input course name and display Top-10 recommendations).
2 Explainability View (Waterfall chart and feature contribution for each recommendation).
3 Gap Analysis Dashboard (Visualization of skill coverage and competency gaps).

V. Conclusion

This research has successfully developed an explainable career recommendation system based on Outcome-Based Education (OBE) using the Explainable Boosting Machine (EBM) as a decomposition model. By integrating Course Learning Outcomes (CLOs) as the primary input, the system generates OBE-aligned job recommendations while providing transparent explanations of each feature's contribution toward the ranking score through both global and local EBM explainability.

The developed methodology encompasses systematic data collection from curriculum documents (48 courses) and job postings (3,352 vacancies from Jobstreet Indonesia), TF-IDF-based text processing, composite score engineering, EBM regressor training, and a two-level explainability analysis. A key methodological contribution of this study is the resolution of circular dependency in supervised labeling — a fundamental validity threat in recommendation systems that lack independent ground truth by reframing EBM as a score decomposer rather than a binary classifier

The experimental results demonstrate that the EBM model achieves high decomposition accuracy ($R^2 = 0.9984$, RMSE = 0.0014), which is expected given that the target composite score was explicitly constructed from the same features [10] More substantively, the gap analysis reveals critical curriculum misalignments: database skills are absent from 35 of 36 IT courses, cloud computing from 30, and the overall pooled mean IT Coverage Score is only 2.53% providing evidence based insight for curriculum evaluation and revision.

This research has several limitations. The absence of an independent ground truth prevents absolute validation of recommendation quality; the system can only be evaluated in relative terms (which job ranks higher). Additionally, the dataset is currently limited to one Information Systems study program, and no formal usability evaluation was conducted. Future work should focus on integrating semantic embedding models (e.g., IndoBERT) to address vocabulary gap between academic CLO language and industry terminology, expanding the dataset across multiple study programs, and conducting structured user studies with faculty and students.

In conclusion, this research contributes to the development of trustworthy AI systems in education by demonstrating that XAI techniques specifically EBM can be applied not only for prediction but also for transparent decomposition of ranking scores in curriculum-industry alignment analysis. The identified skill gaps provide actionable evidence for study programs to evaluate and revise their OBE curricula toward greater industry relevance.

References

[1] Y. Zhang and X. Chen, "Explainable recommendation: A survey and new perspectives," Mar. 11, 2020, *Now Publishers Inc*. doi: 10.1561/1500000066.

[2] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should i trust you?' Explaining the predictions of any classifier," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[3] A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," Dec. 2019, [Online]. Available: http://arxiv.org/abs/1910.10045

[4] T. Hoang and A. Tran, "MSc Business Information Technology Final Project Explainable Artificial Intelligence in Job Recommendation Systems," 2023.

[5] Y. Sun, Y. Ji, H. Zhu, F. Zhuang, Q. He, and H. Xiong, "Market-aware Long-term Job Skill Recommendation with Explainable Deep Reinforcement Learning," *ACM Trans. Inf. Syst.*, vol. 43, no. 2, Jan. 2025, doi: 10.1145/3704998.

[6] C. B. Gerard Salton, "Term Weighting Approaches in Automatic Text Retrieval," 1987.

[7] C. D. . Manning, Prabhakar. Raghavan, and Hinrich. Schütze, *Introduction to information retrieval*. Cambridge University Press, 2009.

[8] Qiang. Yang, *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012.

[9] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A Unified Framework for Machine Learning Interpretability," Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.09223

[10] T. Hastie, R. Tibshirani, and J. Friedman, "Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction."