Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 8 No. 3, November 2023, pp. 298 – 309
DOI:

Prediction of Sleep Disorders Based on Occupation and Lifestyle: Performance Comparison of Decision Tree, Random Forest, and Naïve Bayes Classifier

Heru Lestiawan[1], Cahaya Jatmoko[2], Feri Agustina[3], Daurat Sinaga[4]
[1,2,3,4]Study Program in Informatics Engineering, University of Dian Nuswantoro, Imam Bonjol No.207, Pendrikan Kidul, Semarang, Jawa Tengah 50131, (024) 3517261
E-mail : herul@dosen.dinus.ac.id[1], jatmoko14@dsn.dinus.ac.id[2], feri.agustina@dsn.dinus.ac.id[3], dauratsinaga@dsn.dinus.ac.id[4]

Lalang Erawan[5]
[5]Study Program in Information System, University of Dian Nuswantoro, Imam Bonjol No.207, Pendrikan Kidul, Semarang, Jawa Tengah 50131, (024) 3517261
E-mail : lalang@dosen.dinus.ac.id[5]

Abstract – Health is a very important thing in life. Therefore, to maintain health, we need adequate rest. Without adequate rest, the body will not be healthy and fit. In this study, a person's sleep disorder prediction will be made based on their lifestyle and work. The predictions made will classify sleep disorders that are absent, sleep apnea and insomnia from certain lifestyles and work. The methods used to make predictions are decision tree classifier, random forest classifier and naïve Bayes classifier. The test was carried out using a total of 375 data which was broken down into 70% training data and 30% testing data. The results obtained after testing with test data are by using the decision tree classifier algorithm to get an accuracy of 89.431%, using the random forest classifier algorithm to get an accuracy of 90.244% and by using the naïve Bayes classifier algorithm to get an accuracy of 86.992%.

Keywords – sleep disturbance, decision tree, random forest, naïve bayes, accuracy

1. INTRODUCTION

Health is a valuable asset and works like a machine to be able to carry out activities [6]. Health plays an important role in determining a person's productivity in carrying out daily activities [2]. In order for the body to stay healthy, it is a must to be able to do physical activity every day. Physical activity is something that has great benefits for physical and mental health [1]. If our body is healthy, then our productivity will also increase. With high performance and productivity, it can affect career direction [7], therefore it will be easy to get the opportunity to be promoted to a higher position [3]. Performance and performance are needed for someone to be able to carry out responsibilities so that an existing goal is achieved from the job [8]. The performance produced by employees is also influenced by their ability and motivation [4]. If the body is not fit, it can affect the mind so that it can cause a decrease in productivity at work [10] and can cause stress. Stress greatly affects a person's performance because it can cause a lack of performance or even a loss of performance [5]. So, maintaining both physical and mental health is very important, because health has a significant impact and affects productivity [9].

Sleep disturbance is a very common disease, which disrupts the body's resting rhythm and has a negative impact on psychological as well as physical well-being [11]. The most common types of sleep disorders are insomnia and sleep acne [18]. Pavlova et al [17], suggested that sleep disorders are characterized by lack of sleep, excessive amount of sleep or abnormal movements that occur during sleep. Sleep disorders greatly affect the quantity of a person's sleep [12]. Due to poor sleep quantity, it can affect sleep patterns, which can affect health and productivity [13]. These impacts can affect life both in the short and long term [14]. Sleep disorders can also affect psychological health which affects performance and professionalism [15]. Wang et. al [16], suggested that sleep disturbance, especially insomnia, is a core symptom of depression. Because of this, we should maintain our sleep patterns to stay healthy, so that we can continue to be more productive.

Machine learning or machine learning is a field that exists in artificial intelligence or often referred to as AI, which is a solution for solving problems digitally [19]. Machine learning aims to enable computers to be able to do their own learning without having to be programmed directly [20]. In machine learning, models will be built based on the data and algorithms provided. Appadoo et. al [21], suggested that the model built from the machine learning algorithm can be used for both classification and regression. Classification is a technique in machine learning to be able to predict or classify the data provided [22]. Sarker [23], explained that the algorithms that are usually used for classification are Naïve Bayes, Linear Discriminant Analysis (LDA), Logistic regression (LR), K-nearest neighbors (KNN), Support vector machine (SVM), Decision tree (DT), Random Forest (RF), Stochastic gradient descent (SGD) and Rule-based classification. In this research, the Decision Tree, Random Forest, and Naïve Bayes algorithms will be used to classify data.

Decision tree or can be called a decision tree is a method commonly used to classify data. In the field of machine learning, learning that is carried out using decision tree models is a popular technique used to be able to classify and build predictive models [26]. Yadav et. al [25], explained that a decision tree is used to be able to extract hidden patterns from existing data. Decision trees work by changing complex decision making into simple ones so that solutions can be found that are easy to understand and visualize [24]. Therefore, using the decision tree method will produce a model which has a simple and easy-to-understand structure and a short running process [27].

The random forest classifier is a machine learning method that is popular enough to be used to develop prediction-based models [28]. Shah et. al suggested that random forest is used to be able to build a predictive model to solve regression and classification problems. Random forest is included in the supervised learning method [30] and based on ensemble learning [29]. It is called ensemble learning because the random forest model is generated from several decision trees which are combined into one so that it forms a forest [30].

Naïve Bayes or NB is the most popular method for classifying [31]. Naïve Bayes works by calculating probabilities so that further analysis can be carried out from the data provided based on the probability results obtained [32]. Sigh et. al [33], stated that Naïve Bayes is a classification algorithm with statistical methods based on Bayes' theorem to find the probability of occurrence of events based on the probability of occurrence of each existing event.

In a study conducted by Dimitriadis et. al [34], discussed the detection of sleep disturbances based on combining EEG cross-frequency and random forest models. This research aims to build a machine learning model that can automatically classify sleep disorders. The results obtained from this study are that the model built has an accuracy of 74% for classifying based on 7 types of sleep disorders and not experiencing sleep disturbances.

Research conducted by Rohan et. al [35], discussed the classification of sleep apnea using a decision tree classifier. The purpose of this study is to be able to build a model that can classify based on data obtained from ECG signals. The results obtained from this study are classification using the decision tree algorithm to get an average accuracy of 98.53%, sensitivity 98.39%, specificity 96.86%, precision 90% and overall F-Score 93.2%.

2. RESEARCH METHOD

2.1. Dataset

Sleep disturbance is a very common disease, which disrupts the body's resting rhythm and has a negative impact on psychological as well as physical well-being [11]. The most common types of sleep disorders are insomnia and sleep acne [18]. Pavlova et al [17], suggested that sleep disorders are characterized by lack of sleep, excessive amount of sleep or abnormal movements that occur during sleep. Sleep disorders greatly affect the quantity of a person's sleep [12]. Due to poor sleep quantity, it can affect sleep patterns, which can affect health and productivity [13].

In this study, we will use the Sleep Health and Lifestyle Dataset obtained from the Kaggle.com website. The total amount of data used is 375 records, with a total of 13 columns where 12 columns are used as predictor variables and 1 column is used as a target. For data that is used as a predictor or influencing data, there is data on Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Heart Rate, Daily Steps, Systolic BP, Diastolic BP. Where the data is lifestyle data and work which will later determine the target variable. For data that is used as a target or affected data, namely Sleep Disorder data which has 3 classes, namely None, Sleep Acnea. For predictor data, the data type of each record can be seen in the Table I below.

Table 1. Data Predictor

No	Record	Tipe Data
1	Gender	Kategorikal {'Male', 'Female'}
2	Age	Numerikal (Non negative)
3	Occupation	Kategorikal { 'Software Engineer', 'Doctor', 'Sales Representative', 'Teacher', 'Nurse', 'Engineer', 'Accountant', 'Scientist', 'Lawyer', 'Salesperson', 'Manager' }
4	Sleep Duration	Numerikal (Non negative)
5	Quality of Sleep	Numerikal (Non negative)
6	Physical Activity Level	Numerikal (Non negative)
7	Stress Level	Numerikal (Non negative)
8	BMI Category	Kategorikal { 'Overweight', 'Normal', 'Obese', 'Normal Weight' }
9	Heart Rate	Numerikal (Non negative)
10	Daily Steps	Numerikal (Non negative)
11	Systolic BP	Numerikal (Non negative)
12	Diastolic BP	Numerikal (Non negative)

After the data has been inputted, a transformation will be carried out from records with categorical values to numerical. It is intended that the value of the predictor can be normalized, so that there are no data gaps. From a total of 375 existing data, it will be split into 70% training data and 30% testing data. With the target data used both for the training process and the testing process has 3 classes, namely None, Sleep Acne, and Insomnia. The training data is used to conduct training on the model using Decision Tree, Random Forest, and naïve Bayes algorithms. While the test data will be used to carry out tests from the results of model training, so that after testing, accuracy and classification results can be calculated.

2.2. Normalization

Normalization is the process of transforming data into a certain range. The purpose of normalization is so that the distribution of data has an even value and there are no data gaps. Usually in the normalization stage, data is converted into a range of values 0 and 1 [36].

Therefore, in this study, the data will be normalized into a range of 0 and 1 values for all predictor variables. Which, before normalization is carried out, records that have categorical values are first transformed into numerical. The normalization formula can be seen below.

$$Scaled\ Data = \frac{N - N.min}{N.max - N.min} \quad (1)$$

2.3. Decision Tree

Decision tree or can be called a decision tree is a method commonly used to classify data. Yadav et. al [25], explained that a decision tree is used to be able to extract hidden patterns from existing data. Decision trees work by changing complex decision making into simple ones so that solutions can be found that are easy to understand and visualize [24]. The decision tree algorithm works by dividing the existing data into subsets based on input features, so that in the process it will build a tree structure accompanied by a series of decision results.

2.4. Random Forest

Random forest or random forest is a method in machine learning that can be used to solve data classification problems. Shah et. al suggested that random forest is used to be able to build a predictive model to solve regression and classification problems. Random forest is included in the supervised learning method [30] and based on ensemble learning [29]. It is called ensemble learning because the random forest model is generated from several decision trees which are combined into one so that it forms a forest [30]. The formula for the random forest can be seen below.

$$Prediction(Z) = \frac{(decisionTree\_1(Z) + decisionTree\_2(Z) + \ldots + decisionTree\_N(Z))}{N} \quad (2)$$

2.5. Naïve Bayes

Naïve Bayes is a method that can be used to classify data. Naïve Bayes works by calculating the probability of the existing data so that further analysis can be carried out from the data provided based on the probability results obtained [33]. Sigh et. al [34], stated that Naïve Bayes is a classification algorithm with statistical methods based on Bayes' theorem to find the probability of occurrence of events based on the probability of occurrence of each existing event. The mathematical formula for naïve Bayes can be seen below. The formula shows the calculation of naïve Bayes with several independent features, where Y = {Y_1, Y_2, …., Y_n} are the features used in class z.

$$P(z/Y) = P(Y_1/z) * P(Y_2/z) * \ldots * P(Y_n/z) * \mathrm{P(z)} \quad (3)$$

2.6. Workflow

In the process of classifying sleep disorders based on lifestyle and work, this study will use MATLAB R2022a for program implementation. First of all, the input data will be read, then the data will be preprocessed and normalized so that it becomes data ready for processing. After preprocessing the data, the data will be divided into 70% training data and 30% testing data. The tests carried out are divided into 2 processes, namely training and testing. At the training stage, the built model will be trained using training data, for the model to learn to recognize existing data. Then after the training process is carried out, a testing process is carried out, in which process the model that has been trained will be tested against the test data to evaluate the performance of the model that has been trained. After the testing process is carried out, the performance measurement process of the model can be carried out. For flowchart visualization can be seen in the image below.

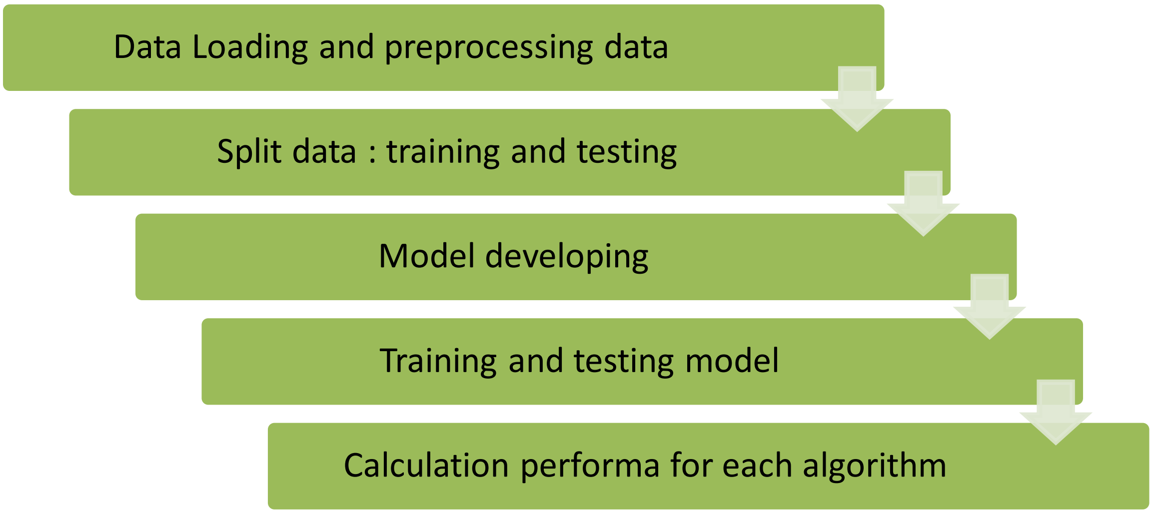

Figure 1. Prediction Scheme

In the data classification process, the first thing that must be done is to input the data used for the training process and model testing. The data used in this study is the Sleep Health and Lifestyle Dataset obtained from the Kaggle.com website. After the data is read, data preprocessing will be carried out, in the form of transformation from categorical data to numeric data and data scaling process. After preprocessing, training and testing of the model is carried out on the data that has been preprocessed. So that later it can be calculated the performance of the model after model training. For an explanation of the process flow, see below.

1. Data reading
2. Data preprocessing
   - Transformation of categorical data into numeric data

     ```
     mapping = containers.Map({'Cat1', 'Cat2', … , 'Cat(n)'}, {Val1, Val2, … , Val(n)});
     Column = data. Column;
     Numeric = cellfun(@(y) mapping(y), Column);
     data. Column = Numeric;
     ```
   - • Scaling the data into a range of 0 and 1 values, using the min max scaler

     ```
     columns = {'Var1', 'Var2', … , 'Var(n)'};
     scale = x_data{:, columns };
     min = min(scale);
     max = max(scale);
     scaled_data = (scale - min) ./ (max - min);
     ```
3. Distribution of preprocessing data into training data and test data with a percentage of 70% training data and 30% testing data.

   ```
   rng('default');
   sizes = 0.33;
   indices = randperm(height(x));split_idx = round(size * height(x));
     X_train = x(indices(split_idx+1:end), :);
   ```

X_test = x(indices(1:split_idx), :);
y_train = y(indices(split_idx+1:end));
y_test = y(indices(1:split_idx));

4. After preprocessing the data and splitting the data, the model development will be carried out. The model built will use the Decision Tree Classifier, Random Forest Classifier and Naïve Bayes Classifier algorithms. The parameters used in each algorithm can be seen in the following Table 2.

Table 2. Decision Tree Parameter

No	**Decision Tree Classifier**	
	Parameter	**Values**
1	MaxNumSplits	4
2	MinLeafSize	1
3	MinParentSize	10
4	SplitCriterion	deviance
5	Surrogate	on
6	Prune	on

Table 3. Random Forest Parameter

No	Random Forest Classifier	
	Parameter	Values
1	NumTrees	100
2	Method	classification
3	MinLeafSize	1
4	FBoot	1.0
5	OOBPrediction	off
6	OOBPredictorImportance	off
7	Surrogate	off
8	PredictorSelection	curvature

Table 4. Naive Bayes Parameter

No	Naïve Bayes Classifier	
	Parameter	Values
1	DistributionNames	normal
2	Prior	empirical
3	ClassNames	unique(y_train)

5. After the model has been built, a training process is carried out using the model that has been built with training data. After training, testing is carried out using test data, which aims to evaluate the model that has been trained.
6. Calculate the performance of the test results that have been carried out. The performance of the training results is calculated based on the accuracy, recall and precision values.

2.7. Calculation Performa

After testing the built model, using either the decision tree classifier algorithm, the random forest classifier or the naïve Bayes classifier, it is possible to calculate the performance of the model test results. In this study, the performance of the model is seen based on the accuracy of the test and the recall value as well as the precision obtained from the classification report of the model test. To calculate accuracy, recall and precision values, values from the confusion matrix can be used, namely true positive (TP), true negative (TN), false positive (FP) and false negative (FN) values. For an explanation of the values used can be seen in the table below.

Table 5. Confusion Matrix

No	Value	Information
1	TP	A true positive is a variable with a positive value that can be predicted exactly
2	TN	A true negative is a variable with a negative value that can be predicted exactly
3	FP	False positive is a negative value variable that is predicted to be positive
4	FN	False negative is a positive value variable that is predicted to be negative

Accuracy is the result of calculating the model's guess compared to the actual target, which means that accuracy can also be said to be able to see how the performance of the engine makes predictions on new data. For accuracy calculations can be seen in equation (4). Recall is a metric used to measure how the performance of the model can predict the positive class well. Recall calculation can be seen in equation (5). Precision is a calculation that aims to see how many positive variable predictions the machine can correctly predict. For the calculation of precision, see equation (6).

$$Accuracy = sum(predict\_y == actual\_y) / len(actual\_y) \quad (4)$$

$$Recall(i) = tp / (tp + fn) \quad (5)$$

$$Precision(i) = tp / (tp + fp) \quad (6)$$

3. RESULTS AND DISCUSSION

This research will use MATLAB R2022a to implement a sleep disorder prediction program based on lifestyle and work. For the process carried out, the first is to input data that will be used for the training process and model testing. The data used in this study is with the .csv extension, which amounts to 375 data and with 12 columns as predictors and 1 column as targets. After the data is input, a conversion process is carried out between records that have categorical data types to be numerical. This aims to facilitate data scaling later. In the process after data conversion, all data that are predictors are normalized so that the data has a value range between 0 and 1. The purpose of scaling this data is to simplify and speed up the time to be able to process data and conduct training or testing with data. After the data scaling process is carried out, then the model will be built according to the algorithm used. The algorithms used in this study are decision tree classifier, random forest classifier and naïve Bayes classifier. After the model is built, the model is trained using the training data. Then the model that has been trained will be tested using the test data. Testing aims to evaluate the results of the model that has been trained. Test evaluation results can be calculated in the form of accuracy and classification reports. In this study, the results of the evaluation that will be used as a comparison are the values of accuracy, recall and precision. So that later it can be seen which model is better for predicting sleep disturbances based on the dataset provided.

3.1. Decision Tree Prediction

In the first test, a decision tree classifier algorithm will be used to be able to predict sleep disturbances. The results obtained from testing the data using the decision tree classifier algorithm are the test accuracy of 89,431%. The results obtained are quite good and can show that the model can predict sleep disturbances well based on the dataset provided.

Classification Report:

	Precision	Recall
none	0.94	0.96
sleep apnea	0.75	0.86
insomnia	0.89	0.71

Figure 2. Precision, and Recall Using Decision Tree

It can be seen in Figure 2, the precision and recall values of each existing class. This value is obtained from the results of testing the model with the decision tree classifier algorithm. The best precision is obtained in the none class with a value of 94%, which means that the model can get many correct guesses in the none class. Meanwhile, the best recall was obtained in the none class with a value of 96%. This shows that the model can perform well in predicting the none class. Meanwhile, the worst precision value was obtained in the sleep apnea class, which was 75%. This shows that there are still many predictions from the model that are not quite right for sleep apnea class data. Then for the worst recall value obtained in the insomnia class of 71%, it shows that the model's performance is still lacking in making predictions in the insomnia class.

3.2. Random Forest Prediction

In the second test, the random forest classifier algorithm will be used to make predictions. The results obtained from this second test are the accuracy of the test obtained by 90.244%. The results obtained are good because the accuracy increases from previous tests and can show that the model can predict sleep disturbances well based on the dataset provided.

Classification Report:

	Precision	Recall
none	0.94	0.96
sleep apnea	0.76	0.90
insomnia	0.94	0.71

Figure 3. Precision, and Recall Using Random Forest

Based on Figure 3, the precision and recall values of each existing class. This value was obtained from the results of testing the model with the Precision random forest classifier algorithm, which was best obtained in the none and insomnia classes with the same value of 94%, which means that the model can get many correct guesses when working in that class. Meanwhile, the best recall was obtained in the none class with a value of 96%. This shows that the model has good performance in predicting the none class. Meanwhile, the worst precision value was obtained in the sleep apnea class, which was 76%. This shows that more data is still incorrectly guessed by the model when predicting sleep apnea classes. Then for the worst recall value obtained in the insomnia class of 71%, it shows that the model's performance is still lacking in making predictions in the insomnia class.

3.3. Naïve Bayes Prediction

In the third test, the naive Bayes classifier algorithm will be used to make predictions. The results obtained from this third test are the test accuracy of 86,992%. The results obtained were quite good but decreased from the accuracy of the previous test. However, this still shows that the model can predict sleep disturbances well based on the given dataset.

```
Classification Report:
                        Precision   Recall
none                       0.94      0.92
sleep apnea                0.75      0.86
insomnia                   0.77      0.71
```

Figure 4. Precision, and Recall Using Naïve Bayes

Based on Figure 4, the precision and recall values of each existing class. This value is obtained from the results of model testing with the naïve Bayes classifier algorithm. The best precision is obtained in the none class with a value of 94%, which means that the model can get many correct guesses when making predictions in the none class. Meanwhile, the best recall was obtained in the none class with a value of 92%. This shows that the model has good performance in predicting the none class. Meanwhile, the worst precision value was obtained in the sleep apnea class, which was 75%. This shows that there is still a lot of data that is incorrectly guessed by the model when making predictions in the sleep apnea class. Then for the worst recall value obtained in the insomnia class of 71%, it shows that the model's performance is still lacking in making predictions in the insomnia class.

4. CONCLUSION

After testing and analysis related to the method used, namely the decision tree classifier, random forest classifier and naïve Bayes classifier to predict sleep disturbances from the Sleep Health and Lifesytle Dataset, the results show that the method used is effective enough to be able to predict sleep disturbances by get an average accuracy of all models built that is equal to 88.889%. The results obtained from each test carried out are predictions using the decision tree classifier algorithm to obtain a testing accuracy of 89,431%, predictions using the random fores classifier algorithm to obtain a testing accuracy of 90,244% and predictions using the Naïve Bayes classifier algorithm to obtain a testing accuracy of 86,992%. . From these results, the best accuracy is obtained when using the random forest classifier algorithm, which is equal to 90.244%. So it can be concluded that the random forest classifier algorithm can effectively predict sleep disturbances compared to the random forest classifier and naïve Bayes classifier algorithms.

For further research, it is hoped that it can add to the methods used for data classification such as using LDA (Linear Discriminant Analysis), SVM (Support Vector Machine), or other methods that can be used for data classification. And it is also expected to be able to add or edit the parameters used to build models, either decision tree classifier, random forest classifier or naïve Bayes classifier.

REFERENCES

[1] Engelen, L. (2020). Does active design influence activity, sitting, wellbeing and productivity in the workplace? A systematic review. In *International Journal of Environmental Research and Public Health* (Vol. 17, Issue 24, pp. 1–15). MDPI AG. https://doi.org/10.3390/ijerph17249228

[2] Lukwa, A. T., Mawoyo, R., Zablon, K. N., Siya, A., & Alaba, O. (2019). Effect of malaria on productivity in a workplace: The case of a banana plantation in Zimbabwe. *Malaria Journal*, *18*(1). https://doi.org/10.1186/s12936-019-3021-6

[3] Niati, D. R., Siregar, Z. M. E., & Prayoga, Y. (2021). The Effect of Training on Work Performance and Career Development: The Role of Motivation as Intervening Variable. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, *4*(2), 2385–2393. https://doi.org/10.33258/birci.v4i2.1940

[4] Yusuf Iis, E., Wahyuddin, W., Thoyib, A., Nur Ilham, R., & Sinta, I. (2022). THE EFFECT OF CAREER DEVELOPMENT AND WORK ENVIRONMENT ON EMPLOYEE PERFORMANCE WITH WORK MOTIVATION AS INTERVENING VARIABLE AT THE OFFICE OF AGRICULTURE AND LIVESTOCK IN ACEH. *International Journal of Economic, Business, Accounting, Agriculture Management and Sharia Administration (IJEBAS)*, *2*(2), 227–236. https://doi.org/10.54443/ijebas.v2i2.191

[5] Asaloei, S. I., Wolomasi, A. K., & Werang, B. R. (2020). Work-related stress and performance among primary school teachers. *International Journal of Evaluation and Research in Education*, *9*(2), 352–358. https://doi.org/10.11591/ijere.v9i2.20335

[6] Merga, H., & Fufa, T. (2019). Impacts of working environment and benefits packages on the health professionals' job satisfaction in selected public health facilities in eastern Ethiopia: Using principal component analysis. *BMC Health Services Research*, *19*(1). https://doi.org/10.1186/s12913-019-4317-5

[7] Pisaniello, M. S., Asahina, A. T., Bacchi, S., Wagner, M., Perry, S. W., Wong, M. L., & Licinio, J. (2019). Effect of medical student debt on mental health, academic performance and specialty choice: A systematic review. In *BMJ Open* (Vol. 9, Issue 7). BMJ Publishing Group. https://doi.org/10.1136/bmjopen-2019-029980

[8] Nowrouzi-Kia, B., Sithamparanathan, G., Nadesar, N., Gohar, B., & Ott, M. (2022). Factors associated with work performance and mental health of healthcare workers during pandemics: a systematic review and meta-analysis. In *Journal of Public Health (United Kingdom)* (Vol. 44, Issue 4, pp. 731–739). Oxford University Press. https://doi.org/10.1093/pubmed/fdab173

[9] Magnavita, N. (2022). Headache in the Workplace: Analysis of Factors Influencing Headaches in Terms of Productivity and Health. *International Journal of Environmental Research and Public Health*, *19*(6). https://doi.org/10.3390/ijerph19063712

[10] Mujan, I., Anđelković, A. S., Munćan, V., Kljajić, M., & Ružić, D. (2019). Influence of indoor environmental quality on human health and productivity - A review. *Journal of Cleaner Production*, *217*, 646–657. https://doi.org/10.1016/j.jclepro.2019.01.307

[11] Zhao, M., Tuo, H., Wang, S., & Zhao, L. (2020). The Effects of Dietary Nutrition on Sleep and Sleep Disorders. In *Mediators of Inflammation* (Vol. 2020). Hindawi Limited. https://doi.org/10.1155/2020/3142874

[12] Kansagra, S. (2020). Sleep disorders in adolescents. *Pediatrics*, *145*(2). https://doi.org/10.1542/PEDS.2019-2056I

[13] Streatfeild, J., Smith, J., Mansfield, D., Pezzullo, L., & Hillman, D. (2021). The social and economic cost of sleep disorders. *Sleep*, *44*(11). https://doi.org/10.1093/sleep/zsab132

[14] Huyett, P., & Bhattacharyya, N. (2021). Incremental health care utilization and expenditures for sleep disorders in the United States. *Journal of Clinical Sleep Medicine*, *17*(10), 1981–1986. https://doi.org/10.5664/jcsm.9392

[15] Simonetti, V., Durante, A., Ambrosca, R., Arcadi, P., Graziano, G., Pucciarelli, G., Simeone, S., Vellone, E., Alvaro, R., & Cicolini, G. (2021). Anxiety, sleep disorders and self-efficacy among nurses during COVID-19 pandemic: A large cross-sectional study. *Journal of Clinical Nursing*, *30*(9–10), 1360–1371. https://doi.org/10.1111/jocn.15685

[16] Wang, X., Cheng, S., & Xu, H. (2019). Systematic review and meta-analysis of the relationship between sleep disorders and suicidal behaviour in patients with depression. In *BMC Psychiatry* (Vol. 19, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s12888-019-2302-5

[17] K. Pavlova, M., & Latreille, V. (2019). Sleep Disorders. In *American Journal of Medicine* (Vol. 132, Issue 3, pp. 292–299). Elsevier Inc. https://doi.org/10.1016/j.amjmed.2018.09.021

[18] Merrill, R. M. (2022). Mental Health Conditions According to Stress and Sleep Disorders. *International Journal of Environmental Research and Public Health*, *19*(13). https://doi.org/10.3390/ijerph19137957

[19] S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[20] Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, *188*(12), 2222–2239. https://doi.org/10.1093/aje/kwz189

[21] Appadoo, K., Soonnoo, M. B., & Mungloo-Dilmohamud, Z. (2020, December 16). Job Recommendation System, Machine Learning, Regression, Classification, Natural Language Processing. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2020*. https://doi.org/10.1109/CSDE50874.2020.9411584

[22] S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[23] Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. In *SN Computer Science* (Vol. 2, Issue 3). Springer. https://doi.org/10.1007/s42979-021-00592-x

[24] Kumar, D. (2020). Biographical notes: Priyanka received her Bachelor of Technology in Computer Science and Engineering (CSE) and Master of Technology in CSE from GJUS&T. In *Int. J. Information and Decision Sciences* (Vol. 12, Issue 3). https://doi.org/10.1504/IJIDS.2020.108141

[25] Yadav, D. C., & Pal, S. (2020). Prediction of thyroid disease using decision tree ensemble method. *Human-Intelligent Systems Integration*, *2*(1–4), 89–95. https://doi.org/10.1007/s42454-020-00006-y

[26] Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2020). BehavDT: A Behavioral Decision Tree Learning to Build User-Centric Context-Aware Predictive Model. *Mobile Networks and Applications*, *25*(3), 1151–1161. https://doi.org/10.1007/s11036-019-01443-z

[27] Lan, T., Hu, H., Jiang, C., Yang, G., & Zhao, Z. (2020). A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification. *Advances in Space Research*, *65*(8), 2052–2061. https://doi.org/10.1016/j.asr.2020.01.036

[28] Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. In *Expert Systems with Applications* (Vol. 134, pp. 93–101). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2019.05.028

[29] Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *Stata Journal*, *20*(1), 3–29. https://doi.org/10.1177/1536867X20909688

[30] Pal, M., & Parija, S. (2021). Prediction of Heart Diseases using Random Forest. *Journal of Physics: Conference Series*, *1817*(1). https://doi.org/10.1088/1742-6596/1817/1/012009

[31] Chen, S., Webb, G. I., Liu, L., & Ma, X. (2020). A novel selective naïve Bayes algorithm. *Knowledge-Based Systems*, *192*. https://doi.org/10.1016/j.knosys.2019.105361

[32] A. M. Rahat, A. Kahir and A. K. M. Masum, "Comparison of Naive Bayes and SVM Algorithm based on Sentiment Analysis Using Review Dataset," 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), Moradabad, India, 2019, pp. 266-270, doi: 10.1109/SMART46866.2019.9117512.

[33] G. Singh, B. Kumar, L. Gaur and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, UK, 2019, pp. 593-596, doi: 10.1109/ICACTM.2019.8776800.

[34] Dimitriadis, S. I., Salis, C. I., & Liparas, D. (2021). An automatic sleep disorder detection based on EEG cross-frequency coupling and random forest model. *Journal of Neural Engineering*, *18*(4). https://doi.org/10.1088/1741-2552/abf773

[35] Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. *ACM International Conference Proceeding Series*, 49–55. https://doi.org/10.1145/3377713.3377722