|  |  |
| --- | --- |
| ***E:\IIETA logo.jpg*** | **Mathematical Modelling of Engineering Problems**  Vol., No., Month, Year, pp. \*\*-\*\*  Journal homepage: http://iieta.org/journals/mmep |

|  |  |  |  |
| --- | --- | --- | --- |
| **Classification of Tuberculosis Disease Based on The Anatomical Location of Lung Disease Using LSTM With K-NN as a Solution for Missing Values** | | |  |
| Eka Mala Sari Rochman1,2, Miswanto1\*, Herry Suprajitno1, Aeri Rachmad2, Mula’ab2 , Iwan Santosa3 | | |  |
| 1 Department of Mathematics, Faculty of Sciences and Technology, Airlangga University, Surabaya, Indonesia  2Departemen of Informatics, Faculty of Engineering, University of Trunojoyo Madura, Bangkalan, Indonesia  3Information Management Department, School of Management, National Taiwan University of Science and Technology, Taiwan  Corresponding Author Email: [miswanto@fst.unair.ac.id](mailto:miswanto@fst.unair.ac.id) | | | |
| https://doi.org/10.18280/mmep.xxxxxx |  | **ABSTRACT** | |
|  |  |  | |
| **Received:**  **Accepted:** |  | One of the common lung diseases experienced by the community is tuberculosis, even this tuberculosis has increased the number of deaths of as many as 98,000 people in Indonesia. Tuberculosis is divided into 2 based on its anatomical location, namely lung for tuberculosis detected in the lung parenchyma tissue and extrapulmonary for tuberculosis detected in organs other than the lungs. However, to detect the anatomical location of tuberculosis, several parameters from laboratory tests and manual diagnoses are needed which take a long time, besides that incomplete test data will certainly affect the diagnostic results. Therefore, there is a need for a fast and accurate classification system for the anatomical location of tuberculosis to help medical personnel determine the anatomical location of tuberculosis. 1This study proposes using the KNN method in handling incomplete data or missing values, the SMOTE method to balance data. As for the classification process of pulmonary and extrapulmonary anatomy, this study uses the LSTM method in which its performance will be tested with several different models, namely Naïve Bayes, SVM, LSTM, and Backpropagation. The four of classification methods proposed in this study, all four have a very good level of accuracy. 3However, the LSTM method is superior to the others based on its accuracy, namely getting 100% compared to Naive Bayes 99.49%, SVM 99.36% and backpropagation 99.78%. These results were obtained by applying the imputation and class balancing stages and using the optimal features of LSTM namely activation function *tanh*, learning rate 0.01, lstm units 100 and optimizer adam. | |
| ***Keywords:***  *tuberculosis, missing value, KNN, classification, Naive Bayes, LSTM, Backpropagation, SVM.* |  |

# Introduction

Tuberculosis (TB) is an infectious disease caused by the bacterium Mycobacterium tuberculosis and is the highest cause of death that occurs at the age of 15-50 [1] and [2]. Air pollution from the growth of industry, cars, and others has reached a very high level, residents cannot be separated from attacks of various types of respiratory diseases, one of which is a very dangerous lung disease, Tuberculosis. Based on Indonesian TB data in 2020, cases of death due to tuberculosis increased to more than 98,000 people, in Indonesia alone this disease is one of the 10 diseases that cause death.

From the data from the World Health Organization (2019), Indonesia is a country that ranks third in the number of cases of TB disease with a total of 842,000 or 46 percent of the total cases. In Indonesia, East Java province is the second highest province in TB cases with a total of 57,014 [3]. The increase in TB cases from year to year in various regions is due to the lack of socialization and providing information about the dangers of the disease as well as how to overcome and prevent TB. In addition, TB cases in Indonesia are increasing because people always underestimate their health and do not complete TB treatment. Therefore, it is necessary to have a classification of people with TB disease to determine the risk of developing TB disease based on the symptoms experienced, this classification is also useful for equating the existing data on TB medical personnel.

Doing the classification required complete data for accurate results. However, in reality, there are often conditions where there is missing information or data [4]. The condition of the missing data or information is known as the missing value. Several factors cause data to be lost, namely errors on the part of the party collecting the data, limited tools or errors in the program used in data collection, and so on [5][6]. The condition of data loss can be overcome by imputation techniques using existing methods. In this study, the method used to perform the imputation technique is KNN (K-Nearest Neighbors).

In society, many people do not care about the symptoms they experience, even though the faster tuberculosis is treated the less its spread. To diagnose this disease, doctors or hospitals usually use several methods, namely by knowing the age and sex, HIV status, diabetes mellitus status, chest X-ray, Molecular Rapid Test (TCM) of the patient. In the data on patients diagnosed with TB disease for the anatomical location it is indicated by six variables in the classification whether including pulmonary or extrapulmonary, but not all of these data are filled in. Pulmonary TB is an infectious disease, which attacks parts of the lung parenchyma organs. Pulmonary TB can be transmitted through the air, when someone with active TB coughs, sneezes or talks [7]. Extrapulmonary TB is tuberculosis that attacks other organs outside the lungs, based on microbiological examination in a preparation Mycobacterium tuberculosis germs can be found in organs other than the lungs [7].

Classification Data mining is a learning technique that can be used to predict the value of a set of attributes in describing and differentiating data classes or concepts whose purpose is to predict object classes whose class identifiers are unknown. In research to build an expert system in diagnosing TB disease using the Certainty Factor (CF) method [7]. Based on the results of the research conducted, the system accuracy value is 85%. in a similar study regarding the classification of TB disease using the Backpropagation model using input data including age, gender, type of laboratory sample, history (follow-up, diagnostics), HIV status [8]. This model includes 70% of the training and 30% of the test set of the entire data set in which validation scores are achieved with 94% accuracy. Users can predict their TB risk after entering data, history, and display of signs and symptoms. Research on TB using LSTM was also carried out by using input parameters selected based on data obtained from the laboratory [9].

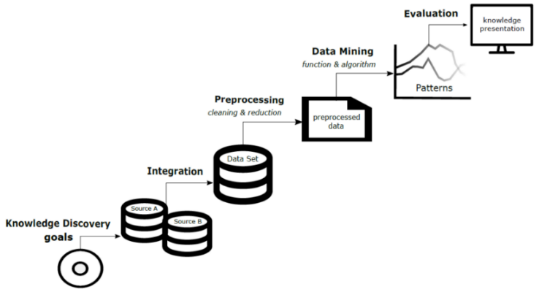
LSTM processor layer interventions, called LSTM memory blocks, detect high-level features in the input layer, analyze the signal, and pass the output to other neurons to give the right answer. The number of neurons in each memory block is determined by the number of patterns being trained, since LSTM requires one memory block per pattern processed. The output of the LSTM network provides an estimate of the likelihood of recurrent active pulmonary tuberculosis.

To minimize the death rate from TB disease and help medical experts reduce misdiagnosis, a technology that can classify and analyze the potential for heart disease in a person using health data from several previous cases is needed. Various methods can be used in the classification process, including Nave Bayes, Support Vector Machines, Artificial Neural Networks, and so on [10],[11], and [12]. Therefore, research is needed regarding the search for which method can provide the most optimal results in classifying data by comparing several existing classification methods.

Classifying tuberculosis in the medical world can help doctors make decisions in diagnosing the disease. Therefore, early diagnosis is very important to reduce the spread of TB in the wider community. Complete and quality data is highly expected during research and data collection. If the research data is incomplete, the observation results cannot be properly analyzed. Handling missing values in medical records is a difficult task. Therefore, the update in this study is to obtain high accuracy values with architectural modeling on LSTM for classification of TB data based on anatomical location by applying the KNN method used to determine Missing values.

1. Materials and Methods

Data mining is the process of extracting, identifying, and analyzing various information to find patterns in data using mathematical techniques, artificial intelligence, statistics, and machine learning. Based on the tasks that can be performed, data mining is grouped into several parts, namely decryption, prediction, estimation, classification, clustering, and association [13]. The term data mining is often referred to as knowledge discovery in database (KDD), the following stages in the data mining process are shown in the figure below.



**Figure. 1**. Stages of Data Mining

In Figure 1 above, there are several stages of the process of data mining. The first stage is the collection of the dataset that will be used, then continued with the pre-processing stage and the third stage is the data mining process using the existing dataset. Then, the last stage is an evaluation to determine the performance results of the method proposed in this study.

**2.1 Data**

The diagnosis of TB is carried out by undergoing a bacteriological examination to confirm TB disease. Bacteriological examination refers to the examination of smears from biological preparations (sputum or other specimens), examination of cultures, and identification of Mycobacterium tuberculosis. Currently, the TCM test with Xpert MTB/RIF is the only molecular test that covers all the necessary reaction elements, including all the reagents needed for the PCR (Polymerase Chain Reaction) process.

The features or criteria used for the classification process are explained respectively as follows:

1. Age, in the form of a numerical feature that contains information about the age of each sample. Based on age group, the older a person is, the higher the risk of developing TB. The older the age, the immune system will also decrease so that it is easy to catch the disease. [14].
2. The sex criterion is a category data feature that contains information on the patient sample's gender. Judging from the analysis, men are also susceptible to TB disease because it is seen from smoking habits that cause TB disease [14].
3. Chest X-ray, in the form of a category feature that contains information about the results of the chest X-ray test. Positive if TB is found and negative if not found. This is done to ensure the coexistence of pulmonary TB [15].
4. HIV status, in the form of a category feature that contains information about the HIV status of each patient. Positive if you have HIV and negative if you don't. Tuberculosis in HIV/AIDS (TB-HIV) patients is often found with a prevalence of 29-37 times more than TB without HIV [8].
5. Diabetes history, in the form of a category feature that contains information about the sample patient's history of diabetes. Diabetes is one of the most common risk factors in pulmonary TB patients. Currently, the prevalence of pulmonary TB is increasing along with the increasing prevalence of diabetes patients. The frequency of diabetes in TB patients is reported to be around 10-15% and the prevalence of this infectious disease is 2-5 times higher in diabetic patients compared to non-diabetics [8].
6. Whereas the TCM result feature is in the form of a category feature that contains information on the results of the TCM test for each individual sample with a value; resistant rif, sensitive rif, and negative. Rif-resistant TB patients are patients who experience resistance to the antibiotic drug rifampicin. Sensitive rif is a test that has confirmed TB but is not resistant to antibiotics. Negative, namely the result that no TB bacteria are found in the lungs.

This research was conducted to be able to classify the types of TB disease based on the anatomical location suffered by individuals by using features in the dataset used to determine the type of TB suffered. Extrapulmonary TB cases are almost always non-infectious, unless the patient also has pulmonary TB because 50-60% of HIV-positive people infected with TB will develop pulmonary TB disease [8]. Meanwhile, to find out if an individual has extrapulmonary TB, it is necessary to carry out a bacteriological test by carrying out a direct microscopic examination, TCM TB [8]. The results of chest x-rays are also needed to be able to classify an individual as a patient with pulmonary TB or extrapulmonary TB.

**2.2. Data Pre-processing**

Data pre-processing starts with data transformation to convert categorical data into numeric so that later the data can be processed. The next process is data imputation to overcome the missing value condition. After the imputation of data, the next step is to normalize the data so that the data used has a value range of 0 to 1.

**2.3. Data Transformation**

Data transformation is a change in the scale of data to other forms so that the data has a distribution as expected. Each data will perform the same mathematical operation on the original data [16]. Changes to all data are intended to keep differences between data relatively constant. If the data exceeds one variable, all variables will be transformed so that the relationship between the data does not change. The data transformation carried out in this study is to change the value in the categorical dataset to be converted to numeric data.

**2.4. Imputation Missing Value**

The missing value is a condition where there is value or information on a subject that is missing. The causes of this missing value vary, it can be due to an error in data collection or the data is not readable in the system so the value is considered missing [6]. There are several methods to overcome the missing value condition, one of which is the imputation process using the KNN (K-Nearest Neighbor) method.

KNN is an imputation method that is based on data by finding the shortest distance to the object data [17]. By implementing this method it aims to determine the value of a new object based on attribute values and training samples besides that because this method classifies based on the closest distance which is implemented in a simple formulation

The closest distance is calculated using the Euclidean formula with the following equation:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Information:

*d(a, b) :* Euclidean distance

*Xi :* first data

*Yi :* second Data

*i :* Attribute *i*

*n* : Number of attributes

**2.5. Data Normalization**

Data normalization is a process to change several variables so that they have the same value range, no data is too large or too small so it will be easier to perform statistical analysis. The method used to normalize tuberculosis data is the Min-Max Normalization method [18]. The equation below is a formula for the min-max normalization method.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Information:

*z :* normalization result*,*

*x :* value *x* (original),

*min(x) :* the minimum value for the variable *x*,

*max(x)* : the maximum value for the variable *x*.

**2.6. Data Mining Process**

In this study, the data mining process involves the process of balancing data and sharing training data and test data. Data balancing is a process to equalize the amount of data in each class to improve the accuracy of the system during the learning process [19]. In this stage, learning the classification model was carried out, namely grouping TB data into pulmonary and extrapulmonary class categories. The number of TB patient data is 985, the extrapulmonary class is 271, while the pulmonary class has a larger number with a total of 714 data. There is no specific size for a dataset that is said to be unbalanced data, but if you look at the data used in this study there are large and very striking differences in the amount of data for each class. So that it can be said for the data used in this study is data with an unbalanced number of classes. On this occasion, the oversampling technique was used to balance the data using the SMOTE (Synthetic Minority Oversampling Technique) method. The way the SMOTE method works is by replicating the data randomly by choosing *k* closest neighbors as the determinant, then the minority data set will be balanced with the majority data.

From this explanation, it can be formulated into the following formula [20].

|  |  |  |
| --- | --- | --- |
|  | | (3) |
|  |  | (4) |

Information:

𝑋𝑖 = vector of features in the minority class

*𝑋knn* = k-nearest neighbors for *𝑋𝑖*

𝛿 = random number between 0 to 1

After balancing the data, the next step is to divide the data set into k to n partitions.Splitting this data is known as K-Fold Cross Validation and is a popular method of solving statistical data where the data is divided into two subsets, namely training data for the learning process and test data for validation or assessment used to assess performance models, methods, or algorithms [19]. K-Fold cross validation can be selected based on dataset size. Usually K-Fold is used to reduce computation time and also maintain the accuracy of the estimation. The size of the shared data depends on the specified K value. In this study, a k-fold value of 10 is used. In each iteration, the original dataset provided is divided randomly by cross-validation into training sets, which are used to train the machine learning algorithm, and tests are determined to evaluate its performance [21] and [22].

In the data mining process, various methods can be used to build a classification model to be applied to a TB disease diagnosis system, including Naïve Bayes, Support Vector Machines, or the application of branches of deep learning methods such as LSTM and Backpropagation.

**2.7. Naive Bayes**

A technique derived from Bayes' theorem for classifying data is the Naive Bayes Classifier. This method can predict future data based on past data or existing data by calculating odds from test data with data stored in the training process [23] and [24]. Its main feature is a very strong (naive) assumption of independence, regardless of any circumstances or events [10]. Bayes theorem is the basis of this method. So, before we get into the explanation of Nave Bayes, let's first explain Bayes' theorem. If there are two separate events in Bayes' theorem (e.g. A and B), then Bayes' theorem can be written using the following formula:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Information:

*A :* Classless Data

*B :* Data Hypothesis

*P(B|A) :* Probability of hypothesis B against condition A

*P(B) :* Probability hypothesis B

*P(A|B) :* Probability of A based on condition of B

*P(A)* : Probability A

The application of the law of total probability Bayes theorem can be developed into the following equation:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

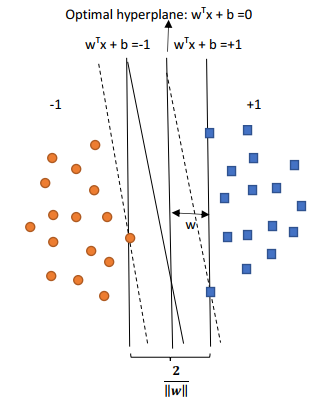
Here, the variables F1 through Fn represent the characteristics of the statements required to complete the classification, while the variable C is a class. Thus, the above formula shows that the probability of including a sample with certain characteristics in class C is equal to the probability of class C occurring (before the sample was included, or what is called a prior) multiplied by the probability of occurrence of the global one Sample characteristics (evidence). Therefore, the equation x above can be formulated using the following equation:

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

The value of evidence is always fixed for each class in one sample. The value of the Posterior will be compared with the Posterior values ​​of other classes to determine to what class a sample will be classified.

**2.8. Support Vector Machine**

SVM (Support Vector Machine) is one method for predicting, both regression and classification cases [18]. The basic principle of SVM uses a linear classifier. A linear classifier is a classification case that is separated linearly, but SVM has evolved to work in non-linear cases by using the kernel concept in high-dimensional workspaces. Hyperplanes are used for high-dimensional spaces that aim to maximize the distance (margin) between data classes [25]. The best separator function (hyperplane) must be found among the unlimited number of other hyperplanes in order to find the optimal separator function (classifier) and separate two different classes. The best hyperplane is the one that is right between two sets of objects of two classes.A hyperplane can be said to be the best if it is located right between two sets of objects from two classes. Figure 2 shows how SVM maximizes the distance between two different sets of classes (margins) by determining the best hyperplane.



Figure*.* 2*.* SVM finds the best hyperplane to separate class-1 from class+1

Figure 2 shows how the hyperplane is used as a separator of two different classes in the classification process to achieve good results, by measuring the hyperplane margin and determining the maximum point. The distance between the hyperplane and the closest pattern of each class is called the margin [26]. In the figure above, the dotted line represents the pattern closest to the hyperplane, which is called the support vector. A hyperplane will be used as the largest class separator by the SVM algorithm. The dividing line between the two classes forms an equation:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

From these equations, it can be seen that the data that falls into the first class category is data that has a larger equation value, while the data that enters the second class category is data that has a smaller equation value. The value of the margin or the value of the distance between the boundary planes based on the formula for the distance of the line to the center point can be represented by the equation below:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

Then this margin value will be maximized by minimizing the value of ||w|| as the denominator. This can be formulated as Quadratic Programming. If the two boundary planes are represented as equation 9, the search for the dividing line with the largest boundary can be formulated as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (10) |
|  |  | (11) |

This problem can be solved by existing computational techniques, such as the Lagrange multiplier expressed in Equation 12.

|  |  |  |
| --- | --- | --- |
|  | +..+b)-1)) | (12) |
|  |  | (13) |
|  | where | (14) |

Information:

*w* : weight vector

*x* : attribute input value

*b* : bias

: support vector

: data class

If the data cannot be separated linearly, it can be said that the classes in the input space cannot be separated perfectly. This causes the constraint in equation 9 is not met, so the optimization cannot be done. To overcome these problems, SVM is formulated using the soft margin technique, the following is the mathematical equation:

|  |  |  |
| --- | --- | --- |
|  |  | (15) |

Thus, equation 15 is converted into an equation as below.

|  |  |  |
| --- | --- | --- |
|  |  | (16) |

Information:

*w* = weight vector

*x* = attribute input value

*b* = bias

= error value

*c* = constant value

In general, the problems that exist rarely have linear separable data and most of them are non-linear. Solving the non-linear SVM problem is to map the x data by a function (x) to a vector space with a higher dimension. In the new vector space, a hyperplane that separates the two classes can be constructed. Then do the dot product calculation from the data that has been transformed in a higher dimensional space, namely (xi). (xi). However, in general, the transformation is difficult to understand, therefore the dot product value search can be replaced with a kernel function as in the following equation.

|  |  |  |
| --- | --- | --- |
|  |  | (17) |

The following are kernel functions that are commonly used to classify non-linear data(Liu & Xu, 2013):

1. Linear Kernel
2. Kernel Polynomial

*K() =*

1. Kernel RBF or Radial Based Function
2. Sigmoid Kernel

**2.9. Backpropagation**

Backpropagation is a supervised learning algorithm that uses multiple layers to change the weights connected to the neurons in the hidden layer [27]. The Backpropagation algorithm minimizes errors in the output generated by the network by changing the value of its weights in the backward direction using the output error. To get the error output, the forward step must be done first. In the Backpropagation algorithm, the training process is carried out in two phases, namely the forward propagation and backward propagation stages [27]. The following is the algorithm flow of backpropagation in each phase.

Phase 1: Forward Propagation

1. Initialize the weight with a small random value, maximum Epoch value, error value, and learning rate
2. Perform the steps below when the epoch value is smaller than the maximum value.
3. Each input unit receives the signal and forwards it to the hidden unit
4. Calculate all outputs on hidden units (j = 1,2, … ,p)

|  |  |  |
| --- | --- | --- |
|  |  | (24) |
|  |  | (25) |

1. Calculate all network outputs in units of (k = 1,2, … , m)

|  |  |  |
| --- | --- | --- |
|  |  | (26) |
|  |  | (27) |

Phase 2: Backward Propagation

1. Calculate the output unit factor based on the error value for each output unit (k = 1,2,..,m)

|  |  |  |
| --- | --- | --- |
|  |  | (28) |

Calculate the weight correction using the formula below.

|  |  |  |
| --- | --- | --- |
|  |  | (29) |

1. Calculate the weight correction using the formula below (k = 1,2, .., m)

|  |  |  |
| --- | --- | --- |
|  |  | (30) |
|  |  | (31) |

Calculate the weight correction using the formula below.

|  |  |  |
| --- | --- | --- |
|  |  | (32) |

1. Add up the change in weight of the line leading to the output unit.

|  |  |  |
| --- | --- | --- |
|  |  | (33) |

1. Add up the change in weight of the line leading to the hidden unit.

|  |  |  |
| --- | --- | --- |
|  |  | (34) |

Information:

*𝑤0𝑘* : bias weight on output unit 𝑦𝑘

𝑣0𝑗 : bias weight on hidden unit 𝑧𝑗

𝑣𝑖𝑗 : line weight from unit to hidden unit 𝑧𝑗

𝑤𝑗𝑘 : line weight from 𝑧𝑗 to output unit 𝑦𝑘

𝑓𝑘 : activation function on the hidden unit to output

𝑓𝑗  : activation function on input unit to hidden

𝑧𝑗 : j-th hidden unit

𝑦𝑘 : k-th output unit

𝑖 : 1, … , 𝑛

𝑗 : 1, … , 𝑝

𝑘 : 1, … , 𝑙

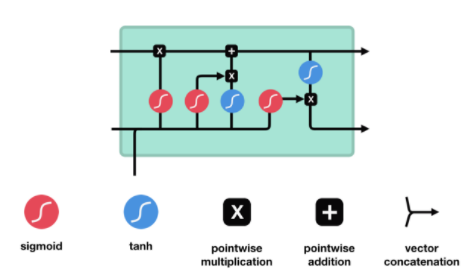
𝑛 : number of input units

𝑝 : number of hidden units

𝑙 : number of output units

**2.10. Long Term Short Memory (LSTM)**

Long Term Short Memory (LSTM) is a development of a neural network that is used to model data in a certain time series. This method can also overcome long-term dependence on the input. Cells in the LSTM store a value, either for a long or a short period of time. The memory block in the LSTM determines which value will be selected as the relevant output for the given input [12]. This LSTM method is proposed as a solution to overcome the vanishing gradient problem on the RNN when processing long sequential data.



Figure*.* 3*.* LSTM Architecture

Fig. 3. Illustrates the architecture of the LSTM. There are four activation function processes at each input to neurons, hereinafter referred to as gate units. The gate units are forgotten gates, input gates, cell gates, and output gates [9]:

1. Forgot Gates

The input data will be processed and it will be chosen which one will be used or discarded in the memory cells. The sigmoid activation function is used in this gate, which results in a value between 0 and 1 [16]. If the output is 1 then all data will be stored and vice versa if the output is 0 then all data will be discarded. With the following formula [28]:

|  |  |  |
| --- | --- | --- |
|  |  | (18) |

Information:

*ft* : forget gate

σ : sigmoid activation function

*Wf* : weights forget gate

*xt* : input cell

ht*−*1 : output cell previously

*bf* : bias forget gate

1. Input Gates

The 2 gates are in the input state, which must first decide which value to replace using the sigmoid function. Then, the tanh activation function will create a new value vector and store it in the memory cell. With the following formula:

|  |  |  |
| --- | --- | --- |
|  |  | (19) |
|  |  | (20) |

Information:

*it* : input gate

*wi* : weights input

*bi* : bias input gate

*Ct* : candidate

*tanh* : tanh activation function

*Wc* : weights candidate

1. Cell Gates

The cell gates replace the value in the previous memory cell with the new memory cell value. This value is obtained by combining the values ​​contained in the forget gate and the input gate. With the following formula:

|  |  |  |
| --- | --- | --- |
|  |  | (21) |

Information:

ct : cell state

ct−1 : cell state previously

1. Output Gates

Two gates must be implemented at the output gates. First, the sigmoid activation function is used to decide which part of the memory cell value to output. Next, the value is inserted into the memory cell using the activation function tanh. Finally, the two gates are multiplied to give the value to be output. With the following formula:

|  |  |  |
| --- | --- | --- |
|  |  | (22) |
|  |  | (23) |

Information:

*Ot* : output gate

*Wo* : weights output gate

*b*o : bias output gate

*ht* : the hidden layer

**2.11. Evaluation**

In modeling the classification process, it is necessary to evaluate the performance of the system to measure how well the method used is [29]. The method commonly used in evaluating the system is the confusion matrix. A confusion matrix is ​​one method of evaluating the calculation of the level of accuracy, precision results, recall, and f-measure of the algorithm used in research and is measured from the results of testing data that have been predicted [30].

Accuracy is the value of the effectiveness of the overall results of the classification process, following the formula:

1. Accuracy is the value of the effectiveness of the overall results of the classification process, following the formula:

|  |  |  |
| --- | --- | --- |
|  |  | (35) |

1. Precision is the result of the percentage of positive classification data labels, following the formula:

|  |  |  |
| --- | --- | --- |
|  |  | (36) |

1. The recall is the result of the effectiveness of the classification process for the identification of positive labels, the following formula is:

|  |  |  |
| --- | --- | --- |
|  |  | (37) |

1. F-measure is the result of mean recall and precision, where the range of f-measure itself is 0-1.

|  |  |  |
| --- | --- | --- |
|  |  | (38) |

Information:

*TP*: the number of data correctly predicted positive,

*TN*: the number of data with the original class is positive but the prediction result is negative,

*FN*: the number of data correctly predicted negative,

*FP*: the number of data with the original class is negative but the prediction result is positive.

1. RESULT AND DISCUSSION

3.1. Data Collection

The data used in this classification process is Bangkalan residents' tuberculosis data which consists of 985 records and 6 attributes, namely age, gender, chest X-ray, HIV status, history of diabetes, and TCM results. In the data used, there is a missing value which will be overcome by implementing the imputation process using the KNN method [4] and [31]. The mission value in the dataset can be seen in Table 1 below.

Table 1 Number of Blank Data on Each Attribute

|  |  |
| --- | --- |
| **Attribute** | **Number of Blank Data** |
| Age | 0 |
| Sex | 0 |
| Thorax X-Ray | 28 |
| HIV | 261 |
| Diabetes | 7 |
| TCM | 439 |

**3.2. Analysis**

Figure 4. describes the IPO diagram as follows:

1. Data Input Process

In this process, the input data used for the classification process is data on patients with TB disease with 6 features.

1. Data Pre-processing

At this stage, the SMOTE process is carried out to balance the data so that the amount of data in each class is balanced. Balancing the data is done by oversampling the minority class, namely the data with the negative class so that the number is balanced with the positive class. The amount of data in the pulmonary class is 596, while the extra lung is 389. So a method is needed to balance the data [19]. This method works by replicating the data randomly by choosing KNN as the determinant, then the minority data set will be balanced with the majority data. The value of k used in this study is *k* = 5.

1. Data Balancing Process

At this stage, the SMOTE process is carried out to balance the data so that the amount of data in each class is balanced. Balancing the data is done by oversampling the minority class, namely the data with the negative class so that the number is balanced with the positive class. This method works by replicating the data randomly by choosing *k* closest neighbors as the determinant, then the minority data set will be balanced with the majority data. The value of k used in this study is *k* = 5.

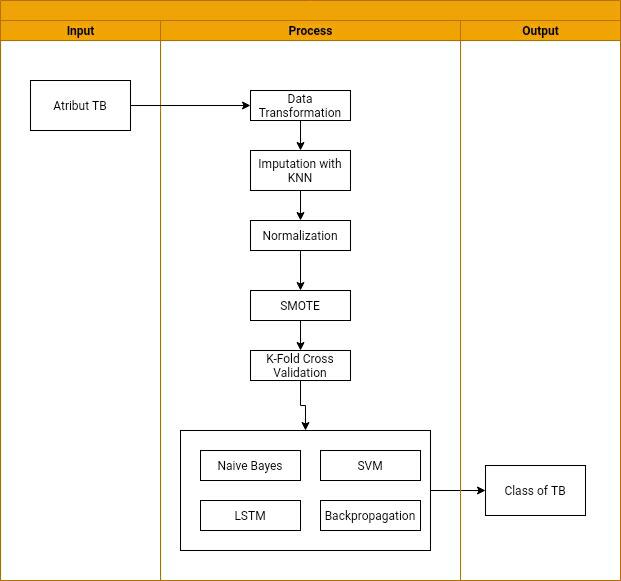
1. Data Sharing Process

The k-fold cross validation method is used to divide training data and test data by k-fold 10. Training data is used in building models with a certain amount of data, while test data is the remaining data that is not used during training to test model performance who have been trained

1. Classification Process

At this stage, the learning process is carried out to obtain a classification model using several different methods, namely Naïve Bayes, Support Vector Machine, Backpropagation, and LSTM. The different classification models will be compared to the accuracy results to obtain optimal classification results.

1. Output

****After the entire process is run, it will produce an output in the form of class predictions from TB attributes based on modelling with the method proposed in this study.

Figure*.*4*.*  Process Flowchart

**3.3. Naive Bayes**

The results obtained by applying the TB dataset using the Naïve Bayes algorithm combined with KNN for handling incomplete data or missing values, as well as using the SMOTE method and without SMOTE will then be tested using a different number of k cross-validations (k), namely k -fold which amounts to 1 to 10.

Table 2 Evaluation Results Using the Naive Bayes Method in percent

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fold** | **With SMOTE** | | | | **Without SMOTE** | | | |
|  | **Acct** | **Prcs** | **Rcl** | **F1-S** | **Prcs** | **Rcl** | **F1-S** | **Prcs** |
| **1** | 98.98 | 96.4 | 100 | 98.2 | 98.9 | 95.2 | 100 | 97.5 |
| **2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **3** | 100 | 100 | 100 | 100 | 98.9 | 97.2 | 100 | 98.6 |
| **4** | 100 | 100 | 100 | 100 | 98.9 | 96.6 | 100 | 98.3 |
| **5** | 98.98 | 97.0 | 100 | 98.5 | 100 | 100 | 100 | 100 |
| **6** | 97.95 | 93.5 | 96.6 | 100 | 100 | 100 | 100 | 100 |
| **7** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **8** | 98.97 | 99.3 | 98.3 | 98.8 | 100 | 100 | 100 | 100 |
| **9** | 98.97 | 94.7 | 100 | 97.3 | 98.9 | 95.8 | 100 | 97.8 |
| **10** | 100 | 100 | 100 | 100 | 98.9 | 97.1 | 100 | 98.5 |
| **Avg** | **99.49** | 98.2 | 100 | 99.1 | 99.5 | 98.2 | 100 | 99.0 |
| **RT** | 5.87(s) | | | | 2.80(s) | | | |

Acct : Accurate

Prcs : precission

Rcl : Recall

F1-S : F1-Score

Avg : average

RT : Running Time

Table 2 shows that the average accuracy is 99.49% with a computation time of 5.87 seconds when the SMOTE method is applied, while without SMOTE the average accuracy is also 99.49% with a computation time of 2.08 seconds.

The advantages of Naive Bayes can be used for both quantitative and qualitative data. No need to do a lot of data training. If there is a missing value, it can be ignored in the calculation, and the calculations are fast and efficient so that they are easy to understand. While the weakness of naive Bayes is assuming that each variable is independent so that it reduces accuracy because usually there is a correlation between one variable and another.

**3.4. Support Vector Machine**

At this stage, the research is carried out by making trial scenarios to draw the right conclusions after the research process is carried out. The trial scenarios carried out in this study are described in Table 3:

Table 3 Testing Scenarios

|  |  |  |
| --- | --- | --- |
| **No.** | **Trial Scenario** | **Parameter** |
| 1 | K-Fold Value | K= 1 to K=10 |
| 2 | Kinds of Kernel | Linear, Polynomial, RBF, Sigmoid |
| 3 | C value | 0,5; 1; 10; 100 |
| 4 | Gamma Scales | Scale; 1; 0,1; 0,01; 0,001; 0,0001 |

Based on the parameters in the test scenario, the best parameters are then searched for by optimizing the parameters using Grid Search. The use of grid search in the pilot scenario for parameter optimization is intended so that the parameter optimization process does not take much time. The results of the experiments were carried out using the best parameters from the results of parameter tuning using the Grid Search method. The best parameters obtained from the SVM model include:

Table 4 Best Parameters of Gird Search Results

|  |  |
| --- | --- |
| **Kernel Coefficient (*gamma*)** | **Scale)** |
| Regulation Value (C) | 0.5 |
| Kernel | RBF (R*adial Basis Function*) |

From the best parameters obtained, it will be applied to the classification process with the SMOTE method and without SMOTE and then will be tested using a different number of k cross-validations (k), namely k-fold which amounts to 1 to 10. The test results can be seen in Table 5. below this.

Table 5 Evaluation Results Using SVM in percent

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fold** | **With SMOTE** | | | | **Without SMOTE** | | | |
| **Acct** | **Prcs** | **Rcl** | **F1-S** | **Acct** | **Prcs** | **Rcl** | **F1-S** |
| **1** | 98.6 | 98.7 | 98.5 | 98.6 | 98.9 | 99.3 | 98.0 | 98.6 |
| **2** | 100 | 100 | 100 | 100 | 98.9 | 99.3 | 98.0 | 98.6 |
| **3** | 98.3 | 99.33 | 99.3 | 99.3 | 98.9 | 99.3 | 98.0 | 98.6 |
| **4** | 98.6 | 98.48 | 98.7 | 98.6 | 97.9 | 98.6 | 96.6 | 97.5 |
| **5** | 99.3 | 99.32 | 99.3 | 99.3 | 100 | 100 | 100 | 100 |
| **6** | 99.3 | 99.38 | 99.2 | 99.3 | 100 | 100 | 100 | 100 |
| **7** | 100 | 100 | 100 | 100 | 98.9 | 99.3 | 98.1 | 98.7 |
| **8** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **9** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **10** | 98.59 | 98.70 | 98.5 | 98.6 | 100 | 100 | 100 | 100 |
| **Avg** | **99.36** | 99.3 | 99.3 | 99.3 | 99.3 | 99.5 | 98.8 | 99.2 |
| **RT** | 0.18(s) | | | | 0.08(s) | | | |

Table 5 shows that the average accuracy is 99.36% with a computing time of 0.18 seconds when the SMOTE method is applied, while without SMOTE the average accuracy is 99.38% with a computation time of 0.08 seconds.

When the data are not linearly separated, the kernel RBF is a kernel function that is frequently employed in analysis. Gamma and Cost are the two parameters that make up the RBF kernel. Parameter The parameter cost, often known as C, is used as an SVM optimization to prevent misclassification in each sample in the training dataset. A low value for the Gamma parameter indicates "far" influence while a high value indicates "close" influence of a sample training dataset. Low gamma makes it reasonable to take into account locations that are far from the separator line while determining the separator line. When the gamma is large, the points should be taken into account in the computations because they are near the line.

The benefit of the SVM method is that it can function reasonably well when there is a distinct line of demarcation between classes, saves more memory because it uses training points from the decision function (support vector), and is effective when the number of dimensions is greater than the number of samples. This method's difficulty in using it to sets with a lot of dimensions is its flaw.

**3.5. Backpropagation**

The use of the Backpropagation method is done by making a test scenario to draw the right conclusions after the research process is carried out. The trial scenarios carried out in this study are described in Table 8

Table 8 Testing Scenarios on the Backpropagation Method

|  |  |  |
| --- | --- | --- |
| **No.** | **Trial Scenario** | **Parameter** |
| 1 | Fold | K=1 to K=10 |
| 2 | Number of Neurons | 60-100 |
| 3 | Optimization | Adam Optimizer, SGD |
| 4 | Activation Functions | Sigmoid, relu, tanh |
| 5 | Learning Rate | 0.001 |
| 6 | Epoch | 25, 50,100, 1000 |

The test uses the sigmoid activation function with the number of neurons being 60 and without using optimization. Experiments using k-fold 10 and a learning rate of 0.001 with many iterations (epochs) of 25. The following Table 20 is the result of tests carried out on data with SMOTE and without SMOTE.

Table 9 Test Results Using Backpropagation With Epoch = 25 percent

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fold** | **With SMOTE** | | | | **Without SMOTE** | | | | | |
| **Acct** | **Prcs** | **Rcl** | **F1-S** | **Acct** | **Prcs** | **Rcl** | | | **F1-S** |
| **1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | 100 |
| **2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | 100 |
| **3** | 100 | 100 | 100 | 100 | 98.9 | 99.33 | 98.0 | | | 98.6 |
| **4** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | 100 |
| **5** | 99.34 | 99.26 | 99.3 | 99.3 | 100 | 100 | 100 | | | 100 |
| **6** | 100 | 100 | 100 | 100 | 98.9 | 99.34 | 97.8 | | | 99.85 |
| **7** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | 100 |
| **8** | 99.31 | 99.31 | 99.3 | 99.3 | 99.3 | 99.31 | 99.3 | | | 99.30 |
| **9** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | | | 100 |
| **10** | 99.18 | 99.39 | 99.2 | 99.2 | 98.9 | 99.29 | 99.3 | | | 98.76 |
| **Avg** | **99.78** | 99.79 | 99.7 | 99.7 | 99.6 | 99.72 | | 99.4 | 99.65% | |
| **RT** | 38.61(s) | | | | 31,75 (s) | | | | | |

|  |  |
| --- | --- |
| Figure 21 Accuracy Results at Epoch=25 with SMOTE | Figure 22 Accuracy Results at Epoch=25 without SMOTE |

Table 9 shows that the average accuracy is 99.78% with a computation time of 38.61 seconds when the SMOTE method is applied, while without SMOTE the average accuracy is also 99.62% with a computation time of 31.75 seconds. The most prominent advantages of the Backpropagation algorithm are Fast, simple, and easy to program. Has no tuning parameters other than the number of inputs. Flexible because it does not require prior knowledge of the network. while the weakness of Backpropagation is that it takes a long time in the learning process.

**3.6. LSTM**

The use of the LSTM method is carried out by making a test scenario to draw the right conclusions after the research process is carried out. In this study, two kinds of tests were carried out, namely, the parameter values ​​to be used and testing for the removal of features that had the greatest missing value to see the effect of the accuracy results obtained. The test scenario for the parameters used is described in Table 6 and the test for the effect of the feature that has the largest number of missing values ​​can be seen in Table 7:

Table *5* LSTM Testing Scenarios

|  |  |  |
| --- | --- | --- |
| **No.** | **Trial Scenario** | **Parameter** |
| 1 | Fold | K=1 to K=10 |
| 2 | Number of Neurons | 60-100 |
| 3 | Optimization | Adam Optimizer, SGD, |
| 4 | Activation Functions | Sigmoid, Tanh |
| 5 | Learning Rate | 0.01, 0.001 |
| 6 | Epoch | 25 |

Table 6 Best Parameters of Gird Search Results

|  |  |
| --- | --- |
| Number of Neurons | 100 |
| Optimization | Adam Optimizer |
| Activation Functions | Tanh |
| Learning Rate | 0.01 |

Based on the parameters in the test scenario, the best parameters are then searched for by optimizing the parameters using Grid Search. The test uses the Tanh activation function with the number of neurons being 100 and using the Adam optimizer. Experiments using k-fold 10 and a learning rate of 0.001 with many iterations (epochs) of 25. The following Table 7 is the results of tests carried out on data with SMOTE and without SMOTE

Table 7 Test Results Using LSTM With 100 Neurons and Epoch = 25 percent

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Fold** | **With SMOTE** | | | | **Without SMOTE** | | | |
| **Acct** | **Prcs** | **Rcl** | **F1-S** | **Acct** | **Prcs** | **Rcl** | **F1-S** |
| **1** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **2** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **3** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **4** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **5** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **6** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **7** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **8** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **9** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **10** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **Avg** | **100** | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| **RT** | 109.08 (s) | | | | 104.90(s) | | | |

|  |  |
| --- | --- |
| Figure 13 Accuracy Results at Epoch=25 with SMOTE | Figure 14 Accuracy Results at Epoch=25 without SMOTE |

Table 7 shows that the average accuracy is 100% with a computation time of 109.08 seconds when the SMOTE method is applied, while without SMOTE the average accuracy is also 100% with a computation time of 104.9 seconds.

LSTM helps in the classification process unlike ordinary neural network processes which are all based on activation and weights, the LSTM cell state plays an important role in providing relevant classification results.

# Conclusion

Based on the analysis and discussion that has been carried out on training and testing data using 10-fold where there are 985 records and 6 attributes, namely age, gender, chest X-ray, HIV status, history of diabetes, TCM (Rapid Molecular Test) results, it is concluded that :

1. By using this research proposed model, it can overcome the problem of missing value in TB data at Syarifah Ambami Rato Ebu Bangkalan Hospital using the KNN method
2. Based on the test results that have been obtained the use of SMOTE provides good performance but does not provide a significant increase in accuracy when compared to the application of datasets without SMOTE in the entire test, this can be seen from the computation time given. With the use of SMOTE, the execution time required becomes more due to the addition of synthesis data to balance the minority data class
3. From the results of this study, it can be concluded that LSTM is very capable of being used to classify TB disease. By applying the GirdSearch method the best accuracy results from several test scenarios, an accuracy value of 100% is obtained stably in experiments using k-fold 10 and a learning rate of as well as the use of the Tanh activation function on the number of neurons 100 using and without SMOTE.
4. Based on the test results to get an accuracy value, the LSTM method is superior to other methods, namely getting 100% compared to Naive Bayes 99.49%, SVM 99.36%, and backpropagation 99.78%.
5. Based on the results of trials conducted, a classification of TB disease was obtained based on the anatomic location of the lungs and extrapulmonary using the LSTM method by improving the performance of the algorithm to obtain optimal accuracy
6. This study succeeded in developing research by creating a classification model so that later it can be applied to TB data and provide recommendations to hospitals regarding pulmonary and extrapulmonary classes and important features of TB disease

Based on the conclusions, the authors provide suggestions for further development and research, namely:

1. The next system development is expected to be able to build a classification model using another imputation method that can combine two different data types
2. The next research is expected to be able to add the amount of data and features used to build a classification model, this can certainly support the performance of the model to be even better by paying attention to the number of missing values and class balance in the dataset used.

# acknowledgment

This research would like to thank the support of Airlangga University and the University of Trunojoyo Madura

**REFERENCES**

1. Saputra, R. A., Noviyani, A., Nopsopon, T. & Pongpirul, K. (2021). Variation of tuberculosis prevalence across diagnostic approaches and geographical areas of Indonesia. PLOS ONE, 16(10): 1-12. <https://doi.org/10.1371/journal.pone.0258809>
2. Gill, C. M., Dolan, L., Piggot, L. M. & McLaughlin, A. M.. (2021). New developments in tuberculosis diagnosis and treatment. Breathe, 8(1): pp. 1-15. <https://doi.org/10.1183/20734735.0149-2021>
3. Nasution, M.K.M. Et al.. (2020). Identification of Tuberculosis (TB) Disease Based on Lung X-Rays using Extreme Learning Machine. Journal of Physics: Conference Series, 1566 (012085): 1-8. <https://doi:10.1088/1742-6596/1566/1/012085>
4. E. M. S. Rochman, Miswanto and H. Suprajitno,. (2022). Overcaming Missing Value Imputation Methods in The Classification Of Tuberculosis. Communications in Mathematical Biology and Neuroscience. 2022 (2022): 1-19. <https://doi.org/10.28919/cmbn/7538>
5. Emmanuel, T. et al.. (2021). A survey on missing data in machine learning. Journal of Big Data, 8(140), 1-37. <https://doi.org/10.1186/s40537-021-00516-9>
6. Kang, H. (2013). The prevention and handling of the missing data. Korean journal of anesthesiology, 64(5): 402-406.

<https://doi.org/10.4097/kjae.2013.64.5.402>

1. Aini,T. Et al. (2017). Sistem Pakar Pendiagnosa Penyakit Tuberkulosis. Jurnal Informatika Mulawarman. 12(1), 56-62 <http://dx.doi.org/10.30872/jim.v12i1.224>
2. Khan, M. T. Et al. (2019). Artificial neural networks for prediction of tuberculosis disease. Frontiers in microbiology, 10: 395. <https://doi.org/10.3389/fmicb.2019.00395>
3. Venkataramana, L. Et al. (2022). Classification of COVID-19 from tuberculosis and pneumonia using deep learning techniques.Medical & Biological Engineering & Computing. 60:2681–2691 <https://doi.10.1007/s11517-022-02632-x> .
4. Berrar, D. (2019). Bayes’ theorem and naive Bayes classifier. Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, 1: 403. [10.1016/B978-0-12-809633-8.20473-1](https://doi.org/10.1016/B978-0-12-809633-8.20473-1)
5. Liu, Z., & Xu, H. (2014). Kernel parameter selection for support vector machine classification. Journal of Algorithms & Computational Technology, 8(2), 163-177. <https://doi.org/10.1260/1748-3018.8.2.1>
6. Lees, T., Reece, S., Kratzert, F., Klotz, D., Gauch, M., De Bruijn, J., ... & Dadson, S. J. (2022). Hydrological concept formation inside long short-term memory (LSTM) networks. Hydrology and Earth System Sciences, 26(12), 3079-3101. <http://dx.doi.org/10.5194/hess-26-3079-2022>
7. Beniwal, S., & Arora, J. (2012). Classification and feature selection techniques in data mining. International journal of engineering research & technology (ijert), 1(6), 1-6. <http://doi.10.17577/IJERTV1IS6124>
8. Marcoa, R. Et al. (2018). Tuberculosis and gender – Factors influencing the risk of tuberculosis among men and women by age group. Pulmonology. 24 (3): 199-202. <https://doi.org/10.1016/j.pulmoe.2018.03.004>
9. Majdawati, A. (2010). Uji Diagnostik Gambaran Lesi Foto Thorax pada Penderita dengan Klinis Tuberkulosis Paru Diagnostic Test for Chest Radiography in Clinical Lung Tuberculose Patients Mutiara Med., 10(2), 180–188,. <https://doi.org/10.18196/mmjkk.v10i2.1582>
10. Lee, D. K. (2020). Data transformation: a focus on the interpretation. Korean journal of anesthesiology, 73(6), 503-508. [10.4097/kja.20137](https://doi.org/10.4097%2Fkja.20137)
11. Anggoro, D. A., & Aziz, N. C. (2021). Implementation of K-Nearest Neighbors Algorithm for Predicting Heart Disease Using Python Flask. Iraqi Journal of Science, 3196-3219. <https://doi.org/10.24996/ijs.2021.62.9.33>
12. Gheorghe, M., & Petre, R. The importance of normalization methods for mining medical data. International journal of computers & technology, 14(8). <https://doi.org/10.24297/ijct.v14i8.1855>
13. Fithriasari, K., Hariastuti, I., & Wening, K. S. (2020). Handling imbalance data in classification model with nominal predictors. (IJCSAM) International Journal of Computing Science and Applied Mathematics, 6(1), 33-37. <http://dx.doi.org/10.12962/j24775401.v6i1.6643>
14. Sain, H., & Purnami, S. W. (2015). Combine sampling support vector machine for imbalanced data classification. Procedia Computer Science, 72, 59-66. <https://doi.org/doi:10.1016/j.procs.2015.12.105>
15. Normawati, D., & Ismi, D. P. (2019). K-fold cross validation for selection of cardiovascular disease diagnosis features by applying rule-based datamining. Signal and Image Processing Letters, 1(2), 62-72. <http://dx.doi.org/10.31763/simple.v1i2.3>
16. Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation. Inter. J. Info. Technol. Comp. Sci., 13, 61-71. <http://dx.doi.org/10.5815/ijitcs.2021.06.0>
17. Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm. Int. J. Eng. Emerg. Technol, 2(1), 53. [https://doi.org/10.24843/IJEET.2017.v02.i01](http://garuda.ristekbrin.go.id/journal/view/16449#!)
18. Harahap, F., Harahap, A. Y. N., Ekadiansyah, E., Sari, R. N., Adawiyah, R., & Harahap, C. B. (2018, August). Implementation of Naïve Bayes classification method for predicting purchase. In 2018 6th International Conference on Cyber and IT Service Management (CITSM) (pp. 1-5). IEEE. <http://dx.doi.org/10.1109/CITSM.2018.8674324>.
19. Rodríguez-Pérez, R., & Bajorath, J. (2022). Evolution of Support Vector Machine and Regression Modeling in Chemoinformatics and Drug Discovery. Journal of Computer-Aided Molecular Design, 1-8. <https://doi.org/10.1007/s10822-022-00442-9>
20. Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press. <https://doi.org/10.1017/CBO9780511801389>
21. Simamora, W.S. Et al. (2019). A Classification: using Back Propagation Neural Network. Algorithm to Identify Cataract DiseaseJournal of Physics: Conference Series 1566 (2020) 012037: 1-7. doi:10.1088/1742-6596/1566/1/012037.
22. Sundararajan, G. Et al. (2022). LSTM Recurrent Neural Network-Based Frequency Control Enhancement of the Power System with Electric Vehicles and Demand Management. International Transactions on Electrical Energy Systems. 2022:1-23. <https://doi.org/10.1155/2022/1281248>
23. Mehdiyev, N., Enke, D., Fettke, P., & Loos, P. (2016). Evaluating forecasting methods by considering different accuracy measures. Procedia Computer Science, 95, 264-271. <https://doi.org/10.1016/j.procs.2016.09.332>
24. Chen, R. C., Dewi, C., Huang, S. W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. Journal of Big Data, 7(1), 1-26. <https://doi.org/10.1186/s40537-020-00327-4>
25. Rochman, E. M. S., & Suprajitno, H. (2022). Comparison of clustering in tuberculosis using fuzzy c-means and k-means methods. Commun. Math. Biol. Neurosci., 2022, Article-ID. <https://doi.org/10.28919/cmbn/7>