Jurnal Riset Sains dan Teknologi
Volume 10 No. 1, March 2026, (A.01 -A.10)
e-ISSN: 2549-9750
p-ISSN: 2579-9118

Cyberloafing Analytics: Predicting Causes Using Machine Learning Models

Analisis Cyberloafing: Prediksi Penyebab Menggunakan Model Machine Learning

Gilang Ferdiansah[1]*, Imam Yuadi[2]
[1]Master's Program Human Resource Development, Graduate School, Airlangga University
[2]Information and Library Science, Faculty of Social and Political Sciences, Airlangga University
[1,2]Jl. Airlangga 4-6, Surabaya 60286, Indonesia

*Corresponding author: **gilang.ferdiansah-2024@pasca.unair.ac.id**

DOI; 10.30595/jrst.25997

Article information:
Received: 26/02/2025

Revised: 24/09/2025

Accepted: 28/10/2025

ABSTRACT

Cyberloafing refers to the practice of employees utilizing internet access for non-job-related activities during work hours. Cyberloafing poses a dilemma for organizations, as it is deemed aberrant conduct that might impact overall performance. Consequently, organizations must ascertain the determinants of cyberloafing. This study seeks to identify a suitable predictive model for the determinants of cyberloafing behavior in the workplace using a machine learning methodology. The employed methodology utilizes the conventional data mining cycle, namely the Cross-Industry Standard Process for Data Mining (CRISP-DM), with Orange Data Mining as the application tool. The findings indicate that Logistic Regression is the most effective model for forecasting cyberloafing. Logistic Regression yields performance scores of 90.5% Precision and 88.9% Recall. Conversely, the Naïve Bayes model had the lowest metrics, with a Precision of 64.8% and a Recall of 51.9%. This study serves as a reference demonstrating that Logistic Regression effectively predicts cyberloafing. This study enables firms to examine the factors contributing to cyberloafing, facilitating the development of policies aimed at mitigating its adverse effects.

Keywords: *Cyberloafing, Machine Learning, Orange Data Mining, Boredom, Fatigue*

ABSTRAK

Cyberloafing merupakan aktivitas pegawai yang menggunakan akses internet untuk kegiatan yang tidak berhubungan dengan pekerjaan selama jam kerja. *Cyberloafing* menjadi tantangan bagi perusahaan karena dianggap sebagai perilaku menyimpang yang dapat mempengaruhi kinerja perusahaan. Oleh karena itu, perusahaan perlu melakukan identifikasi faktor *Cyberloafing*. Penelitian ini bertujuan untuk menentukan model prediksi yang tepat untuk penyebab perilaku cyberloafing di tempat kerja melalui pendekatan *machine learning*. Metode yang digunakan dengan menerapkan siklus standar data mining yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dan aplikasi alat yang digunakan adalah Orange Data Mining. Hasil penelitian menunjukkan bahwa Logistic Regression menjadi model terbaik untuk prediksi *Cyberloafing*. Melalui metriks evaluasi, Logistic Regression memberikan performa nilai Precision 90.5% dan Recall 88.9%. Sedangkan nilai terendah ditunjukkan oleh model Naïve Bayes dengan nilai Precission 64.8% dan Recall 51.9%. Penelitian ini dapat menjadi rujukan bahwa Logistic Regression mampu melakukan prediksi *Cyberloafing* dengan baik. Secara keseluruhan, penelitian ini dapat membantu perusahaan dalam mengenalisis penyebab *Cyberloafing* guna membuat kebijakan untuk mengurangi dampak negatif dari *Cyberloafing*.

Kata Kunci: *Cyberloafing, Machine Learning, Orange Data Mining, Boredom, Fatigue*

1. INTRODUCTION

Digitalization in the Industrial world has become a common technological development today. The use of digitalization tools provides several benefits, such as good communication between lines, efficient processes, and the ability to get the latest information via the Internet. However, this technological development brings challenges, namely the cyberloafing phenomenon. According to Divya & Narwal (2023), cyberloafing is an act committed by employees who use company internet access for personal purposes during working hours and is not related to work. Activities unrelated to work that use the internet during working hours usually include sending personal emails, social media activities, and playing games (Demirboga et al., 2025; Henle & Kedharnath, 2012). Meanwhile, Lizarte Simón et al. (2024) show that work pressure and obstacle stressors influence cyberloafing, and smartphone dependency reinforces this. The phenomenon of cyberloafing can occur in any industrial company, such as in power generation companies that manage hydroelectric power plants in the Central Java region of Indonesia.

The power generation company has three divisions, namely operations & maintenance, engineering, and administration, with a total of 131 employees. Almost all employees have gadgets, and the company provides Wi-Fi access to all employees. This arrangement has triggered the phenomenon of cyberloafing. The phenomenon of cyberloafing is a challenge for power generation companies. Cyberloafing can be considered deviant behavior in the workplace that can affect company performance (Jamaluddin et al., 2023). To create a conducive work environment, we must address issues related to employees (Chung et al., 2023). Cyberloafing behavior may arise as a response to several contributing factors. Therefore, the human resource managers of power generation companies need to analyze and address cyberloafing. A cyberloafing prediction model is needed to provide an overview of the factors causing cyberloafing, and the results can be used as input for management policy-making.

Exploratory research on cyberloafing has been conducted extensively. For example, Krishna & Agrawal (2023) stated that cyberloafing behavior affects psychological well-being and has a negative impact. Research related to cyberloafing in the workplace as counterproductive work behavior (CWB) states that there is a relationship between workload, boredom, and cyberloafing (Pindek et al., 2018). Observational studies on cyberloafing often use samples of students or workers in a particular sector. These studies identify cyberloafing using quantitative and qualitative methods, but few use a machine learning approach to identify and classify the causes of it. Therefore, we need to conduct research on predictive models of the causes of cyberloafing in the workplace.

Numerous research studies in staff management within human resource management have commenced utilizing machine learning methodologies. Similar to the study of Novia & Yuadi (2023), they performed employee potential predictions via probationary evaluations utilizing five techniques in Orange Data Mining: Logistic Regression, Naïve Bayes, k-NN, SVM, and Decision Tree. The study reported that Logistic Regression provided the highest accuracy in classifying talent, at 90%. To expand the algorithm model study, this research added Random Forest and Neural Network. Several studies in the medical field, malware classification, and data analysis mention that Random Forest has high accuracy in classification tasks (Alabadee & Thanon, 2021; Maindola et al., 2024; Raposo et al., 2020). Meanwhile, Neural Network provides advantages in terms of flexibility and solving complex classifications (X. Li et al., 2023; Shobika et al., 2023). Therefore, this study uses 7 tools in Orange Data Mining, namely Logistic Regression, SVM, k-NN, Naïve Bayes, Decision Tree, Random Forest, and Neural Network. We expect the addition of more tools to provide a broader picture of the optimal model that companies can adopt.

This study aims to provide an appropriate model for predicting the causes of cyberloafing behavior in the workplace using the Orange Data Mining application. This study uses an approach of seven tools, including Logistic Regression, SVM, k-NN, Naïve Bayes, Decision Tree, Random Forest, and Neural Network. We compare all these tools using the Orange Data Mining application's evaluation feature, the confusion matrix. The confusion matrix will yield the accuracy value of each model. The highest accuracy value can be used as a reference for the best model. This research is expected to serve as a reference for prediction and a basis for decision-making for companies in modifying their human resource management policies. As explained by Song et al. (2021), 30–50% of employees use the internet during work hours for other activities. Such usage can result in lost work time, which in turn can reduce work output and efficiency.

2. METHODS

This research used the Cross Industry Standard Process for Data Mining (CRISP-DM) as its analytical technique. This strategy was selected due to its methodical approach for resolving HR challenges, such as cyberloafing, through machine learning techniques. This strategy ensures that the analytical procedures are executed systematically and structurally. The CRISP-DM flowchart is shown in **Figure 1** below

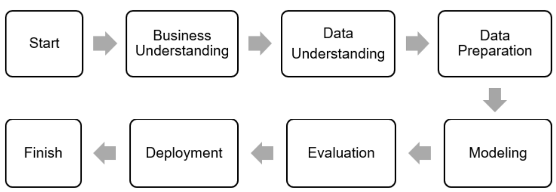

Figure 1. Research Design Flow

Business Understanding. HR managers have discovered the phenomenon of cyberloafing in their companies. Research by Krishna & Agrawal (2023) and Pindek et al. (2018) states that cyberloafing in the workplace can affect company performance. Therefore, HR managers need to analyze the factors that cause cyberloafing using a predictive model, which will serve as input for developing future strategic HR management plans.

Data Understanding. Researcher obtained data from HR managers in the form of company employee databases. Companies use Enterprise Resource Planning (ERP) applications to manage their employee databases. These tools made it a breeze for researchers to retrieve data, as employee data retrieval involved a process of downloading and extracting data. In addition, HR managers also had data from interviews related to cyberloafing. The cyberloafing data includes target attributes (boredom and fatigue) that have been validated by the HR manager. Therefore, instead of conducting direct observations, the researcher relied on data already available in the HR division. The personnel data used consists of information on 131 active employees for the period of November 2024. **Table 1** provides a description of the data used in this study.

Data Preparation. Researchers conducted a comprehensive examination of the data. The inspection included data cleaning and standardization of the writing format. The personnel data sourced from downloads and data extraction in Excel format still contained inconsistent data formats. In addition, the personnel database also contained data that was not relevant to the research needs, such as ID card numbers, tax identification numbers, religion, place of birth, uniform sizes, and shoe sizes. Retired employee data was also still present in the database. Therefore, data cleaning was carried out in two stages: the initial stage was performed manually in Microsoft Excel, and the subsequent stage was performed in the Orange Data Mining Application. The cleaning steps included deleting data that was not relevant to the research requirements and adjusting the data writing format.

Modeling. Researchers use the Orange Data Mining application to perform model building as part of the data processing process. This research will build a prediction model on factors contributing to workplace cyberloafing. As shown in **Figure 2**, the Cause data becomes the target attribute. Meanwhile, the attributes Number, ID Number, Name, and Position (Job) are ignored because they do not relate to the prediction model being developed. Furthermore, making a prediction model requires 2 data sets: training data and testing data. Researchers compared 80% of the data for the training model and 20% of the data for the testing model. This research uses seven supervised learning algorithms to test, evaluate, and identify the right model for this case.

Table 1. Description Data

Attribute	Description	Type
id	Employee identity number	Numeric
Name	Employee name	Text
Position	Current job title	Categorical
Gender	Male or female	Categorical
Level	Job Grade	Categorical
Location	Workplace of employee	Categorical
Position Type	Filed of Employee	Categorical
Latest Education	Highest completed education level	Categorical
Age	Employee's current age	Numeric
Most Visited	Frequently accessed website during work	Categorical
Performance 2024 Semester 1	Work achievement rating	Categorical
Performance 2023 Semester 2	Work achievement rating	Categorical
Tenure	Length of service years	Categorical
Cause	Reason of cyberloafing	Categorical

Evaluation. The prediction model formed based on the supervised learning algorithm is then compared and evaluated by Confusion Matrix in Orange Data Mining. This evaluation is processed by comparing the prediction of the machine learning model against the actual value. The terms in the confusion matrix are as follows:true positives are known as TP, true negatives are known as TN, false positives are known as FP, and true negatives are known as TN. True positives are the number of model predictions that are positive, and the actual data is positive (TP). True negatives are the number of predictions that produce negative values, and the actual data is negative (TN). False positives are the number of predictions that are positive, but the actual data is negative (FP). A false negative is the number of predictions that produce negative values, but the actual data is positive (FN).

In addition, evaluation can also be done by looking at the calculation of the Classification Accuracy value known as CA, Precision value known as Prec, Recall value known as Recall, Area Under the Curve (AUC) value, F1 Score value known as F1, and Matthews Correlation Coefficient value known as (MCC). The CA value is the ratio value shown based on the correct predictions of all predictions made. Prec value is the ratio value for correct predictions out of all positive prediction values. Recall value is the ratio of correct predictions of all positive actual values. The F1 value is the balance value of precision and recall. The AUC value is the ability shown by the model to differentiate between negative values and positive values. The MCC value is a matrix form that takes into account all elements of the confusion matrix, including TP, TN, FP, and FN.

Deployment. The best model is submitted to the HR manager. Furthermore, the HR manager can implement it in the company. This individual can provide input to management to adjust Human Resources Management strategic policies in the future.

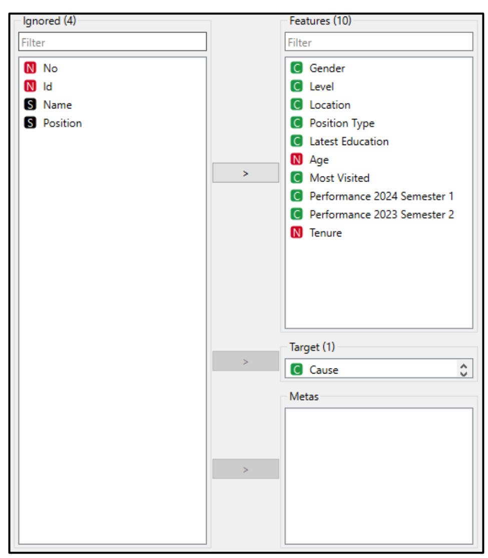

Figure 2. Attribute Data

2.1 Literature Review

Cyberloafing is something that human resource management and organizational behavior studies are paying more and more attention to. This trend is especially true now that workers are more reliant on digital technology at work. Researchers think this behavior is crucial to research since it can hurt productivity and help employees recuperate mentally. As data analytics technologies have improved, cyberloafing study has moved beyond just looking at its conceptual and psychological components and is now starting to apply computational methods to find out what causes it. In this situation, supervised learning algorithms have become a useful model because they can sort behavior patterns based on measurable predictor variables. This helps managers make decisions based on data.

a. Cyberloafing

Cyberloafing is the use of digital resources for non-work-related purposes during working hours (Metin-Orta & Demirtepe-Saygılı, 2023). Research by O'Neill et al. (2014) explains something similar, defining cyberloafing as the use of the internet at work by an employee for personal interests during working hours. The phenomenon of cyberloafing occurs due to several factors in the workplace, mainly boredom and fatigue. Boredom is described as a state in which employees feel underutilized and experience monotonous activities at work (Pindek et al., 2018). Fatigue is usually caused by a high workload. Therefore, some employees refer to cyberloafing as an attempt to restore cognitive function and creativity (Zhong et al., 2022).

b. Logistic Regression

Logistic regression is a probabilistic model utilized for binary or categorical classification in statistics or machine learning, owing to its capacity to balance predicted efficacy with interpretability (Tay et al., 2023). Sun et al. (2019) emphasize that this technique is straightforward and efficient for many applications, particularly in predicting event probabilities. Consequently, logistic regression is valuable for comprehending phenomena like employee behavior.

c. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a tool for solving machine learning problems based on statistical theory with a simple structure and effective generalization and optimization (Hu, 2020). SVM has advantages including flexibility, robust performance, and feature selection (Chidambaram & Srinivasagan, 2016).

d. K-Nearest Neighbors (k-NN)

K-Nearest Neighbors (k-NN) is an algorithm that operates by identifying the closest data points in feature space to a given input and making predictions based on neighbors. This algorithm works by classifying unknown data points by considering the distance to the closest majority class (Hoque et al., 2021). K-NN is effective for solving multi-class problems and can be adapted to include weighted distances between neighbors. Common distance metrics include Euclidean, Manhattan, Minkowski, and Mahalanobis distances (Kalra et al., 2022).

e. Naïve Bayes

Naïve Bayes is a probabilistic algorithm based on Bayes' theorem, mainly used for classification. Naïve Bayes has been effectively applied to text classification due to the algorithm's ability to handle high-dimensional data(Wu et al., 2015). Naïve Bayes is capable of handling high-dimensional data simply and efficiently (Pajila et al., 2023).

f. Decision Tree

Decision Tree is a decision model that is likened to a tree structure, where each internal node represents a test on an attribute, each branch represents the test result, and each leaf node represents a class label or target value (C. Li et al., 2021). Decision Trees have the ability to interpret models that are easy to understand or do not require assumptions about data distribution (Trabelsi et al., 2019).

g. Random Forest

Random Forest is a machine learning algorithm that works by building several decision trees to improve prediction accuracy and reduce the potential for overfitting (Karabadji et al., 2023). Random forest has the advantage of being able to measure variables that can understand the impact of each feature on a prediction (Pious et al., 2024).

h. Neural Network

Neural Networks are algorithm models inspired by the structure and network of biological nerves, or the human brain. This network consists of layers of interconnected nodes that process data through connection weights and activation functions (Shobika et al., 2023). Neural Networks have the ability to model complex and non-linear relationships between inputs and outputs, making them effective for traditional models (Schmidhuber, 2015).

3. RESULTS AND DISCUSSIONS

The model design produced through the supervised learning approach using Orange Data Mining is shown in **Figure 3**.

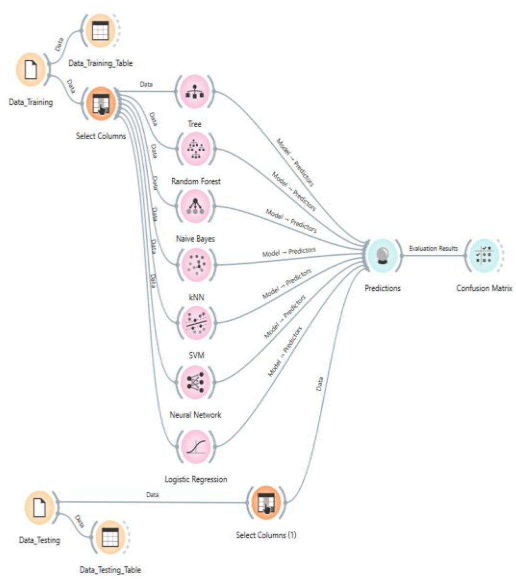

Figure 3. Design Model

This study uses the Orange Data Mining application with seven algorithms to make predictions, and the results will be compared to obtain the best model. Initially, the data is divided into two types: training data and testing data, with a ratio of 80%:20%. Next, as shown in **Figure 4**, predictions are made to provide an overview of the model for each algorithm.

	Tree	Random Forest	Naive Bayes	kNN	SVM	Neural Network	Logistic Regression	Cause	Gender
1	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.60 : 0.40 → BOREDOM	0.49 : 0.51 → FATIGUE	0.84 : 0.16 → BOREDOM	0.47 : 0.53 → FATIGUE	FATIGUE	Male
2	0.00 : 1.00 → FATIGUE	0.22 : 0.78 → FATIGUE	0.45 : 0.55 → FATIGUE	0.40 : 0.60 → FATIGUE	0.21 : 0.79 → FATIGUE	0.08 : 0.92 → FATIGUE	0.26 : 0.74 → FATIGUE	FATIGUE	Male
3	1.00 : 0.00 → BOREDOM	0.76 : 0.24 → BOREDOM	0.98 : 0.02 → BOREDOM	1.00 : 0.00 → BOREDOM	0.73 : 0.27 → BOREDOM	0.98 : 0.02 → BOREDOM	0.94 : 0.06 → BOREDOM	BOREDOM	Male
4	1.00 : 0.00 → BOREDOM	0.88 : 0.12 → BOREDOM	1.00 : 0.00 → BOREDOM	1.00 : 0.00 → BOREDOM	0.83 : 0.17 → BOREDOM	0.97 : 0.03 → BOREDOM	0.94 : 0.06 → BOREDOM	BOREDOM	Female
5	0.00 : 1.00 → FATIGUE	0.11 : 0.89 → FATIGUE	0.58 : 0.42 → BOREDOM	0.40 : 0.60 → FATIGUE	0.18 : 0.82 → FATIGUE	0.08 : 0.92 → FATIGUE	0.31 : 0.69 → FATIGUE	FATIGUE	Male
6	1.00 : 0.00 → BOREDOM	0.88 : 0.12 → BOREDOM	0.98 : 0.02 → BOREDOM	0.80 : 0.20 → BOREDOM	0.89 : 0.11 → BOREDOM	1.00 : 0.00 → BOREDOM	0.95 : 0.05 → BOREDOM	BOREDOM	Male
7	0.00 : 1.00 → FATIGUE	0.20 : 0.80 → FATIGUE	0.47 : 0.53 → FATIGUE	0.40 : 0.60 → FATIGUE	0.30 : 0.70 → FATIGUE	0.08 : 0.92 → FATIGUE	0.23 : 0.77 → FATIGUE	FATIGUE	Male
8	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.60 : 0.40 → BOREDOM	0.49 : 0.51 → FATIGUE	0.83 : 0.17 → BOREDOM	0.43 : 0.57 → FATIGUE	FATIGUE	Male
9	0.00 : 1.00 → FATIGUE	0.30 : 0.70 → FATIGUE	0.26 : 0.74 → FATIGUE	0.40 : 0.60 → FATIGUE	0.24 : 0.76 → FATIGUE	0.02 : 0.98 → FATIGUE	0.10 : 0.90 → FATIGUE	FATIGUE	Male
10	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.60 : 0.40 → BOREDOM	0.49 : 0.51 → FATIGUE	0.83 : 0.17 → BOREDOM	0.43 : 0.57 → FATIGUE	FATIGUE	Male
11	0.00 : 1.00 → FATIGUE	0.07 : 0.93 → FATIGUE	0.42 : 0.58 → FATIGUE	0.20 : 0.80 → FATIGUE	0.13 : 0.87 → FATIGUE	0.03 : 0.97 → FATIGUE	0.21 : 0.79 → FATIGUE	BOREDOM	Male
12	0.00 : 1.00 → FATIGUE	0.22 : 0.78 → FATIGUE	0.22 : 0.78 → FATIGUE	0.20 : 0.80 → FATIGUE	0.22 : 0.78 → FATIGUE	0.01 : 0.99 → FATIGUE	0.07 : 0.93 → FATIGUE	FATIGUE	Male
13	1.00 : 0.00 → BOREDOM	0.92 : 0.08 → BOREDOM	1.00 : 0.00 → BOREDOM	0.80 : 0.20 → BOREDOM	0.79 : 0.21 → BOREDOM	0.84 : 0.16 → BOREDOM	0.85 : 0.15 → BOREDOM	BOREDOM	Female
14	0.00 : 1.00 → FATIGUE	0.25 : 0.75 → FATIGUE	0.17 : 0.83 → FATIGUE	0.20 : 0.80 → FATIGUE	0.19 : 0.81 → FATIGUE	0.02 : 0.98 → FATIGUE	0.11 : 0.89 → FATIGUE	FATIGUE	Male
15	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.60 : 0.40 → BOREDOM	0.48 : 0.52 → FATIGUE	0.82 : 0.18 → BOREDOM	0.45 : 0.55 → FATIGUE	FATIGUE	Male
16	1.00 : 0.00 → BOREDOM	0.95 : 0.05 → BOREDOM	1.00 : 0.00 → BOREDOM	0.80 : 0.20 → BOREDOM	0.86 : 0.14 → BOREDOM	1.00 : 0.00 → BOREDOM	0.93 : 0.07 → BOREDOM	BOREDOM	Female
17	1.00 : 0.00 → BOREDOM	0.59 : 0.41 → BOREDOM	0.67 : 0.33 → BOREDOM	0.60 : 0.40 → BOREDOM	0.61 : 0.39 → BOREDOM	0.82 : 0.18 → BOREDOM	0.45 : 0.55 → FATIGUE	FATIGUE	Male
18	1.00 : 0.00 → BOREDOM	0.88 : 0.12 → BOREDOM	0.98 : 0.02 → BOREDOM	0.60 : 0.40 → BOREDOM	0.72 : 0.28 → BOREDOM	0.98 : 0.02 → BOREDOM	0.70 : 0.30 → BOREDOM	BOREDOM	Female
19	0.00 : 1.00 → FATIGUE	0.17 : 0.83 → FATIGUE	0.66 : 0.34 → BOREDOM	0.20 : 0.80 → FATIGUE	0.31 : 0.69 → FATIGUE	0.73 : 0.27 → BOREDOM	0.45 : 0.55 → FATIGUE	FATIGUE	Male
20	0.00 : 1.00 → FATIGUE	0.08 : 0.92 → FATIGUE	0.26 : 0.74 → FATIGUE	0.20 : 0.80 → FATIGUE	0.13 : 0.87 → FATIGUE	0.00 : 1.00 → FATIGUE	0.07 : 0.93 → FATIGUE	FATIGUE	Male
21	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.40 : 0.60 → FATIGUE	0.48 : 0.52 → FATIGUE	0.81 : 0.19 → BOREDOM	0.42 : 0.58 → FATIGUE	FATIGUE	Male
22	1.00 : 0.00 → BOREDOM	0.53 : 0.47 → BOREDOM	0.62 : 0.38 → BOREDOM	0.40 : 0.60 → FATIGUE	0.41 : 0.59 → FATIGUE	0.32 : 0.68 → FATIGUE	0.39 : 0.61 → FATIGUE	FATIGUE	Male
23	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.40 : 0.60 → FATIGUE	0.48 : 0.52 → FATIGUE	0.79 : 0.21 → BOREDOM	0.44 : 0.56 → FATIGUE	BOREDOM	Male
24	0.50 : 0.50 → BOREDOM	0.65 : 0.35 → BOREDOM	0.65 : 0.35 → BOREDOM	0.40 : 0.60 → FATIGUE	0.48 : 0.52 → FATIGUE	0.79 : 0.21 → BOREDOM	0.44 : 0.56 → FATIGUE	FATIGUE	Male
25	0.50 : 0.50 → BOREDOM	0.43 : 0.57 → FATIGUE	0.94 : 0.06 → BOREDOM	0.40 : 0.60 → FATIGUE	0.51 : 0.49 → BOREDOM	0.76 : 0.24 → BOREDOM	0.38 : 0.62 → FATIGUE	FATIGUE	Female
26	0.00 : 1.00 → FATIGUE	0.10 : 0.90 → FATIGUE	0.45 : 0.55 → FATIGUE	0.40 : 0.60 → FATIGUE	0.20 : 0.80 → FATIGUE	0.14 : 0.86 → FATIGUE	0.24 : 0.76 → FATIGUE	BOREDOM	Male
27	0.00 : 1.00 → FATIGUE	0.22 : 0.78 → FATIGUE	0.45 : 0.55 → FATIGUE	0.20 : 0.80 → FATIGUE	0.21 : 0.79 → FATIGUE	0.07 : 0.93 → FATIGUE	0.22 : 0.78 → FATIGUE	FATIGUE	Male

Figure 4. Model Prediction Results

The dataset was created based on the personnel database, which contains a total of 131 employees. The distribution of the target variables is 60% fatigue and 40% boredom. The data imbalance is moderate, so oversampling was not performed. The work was done to maintain the original class distribution so that it represents the actual HR operational conditions and to avoid overfitting if oversampling was performed. As a mitigation, this study included various algorithm models capable of handling data imbalance, such as Random Forest and SVM. Through a comparison of 80% for training data and 20% for testing data, a total of 27 employees were predicted for the "Cause" attribute. We compared each model's prediction result for the "Cause" attribute with the actual conditions. The Best Model is the one that predicts values that are the same or close to the actual conditions, according to the prediction model results. Refer to **Figure 5** for further details.

Model	AUC	CA	F1	Prec	Recall	MCC
Logistic Regression	0.812	0.889	0.882	0.905	0.889	0.756
SVM	0.756	0.815	0.812	0.811	0.815	0.574
kNN	0.778	0.704	0.710	0.723	0.704	0.373
Random Forest	0.765	0.630	0.639	0.711	0.630	0.316
Tree	0.762	0.593	0.600	0.691	0.593	0.267
Neural Network	0.812	0.593	0.600	0.691	0.593	0.267
Naive Bayes	0.802	0.519	0.519	0.648	0.519	0.167

Figure 5. Value of Each Model

An evaluation matrix is used to measure how well a model performs prediction or classification tasks (Muntean & Militaru, 2023). All models provide their performance values in several indicators, including Area Under the Curve (AUC), Classification Accuracy (CA), F1 Score (F1), Precision (Prec), Recall (Recall), and

Matthews Correlation Coefficient (MCC). In the context of this study, the HR Team analyzed the causes of cyberloafing behavior at work, namely “Boredom” (psychological) and “Fatigue” (physical). The data was unbalanced because there was more "Fatigue" data than "Boredom" data. Therefore, the model performance parameters of concern for unbalanced data are Precision and Recall (Riyanto et al., 2023; Wong, 2022). Precision is necessary to ensure accurate predictions regarding whether employees engage in cyberloafing behavior due to psychological aspects of boredom or physical aspects of fatigue. Meanwhile, Recall supports minimizing prediction errors. Therefore, the best model selection in the context of this study is the model that provides the highest value for the Precision and Recall parameters.

Based on these considerations, the best model for predicting cyberloafing in power generation companies is Logistic Regression. The Logistic Regression model provides the highest value for Precision at 0.905, or 90.5%, and Recall at 0.889, or 88.9%. The values in this model are higher than those in other models. Meanwhile, the model that provides the lowest values for Precision at 0.648, or 64.8%, and Recall at 0.519, or 51.9%, is Naïve Bayes. Furthermore, for more detailed information on each model, see the Confusion Matrix in **Table 2** below.

Table 2. Model Comparison based on Confusion Matrix

No	Model	TP	TN	FP	FN
1	Logistic Regression	6	18	0	3
2	SVM	6	16	2	3
3	k-NN	6	13	5	3
6	Random Forest	7	10	8	2
5	Decision Tree	7	9	9	2
7	Neural Network	7	9	9	2
4	Naïve Bayes	7	7	11	2

Based on **Table 2**, the Logistic Regression model, as the best model, was able to produce a TP value of 6 and a TN value of 18. In terms of the total TP and TN values, the Logistic Regression model was higher than the other models. This shows that the Logistic Regression model was able to provide more correct prediction values. Furthermore, Logistic Regression produced an FP value of 0 and an FN value of 3. In terms of the total FP and FN values, the Logistic Regression model was smaller than the other models. This shows that the Logistic Regression model provides fewer incorrect predictions. Logistic Regression is a simple model that is straightforward to implement and requires few computational resources, making it suitable for small datasets (Abapihi et al., 2021; Mojsilovic, 2005). Furthermore, even with a small dataset, Logistic Regression is still capable of providing high accuracy. As in related studies on health (diabetes prediction) and plants (rice seed classification), logistic regression performs well in various applications (kumar & Kumar, 2019; Mara et al., 2025)

4. CONCLUSIONS

This study shows that Logistic Regression is an appropriate model for accurately predicting the causes of cyberloafing behavior in the workplace at power generation companies. Logistic Regression provides a Precision value of 90.5% and a Recall value of 88.9%. This indicates that the Logistic Regression model can classify “Boredom” or “Fatigue” based on related factors within the company. Boredom is closely related to psychological factors, while fatigue is related to physical factors. Therefore, this model should be considered as an initial screening tool to detect employees who are at risk of decreased performance due to cyberloafing.

In practical terms, this can help HR teams formulate strategic company policies for the future. This is especially true for companies with high talent mobility. Adjustments to factors that shape work activities, workloads, environments, and employee backgrounds may be necessary so that cyberloafing does not affect company performance. Furthermore, in terms of theoretical implications, this study reinforces the view that machine learning approaches can be applied to human resource issues in companies, such as cyberloafing. This is through the best Logistic Regression model, which is capable of predicting the causes of cyberloafing.

This study certainly has limitations, namely the availability of datasets. The datasets in the target labels exhibit an imbalanced distribution. However, the imbalance in the dataset is moderate and insignificant. This situation reflects the company's actual circumstances and replicates the operational conditions that exist. Therefore, the evaluation metrics for each model are presented as a whole.

Further research could be developed on large datasets with various attributes that constitute cyberloafing at work. Research on the relationship between cyberloafing at work and intention to quit may also help some companies manage their employees, given that internet use is now widespread anywhere and anytime. In addition, the use of similar platforms such as Knime, Weka, and so on is also recommended to provide certain advantages of each platform.

Thus, this study not only presents the predictive capabilities of machine learning models but also provides practical recommendations for HR to translate the findings into company programs. In addition, it opens up opportunities for the development of machine learning approaches in human resource science, especially related to cyberloafing.

REFERENCES

Abapihi, B., Mukhsar, Adhi Wibawa, G. N., Baharuddin, Lumbanraja, F. R., Faisal, M. R., & Sani, A. (2021). Parameter estimation for high dimensional classification model on colon cancer microarray dataset. *Journal of Physics: Conference Series*, *1899*(1), 012113. https://doi.org/10.1088/1742-6596/1899/1/012113

Alabadee, S., & Thanon, K. (2021). Evaluation and Implementation of Malware Classification Using Random Forest Machine Learning Algorithm. *2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM)*, 112–117. https://doi.org/10.1109/ICCITM53167.2021.9677693

Chidambaram, S., & Srinivasagan, K. G. (2016). *Optimization Approach for Feature Selection and Classification with Support Vector Machine* (Vol. 410, pp. 103–111). Springer Verlag. https://doi.org/10.1007/978-81-322-2734-2_11

Chung, D., Yun, J., Lee, J., & Jeon, Y. (2023). Predictive model of employee attrition based on stacking ensemble learning. *Expert Systems with Applications*, *215*, 119364. https://doi.org/10.1016/j.eswa.2022.119364

Demirboga, S., Baykara, M., & Kılınç, A. (2025). Detection and Analysis of Cyberloafing Behaviors in The Workplace: A Multi-Method Literature Review. *2025 9th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, 1–7. https://doi.org/10.1109/ISAS66241.2025.11101964

Divya, & Narwal, M. (2023). Digital Distractions in the Workplace: Exploring Cyberloafing Impact on Employee Behaviour and Innovation. *Virtual Economics*, *6*(4). https://doi.org/10.34021/ve.2023.06.04(1)

Henle, C. A., & Kedharnath, U. (2012). Cyberloafing in the Workplace. In *Encyclopedia of Cyber Behavior* (Vol. 1, pp. 560–573). IGI Global. https://doi.org/10.4018/978-1-4666-0315-8.ch048

Hoque, N., Bhattacharyya, D. K., & Kalita, J. K. (2021). *KNN-DK: A Modified K-NN Classifier with Dynamic k Nearest Neighbors* (Vol. 1319, pp. 21–34). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-33-6919-1_2

Hu, Z. (2020). Economic Forecast of Coastal Regions Based on Support Vector Machine Model. *Journal of Coastal Research*, *110*(sp1). https://doi.org/10.2112/JCR-SI110-050.1

Jamaluddin, H., Ahmad, Z., & Wei, L. T. (2023). Exploring cyberloafing as a coping mechanism in relation to job-related strain and outcomes: A study using the mediational model of stress. *Cogent Social Sciences*, *9*(2). https://doi.org/10.1080/23311886.2023.2278209

Kalra, V., Kashyap, I., & Kaur, H. (2022). Effect of Distance Measures on K-Nearest Neighbour Classifier. *2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 1–7. https://doi.org/10.1109/ICCSEA54677.2022.9936314

Karabadji, N. E. I., Amara Korba, A., Assi, A., Seridi, H., Aridhi, S., & Dhifli, W. (2023). Accuracy and diversity-aware multi-objective approach for random forest construction. *Expert Systems with Applications*, *225*, 120138. https://doi.org/10.1016/j.eswa.2023.120138

Krishna, S. M., & Agrawal, S. (2023). Cyberloafing: Exploring the Role of Psychological Wellbeing and Social Media Learning. *Behavioral Sciences*, *13*(8). https://doi.org/10.3390/bs13080649

kumar, S., & Kumar, Mr. N. (2019). Diagnosis of Diabetes by using Data Mining Techniques. *International Journal of Innovative Technology and Exploring Engineering*, *9*(1), 2368–2372. https://doi.org/10.35940/ijitee.L3174.119119

Li, C., Yang, L., Lu, P., & Wang, Z. (2021). Application of decision tree model to surgical action analysis. *2021 2nd International Conference on Artificial Intelligence and Education (ICAIE)*, 163–166. https://doi.org/10.1109/ICAIE53562.2021.00041

Li, X., Xu, K., Men, X., Lin, Z., Zhao, Y., & Wei, D. (2023). Application Research of Neural Network-based Data Assessment and Prediction Method in Statistical Analysis. *2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP)*, 1185–1189. https://doi.org/10.1109/ICSP58490.2023.10248747

Lizarte Simón, E. J., Khaled Gijón, M., Galván Malagón, M. C., & Gijón Puerta, J. (2024). Challenge-obstacle stressors and cyberloafing among higher vocational education students: the moderating role of smartphone addiction and Maladaptive. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1358634

Maindola, M., Al-Fatlawy, R. R., Kumar, R., Boob, N. S., Sreeja, S. P., Sirisha, N., & Srivastava, A. (2024). Utilizing Random Forests for High-Accuracy Classification in Medical Diagnostics. *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, 1679–1685. https://doi.org/10.1109/IC3I61595.2024.10828609

Mara, M. N., Hidayat, S. S., Putri, F. T., Rahmawati, D., Wahyuni, S., Prabowo, M. C. A., Kabir, N. N. S., & Indra, R. T. (2025). Overcoming imbalanced rice seed germination classification: enhancing accuracy for effective seedling identification. *IAES International Journal of Artificial Intelligence (IJ-AI)*, *14*(1), 62. https://doi.org/10.11591/ijai.v14.i1.pp62-72

Metin-Orta, I., & Demirtepe-Saygılı, D. (2023). Cyberloafing behaviors among university students: Their relationships with positive and negative affect. *Current Psychology*, *42*(13), 11101–11114. https://doi.org/10.1007/s12144-021-02374-3

Mojsilovic, A. (2005). A Logistic Regression Model for Small Sample Classification Problems with Hidden Variables and Non-Linear Relationships: An Application in Business Analytics. *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, 329–332. https://doi.org/10.1109/ICASSP.2005.1416307

Muntean, M., & Militaru, F.-D. (2023). *Metrics for Evaluating Classification Algorithms* (Vol. 321, pp. 307–317). https://doi.org/10.1007/978-981-19-6755-9_24

Novia, A., & Yuadi, I. (2023). Forecasting Employee Potential through Probationary Assessment. *JBMP (Jurnal Bisnis, Manajemen Dan Perbankan)*, *9*(2), 131–149. https://doi.org/10.21070/jbmp.v9i2.1718

O'Neill, T. A., Hambley, L. A., & Chatellier, G. S. (2014). Cyberslacking, engagement, and personality in distributed work environments. *Computers in Human Behavior*, *40*, 152–160. https://doi.org/10.1016/j.chb.2014.08.005

Pajila, P. J. B., Sheena, B. G., Gayathri, A., Aswini, J., Nalini, M., & R, S. S. (2023). A Comprehensive Survey on Naive Bayes Algorithm: Advantages, Limitations and Applications. *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)*, 1228–1234. https://doi.org/10.1109/ICOSEC58147.2023.10276274

Pindek, S., Krajcevska, A., & Spector, P. E. (2018). Cyberloafing as a coping mechanism: Dealing with workplace boredom. *Computers in Human Behavior*, *86*, 147–152. https://doi.org/10.1016/j.chb.2018.04.040

Pious, I. K., Rajalakshmi, A., R, P. K., C M, V., Nalini, M., & R, S. S. (2024). Enhancing Prediction Accuracy Through Random Forest in Classification and Regression. *2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG)*, 1–6. https://doi.org/10.1109/ICSTSDG61998.2024.11026358

Raposo, L. M., Rosa, P. T. C. R., & Nobre, F. F. (2020). *Random Forest Algorithm for Prediction of HIV Drug Resistance* (pp. 109–127). https://doi.org/10.1007/978-3-030-38021-2_6

Riyanto, S., Sitanggang, I. S., Djatna, T., & Atikah, T. D. (2023). Comparative Analysis using Various Performance Metrics in Imbalanced Data for Multi-class Text Classification. *International Journal of Advanced Computer Science and Applications*, *14*(6). https://doi.org/10.14569/IJACSA.2023.01406116

Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. In *Neural Networks* (Vol. 61, pp. 85–117). Elsevier Ltd. https://doi.org/10.1016/j.neunet.2014.09.003

Shobika, G. V., Siddharth, S. P., Vishwa, K., Hemachandran, K., & Rege, M. (2023). ANN Model for Analytics. In *Artificial Intelligence and Knowledge Processing* (1st ed., pp. 195–207). CRC Press. https://doi.org/10.1201/9781003328414-17

Song, M., Ugrin, J., Li, M., Wu, J., Guo, S., & Zhang, W. (2021). Do Deterrence Mechanisms Reduce Cyberloafing When It Is an Observed Workplace Norm? A Moderated Mediation Model. *International Journal of Environmental Research and Public Health*, *18*(13), 6751. https://doi.org/10.3390/ijerph18136751

Sun, Y., Zhang, Z., Yang, Z., & Li, D. (2019). Application of Logistic Regression with Fixed Memory Step Gradient Descent Method in Multi-Class Classification Problem. *2019 6th International Conference on Systems and Informatics (ICSAI)*, 516–521. https://doi.org/10.1109/ICSAI48974.2019.9010220

Tay, J. K., Narasimhan, B., & Hastie, T. (2023). Elastic Net Regularization Paths for All Generalized Linear Models. *Journal of Statistical Software*, *106*(1). https://doi.org/10.18637/jss.v106.i01

Trabelsi, A., Elouedi, Z., & Lefevre, E. (2019). Decision tree classifiers for evidential attribute values and class labels. *Fuzzy Sets and Systems*, *366*, 46–62. https://doi.org/10.1016/j.fss.2018.11.006

Wong, T.-T. (2022). Linear Approximation of F-Measure for the Performance Evaluation of Classification Algorithms on Imbalanced Data Sets. *IEEE Transactions on Knowledge and Data Engineering*, *34*(2), 753–763. https://doi.org/10.1109/TKDE.2020.2986749

Wu, J., Pan, S., Zhu, X., Cai, Z., Zhang, P., & Zhang, C. (2015). Self-adaptive attribute weighting for Naive Bayes classification. *Expert Systems with Applications*, *42*(3), 1487–1502. https://doi.org/10.1016/j.eswa.2014.09.019

Zhong, J., Chen, Y., Yan, J., & Luo, J. (2022). The mixed blessing of cyberloafing on innovation performance during the COVID-19 pandemic. *Computers in Human Behavior*, *126*, 106982. https://doi.org/10.1016/j.chb.2021.106982