Jurnal Informatika dan Rekayasa Perangkat Lunak ISSN 2656-2855
Vol. 8, No. 1, March 2026, Page. 1-8 e-ISSN 2685-5518

Enhancing the Accuracy of Competency Portfolio Assesments using Machine Learning: a Comparative Analysis of Predictive Models

Aditya Cahya Saputra[1*], **Imam Yuadi**[2]
[1]Magister of Human Resource Development, Postgraduate School, Airlangga University, Surabaya
[2]Department of Information and Library Science, Faculty of Social and Political Sciences, Airlangga University, Surabaya
Email: [1*]adityacahyasaputra17@gmail.com, [2]imam.yuadi@fisip.unair.ac.id

Abstract

This study elaborates the application of various machine learning (ML) models to measure competency portfolio assessments for job grade conversion needed of employees. The purpose is choosing the best ML models to enhance the accuracy, scalability, and fairness. Logistic regression and support vector machines is two traditional methods were evaluated together with random forest and gradient boosting as ensemble models and neural network as deep learning models. This study taken data of 117 employees invited to join on the competency portfolio assessment event on November 2024, all models were measured through cross-validation on parameters such as accuracy, precision and recall by Orange Data Mining. The best performance model in this study is Random Forest, achieving the highest score on Precision and Recall parameters. While Neural Networks demonstrated potential performance that almost has the same result with logistic regression. Based on this research, Random Forest can be prioritized and implemented to help the company to enhance the accuracy of competency portfolio results that needed to develop employees career, eligible competencies, and help decision making of job grade conversion assessment.

***Keywords**: Comparative Analysis, Competency Portfolio Assessment, Machine Learning*

INTRODUCTION

Competency portfolio assessments are a part of the way to evaluate and measure the hard and soft competencies of employees. It is increasingly utilized across industries and academia to evaluate an individual's skills, qualifications, and readiness for specific roles or certifications to move to the next job grade or career level. Accurate and consistent assessment of these projects is therefore crucial (Kamper et al., 2023). In the past, these assessments are measured on manual evaluation methods that are time-intensive, subjective, and potential to get the inconsistencies results. The growing complexity of skills required and various indicators measured in today's dynamic environment need more efficient and accurate approach to determining eligibility for competency portfolios in the future. Machine learning or ML models have emerged as a tool advance solution that offering automation and data-driven methods to predict outcomes and results with high precision. By analyzing large datasets, ML can capture patterns and insights that traditional methods often overlook, enhancing the reliability of portfolio assessments. Machine learning algorithms are computational techniques that enable computers to learn and make predictions or decisions without explicit programming. They form the basis of modern data analytics and predictive modeling (Patil et al., 2023). However, the next challenges remain in selecting the best and the most suitable prediction models to ensure the fairness and robustness as a tool for assessments. Handling these issues is critical to get the most effective of ML models in competency portfolio assessments, paving the way for more equitable and efficient decision-making process (Kamper et al., 2023).

Previous research on competency portfolio assessment using ML models has showed the transformative potential of predictive analysis in streamlining evaluation process. Early works in the past focused on applying linear regression and logistic regression models as traditional methods algorithms to predict eligibility outcomes, laying the foundation for integrating data-driven approaches into assessments. The development of decision tree based methods like random forests and gradient boosting, were later adopted for their ability and capacity to handle categorical and continuous data effectively, but achieving higher accuracy in predicting competency (Patil et al., 2023). And then in deep learning as advanced algorithm, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have further enhanced the ability to analyze unstructured data like portfolio narratives and performance reviews. Researchers have also explored hybrid approaches through combining natural language processing (NLP) techniques with ML models to extract insights from textual data or often we called 'text mining'. However, the significant challenges remain, including to ensure model interpretability, handling bias in training data, and maintaining fairness in decision-making processes. The integration of explainable Artificial Intelligence or AI methods and ethical frameworks into ML models for competency assessments is increasingly emphasized to handle and solve these concerns.

The objective of this research is to explore the most suitable application of ML models in competency portfolio assessment, focused on improving accuracy, fairness, and scalability. The study addresses the most effective and suitable ML models to predict the competency portfolio eligibility across various datasets and contexts. In other that, the study also addresses the way of ML techniques enhance the interpretability and reliability of eligibility assessments.

To achieve this goal, this study will evaluate and compare the performance metrics of traditional ML algorithms and advanced techniques, such as ensemble techniques (random forest and gradient boosting) and deep learning (neural network) in analyzing portfolio datasets. Furthermore, the study seeks to achieve the goal by using the best algorithm model to predict the results of the competency portfolio assessment used to predict the costs that the company will incur if employees pass the assessment and experience an increase in person grade.

The methodology section emphasizes the research design, including dataset selection, ML models utilization, and metrics evaluation. The results and discussion section identifies the findings, comparing model performance and addressing ethical and interpretability considerations. Finally, the conclusion and future work section summarizes the research contributions, discusses practical implications, and suggests directions to determine decision making and for the future research in this topic. Through these sections, this article aims to provide a comprehensive analysis of how ML models as algorithms can transform and predict the results competency portfolio assessments.

LITERATURE REVIEW

The research on the portfolio assessments have always developed to enhance the accuracy and efficiency of ML. Early studies that utilized the traditional ML methods, such as logistic regression and decision trees, demonstrating significant improvements in consistency compared to manual evaluations aimed at predicting portfolio outcomes (Patil et al., 2023).

Traditional machine learning such as Logistic Regression is an algorithm for supervised learning. It models the link between a dependent variable that is binary and one or more factors using logistic functions. It is mostly used to classify whether a particular event will occur or not (Patil et al., 2023).

The other traditional machine learning algorithms is Support vector machine (SVM). SVM is an algorithm for two group classification problem. The algorithms conceptually implement input vector are non-linearly mapped to a very high dimension feature space (Cortes & Vapnik, 1995). The Support Vector Machine Learning Algorithm can be used for both classification and regression (Aruna et al., 2022). However, these models have limited ability to address the complexity and unstructured data, which is common in portfolio submissions.

To address these limitations, ensemble techniques such as random forest and gradient boosting algorithms can be solutions that can be offered to obtain higher predictive accuracy and robustness by combining multiple models. Ensemble methods build many base models and then merge them into one to achieve better prediction results than using a single base model (AlShourbaji et al., 2023).

Such as Random Forest algorithms is a learning process that includes multiple decision tree to increase the accuracy of predictions and reduce competition. Each tree in the forest is trained with different materials and properties (Patil et al., 2023).

After that, the family of boosting methods depends on different constructive strategies of ensemble formation, called Gradient Boosting

Machine (GBM) is included as ensemble methods. The GBM can be considered an optimization model aiming to train a series of weak-learner models, which sequentially minimizes a pre-defined loss function (AlShourbaji et al., 2023).

With the increase of computing power, deep learning algorithms have been widely used. Compared with traditional intelligent algorithms, deep learning can use multi-layer neurons to map data to higher dimensions and automatically extract features (He et al., 2022). Further, Neural Network is other data analysis techniques such as natural language processing (NLP) can be applied to the textual data coming from the resumes and cover letter to get a deeper and more comparative and quantitative analysis about the candidate's suitability for the position (Sharma & Sohal, 2024).

In previous research, text mining has been primarily used to analyze sentiment in HR management or assist in the employee recruitment process. For example, Lee & Song (2024) conducted sentiment analysis using text mining techniques to assess employee reviews and extract different factors of positive experiences to increase employee satisfaction, engagement, and productivity. Tikhonova's (2020) study used text mining to propose an automated information extraction procedure from text documents, specifically from candidate CVs.

This study will examine the most appropriate algorithm modeling for applying text mining to oral-test-based portfolio competency assessments. The output of this research will be a prototype application that can be used to assist assessors in analyzing employee competency through oral tests during employee competency assessments. This will strengthen assessors' decisions in classifying employees as competent or incompetent in their field of work or job grade. Consequently, this research is expected to help improve employee competency and enhance company performance.

METHODS

The study on competency portfolio assessments using machine learning started with traditional methods like logistic regression to improve evaluation accuracy. To address the problems in handling complex and unstructured data, the ensemble techniques such as random forests and gradient boosting machines have been introduced. Additionally, natural language processing (NLP) techniques has been applied to analyze textual data in portfolios. As the field advances, deep learning methods as a part of machine learning like convolutional neural networks (CNN) and recurrent neural networks (RNN) are employed to assess visual and temporal data.

The flowchart of this research methods has three main steps: (1) Data collection and preprocessing with a focus on gathering portfolio data and conditioning it; (2) Model selection involving testing various ML algorithms such as logistic regression, SVM, random forest, gradient boosting, and RNN; (3) Evaluation with metrics to obtain the accuracy, fairness, and interpretability. The flowchart is visually clear as showed by arrows connecting the steps sequentially.

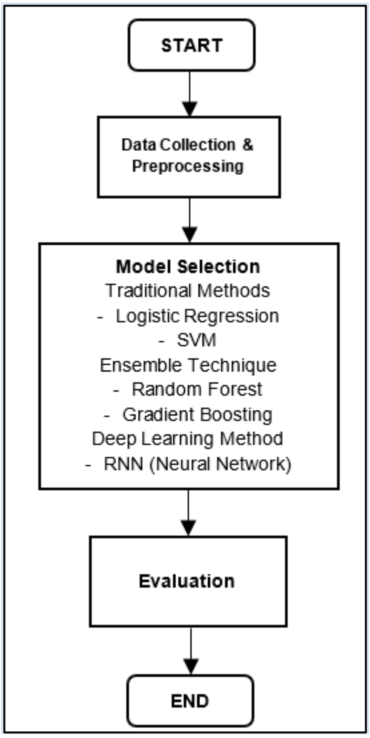

Figure 1. Flowchart of Research Methods

3.1 Data Collection and Preprocessing

This step focuses on preparing good data for analysis needs by collecting relevant portfolio data and ensuring that the data is clean, clear, and well-prepared, thus taking longer than other steps. Data quality is a very important part of the performance of machine learning models, because errors or missing data can lead to biased or inaccurate predictions. Data collection usually includes collecting job titles, job levels, professional branches, competency names, and competency areas. Preprocessing is very important and crucial because it plays a role in reducing noise and eliminating inconsistencies from the data set so that it can improve model accuracy. Preprocessing techniques such as cleaning data, normalizing data sets, and transforming data to prepare for the next step in the modeling process.

The object of this research is 117 employees of PT PLN (Persero) who join the competency portfolio assessment through interview technique. The research measure various attributes like name positions, position levels, branch of professions, field competency, category, test field compatibility, description of the test results and conclusions of the results. All attributes are categorical data except name, id employee, work unit and description of the test results is textual data. Conclusions of the results is to be attribute target while the position level, branch of profession, category, test field of compatibility, position and field competency are to be attributes features. The results of assessment found an imbalance in the data, consisting of 86 employees who eligible and 31 employees who ineligible.

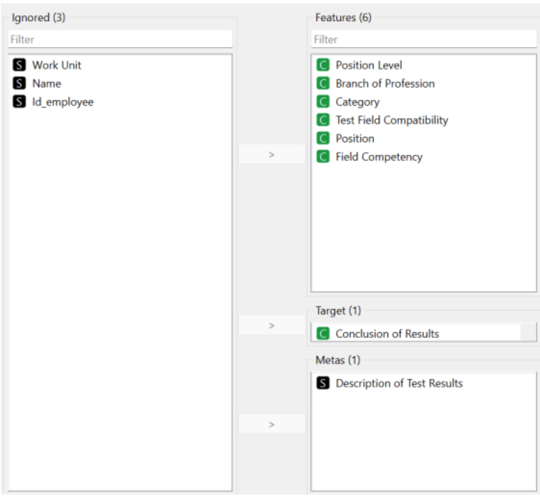

Figure 2. Attribute Data

3.2 Model Selection and Training

This step selects and trains various machine learning algorithm models to select the best model for a given task. This step focuses on training the model using data that has been preprocessed previously. After that, various algorithms are assessed based on their performance metrics. This is important to determine the right model selection in analyzing the existing data set so that it can provide accurate predictions regarding the results of the competency portfolio assessment. In this study, several machine learning algorithm models will be tested for their application. First, traditional machine learning models, namely logistic regression and support vector machines (SVM) which are well-known for solving classification tasks, with logistic regression being efficient for binary classification and SVM being superior in high-dimensional spaces (Cortes & Vapnik, 1995). Second, ensemble technique machine learning models, namely random forest and gradient boosting, which work by combining several models to improve prediction performance by reducing variance (random forest) and bias (gradient boosting) (Breiman, 2001; Friedman, 2001).Third, deep learning methods, such as recurrent neural networks (RNNs) are known to be very effective in modeling sequential data, such as time series or natural language, by capturing dependencies across time steps (Hochreiter & Schmidhuber, 1997). Different methods have their own advantages and will be selected based on business understanding and research objectives.

3.3 Evaluation

This step is carried out to re-test the selected model by evaluating the robustness and effectiveness of the selected model. This step involves comparing five algorithm models, namely Logistic Regression, SVM, Random Forest, Gradient Boosting and Neural Network using accuracy metrics. The results of the model comparison will be analyzed to identify which model has the highest accuracy metrics such as precision, recall, and F1 score. These accuracy metrics can help measure how well the model performs in analyzing the data set so that it can be used in predictive modeling.

RESULTS AND DISCUSSION

The following design model is generated by using Orange Data Mining with five algorithms to predict the results of the competency portfolio assessment.

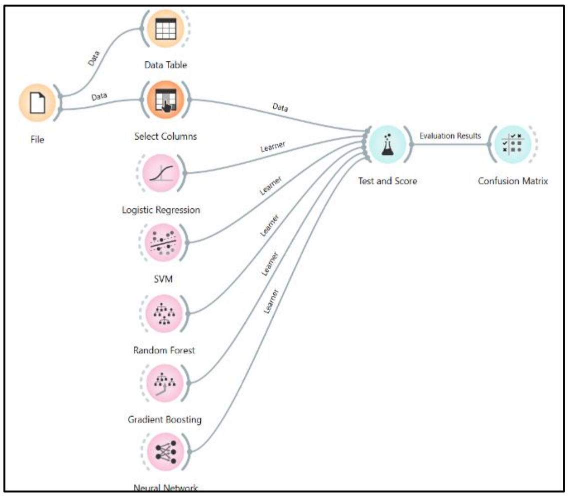

Figure 3. Design Model

This design model illustrates the results of machine learning modeling applied to a dataset using cross-validation. The data is divided into training and testing subsets. Then, 10-fold stratified cross-validation is performed to ensure and maintain the consistency of class distribution across folds. This process is important in determining the best model selection through training steps to assess the performance of different algorithms systematically. The goal of this approach is to prevent overfitting and provide reliable estimates of model performance on unseen data. The target attribute data in this study is categorical outcome inference based on qualified and unqualified values. The five models compared in this evaluation are Logistic Regression and Support Vector Machine (SVM) as traditional machine learning methods, Random Forest and Gradient Boosting as ensemble techniques and Neural Network as a subset of deep learning methods. The five models are evaluated on several performance metrics, such as AUC (Area Under Curve), Classification Accuracy (CA), F1 Score, Precision (Prec), Recall, and MCC (Matthews Correlation Coefficient). Because the dataset used is an imbalanced, the performance metric assessment will be focused on the F1 Score and MCC results because they are more relevant to use than other metrics.

Based on data the amount of value “Eligible” is greater than “Ineligible’ data, so it is not balance data. Therefore, the concern of model performance parameters used are precision and recall parameters. Precision is needed to ensure the prediction that employees are really eligible or ineligible on the competency assessment. Meanwhile, recall supports minimizing the error prediction.

Table 1 .Comparison of Model Results-1

Model	AUC	CA	F1
Random Forest	0.798	0.821	0.803
Neural Network	0.778	0.786	0.785
Logistic Regression	0.803	0.786	0.761
Gradient Boosting	0.763	0.786	0.765
SVM	0.797	0.761	0.708

Table 2. Comparison of Model Results-2

Model	Prec	Recall	MCC
Random Forest	0.815	0.821	0.495
Neural Network	0.784	0.786	0.446
Logistic Regression	0.772	0.786	0.381
Gradient Boosting	0.771	0.786	0.388
SVM	0.742	0.761	0.263

Random Forest as ensemble machine learning emerged as the best performer model in almost all performance metrics, showed by F1 score of 0.803, precision of 0.815, recall of 0.821 and MCC of 0.495. The model works by combining predictions of multiple decision trees, reducing variance and increasing robustness. Its high recall and precision indicate its ability to identify true positives more effectively and minimize false positives, making it ideal for datasets with complex interactions and non-linear patterns.

Neural Networks as deep learning model achieved F1 score of 0.785, precision of 0.784, recall of 0.786 and MCC of 0.446 that placing it above logistic regression but slightly below random forest. Neural networks are powerful for capturing complex non-linear relationships in data, but their performance can be sensitive to hyper parameter tuning, architecture and data size. In this case, while Neural Networks give reasonable results but still below the performance of random forest.

Logistic Regression as a traditional machine learning achieves F1 of 0.761, precision of 0.772, recall of 0.786 and MCC of 0.381

reflecting its efficiency in binary classification tasks. As a traditional model, it works computationally efficient and interpretable, making it suitable for simple datasets or priority data to be explained. However, its performance is surpassed by random forest and neural network which leverage the power of multiple models to improve predictions.

Gradient Boosting as ensemble model, with F1 of 0.765, precision of 0.771, recall of 0.786 and MCC of 0.388 performs better than SVM but below random forest and neural network. This technique sequentially builds models to correct the errors of the previous model, reducing bias and improving accuracy.

Support Vector Machine (SVM) as traditional model showed the weakest performance in this evaluation with F1 of 0.708, precision of 0.742, Recall of 0.761, and MCC of 0.263. Although effective in high-dimensional space, these results suggest that SVM may not be suitable for this particular dataset, likely due to the nature of the data or lack of hyper parameter optimization. SVM remains valuable for certain tasks but struggles compared to ensemble models in this case.

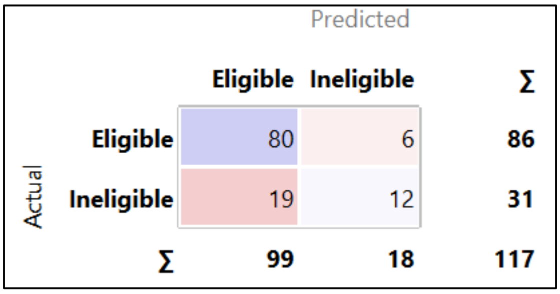

Figure 4. Confusion Matrix of Random Forest

Based on Random Forest model provides True Positive (TP) is 82, True Negative (TN) is 14, False Positive (FP) is 17, and False Negative (FN) is 4.

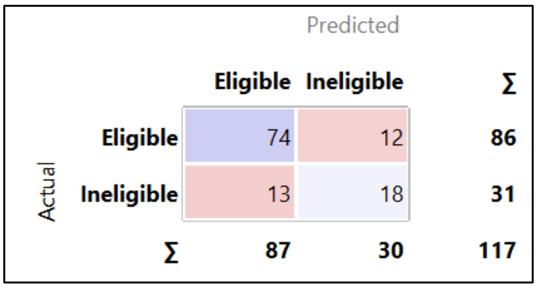

Figure 5. Confusion Matrix of Neural Network

Based on Neural Network model provides True Positive (TP) is 74, True Negative (TN) is 18, False Positive (FP) is 13, and False Negative (FN) is 12.

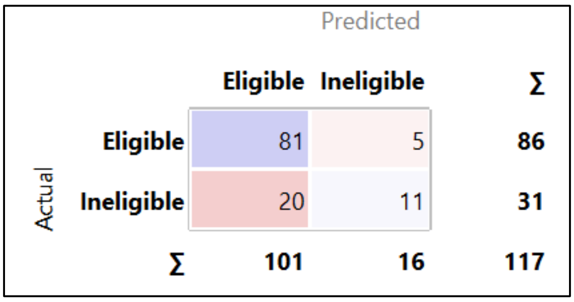

Figure 6. Confusion Matrix of Logistic Regression

Based on Logistic Regression model provides True Positive (TP) is 81, True Negative (TN) is 11, False Positive (FP) is 20, and False Negative (FN) is 5.

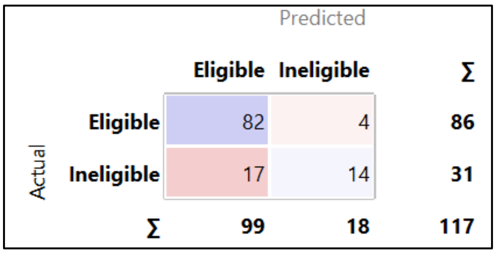

Figure 7. Confusion Matrix of Gradient Boosting

Based on Gradient Boosting model provides True Positive (TP) is 80, True Negative (TN) is 12, False Positive (FP) is 19, and False Negative (FN) is 6.

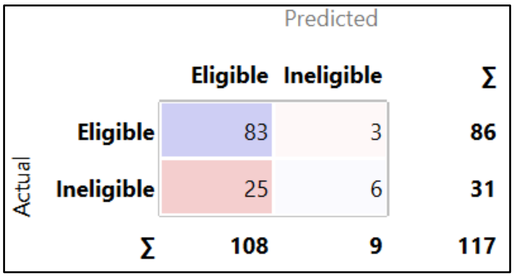

Figure 8. Confusion Matrix of SVM

Based on SVM model provides True Positive (TP) is 83, True Negative (TN) is 6, False Positive (FP) is 25, and False Negative (FN) is 3.

Table 3. Confusion Matrix Comparison

Model	TP	TN	FP	FN	TP+TN	FP+FN
Random Forest	82	14	17	4	96	21
Neural Network	74	18	13	12	92	25
LogRegression	81	11	20	5	92	25
G. Boosting	80	12	19	6	92	25
SVM	83	6	25	3	89	28

The following above is comparison of confusion matrix from five algorithms model used in this research. Based on table of confusion matrix comparison above, describe that Random Forest model is the highest accuracy model with TP+TN is greater than the other models. While the value of FP+FN is smaller than the other models. It shows the Random Forest model provides more little false prediction values.

Based on the results above, the choice of the best model depends on the specific objectives and requirements of the task. Random Forest is the most consistent and reliable performer across precision and recall, making it the ideal choice for general used in this evaluation. However, Neural Networks could be preferred for tasks requiring the second rank of recall and precision, depending on the application.

CONCLUSION

This research highlights the advancements in machine learning (ML) for competency portfolio assessments, demonstrating the effectiveness of ensemble models like Random Forest and Gradient Boosting over traditional methods and Neural Networks. Random Forest has the best performance model in this evaluation with the highest accuracy and lower false predictions based on confusion matrix imagined by Orange Data Mining. Random Forest has the Precision to 81,5% and Recall to 82,1% that shows this model can be used to predict the conclusion of the results of competency portfolio assessment properly.

Based on this research, Random Forest can be prioritized and implemented to help the company to enhance the accuracy of competency portfolio results that needed to develop employees career, eligible competencies, and for job grade conversion assessment.

Future research should prioritize addressing on a large dataset from all employees in a company with more various attributes can be help increasingly validity and reliability of the decision-making by management. This research can also use by foundation to develop the next research about AI implementation to help evaluation and measurement of competency portfolio assessment by using voice record recognition to determine capacity and capability of employees in answering questions from the assessor as consideration to take “eligible” or “ineligible” decision making to employees.

REFERENCES

AlShourbaji, I., Helian, N., Sun, Y., Hussien, A. G., Abualigah, L., & Elnaim, B. (2023). An efficient churn prediction model using gradient boosting machine and metaheuristic optimization. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-41093-6

Aruna, P., Sujatha D, C., Surya Prakash, S., & Prince, R. (2022). An Automated Self Rating Questionnaire Assessment Tool for the evaluation of Entrepreneurial Competencies using SVM Algorithm. *5th International Conference on Inventive Computation Technologies, ICICT 2022 - Proceedings*, 1230–1234. https://doi.org/10.1109/ICICT54344.2022.9850442

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5-32. https://link.springer.com/article/10.1023/A:1010933404324

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273-297.https://link.springer.com/article/10.1007/BF00994018

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189-1232. https://projecteuclid.org/euclid.aos/1013203451

He, Q., Chen, L., Tang, W., He, G., Tan, S., Shu, Y., & Jiang, L. (2022). Application and Comparative Analysis of Traditional Machine Learning and Deep Learning in Transmission Line Fault Classification. *IMCEC 2022 - IEEE 5th Advanced Information Management, Communicates, Electronic and Automation Control Conference*, 1715–1719.

https://doi.org/10.1109/IMCEC55388.2022.10020121

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735-1780. https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory

Kamper, H., Niehaus, C., & Wolff, K. (2023). Using Machine Learning to Understand Assessment Practices of Capstone Projects in Engineering. *2023 IEEE IFEES World Engineering Education Forum and Global Engineering Deans Council: Convergence for a Better World: A Call to Action, WEEF-GEDC 2023 - Proceedings*. https://doi.org/10.1109/WEEF-GEDC59520.2023.10344177

Lee, J., & Song, J. H. (2024). How does algorithm-based HR predict employees' sentiment? Developing an employee experience model through sentiment analysis. *Industrial and Commercial Training*, *56*(4), 273–289. https://doi.org/10.1108/ICT-08-2023-0060

Patil, A., Sarda, T., Shetye, S., Bachchan, S., & Gutte, V. S. (2023). Comparing Logistic Regression and Tree Models on HR Data. *Proceedings - 2023 International Conference on Advanced Computing and Communication Technologies, ICACCTech 2023*, 671–677. https://doi.org/10.1109/ICACCTech61146.2023.00113

Sharma, R., & Sohal, J. (2024). Data-Driven Talent Management: The Impact of Machine Learning on HR Efficiency and Effectiveness. *2024 2nd World Conference on Communication and Computing, WCONF 2024*. https://doi.org/10.1109/WCONF61366.2024.10691955

Tikhonova, M. (2020). Text mining for evaluation of candidates based on their CVs. *Communications in Computer and Information Science*, *1086CCIS*, 184–189. https://doi.org/10.1007/978-3-030-39575-9_19