Jurnal Sistem Informasi Bisnis 01(2025)
 p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/64313

Prediction Analysis of Sleep Disorders Using Machine Learning-Based Techniques

Mega Setiawati, Denise Aldianto, Sulis Sandiwarno*

Faculty of Computer Science, Information Systems, Universitas Mercu Buana, Jakarta, Indonesia

Submitted: June 6th, 2024; Revised: January 20th, 2025;
Accepted: January 22th, 2025; Available Online: February 28th, 2025
DOI: 10.21456/vol15iss1pp89-101

Abstract

Sleep is crucial indicator for an individual. Poor sleep quality has serious implication for health. This condition is often triggered by high work pressure and imbalance between work and rest time. While previous research with similar topic has been conducted, it has not comprehensively elucidated the key factors influencing sleep disorders. Therefore, this study conducts more in-depth analysis of factors contributing to sleep disorders including; gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI, heart rate, and daily steps. Subsequently, we employ Machine Learning (ML) techniques to investigate further sleep disorders. The ML models include: Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Convolutional Neural Network (CNN), dan Long Short-Term Memory Network (LSTM). The objective is to assess the effectiveness of ML model implementation based on information from data and the significance of specific factors in predicting sleep disturbances. The results of this study indicate that the combination of the LR model with Chi-Square achieved the highest average F1 score, which was 84.75%, in sleep disorder classification. The research comprises several stages: (1) Data collection, (2) Pre-processing of the collected data, and (3) Training models capable of processing data for evaluation to understand the contribution of indicators to sleep disorder predictions. The findings of this study provide insights into the effectiveness of the constructed models in predicting sleep disorders

Keywords: Prediction; Sleep Disorder; Machine Learning; Occupation.

1. Introduction

Sleep disorders are a common health issue worldwide and have serious impacts on individuals' physical and mental well-being (Nelson et al., 2022). Globally, 44% of adults participating in the Philips Global Sleep survey reported that their sleep quality has worsened over the past five years. The survey also revealed that 76% of adults worldwide experience at least one condition affecting their sleep quality, including insomnia (37%) and snoring (29%). (O'Reilly, 2019).

In Indonesia, epidemiological research on sleep disorders remains scarce. However, more recent studies indicate that the prevalence of insomnia among individuals aged ≥19 years in Indonesia is 43.7% (Edison & Nainggolan, 2021). Therefore, it is essential to raise awareness of sleep disorder issues in society and take proactive steps to address this problem. Early detection of potential sleep disorders by identifying factors that significantly affect sleep quality can help prevent them. With these efforts, individuals at risk of developing sleep disorders can receive appropriate treatment earlier, reducing the adverse effects on physical and mental health and improving overall quality of life.

In previous research, sleep disorder analysis was conducted using Machine Learning (ML) with the Random Forest (RF) model, which demonstrated good accuracy and effectively addressed overfitting (Hidayat, 2023) However, the study did not highlight the most critical factors contributing to sleep disorders. Additionally, several prior studies using various ML methods have motivated us to explore multiple ML models to identify the most effective one. Therefore, this study employs six ML models: Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Logistic Regression (LR), Convolutional Neural Network (CNN), and Long Short-Term Memory Network (LSTM). These six models will be compared to further evaluate the performance of each. Moreover, this study aims to identify the key factors that contribute most significantly to predicting sleep disorders. By doing so, it is hoped that the models utilized in this research can provide practical solutions for early prediction of sleep disorders.

2. Theoretical Framework

The following are some theories applied in this study:

2.1. Sleep Health

Sleep is a crucial aspect of physical and mental well-being. Good sleep quality plays a vital role in ensuring the body and mind are prepared to handle daily tasks. Sleep disorders such as insomnia or sleep

*)Corresponding author: sulis.sandiwarno@mercubuana.ac.id

Copyright ©2025, JSINBIS, p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/64313

apnea can significantly impact overall quality of life. Each person's sleep needs vary; some, known as long-sleepers, require 9-10 hours of sleep at night, while others may need less than 6 hours. However, sleep duration is not always directly related to sleep disorders (Revaldo et al., 2023).

Sleep is an activity during which the body and brain undergo essential processes, including physical recovery and mental rejuvenation. In today's modern era, where people are busy and under constant pressure, the importance of adequate sleep is often overlooked.

Sleep disorders are a common global health issue and are frequently neglected. Disorders such as insomnia or sleep apnea can affect sleep quality and have serious implications for an individual's health. Chronic sleep deprivation increases the risk of various health problems, including heart disease, obesity, and mental disorders. Sleep disorders can also impact productivity and overall quality of life. Therefore, it is crucial to understand the importance of quality sleep and to strive for more effective detection and management of sleep disorders.

2.2. Machine Learning

Machine learning (ML) is a subfield of Artificial Intelligence (AI) technology and is widely adopted for sentiment analysis and other applications (Sandiwarno et al., 2023a). ML is useful for processing data, which is then analyzed to provide predictive insights automatically (Wardhana et al., 2023). ML leverages computational algorithms to recognize patterns in a dataset, which are then used to make predictions and support decision-making automatically. This study uses four classification models and two neural network models.

Supervised learning refers to ML algorithms that map an input to an output based on input-output pairs from a dataset. The dataset is labeled and divided into two parts: training data and test data. The training data contains output variables that need to be predicted and classified. The patterns learned from the training data are then applied to the test data for prediction and classification (Uddin et al., 2022a). Most of the algorithm models from supervised learning are used in this study, such as NB, SVM, KNN, and Logistic Regression.

Unsupervised learning, ML algorithms that are allowed to find and display interesting structures in the data on their own. These algorithms do not use test data for training. Models are formed by identifying patterns and characteristics in the data. When new data is introduced, the model uses previously learned features to recognize the class of the data (Uddin et al., 2022b). The unsupervised learning algorithm model used is K-Means Clustering.

Reinforcement learning, ML algorithms typically used to optimize a system to achieve complex goals. Reinforcement learning has two types, positive reinforcement learning and negative reinforcement learning (Tineges, 2021).

2.3.1 Naïve Bayes

Naïve Bayes is a probability-based classification algorithm model based on Bayes' theorem. Bayes' theorem is combined with "Naïve," meaning that each attribute is assumed to be independent. Naïve Bayes can be trained and is considered a model in supervised learning. This model has the advantage of requiring only a small set of test data to estimate the parameters needed for classification. Bayes' rule is used to calculate the probability of a class by combining prior probabilities with conditional probabilities into a formula that can be used to compute the likelihood of each possible outcome (Dwi Putra et al., n.d.). The formula used in Bayes' theorem is given in Equation (1).

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \tag{1}$$

A is the class of the data, and B is the test data. The posterior P(A|B) refers to the estimated probability of data belonging to a specific class after observing its feature values. Likelihood P(B|A) refers to the probability of the test data given the class. Prior P(A) and P(B) reflect the initial probabilities of class distributions in the test data. By combining the prior, the likelihood, and the overall information from the margin probability, the result of this model is obtained.

2.3.2 Support Vector Machine

Support Vector Machine (SVM) was developed by Oser, Guyon, and Vapnik in 1992. The classification process using SVM is simply the combination of feature space based on the type of kernel, and its classification is performed using a hyperplane approach, as shown in Figure 1 (Diana Dewi et al., 2023).

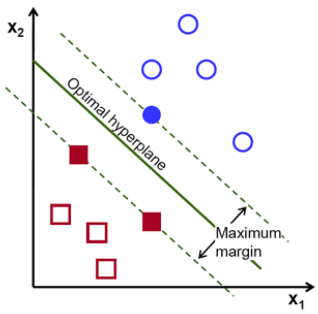

Figure 1. Support Vector Machine

The best hyperplane essentially has the largest margin between the classes being tested. SVM aims to find the optimal separating hyperplane between classes by focusing on the training cases placed in the class descriptors, commonly referred to as support vectors (Rahayu et al., 2023).

2.3.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a supervised learning algorithm where the result of a new query instance is classified based on the majority category in KNN. The class that appears most frequently will become the classification result (Nugroho et al., n.d.). KNN identifies similarities by utilizing information from its nearest neighbors to determine whether a point belongs to a set or not (Qiudandra & Akram, 2022). KNN is performed by finding a group of *k* objects in the training data that are most similar to the object in the test data. To calculate the distance between two points, the Euclidean distance method is used, as given in Equation (2).

$$d_i = \sqrt{\sum_{i=1}^{n}(x_{2i} - x_{1i})^2} \quad (2)$$

Where x_1 represents a sample from the dataset, x_2 represents the test data, i denotes the variables in the data, and d represents the distance between the data points. This algorithm classifies objects based on the closest training data by finding cases using proximity calculations between new and existing cases. This is done by matching the weights of several features. After calculating the squared Euclidean distance of each object to the sample data, these objects are sorted into groups with the smallest Euclidean distances. Once the category of the nearest neighbours with the highest frequency is determined, the value of the query instance can be predicted.

2.3.4 Logistic Regression

Logistic Regression (LR) is an algorithm used to measure the relationship between independent variables and categorical dependent variables, as illustrated in Figure 2. This model has two categories: single logistic regression, where there is one type of input variable, and multiple logistic regression, where there are two or more input variables to predict the value of the output variable (Dwi Yulianto et al., 2023a).

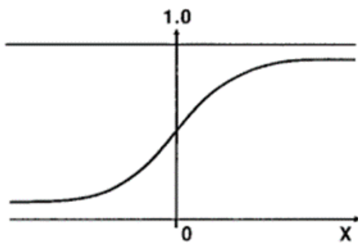

Figure 2. Sigmoid Equation

LR equates the value of Y in the Linear Function with the value of Y in the Sigmoid Function. In the figure above, the line formed by LR represents the probability of an event occurring, ranging from 0 (not occurring/not accurate) to 0.5 (intermediate point) and 1 (occurring/accurate). By measuring the relationship between the target variable and the input variables, probabilities are calculated using the sigmoid function, which transforms the values into 0 or 1(Tri Putra et al., n.d.).

2.3.5 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a variant of artificial neural networks that falls under the category of Deep Learning (DL). CNN is a type of neural network that mimics the visual processing computations that occur in the neural cells of the human brain (Barnes et al., 2022a). CNN consists of a series of layers, including convolution layers, pooling layers, and fully connected layers, as illustrated in Figure 3 (Barnes et al., 2022b).

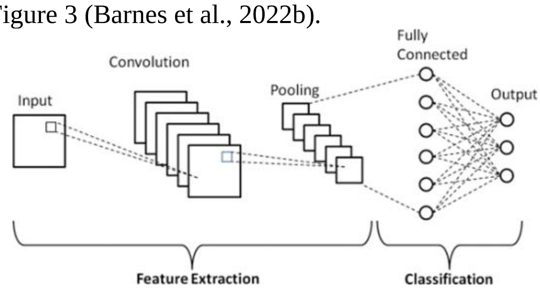

Figure 3. CNN Architecture

The working process of CNN can be observed in Figure 3, where the input, convolutional layers, and pooling layers extract features. Then, the fully connected layer processes the output from the convolutional process to predict different image classes.

2.3.6 Long Short-Term Memory

Long Short-Term Memory (LSTM) is an improvement over the Recurrent Neural Network (RNN) algorithm, designed to address RNN's limitations in handling past data or long-term dependencies (Kashif Bashir et al., n.d.-a). The performance process of LSTM is more efficient in processing, predicting, and classifying data based on its temporal sequence. The structure of LSTM consists of an input layer, an output layer, and a hidden layer, where the hidden layer is composed of memory cells that have several Gates, each with its own function. These include the Input Gate, which functions to manage the amount of information stored in the cell to prevent unnecessary information accumulation; the Forget Gate, which functions to regulate which information is deleted or retained in the memory cell; and the Output Gate, which functions to determine the amount of information produced from the memory cell and generate the output (Kashif Bashir et al., n.d.-b). Below are the gates within the LSTM memory cell as shown in Figure 4.

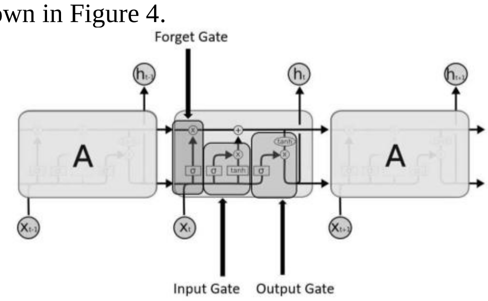

Figure 4. LSTM Gateway

The Forget Gate has an output of 0 or 1 determined by the sigmoid layer called the "forget gate layer," where if the result is 0, the information will be deleted from the memory cell.

$$ft = \sigma(U_f\, x_t + w_f h_{t-1} + b_i) \quad (3)$$

Where f_t is the Forget cell output, w_f and U_f are the weights for the Forget Gate, $ht - 1$ is the previous output, b_i is the bias vector for the Forget Gate, σ is the sigmoid function that transforms the input into values between 0 and 1, and x_t is the input value at time *t*.

The Input Gate functions to input new data from the previous output that passes through the Sigmoid layer, producing values of 0 or 1.

$$i_t = \sigma(w_i\, x_t + U_f h_{t-1} + b_i) \quad (4)$$

Where i_t is the input cell output, w_i and U_f are the weights for the Input Gate, σ is the sigmoid function that transforms the input into values between 0 and 1, b_i is the bias value for the Input Gate, and $ht - 1$ is the previous output.

The Cell State functions as a container to store prior information that can become the output of the LSTM, and $\tilde{C}_t$ represents the updated cell state with prior memory information.

$$C_t = f_t\, x C_{t-1} + i_t\, x\, \tilde{C}_t \quad (5)$$

Where C_t is the candidate cell state, f_t is the output from the Forget Gate, C_{t-1} is the cell state from the previous step, i_t is the output from the Input Gate, and $\tilde{C}_t$ is the candidate for new memory.

The Output Gate regulates how much of the state is passed to the output and operates similarly to other gates, resulting in a new cell state. Below are the formulas for Ot and h_{t-1}.

$$o_t = \sigma(U_o\, W_o h_{t-1} + b_o) \quad (6)$$

$$h_t = O_t\, x\, tanh\,(C_t) \quad (7)$$

Where o_t is the Output Gate, σ is the sigmoid function that transforms the input into values between 0 and 1, U_o and W_o are the weight matrices for the Output Gate, b_o is the bias vector for the Output Gate, h_t is the predicted output, O_t is the output from the Output Gate, and *tanh* is the tangent activation function that limits values to the range between -1 and 1.

2.3. Feature Selection

Feature Selection (FS) is a pre-processing step used to reduce model dimensions and improve classification effectiveness by decreasing the amount of data analyzed and identifying suitable features to consider in the model learning process. FS works by reducing the number of features, eliminating less relevant attributes, and applying appropriate algorithms (Julianto et al., 2022). Before the FS stage, label encoding must be performed to convert categorical labels into numerical values. This step is necessary because FS processes only numerical data. In this study, the authors used two FS algorithms: Chi-square and Mutual Information.

2.3.1 Chi-Square

The chi-square test utilizes statistical theory to examine the independence between terms and their categories (Manfaati Nur & Muntasiroh, 2023). The chi-square method is chosen because it can evaluate the degree of independence of a term or feature with respect to a class label (Ernayanti et al., 2023). The chi-square for each class can be calculated using the equation.

$$x^2(t,c) = \frac{N(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)} \quad (8)$$

Where x^2 represents the chi-square value of feature *t* in class *c*, *N* is the total number of documents. *A* describes the probability of the number of documents in class *c* that contain feature *t*, *B* describes the probability of the number of documents in class *c* that do not contain feature *t*, and *D* is the probability of the number of documents outside class *c* that do not contain feature *t*. The total chi-square value is then calculated using the equation.

$$X^2(t) = \sum_{c=1}^{k} x^2\,(t,c) \quad (9)$$

Where x^2 represents the chi-square value for feature *t*, calculated by summing x^2 (*t*, *c*), the chi-square values of feature *t* across the first class *c* up to *k*, the total number of classes in the label. After determining the chi-square values for each feature, the features are ranked in descending order based on their chi-square values. Features with high chi-square values indicate a high level of dependence on the class label and have a significant impact on the system (Prayoga Permana et al., 2023a).

2.3.2 Mutual Information

Mutual Information (MI) is defined using the concept of Entropy, which indicates how much information is needed to encode a class (AL QODRIN ARUDA, 2022a). The Entropy value serves as an indicator to measure the diversity level of a dataset, with higher dataset diversity resulting in higher Entropy values. MI works by evaluating or assigning weights to continuous attributes, which are then converted into discrete values using their maximum Entropy (AL QODRIN ARUDA, 2022b). MI helps improve the speed and effectiveness of model performance in classification because it eliminates features that do not interpret the classes within the model. In the initial stage, the Entropy value must first be calculated using the equation.

$$Entropy(S) = \sum_{i}^{c} -p_i\, log_2 P_i \quad (10)$$

Where *c* is the number of values in the classification class, and P_i is the proportion of samples in class *i*.After obtaining the Entropy value, the Information Gain can be calculated using the equation.

$$Gain(S,A) = Entropy(S) - \sum_{Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (11)$$

Where *A* is an attribute, *v* is a possible value for attribute *A*, *Values(A)* is the set of possible values for

On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/64313

A, S_v is the number of samples with value v, S is the total number of samples, and Entropy (S_v) is the Entropy of the samples with value v.

2.4. *Confusion Matrix*

The confusion matrix (CM) is used as an evaluation tool to assess the performance of the tested model. This method has four main components: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These four components are used to generate parameters such as accuracy, precision, recall, and f1-score. The formulas for calculating these parameters are presented in the equation (Permana et al., 2023).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

$$Precission = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{TP + FN} \quad (14)$$

$$F1 - Score = \frac{(2 * precission * recall)}{(precission + recall)} \quad (15)$$

3. Method

3.1 *Data Source Description*

This research uses secondary data from a public dataset that can be accessed through Kaggle. The data structure to be used consists of 11 features, namely; gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI, heart rate, daily steps, and 3 classes: Normal, Sleep Apnea, and Insomnia.

3.2 *Data Collection Techniques*

The quantitative methods form the foundation of this research to analyze and interpret the collected data. This method enable the measurement and identification of the data. The following are some of the data collection techniques used in this research.

3.2.1 *Secondary Data Analysis*

Secondary data analysis is the process of collecting and analyzing data that already exists and has been collected by other parties. Secondary data can be accessed publicly for purposes not related to this research.

3.2.2 *Literature Review*

In this method, the author also conducts research from other sources such as journals, books, as well as other sources like articles and documents related to the research topic.

3.3 *Research Flow Diagram*

In this section, the author will describe a series of activities and stages in the research process as shown in Figure 5. Some of the processes include analyzing the problem, determining the problem-solving method, collecting the dataset, identifying the dataset, processing the dataset through a machine learning model, testing the model, and then visualizing the data. The detailed stages of the process will be explained in the flow diagram below.

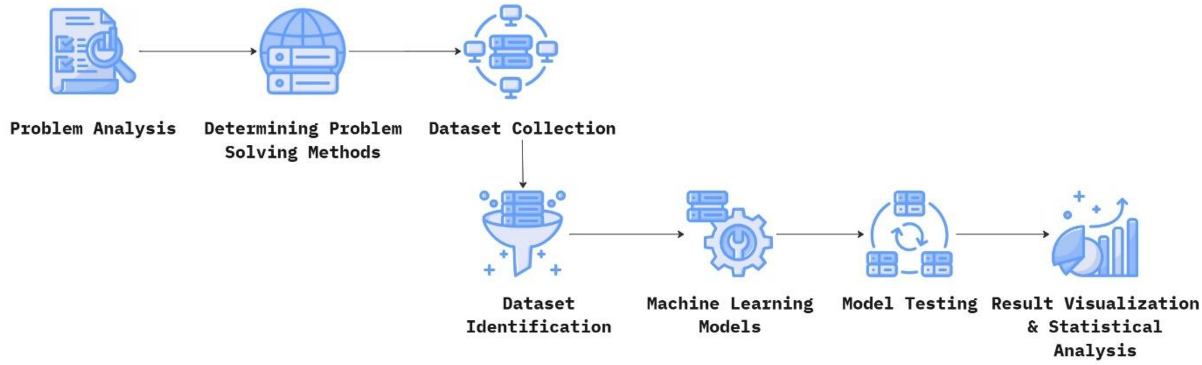

Figure 5. Research Flow Diagram

Here is an explanation of the research flow from the diagram above, as shown in Figure 5.

3.1.1 *Problem Analysis*

In the first stage, the author identifies and analyzes the problem that is the main topic of the research. This stage focuses on gaining a deep understanding of the problem, establishing the research objectives and scope, and conducting a literature analysis of previous research.

3.1.2 *Determining the Method*

After the problem is identified, and the objectives and scope of the research are defined, the author determines the method or approach to solving the problem. This stage involves selecting the appropriate ML model.

3.1.3 *Dataset Collection*

In this stage, the author uses a publicly accessible dataset that has been collected by other parties.

 p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/64313

3.1.4 Dataset Identification

Once the dataset is collected, the next stage involves identifying relevant information, including pre-processing steps.

3.1.5 Pre-Processing

The pre-processing stage is when the data undergoes a series of processes before being used to build the model. The purpose of this stage is to clean and prepare the data so that it becomes more suitable for analysis.

3.1.6 Machine Learning Model Design

In this stage, the author will perform ML modeling using algorithms such as NB, SVM, KNN, LR, CNN, and LSTM. These models are used to predict the potential sleep disorders, and the details of the modeling can be seen in Figure 6.

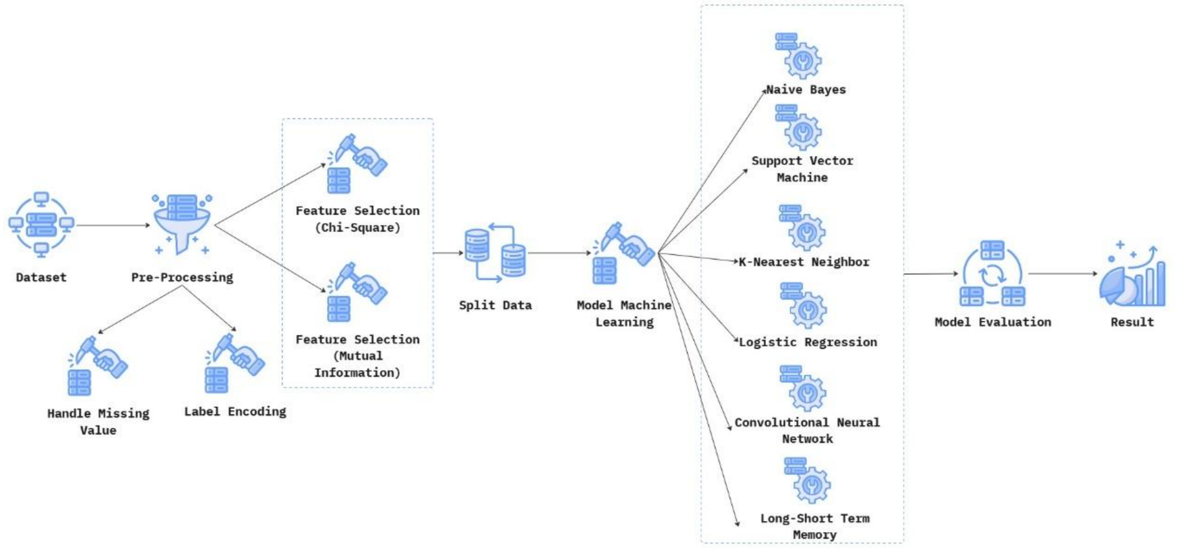

Figure 6. Machine Learning Modeling

The stages in the ML modeling process are as follows:

3.2 Dataset

The first stage involves collecting the data that will be used in the research, which is then prepared for the next stage.

3.3 Pre-Processing

The pre-processing stage is when the data undergoes a series of processes before being used to build the model. The goal of this stage is to clean and prepare the data so that it becomes more suitable for analysis. Some of the actions performed during this stage are as follows:

3.3.1 Handle Missing Value

In this stage, we handle missing or null values in the dataset we collected.

3.3.2 Label Encoding

The next step is performing label encoding on the data, which aims to convert categorical values or labels into numerical values.

3.4 Feature Selection

In the next stage, we perform feature selection, where we choose a subset of features from the available dataset to be used in the model. The purpose of feature selection is to improve model performance by eliminating irrelevant, redundant features, or those that may negatively impact the model. In this process, we use "Sleep Disorder" data as the dependent variable and apply two feature selection algorithms. The selected feature selection methods are chi-square and mutual information.

3.5 Machine Learning Model

This stage involves applying the ML model to the previously processed data

3.6 Model Evaluation

Cross-validation is a statistical method used to evaluate and test the accuracy of a model developed based on a specific dataset (Dwi Yulianto et al., 2023b). To evaluate the model in this research, we use the Confusion Matrix method to determine accuracy, class precision, recall, and other criteria, which will be used to compare the NB, SVM, KNN, LR, CNN, and LSTM models.

3.7 Result

In the final stage, the results of the comparison of several ML algorithms will be visualized to provide recommendations for detecting sleep disorders.

 p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/64313

4. Result and Discussion

4.1 Experimental Setup

This research uses hardware equipped with an Apple M1 Pro processor, 16 GB RAM, and macOS Sonoma 14.0 operating system. The programming language used is Python 3.10 and developed using Jupyter Notebook. Other libraries used in this research can be seen in Table 1.

Table 1. List of Python Libraries

Library	Fuction
Pandas	Python library that allows for data manipulation, cleaning, and analysis.
Numpy	Library that provides mathematical functions and is used to perform various data computations.
Matplotlib	Library for creating data visualizations that provides various types of plots.
TensorFlow	Library that provides a framework for developing and training models.
Scikit Learn	Library that provides various ML algorithms.
Keras	API that facilitates the development, training, and evaluation of Neural Networks models.

4.2 Dataset

The data used in this study comes from a publicly accessible dataset through Kaggle. The dataset contains a total of 374 entries and is structured into 13 columns: personID, gender, age, occupation, sleep duration, quality of sleep, physical activity level, stress level, BMI category, blood pressure, heart rate, daily steps, and sleep disorders, with 3 classes: Normal, Sleep Apnea, and Insomnia. Before the data is used, pre-processing will be conducted first. The pre-processing steps include data cleaning and converting some categorical variables, such as gender, occupation, BMI category, blood pressure, and sleep disorder, into numerical values using label encoding. These steps are necessary to make model development more efficient and consistent. The data distribution by class in the sleep disorder category can be seen in Table 2.

Table 2. Data Distribution by Class

Sleep Disorder	Total
Normal	219
Sleep Apnea	78
Insomnia	77

4.3 Model Training & Evaluation

The trained model will be evaluated using the Confusion Matrix (CM) to assess the model’s performance. This method includes four main components: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These components are used to generate parameters such as accuracy, precision, recall, and F1-score.

Additionally, the model will also be tested using Stratified K-Fold Cross Validation (K-Fold CV). K-Fold CV is a method used to evaluate a model's performance by testing the model on different data subsets (Hafid, 2023). In this study, the data will be divided into 10 different parts, and each model will be tested 10 times, with 20% of the data used as test data and 80% as training data in turn. This method aims to ensure that the model is not only effective on one specific data subset, allowing for more consistent model evaluation.

The parameter settings for deep learning models like CNN and LSTM can be seen in Table 3.

Table 3. Deep Learning Parameter Settings

Parameter Setting	CNN	LSTM
Learning rate	0.001	0.001
Batch size	32	32
Dropout	0,2	0.2
Number of Filters	32	-

Learning rate is one of the hyperparameters used in optimizing ML models, especially deep learning models. It controls the magnitude of the weight changes applied in each iteration during the optimization process. If the learning rate is too small, the training time will be longer. On the other hand, if the learning rate is too large, learning might not be optimal because it progresses too quickly, making the training process unstable (Rochmawati et al., 2021).

A fold is a complete iteration through the entire training data during the model's learning process. The number of epochs chosen for training the model is one hyperparameter that must be defined beforehand. An epoch involves processing the entire training data, and this is repeated to optimize the model (Pitaloka et al., 2024). Batch size refers to the number of data samples that are grouped and processed together in the model. In each iteration, the model receives a batch of data samples and computes the gradients of the loss function with respect to the weights (Pitaloka et al., 2023).

Dropout is a technique used in deep learning to prevent overfitting in the model. Dropout helps improve the model's generalization and makes it more resilient to variations in the data during training. Filters refer to the number of filters applied to perform convolution operations on the input data. During the convolution process, the filters move iteratively across the input data, producing feature maps that represent those features (Sandiwarno et al., 2023b).

4.4 Results of Research

The results from each machine learning and deep learning classification model based on the feature selection algorithms Chi-Square (CS) and Mutual Information (MI) will be compared in terms of precision (Pre), recall (Rec), and F1-score (F1). From Table 4.4, it can be seen that CS provides an average accuracy of 84.75% across all classes in terms of F1. On the other hand, MI provides an average accuracy of 78.59% across all classes in the same context.

Additionally, LR with the CS algorithm outperforms LR with the MI algorithm, with an improvement of 4.42%, 5.13%, and 6.62% across all classes in terms of precision, recall, and F1-score.

Overall, CS provides better results with an average improvement of 4.51% compared to MI. Therefore, in this context, the use of CS for feature selection is more effective in improving the model's performance, especially when used with classification algorithms like NB, SVM (Prayoga Permana et al., 2023b). In this case, LR was evaluated in terms of precision, recall, and F1-score. Below is the detailed model comparison in Table 4.

Table 4. Model Comparison

		Sleep Apnea			Insomnia			Normal		
Models	Metrics	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Chi - Square	Naïve Bayes	74.23	82.62	77.19	78.35	75.48	76.06	94.43	90.85	92.41
	K-Nearest Neighbor	86.99	77.62	79.86	88.69	87.14	81.19	91.28	93.63	92.16
	Support Vector Machine	91.73	79.29	83.34	81.21	80.48	80	90.61	93.73	91.92
	Logisctic Regression	87.96	84.05	84.77	81.13	85.48	82.29	94.67	93.14	93.71
	Convolutional Neural Network	78.92	79.29	77.48	83.44	80.48	81.15	93.81	93.14	93.22
	Long Short-Term Memory Network	87.58	82.62	83.54	78.52	77.14	77.11	91.95	93.14	92.38
Mutual Information	Naïve Bayes	71.71	39.76	48.65	43.77	83.81	56.92	94.19	77.22	84.51
	K-Nearest Neighbor	84.85	80.95	81.22	84.96	87.14	85.06	91.93	90.75	91.06
	Support Vector Machine	88.65	82.62	84.20	84.73	87.14	81.80	94.74	94.80	94.61
	Logisctic Regression	80.59	80.95	79.15	77.65	75.48	75.59	94.12	93.14	93.44
	Convolutional Neural Network	86.65	82.62	83.47	80.01	75.48	76.7	91.45	93.73	92.41
	Long Short-Term Memory Network	70.83	85.71	71.44	76.31	39.52	49.38	71.19	97.12	81.96

From the model comparison, we also visualized the results of the 10-Fold for each model we ran. The 10-Fold results are divided into 3 parts: Precision, Recall, and F1-Score, as illustrated in Figure 7 below.

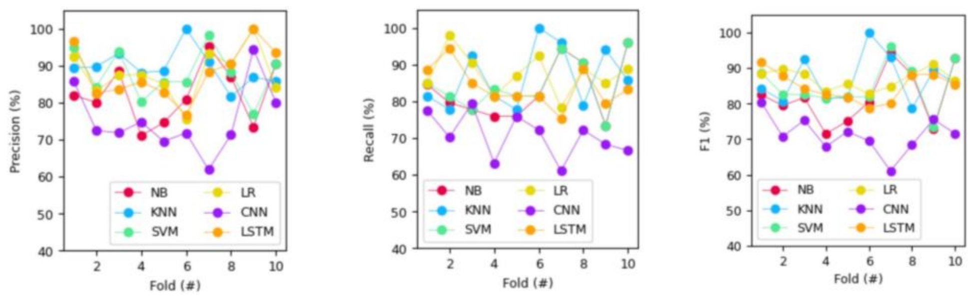

Figure 7. Visualization of 10-Fold

4.5 Statistical Analysis

In this phase, we evaluate the statistical significance using the Analysis of Variance (ANOVA) test. This test aims to determine whether there is a significant difference between the results of the models used with the CS algorithm in terms of F1. In the ANOVA test, Degree of Freedom (DF) is defined as the amount of information in the dataset, Sum of Squares (SS) represents the amount of variation explained by each model, and Mean of Squares (MS) is calculated by dividing the SS by the corresponding DF. Standard Deviation (SD) is a statistic that measures how spread out the data is from the average value. It is calculated as the square root of the variance (Sandiwarno et al., 2023c). A measures the variation of the model's components, and Adjusted Mean of Squares (Adj MS) represents the amount of variation explained by each model term. The F-value is the test statistic used to determine whether a term is related to the response, and the p-value defines the decision on the statistical significance of the model terms.

 p-ISSN: 2502-2377, e-ISSN: 2088-3587
On-line: http://ejournal.undip.ac.id/index.php/jsinbis/article/view/64313

Table 2. ANOVA Analysis on the Model

F1 values of feature modelling techniques						
Source	DF	Adj SS	Adj MS	F-value	p-Value	F-critical
Models	5	1765	353,02	10,37	0	
Error	54	1838	34,05			
Total	59	3604				

Models	Total CV	Mean	SD	95% CI
NB	10	81,89	7,90	(78,19;85,59)
KNN	10	86,76	6,84	(83,06;90,46)
SVM	10	85,09	6,58	(81,39;88,79)
LR	10	86,920	2,75	(83,221;90,619)
CNN	10	71,25	5,22	(67,55;74,95)
LSTM	10	84,89	4,13	(81,19;88,59)
Pooled StDev = 5,83491				

In the ANOVA test results in Table 5, an F-value of 10.37 was obtained for the F1 score of each model, and the F-critical value was 2.68. Based on these results, we can interpret that there is a significant difference in the performance of the six classification models we used (Sandiwarno et al., 2023d). Furthermore, LR based on CS outperforms other models, as shown in Figure 8.

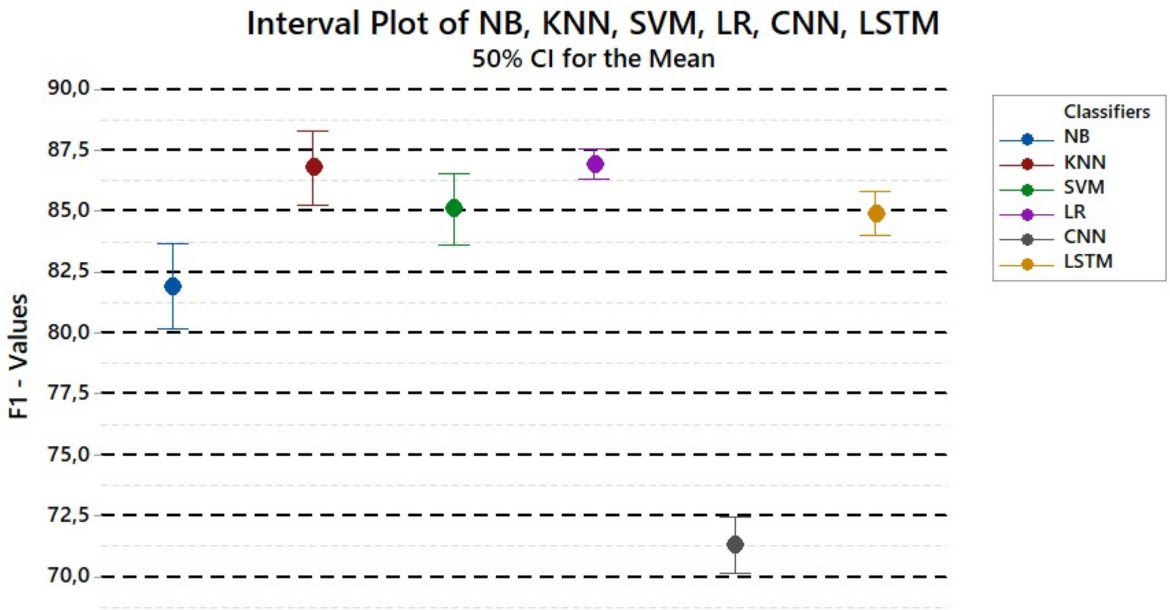

Figure 8. ANOVA Analysis

4.6 Discussion

In this study, we compared several ML algorithms and feature selection algorithms, and successfully identified the best model combination with a high F1 score using LR and CS. The model in this study helped us evaluate the significance of certain factors, such as gender, BMI, and occupation, on sleep disorders.

Table 3. Sleep Disorder Based On Gender, BMI Category, dan Occupation

Variable		Sleep Apnea				Insomnia				Normal			
		PAL		BP		PAL		BP		PAL		BP	
		<75	>75	<120/80	>120/80	<75	>75	<120/80	>120/80	<75	>75	<120/80	>120/80
Gender	Male	56,82	-	-	130/85	48,41	-	-	131/85	-	72,62	120/79	-
	Female	-	77,75	-	139/94	45,00	-	-	133/88	50,15	-	119/79	-
BMI	Normal	54,38	-	-	125/83	45,00	-	-	127/83	55,80	79,59	-	123/81
	Overweight	59,79	-	-	134/89	56,63	-	-	132/87	56,27	-	121/78	-
	Obese	59,00	-	-	137/89	50,00	-	-	141/91	-	-	-	-
Occupation	Sales	31,00	-	-	136/88	-	-	-	-	-	-	-	-
	Software Engineer	-	-	-	-	30,00	-	-	140/90	54,00	-	120/81	-
	Teacher	42,50	-	-	127/83	40,65	-	-	135/89	47,02	-	120/79	-
	Accountant	-	-	-	-	47,14	-	-	128/84	55,52	-	117/76	-
	Doctor	68,75	-	-	130/86	-	-	-	136/88	-	75,73	120/76	-

From this study, we attempted to analyze the significance of various factors on sleep disorders such as Sleep Apnea and Insomnia. In Table 6, we calculate the overall averages for each factor such as physical activity level (PAL), blood pressure (BP), BMI, gender, and occupation. We calculated the number of people with positive PAL scores (PAL >75) and negative PAL scores (PAL <75), based on the recommendation for adults to achieve at least 75 minutes of physical activity per week for a healthy lifestyle (*How Much Physical Activity Do Adults Need?*, 2022) From the table, it can be seen that for the sleep disorders Sleep Apnea and Insomnia, the number of people with a negative PAL (<75) is greater than those with a positive PAL (>75). This indicates that those with low PAL are more susceptible to sleep disorders due to poor sleep quality (Lakshmi Narayana et al., 2023).

Other factors such as body mass index (BMI) and occupation also affect whether a group experiences sleep disorders. For example, in the sleep apnea group, those with normal BMI and negative PAL represent 14.5%, overweight 15.94%, and obese 15.73%, without having a positive PAL. For the insomnia group, those with normal BMI and negative PAL represent 12%, overweight 15.01%, obese 13.33%, with no positive PAL score. For the group without sleep disorders, those with normal BMI and negative PAL represent 14.88%, overweight 15.01%, and positive PAL 21.28%. When we sum the total of negative PAL for the groups with sleep disorders, we get 26.5% for normal BMI, 30.95% for overweight, and 29.06% for obese. From these results, we conclude that the overweight and obese groups are more prone to sleep disorders (Amiri, 2023).

In addition to BMI, the sleep apnea group employed as salespersons with negative PAL is 8.29%, teachers 11.36%, and doctors 18.38%. For the insomnia group, software engineers with negative PAL represent 8.02%, teachers 10.87%, doctors 11.59%, and accountants 12.60%. For the normal group, salespeople with negative PAL represent 14.44%, teachers 12.57%, and accountants 14.84%, while other occupations like doctors with positive PAL represent 20.25%. We then summed the total negative PAL for the sleep apnea and insomnia groups across each occupation category: software engineer 8.02%, sales 8.29%, accountant 12.60%, doctor 18.38%, teacher 22.23%. From this analysis, we conclude that teachers are more vulnerable to sleep disorders. This may be influenced by other factors, such as work pressure leading to high stress (Schmidt et al., 2023).

Another factor, such as blood pressure (BP), also plays a significant role in sleep disorders. We used the guidelines set by the American Heart Association (AHA) to determine positive and negative BP values. Positive BP is defined as <120/80, classified as "Normal," and negative BP is defined as >120/80, classified as "Elevated" or pre-hypertension (American Heart Association, 2023). From the table, it can be observed that those with sleep disorders, whether sleep apnea or insomnia, generally have negative BP >120/80, averaging 133/87. On the other hand, the group without sleep disorders (normal) falls within the positive BP range <120/80, averaging 120/79, which is considered normal according to AHA guidelines.

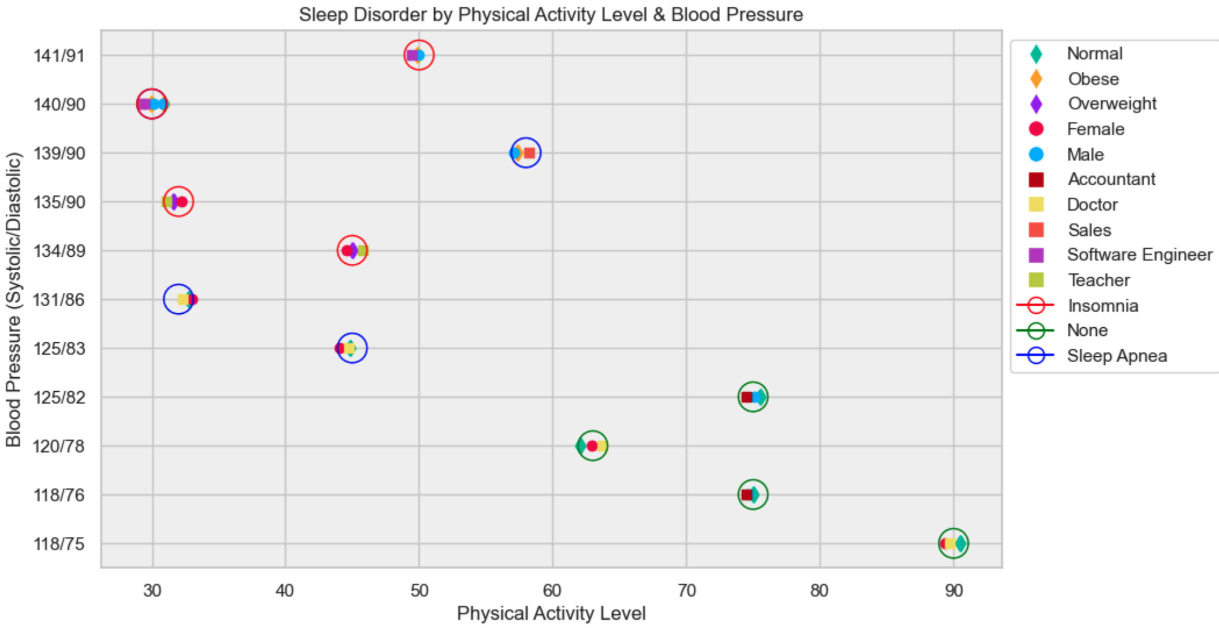

Figure 9. Scatter Plot of Sleep Disorder by Physical Activity Level & Blood Pressure

In Figure 9, we attempt to illustrate the data distribution concerning sleep disorders, with sleep apnea marked in blue, insomnia in red, and green for None (no sleep disorder). In the sleep apnea and insomnia groups, it can be observed that the data distribution across various categories such as BMI, gender, and occupation clusters in an area with high BP ranges, between 125/83 - 141/91, and low PAL, between 30-58. In contrast, for the 'Normal' (None) group, the plot shows that this group tends to have relatively low BP, ranging from 118/75 - 125/82, and relatively high PAL, between 65-90.

5. Conclusion and Recommendations

5.*1 Conclusion*

The results of the analysis and testing on sleep disorder data indicate that the combination of LR and CS models aids in classifying sleep disorders such as sleep apnea and insomnia through indicators that provide high significance, such as BMI, occupation, physical activity level, and blood pressure. The classification results demonstrate that the model is not only able to distinguish between individuals with sleep disorders and those without but also provide insights into the indicators that correlate with sleep disorders.

The study results show that LR, using CS, can provide an average F1 score of 84.75% in the context of sleep disorder classification. In contrast, previous studies have shown that the RF model has good accuracy and can handle overfitting in predicting sleep disorders (Hidayat, 2023), In efforts to improve this sleep disorder research, we also conducted an in-depth analysis of the relationships between certain indicators and sleep disorders. The findings indicate a high correlation between certain indicators, such as BMI categories and specific job backgrounds, with sleep disorders, whether sleep apnea or insomnia.

The analysis then showed that groups with overweight and obese BMI categories, as well as those with a teaching occupation background, are more vulnerable to sleep disorders, whether sleep apnea or insomnia. This demonstrates that the model not only performs classification but also provides a deeper understanding of the factors involved in the emergence of sleep disorders.

5.2 Recommendations

Although the combination of LR and CS models has shown good performance in classifying sleep disorders, several recommendations can be considered for future research. First, it is important to expand the dataset by collecting more sleep disorder-related data from various sources. A larger dataset could provide more representative results and reduce the risk of overfitting in the model. Second, in addition to the factors already considered, other factors that may influence sleep disorders, such as eating habits, sleep patterns, stress levels, and sleeping environments, should be taken into account. By conducting further analysis of other indicators and sleep disorders, future research can uncover more complex patterns in the data using more advanced machine learning models.

Acknowledgments

We would like to express our deepest gratitude to the University of Mercu Buana Jakarta for providing support and guiding the authors in conducting this research.

Bibliography

Al Qodrin Aruda, S., 2022. Perbandingan Metode Seleksi Fitur untuk Klasifikasi Pertanyaan Berbahasa Indonesia menggunakan Algoritma Support Machine (SVM). 1–32.

American Heart Association, 2023. Understanding blood pressure readings. Retrieved from https://www.heart.org/en/health-topics/high-blood-pressure/understanding-blood-pressure-readings

Amiri, S., 2023. Body mass index and sleep disturbances: A Systematic Review and Meta-Analysis. Postepy Psychiatrii I Neurologii, 32(2), 96–106. Termedia Publishing House Ltd. https://doi.org/10.5114/ppn.2023.129067

Barnes, L. D., Lee, K., Kempa-Liehr, A. W., & Hallum, L. E., 2022. Detection of Sleep Apnea From Single-Channel Electroencephalogram (EEG) using an Explainable Convolutional Neural Network (CNN). PLOS ONE, 17(9), e0272167. https://doi.org/10.1371/journal.pone.0272167

Diana Dewi, D., Qisthi, N., Lestari, S. S. S., & Putri, Z. H. S., 2023. Perbandingan Metode Neural Network dan Support Vector Machine dalam Klasifikasi Diagnosa Penyakit Diabetes. Cerdika: Jurnal Ilmiah Indonesia, 3(09), 828–839. https://doi.org/10.59141/cerdika.v3i09.662

Dwi Putra, H., Khairani, L., Hastari, D., & Studi Sistem Informasi, P. (n.d.). Comparison of Naive Bayes Classifier and Support Vector Machine Algorithms for Classifying Student Mental Health Data. Sentimas: Seminar Nasional Penelitian dan Pengabdian Masyarakat. Retrieved from https://journal.irpi.or.id/index.php/sentimas

Dwi Yulianto, L., Hermaliani, H. E., & Kurniawati, L., 2023. Penerapan Machine Learning dalam Analisis Stadium Penyakit Hati untuk Proses Diagnosis dan Perawatan. Resolusi: Rekayasa Teknik Informatika dan Informasi, 3(4), 144–154. Retrieved from https://djournals.com/resolusi

Edison, H., & Nainggolan, O., 2021. Hubungan Insomnia dengan Hipertensi. Buletin Penelitian Sistem Kesehatan, 24(1), 46–56. https://doi.org/10.22435/hsr.v24i1.3579

Ernayanti, T., Mustafid, M., Rusgiyono, A., & Hakim, A. R., 2023. Penggunaan Seleksi Fitur Chi-Square dan Algoritma Multinomial Naïve Bayes untuk Analisis Sentimen Pelangggan Tokopedia. Jurnal Gaussian, 11(4), 562–571. https://doi.org/10.14710/j.gauss.11.4.562-571

Hafid, H., 2023. Penerapan K-Fold Cross Validation untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor Pada Data Kasus Covid-19 di Indonesia. Journal of Mathematics Computations and Statistics, 6(2), 161–168.

Hidayat, I. A., 2023. Classification of Sleep Disorders using Random Forest on Sleep Health and Lifestyle Dataset. Journal of Dinda Data Science, Information Technology, and Data Analytics, 3(2), 71–76. Retrieved from http://journal.ittelkom-pwt.ac.id/index.php/dinda

Centers for Disease Control and Prevention, 2022. How Much Physical Activity Do Adults Need? Retrieved from https://www.cdc.gov/physicalactivity/basics/adults/index.htm

Julianto, I. T., Kurniadi, D., Nashrulloh, M. R., Mulyani, A., & Komputer, J. I., 2022. Comparison of Classification Algorithm and Feature Selection in Bitcoin Sentiment Analysis. Jurnal Teknik Informatika (JUTIF), 3(3). https://doi.org/10.20884/1.jutif.2022.3.3.343

Kashif Bashir, A., Bingfeng, L., Haifeng, L., Feng, H., Sun, L., Zhuang, L., Dai, M., & Zhou, Y. (n.d.). Intelligent automatic sleep staging model based on CNN and LSTM.

Lakshmi Narayana, Y., Kumar, T. S., & Yaswanthi, T., 2023. A Cross Sectional Study on Effect of Physical Activity on Improving Sleep Quality Among Young Adults. International Journal of Health Sciences and Research, 13(2), 50–62. https://doi.org/10.52403/ijhsr.20230209

Manfaati Nur, I., & Muntasiroh, L., 2023. Implementasi Adaptive Synthetic Bernoulli Naïve Bayes untuk Mengatasi Imbalance Class Data Pada Kasus Mental Health di Indonesia. Jurnal Data Informasi, 1(1). Retrieved from http://journalnew.unimus.ac.id/index.php/jodi

Nelson, K. L., Davis, J. E., & Corbett, C. F., 2022. Sleep quality: An Evolutionary Concept Analysis. Nursing Forum, 57(1), 144–151. https://doi.org/10.1111/nuf.12659

Nugroho, A., Hendric, H. L., Warnars, S., Gaol, F. L., & Matsuo, T. (n.d.). Trend of Stunting Weight for Infants and Toddlers using Decision Tree.

O'Reilly, K., 2019. Philips Global Sleep Survey Shows We Want Better Sleep, But Only if it Comes Easily. Retrieved from https://www.philips.com/a-w/about/News/Archive/Standard/News/Press/2019/20190307-Philips-Global-Sleep-Survey-Shows-We-Want-Better-Sleep-but-Only-If-It-Comes-Easily.html

Permana, A. P., Chamidy, T., & Crysdian, C., 2023. Klasifikasi Ulasan Fasilitas Publik menggunakan Metode Naïve Bayes dengan Seleksi Fitur Chi-Square. JISKA (Jurnal Informatika Sunan Kalijaga), 8(2), 112–124. https://doi.org/10.14421/jiska.2023.8.2.112-124

Pitaloka, E., Bagus Agung Hartanto, T., & Sandiwarno, S., 2023. Penerapan Machine Learning untuk Prediksi Bencana Banjir. In Jurnal Sistem Informasi Bisnis.

Prayoga Permana, A., Chamidy, T., & Crysdian, C., 2023a. Klasifikasi Ulasan Fasilitas Publik menggunakan Metode Naïve Bayes dengan Seleksi Fitur Chi-Square. In Jurnal Informatika Sunan Kalijaga) (Vol. 8, Issue 2). Mei.

Qiudandra, E., & Akram, R., 2022. Sistem Pakar Diagnosa Penyakit Osteoarthritis dengan menggunakan Metode K-Nearest Neighbor. Methotika, 2(2). Retrieved from http://ojs.fikom-methodist.net/index.php/METHOTIKA

Rahayu, S., Nugroho, A., Dwika Putra, E., Purba, M., Setiawan, H., & Sulis Sandiwarno, 2023. Comparison of HSV, LAB and YCrCb Color Feature Extraction Results in the SVM Algorithm for Bird Species Classification. JSAI (Journal Scientific and Applied Informatics), 6(3), 451–456. https://doi.org/10.36085/jsai.v6i3.5920

Revaldo, A., Yupianti, Y., & Beti, I. Y., 2023. Sistem Pakar Diagnosa Penyakit Gangguan Tidur dengan Metode Forward Chaining Berbasis Web (Studi Kasus : UPTD Puskesmas Telaga Dewa Kota Bengkulu). Jurnal Media Infotama, 19(1), 44–51. https://doi.org/10.37676/jmi.v19i1.3314

Rochmawati, N., Hidayati, H. B., Yamasari, Y., Tjahyaningtijas, H. P. A., Yustanti, W., & Prihanto, A., 2021. Analisa Learning Rate dan Batch Size pada Klasifikasi Covid Menggunakan Deep Learning dengan Optimizer Adam. Journal of Information Engineering and Educational Technology, 5(2), 44–48. https://doi.org/10.26740/jieet.v5n2.p44-48

Sandiwarno, S., Niu, Z., & Nyamawe, A. S., 2023. A Novel Hybrid Machine Learning Model for Analyzing E-Learning Users' Satisfaction. International Journal of Human-Computer Interaction. https://doi.org/10.1080/10447318.2023.2209986

Schmidt, L. I., Steenbock, L. M., & Sieverding, M., 2023. Improving Sleep Among Teachers: an Implementation-Intention Intervention. International Journal of Behavioral Medicine, 30(1), 49–61. https://doi.org/10.1007/s12529-022-10069-7

Tineges, R., 2021. Wajib tahu, 3 tipe algoritma machine learning. Retrieved from https://dqlab.id/3-tipe-algoritma-machine-learning

Tri Putra, B., Yulianingsih, E., & Bina Darma, U. (n.d.). Analisis Tingkat Akurasi Prediksi Gejala

COVID-19 dengan menggunakan Metode Logistic Regression dan Support Vector Machine. Retrieved from http://github.com/nshomron/covidpred

Uddin, S., Haque, I., Lu, H., Moni, M. A., & Gide, E., 2022. Comparative Performance Analysis Of K-Nearest Neighbour (KNN) Algorithm and its Different Variants for Disease Prediction. Scientific Reports, 12(1), 6256. https://doi.org/10.1038/s41598-022-10358-x

Wardhana, R. G., Wang, G., & Sibuea, F., 2023. Penerapan Machine Learning dalam Prediksi Tingkat Kasus Penyakit di Indonesia. Journal of Information System Management (JOISM), 5(1), 40–45. https://doi.org/10.24076/joism.2023v5i1.1136