

# Artikel\_ICAISD\_Hilda\_Amalia.pdf

*by*

---

**Submission date:** 20-Jul-2020 01:55PM (UTC+0700)

**Submission ID:** 1359841920

**File name:** Artikel\_ICAISD\_Hilda\_Amalia.pdf (336.44K)

**Word count:** 2147

**Character count:** 10646

# Determination Wart Treatment Method Using Data Mining Technique

Hilda Amalia<sup>1\*</sup>, Yunita<sup>2</sup>, Achmad Baroqah Pohan<sup>3</sup>, Ari Puspita<sup>4</sup>,  
Ade Fitria Lestari<sup>5</sup>, Tri Retnasari<sup>6</sup>

Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Sastra Inggris, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, Jakarta, Indonesia

Teknologi Informasi, Sekolah Tinggi Ilmu Manajemen dan Komputer, Jakarta, Indonesia

E-mail: hilda.ham@bsi.ac.id

**Abstract.** Wart disease is a benign tumor that can grow on all parts of human skin. Although this wart disease belongs to a type of benign tumor, this wart disease can interfere with physical appearance and disturb the psychological sufferers if the infected part of the skin is the area around the face. This wart disease will also be very dangerous if there is infection in vital parts of the body such as the genitals. For treatment or treatment of wart harm, two methods are used: immunotherapy and cryotherapy. However, until now it is still unknown which treatment is better for patients with this wart disease. It has long been known that data mining is a method that is widely used in the health world to obtain valuable knowledge from medical data sets. Data Mining is used to classify and detect early diseases and choose the optimal type of treatment for warts. Data mining is able to find patterns and predict the trends of events. In this study four data mining methods will be used, namely the neural network, naive Bayes, SVM and K-NN. from the results of the study obtained the best method for processing the wart dataset is neural network

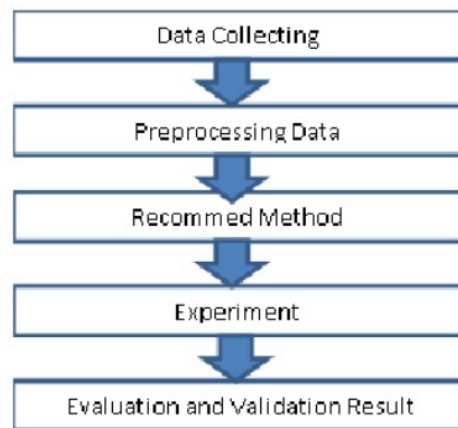
## 1. Introduction

In the world of health data mining has been widely used to provide alternative decision-making for various problems in the medical world [1]. Data Mining is an important way to find a new trend from uncovering data in health care[2]. Data Mining is a method of extracting knowledge from a set of among data. In the world of health, there is much data is only stored and not used anymore. These data only become obsolete data collisions. Data mining is able to find valuable knowledge from these obsolete data stacks. The use of data mining can provide intelligent solutions to problems faced by the medical world. One of data mining task is doing classification and it is very important task, target of classification is use to find out classification models [3]. Naïve Bayes, K-nearest neighbors, neural network and C4.5 are most popular classification techniques of data mining [4]. Naïve Bayes, K-nearest neighbors, SVM are fours of top ten algorithms identified by IEEE international Conference of Data Mining (ICDM) and present as most influence algorithms for classification, clustering, statistical learning and association analysis [5] Data mining change big data to smart data produce new pattern, new trend and new knowledge for various fields. In the world of data mining health capable of managing past

data from a disease into a valuable knowledge often produces intelligent systems that are able to detect an illness early or choose an optimal treatment model for patients. The knowledge generated by mining this data in the world of health has been widely felt that is also able to improve the quality of service to patients, the selection of better types of therapy, anticipating future patient behavior [6] Produce applications in solving problems in the world of media and used in hospital management, pharmaceutical industry and medical device industry [7] Warts disease is caused by the Human papillomavirus virus (HPV) [8] Warts can be present in every part of skin. There is a wart called genital warts that Small amounts of skin warts can be reported can develop into cancer specially if the lesion persists for long time in vagina and cervix, penis, anus [9] Genital warts can lead become a cervical cancer [10]. Treatment can be done in several ways, namely immunotherapy and cryotherapy [11]. Both of these methods are the most preferred choice by doctors to treat skin warts. For that, it is necessary to choose the optimal treatment or treatment for this wart disease. Previous research for the selection of treatment for wart disease was done by separating immunotherapy data from the data of patients who performed cryotherapy, the method used was the method of cloud computing that uses the Fuzzy method [12]. In this study four data mining methods were used, namely neural network, naive Bayes, and SVM and k-nearest network (K-NN) to select the treatment method for wart disease. Neural network, naive Bayes, SVM and K-NN are data mining methods in the scope of prediction and classification work and include supervised learning groups. The four methods are very good to use for data whose records are in numeric form.

## 2. Method

In this study several research phases are illustrated in the figure below:



**Figure 1.** Step of Research .

Figure 1 explains the stages of the research carried out in this study, the initial stage starts from data collection, the data that has been collected is processed in the initial data processing stages, after initial data processing is obtained, a dataset is ready to be used for this study. The next stage is to design a proposed method for processing this wart disease dataset, the next step is to conduct an experiment in processing wart disease training data using some of the latest data mining methods to obtain results.

### 2.1. Data Collection

In this study data obtained from data provider sites that have been widely used for research purposes, namely from the page uci repository. Data can be accessed on the page uci repository. The data presented in the collection of immunotherapy data is 91 records with eight attributes and in the cryotherapy data collection as many as 91 records with seven attributes. The following table attributes are used:

**Table 1.** Attributes Used.

Name	Data Type
3 Sex	Binominal.
Age	Numeric.
Time	Real.
Number of wart	Numeric.
Type	Numeric.
Induration Diameter	Numeric.
Result	Binominal/label.

### 2.2. Data Preprocessing

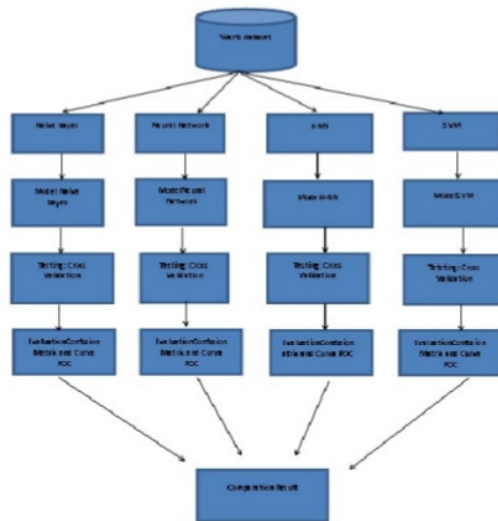
The next step after collecting the data needed for research is data preprocessing stage. The major step involved data preprocessing, namely data cleaning, data integration, data reduction and integration[13]. In this study the data that has been collected has been validated so that the number of records used is 91 records from the cryotherapy data collection and 91 from the immunotherapy data collection. Data collected is selected by integrating. Data from several types are changed and put together to produce a more meaningful dataset. In this study a combination of cryotherapy and immunotherapy datasets was carried out. So that obtained a total of 182 data records for processing wart treatment data. From 182 available data, the data is labeled yes for immunotherapy and yes for cryotherapy. Data with labels or results not for each dataset are removed because in this study a more effective selection of wart handling methods will be conducted. So that from 182 records only 162 records remains from the data integration process.

## 3. Results and Discussion

### 3.1. Recommended Model

At this stage the stages of the proposed model are made to process the training data for wart disease. Recommended models are shown in Figure 2

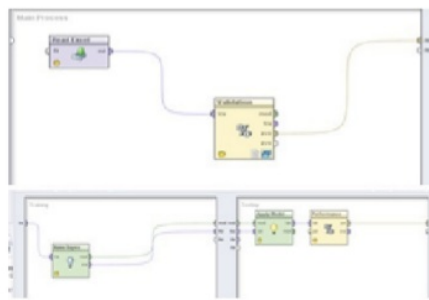
The next step is the making of a model for processing wart treatment data. Figure 2 presents the proposed method for conducting research on wart disease data processing. The wart disease dataset is treated or subjected to a trial of wart disease training data processing with four data mining methods namely naïve bayes, Neural Netowk, K-NN and Suppoert Vector Machince (SVM). After that testing is done using Cross Validation for each data mining method, the next step is to evaluate the models obtained from each data mining method using confusion matrix and ROC curves. Next, a confusion matrix value and the value obtained are compared. from the ROC curva.



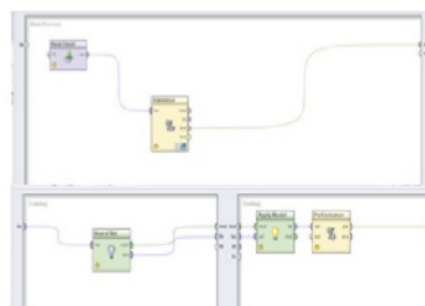
**Figure 2.** The Recommended Method.

### 3.2. Experiment

In this study four data mining methods were used which will be tested using a wart disease dataset using RapidMiner tools. The stages of wart disease dataset processing by RapidMiner for the naïve bayes method are presented in Figure 3, the wart disease dataset processing using the neural network method is presented in Figure 4. The wart disease dataset processing using the SVM method is presented in Figure 4 and finally Figure 5 is a disease dataset processing warts using the K-NN method.



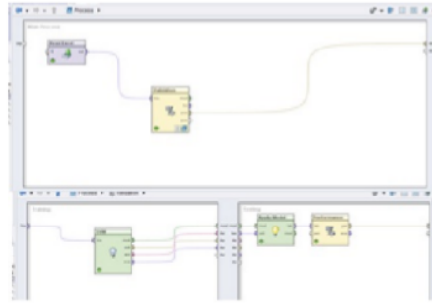
**Figure 3.** Data Processing using Naive Bayes Method.



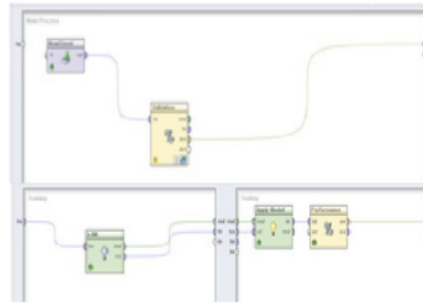
**Figure 4.** Data Processing using Neural Network Method.

In Figure 3, the modules used to do the data processing experiments are shown using the Naïve Bayes method, which is by using the Read Excel module, the module used to place the dataset to be processed in conjunction with the Validation module, this module is the place to put the method to be used. In the Validation module there is a Naive Bayes module placed in the training section, and for testing it uses the apply model module and the performance module.

The second experiment was conducted using the Neural Network method. The following modules are used to conduct data processing experiments using the neural network method



**Figure 5.** Data Processing using SVM Method.



**Figure 6.** Data Processing using K-NN Method.

Figure 4 presents data processing using the neural network method with RapidMiner tools. In the same way as processing using the previous method of Naive Bayes. The only difference is that the validation module is changed to a neural network module.

The next experiment used the Support Vector Machine (SVM) method with the same tools, RapidMiner. The modules used are presented in Figure 5

Figure 5 shows the module used, the read excel module to hold the dataset to be processed. The Validation Module which contains the method used, namely SVM, is associated with the Apply Model and Performance Module. K-NN

Finally, the processing of the wart dataset is processed using the K-NN method. The processing stages are presented in Figure 6

Figure 6 shows the steps in processing the wart database using the K-NN method. The processing stages of all data mining methods are almost the same, using the Read Excel module that is linked to the Validation module. In the validation module there is a method to be used in Figure 6. The method used is K-NN, connected with the Apply Model Module and the Performance Module. Following are the results of comparison of accuracy and AUC values for all methods:

**Table 2.** Comparative table of data mining method accuracy values.

Method	Accuracy	AUC
Naive Bayes	68,3%	0,74.
neural Network	76,52%	0,79.
K-NN	73,11%	0,73.
SVM	87,50%	0,50.

#### 4. Conclusion

Based on the experimental results of the Wart dataset processing the accuracy value for the Naïve Bayes method is 68.03% with the AUC value of 0.749. Warts dataset processing using the

Neural Network method obtained an accuracy value of 76.52% with an AUC value of 0.79. wart dataset processing using the SVM method produces an accuracy value of 87.50 with an AUC of 0.50. dataset processing using the K-NN method produces an accuracy value of 73.11% with an AUC value of 0.73%. Based on the results of experiments that have been carried out using four data mining methods it is known that the highest accuracy value is generated by the SVM method but the highest AUC value is generated by the Neural Network method which is 0.79. So it can be concluded that the prediction of the Warts Disease dataset that works well is the Neural Network Method because the accuracy value is directly proportional to the AUC value

## 5. References

- [1] S Patel and H Patel 2016 *Int. J. Inf. Sci. Tech.* **6** 53-60
- [2] V Rogeith and S Magesh 2017 *Int. J. Pure Appl. Math.* **117** 665-72
- [3] Q Dai, C Zhang, and H Wu 2016 *Int. J. Database Theory Appl.* **9** 1-8
- [4] M Karim and R M Rahman 2013 *J. Softw. Eng. Appl.* **6** 196-206
- [5] N Settouti, M E A Bechar, and M A Chikh 2016 *Int. J. Interact. Multimed. Artif. Intell.* **4** 419
- [6] K Srinivas, K Rani, and A D Gordhdhan 2010 *Int. J. Data Min. Tech. Appl.* **2** 250-5
- [7] R Ray 2018 *Int. Res. J. Eng. Technol.* 3378-3742
- [8] D D P D J Shah 2015 *Int. J. Sci. Res.* **4** 87-9
- [9] S S and P AK 2015 *Int. J. Clin. Dermatology Res.* **3** 71-5
- [10] S Jain, A Diwan and S Sardana 2015 *Genit. warts Hum. papillomavirus An Updat.* **6** 16-26
- [11] R S Sefcik and C G Burkhart 2017 *Open Dermatol. J.* **1** 30 - 34
- [12] F Khozeimeh, R Alizadehsani, M Roshanzamir, A Khosravi, P Layegh, and S Nahavandi 2017 *Comput. Biol. Med.* **81** 167 - 75
- [13] R S Sefcik and C G Burkhart 2017 *Open Dermatol. J.* **1** 30 - 34

ORIGINALITY REPORT

---

3%

SIMILARITY INDEX

1%

INTERNET SOURCES

2%

PUBLICATIONS

2%

STUDENT PAPERS

---

PRIMARY SOURCES

---

1

[pt.scribd.com](https://pt.scribd.com)

Internet Source

1%

2

[www.research-collection.ethz.ch](http://www.research-collection.ethz.ch)

Internet Source

1%

3

Submitted to University of Pittsburgh

Student Paper

<1%

4

Submitted to UT, Dallas

Student Paper

<1%

5

Herlambang Brawijaya, Samudi Samudi, Slamet Widodo. "Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes Pada Pengobatan Penyakit Kutil Menggunakan Cryotherapy", JUITA : Jurnal Informatika, 2019

Publication

<1%

6

L. S. Klivans. "Computer Controlled Launch Control and Checkout of Operational Satellite Systems", IEEE Transactions on Aerospace, 1963

Publication

<1%

---



---

Exclude quotes On

Exclude matches < 5 words

Exclude bibliography On