Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

TUGAS AKHIR

Hilwa Lelisa
41518010163

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Tugas Akhir

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:
Hilwa Lelisa
41518010163

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2022

LEMBAR PERNYATAAN ORISINALITAS

LEMBAR PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

NIM	: 41518010163
Nama	: Hilwa Lelisa
Judul Tugas Akhir	: Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Menyatakan bahwa Laporan Tugas Akhir saya adalah hasil karya sendiri dan bukan plagiat. Apabila ternyata ditemukan didalam laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.

Jakarta, 28 Juni 2022

Hilwa Lelisa

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Hilwa Lelisa
NIM	:	41518010163
Judul Tugas Akhir	:	Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Juni 2022

Hilwa Lelisa

SURAT PERNYATAAN LUARAN TUGAS AKHIR

SURAT PERNYATAAN LUARAN TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa	:	Hilwa Lelisa
NIM	:	41518010163
Judul Tugas Akhir	:	Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Menyatakan bahwa :

1. Luaran Tugas Akhir saya adalah sebagai berikut :

No	Luaran	Jenis			Status	
1	Publikasi Ilmiah	Jurnal Nasional Tidak Terakreditasi			Diajukan	
		Jurnal Nasional Terakreditasi				
		Jurnal International Tidak Bereputasi			Diterima	v
		Jurnal International Bereputasi	v			
	Disubmit/dipublikasikan di :	Nama Jurnal	: ICONETSI 2022 (International Conference on Engineering and Information Technology for Sustainable Industry)			
		ISSN	: ISBN 978-1-4503-9718-6			
		Link Jurnal	: https://easychair.org/conferences/?conf=iconetsi2022			
		Link File Jurnal Jika Sudah di Publish	:			

2. Bersedia untuk menyelesaikan seluruh proses publikasi artikel mulai dari submit, revisi artikel sampai dengan dinyatakan dapat diterbitkan pada jurnal yang dituju.
3. Diminta untuk melampirkan scan KTP dan Surat Pernyataan (Lihat Lampiran Dokumen HKI), untuk kepentingan pendaftaran HKI apabila diperlukan

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Juni 2022

Hilwa Lelisa

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41518010163
Nama	:	Hilwa Lelisa
Judul Tugas Akhir	:	Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 3 Agustus 2022

(Runi)

(Saruni Dwiasnati, ST, MM, M.Kom)

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41518010163
Nama	:	Hilwa Lelisa
Judul Tugas Akhir	:	Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 3 Agustus 2022

(Achmad Kodar, Drs. MT)

LEMBAR PERSETUJUAN PENGUJI

LEMBAR PERSETUJUAN PENGUJI

NIM	:	41518010163
Nama	:	Hilwa Lelisa
Judul Tugas Akhir	:	Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 3 Agustus 2022

(Wawan Gunawan, S.Kom, MT)

LEMBAR PENGESAHAN

NIM : 41518010163

Nama : Hilwa Lelisa

Judul Tugas Akhir : Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 3 Agustus 2022

Menyetujui,

(Rahmat Budiarto, Dr. Prof)
Dosen Pembimbing

Mengetahui,

(Wawan Gunawan, S.Kom, MT)	(Ir. Emil R. Kaburuan, Ph.D., IPM.)
Koord. Tugas Akhir Teknik Informatika	Ka. Prodi Teknik Informatika

KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah SWT atas rahmat dan hidayah-Nya, sehingga tugas Akhir yang berjudul "Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim" dapat diselesaikan dalam jangka waktu yang sudah ditentukan. Laporan Tugas Akhir ini dibuat sebagai syarat untuk LULUS sebagai sarjana Ilmu Komputer dari Universitas Mercu Buana.

Dalam menyelesaikan Laporan Tugas Akhir ini penulis mendapatkan bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Kedua orang tua, Mama, Ayah, Abang, dan Adik, yang selalu memberikan dukungan dan semangat agar saya dapat menyelesaikan kuliah dengan baik, juga tak pernah luput untuk mendoakan setiap proses untuk meraih gelar sarjana bagi saya.
2. Bapak Boy Yuliadi dan Bapak Achmad Kodar, Drs. MT selaku dosen Pembimbing Akademik yang telah membimbing saya selama perkuliahan.
3. Bapak Rahmat Budiarto, Dr. Prof selaku dosen Pembimbing Tugas Akhir yang telah membimbing dan meluangkan waktu untuk memberikan masukan dalam penyusunan tugas akhir hingga selesai.
4. Seluruh Dosen Program Studi Teknik Informatika yang telah memberikan ilmu yang bermanfaat selama perkuliahan.
5. Dinas Kesehatan Kabupaten Muara Enim.
6. Seluruh teman baik saya, baik yang dikenal secara langsung maupun secara virtual, angkatan Teknik Informatika 2018, mi casa, BTS dan ARMY. Terkhusus kepada Ismawati yang menjadi partner dari awal proses perkuliahan hingga akhir.
7. Seluruh keluarga besar dan pihak-pihak yang tidak dapat saya sebutkan satu persatu, yang mendukung dan memberikan semangat kepada saya untuk menyelesaikan perkuliahan.

Akhir kata, penulis berharap Tugas Akhir ini menjadi referensi untuk dapat dikembangkan, penulis menyadari bahwa laporan Tugas Akhir ini masih jauh dari sempurna dan masih terdapat banyak kekurangan. Untuk itu, kritik dan saran yang membangun sangat dihargai dan diharapkan. Semoga Tugas Akhir Ini memberikan manfaat bagi para pembaca.

Jakarta, 28 Juni 2022
Penulis

DAFTAR ISI

NASKAH JURNAL

Analysis of Tuberculosis Disease Spreading Pattern in Muara Enim District using KNN Algorithm

Analysis of TB Disease Spreading Pattern using KNN Algorithm

Hilwa Lelisa

Informatics Dept., Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia,
41518010163@student.mercubuana.ac.id

Yaya S. Triana

Informatics Dept., Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia,
yaya.sudarya@mercubuana.ac.id

Rahmat Budiarto

Informatics Dept., Faculty of Computer Science, Universitas Mercu Buana, Jakarta, Indonesia,
rahmat.budiarto@mercubuana.ac.id

Tuberculosis (TB) is a type of infectious disease caused by Mycobacterium tuberculosis, which not only attacks the lungs, but can also attack the bones, intestines, or glands. During the Covid-19 pandemic, TB cases in Indonesia also increased. TB and Covid-19 had the similar symptoms such as cough, fever, and breathing difficulty, so that TB sufferers must be given serious treatment to avoid Covid-19. In predicting a disease, it is important for health workers to make decisions, thus it is necessary to do an early diagnosis in order to reduce the transmission of TB in the community. There are many algorithm methods used in conducting data analysis, for this study the authors use K-Nearest Neighbor (K-NN) algorithm and Logistic Regression as comparison. Experimental results using available dataset collected from health centers in Muara Enim District of South Sumatra Province show that the K-NN algorithm provides the best accuracy of 89% on dataset with training to testing data ratio of 80%:20%, while the Logistic Regression provides the best accuracy of 96% on 70%:30% ratio. The analysis mechanism discussed in this paper may be considered as tool for the authority to predict and take necessary actions to prevent the TB spreading.

CCS CONCEPTS • Computing methodologies • Machine learning

Additional Keywords and Phrases: TB, Logistic Regression, K-Nearest Neighbor

ACM Reference Format:

First Author's Name, Initials, and Last Name, Second Author's Name, Initials, and Last Name, and Third Author's Name, Initials, and Last Name. 2018. The Title of the Paper: ACM Conference Proceedings Manuscript Submission Template: This is the subtitle of the paper, this document both explains and embodies the submission format for authors using Word. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

1 INTRODUCTION

Tuberculosis (TB) is a type of infectious disease caused by Mycobacterium tuberculosis, which not only attacks the lungs, but can also attack the bones, intestines, or glands. World Health Organization (WHO) has declared TB as a global emergency. TB is still a threat to countries in the world and needs to be eradicated optimally. Symptoms such as fever, cough, chest pain and shortness of breath are things that cause TB disease. In Indonesia, TB is ranked as the second dangerous infectious disease with 845, 000 cases and 98,000 deaths. The Indonesian government attempts to reduce the number of TB sufferers by 2030 by means of prevention, diagnosis, treatment, and providing health services for all TB sufferers. During the pandemic, TB cases in Indonesia increased, TB and Covid-19 had the same symptoms such as cough, fever, and difficulty breathing. So that TB sufferers must be given serious treatment to avoid Covid-19 [1].

Because this disease is contagious, a system is needed to predict the spread of TB cases, one of which is making predictions using data science techniques. In predicting a disease, it is important for health workers to make decisions, thus it is necessary to perform an early diagnosis based on appearing symptoms, in order to

reduce the transmission of TB in the community. There are many algorithms used in conducting data analysis. The K-Nearest Neighbor (K-NN) algorithm is one of the popular methods in data classification. This algorithm is very suitable to be used because it is easy to implement in diagnosing a disease, K-NN is usually applied in data classification based on the value of K, i.e.: the value of the nearest neighbor [5]. Because it is only seen from the value of the nearest neighbor, this algorithm is called lazy learning [6]. Besides, K-NN is also effective if implemented on large training data [7]. Logistic Regression Algorithm is the most famous machine learning algorithm after Linear Regression. The algorithm is also used for classification tasks [8and commonly used in the approach to make a predictive model of an event [9].

A previous study on warts disease analysis using K-NN obtained an accuracy of 90.90% [2]. Another study on heart disease analysis using the K-NN algorithm provided an accuracy of 96.66% [3]. A research work using the Logistic Regression algorithm on cardiovascular disease analysis obtained an accuracy of 87.10% [4].

Based on previous research works mentioned above, the authors are interested in choosing the K-NN algorithm for this study and use the Logistic Regression as comparison to analyze and predict the spread of TB cases.

The rest of the paper is arranged as follows. Section 2 discusses the proposed method along with the dataset for the experiment. Section 3 presents the experimental results and discussion, and lastly, Section 4 provides the conclusion and future works.

2 RESEARCH METHOD

Figure 1 depicts the work flow of the proposed research method. The first step is to construct TB disease spreading dataset. The raw data is obtained from the Unit of Health, Muara Enim District. Then the raw data was cleaned through a data preprocessing. The clean dataset are then splitted into training and testing dataset. Next step is implementation of the KNN and Logistic Regression algorithms followed by experimentation and validation using K-fold Cross Validation. The last step is to report the experimentation results in terms of accuracy and other performance metrics.

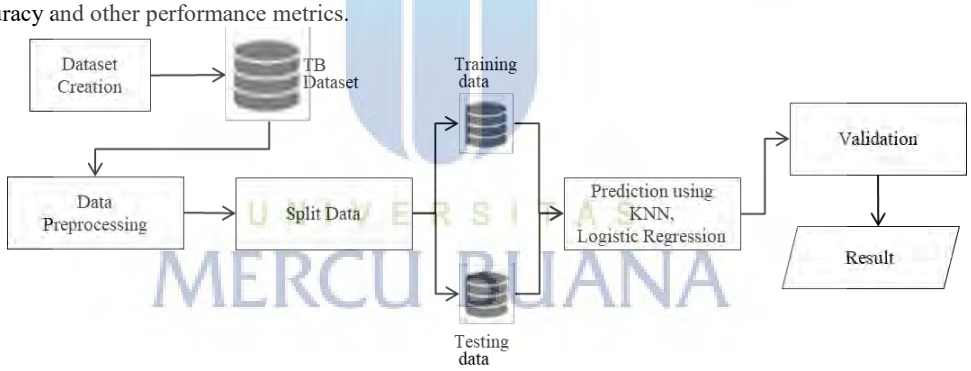


Figure 1. The proposed research method

2.1 Dataset

The data are collected from 40 healthcare centers in the district. After data cleaning process, the dataset consists of 640 rows. Each row has 9 attributes, i.e.: Health center name, category, age, gender, type of patient (new patient, relapse patient, patient with TB disease history other than relapse, patient without treatment history), and number of patients.

2.2 Data Split

Types of data in the dataset are already in numeric format, thus no data conversion is performed during the preprocessing. The preprocessing data involves normalization and completing the missing data only.

Next is dataset description process and determine the dependent variable and independent variables. The dataset is split with three scenarios for the ratio of training and testing data as follows: 90%:10%, 80%:20% and 70%:30%. The dataset description is shown in Figure 2 to Figure 6. The figures show the increase trend of TB disease sufferers.

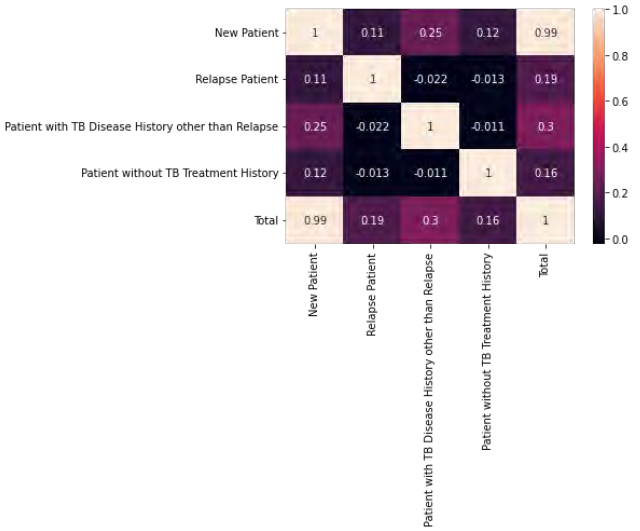


Figure 2. Correlation among features/attributes

Figure 2 sows the correlation test among variables, generated by Heatmap.

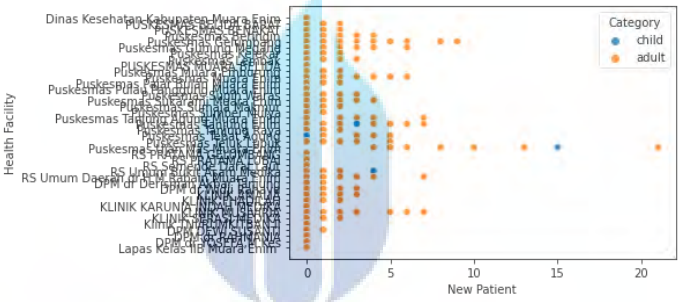


Figure 3. Scatterplot of New Patient

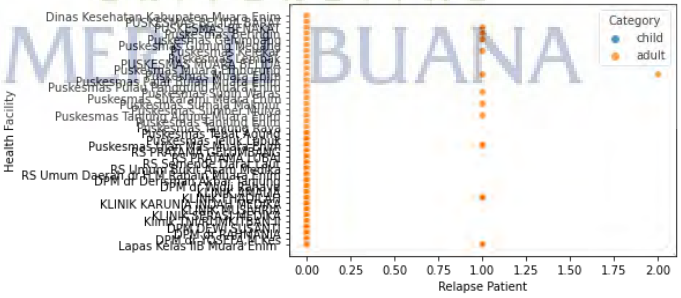


Figure 4. Scatterplot of Relapse Patient

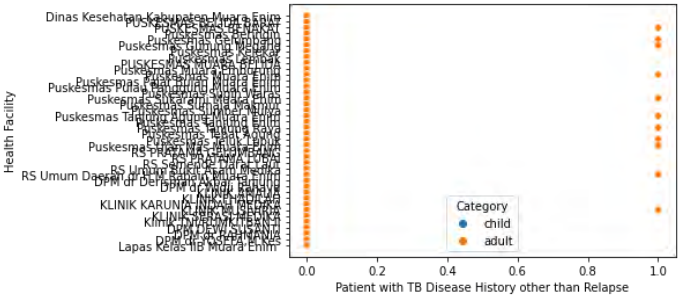


Figure 5. Scatterplot of patient with TB disease history other than relapse

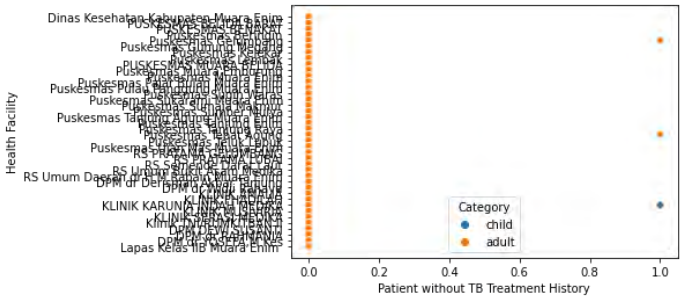


Figure 6. Scatterplot of patient without TB treatment history

2.3 K-Nearest Neighbor

K-Nearest Neighbor (KNN) is a type of classification algorithm by determining the value of the nearest neighbor. The KNN algorithm was chosen because it has advantages such as easy to understand and implement, very fast and effective training if the training data is large enough. One of the weaknesses of this algorithm is that it is necessary to determine the value of the K parameter, namely the number of nearest neighbors. To use this algorithm, it is necessary to initialize the K parameter [12], in this study the K value used is K = 1 to K = 9, and obtain the K value with the best accuracy. Euclidean distance is the most commonly used distance calculation in the KNN algorithm (1).

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (1)$$

x_i = sample data
y_i = testing data
i = variable data
d(x, y) =dissimilarity/distance
n = data dimension

2.4 Logistic Regression

Logistic Regression is a classification algorithm to find the relationship between discrete/continuous features (input) with the probability of certain discrete output results and is part of the regression analysis used when the dependent variable (response) is a dichotomous variable that has a value of 0 or 1 [10]. The result variable is dichotomous which has two possibilities, yes or no, this algorithm can be used to detect TB disease. The equation for Logistic Regression is shown in (2) [16].

$$Ln\left(\frac{p}{1-p}\right) = B_0 + B_1X \quad (2)$$

B_0 = constant
B_1 = coefficient of each variable

The p value or probability (Y =1) can be determined by (3).

$$p = \frac{e^{(B_0 + B_1X)}}{(1 + e^{(B_0 + B_1X)})} \quad (3)$$

2.5 k-Fold Cross Validation

k-Fold Cross Validation is one method that is commonly used in model selection [13]. k-Fold Cross Validation increases the flexibility of the selected model and obtains maximum accuracy results [14]. Its function is to divide the data sample randomly and group the data as many as the k value. In machine learning, the k value that is often used is 10 [15].

3 Result And Discussion

As mentioned in Section 2.1, the experiments use three data split scenarios for training data and testing data. The experiments are carried out on a computer with the following specifications. Processor 11th Gen Intel(R)

Core(TM) i5-1135G7 @ 2.40GHz, 8 GB RAM, running Windows 11 operating system and Python language programming on *Google Colab* platform.

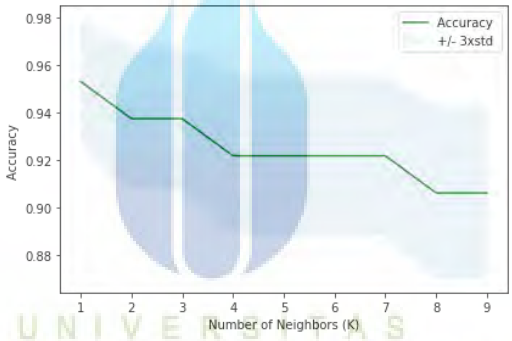
3.1 K-Nearest Neighbor Implementation

The KNN algorithm is implemented using Skicit learn library. Table 2 shows the accuracy for K=4 on training data and testing data. The testing with the value of K = 4 is considered based on the number and size of the data dimensions. The more data used, the lower the value of K. The larger the data dimension the higher the value of K chosen.

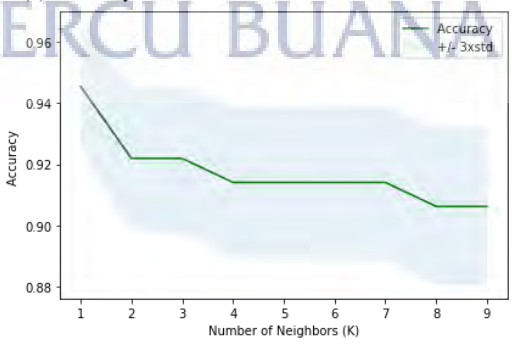
Table 1. Accuracy for K=4

Scenario	Training	Testing
90%:10%	0.97	0.92
80%:20%	0.97	0.91
70%:30%	0.97	0.93

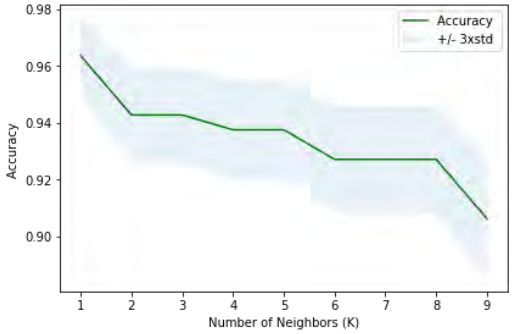
Next, performing experiments for variant values of K (from K=1 to K=9) for each dataset ratio scenario. Accuracy for variant values of K on the testing dataset is shown in Figure 7. The best accuracy for 90% to 10% ratio of dataset is achieved for K=1, i.e.: 0.95 (see Figure 7.a). The best accuracy for 80% to 20% ratio of dataset is also achieved for K=1, i.e.: 0.94 (see Figure 7.b). 70% to 30% ratio of dataset is also achieved for K=1, i.e.: 0.96 (see Figure 7.c). Table 2 summarizes the best accuracy during the testing for each algorithm and for each scenario.



(a) Accuracy for variant values of K on 90%:10% ratio



(b) Accuracy for variant values of K on 80%:20% ratio



(c) Accuracy for variant values of K on 70%:30% ratio

Figure 7. Accuracy for different values of K

Table 2. Accuracy Result of KNN Algorithm

Dataset ratio	The best value of K	Training Accuracy	Testing Accuracy
90%:10%	K=1	1.0	0.95
80%:20%	K=1	1.0	0.94
70%:30%	K=1	1.0	0.96

3.2 Logistic Regression Implementation

Experimental results during testing for Logistic Regression algorithm are shown in Table 3.

Table 3. Accuracy Result of Logistic Regression Algorithm

Dataset ratio	The best value of K	Training Accuracy	Testing Accuracy
90%:10%	K=1	1.0	82.81
80%:20%	K=1	1.0	89.84
70%:30%	K=1	1.0	77.08

3.3 k-Fold Cross Validation Result

The KNN and Logistic Regression algorithms are validated using 10-Fold cross validation method with random state 1 and the results are shown in Table 4. Average score for K-nearest Neighbor and Logistic Regression algorithm is 0.98 and 0.82, respectively.

Tabel 4 Accuracy comparison

Dataset ratio	Logistic Regressiom	K-Nearest Neighbor
90%:10%	82%	95%
80%:20%	89%	94%
70%:30%	77%	96%

From Table 4, we may conclude that Logistic Regression algorithm achieves the best accuracy level of 89% using the 80%:20% dataset ratio. Whereas, K-Nearest Neighbor algorithm achieves the best accuracy level of 96% using the 70%:30% dataset ratio.

4 CONCLUSION

Tuberculosis and Covid-19 have almost similar symptoms, so it is necessary to pay attention to avoid TB patients from Covid-19 infection. Experimental results showed that the accuracy of prediction analysis of Tuberculosis disease spreading pattern in Muara Enim District using KNN algorithm is higher than Logistic Regression. The best accuracy value using Logistic Regression algorithm is 89% while the best accuracy using K-Nearest Neighbor algorithm is 96%.This research provides useful insights regarding the spread of TB during the Covid-19 pandemic and the use of K-Nearest Neighbor algorithm as a reference for further research.

This research only compares the KNN algorithm with Logistic Regression and does not consider existing works, as we use specific dataset. Thus, the authors consider other techniques such as Artificial Neural Network, Support Vector Machine, and Fuzzy inference to be implemented to analyze the dataset.

ACKNOWLEDGMENTS

Authors thanks to Research & Technology Directorate, Universitas Mercu Buana for partially support on this research, Grant no: 02-5/1337/B-SPK/V/2022.

REFERENCES

[1] Y. Arvita and D. Meisak. 2021. Information System for Collecting Data on People Affected by Tuberculosis for Covid-19, (in Bahasa Indonesia), J. Sains dan Inform., vol. 7, no. 2, pp. 202–211, 2021, doi: 10.34128/jsi.v7i2.344.

[2] H. Samudi, and S. Widodo, 2019. Comparison of the K-Nearest Neighbor and Naive Bayes algorithm in the treatment of warts using cryotheraphy), JUITA J. Inform., vol. 7, no. 2, pp. 93–99, 2019.

[3] R. Yunus, U. Ulfa, and M. D. Safitri. 2021. Application of the K-Nearest Neighbors (K-NN) algorithm for classification of heart failure, J. Appl. Intell. Syst., vol. 6, no. 1, pp. 1–9, 2021.

[4] A. G. B. Ganesh, A. Ganesh, C. Srinivas, Dhanraj, and K. Mensinkal. 2022. Logistic Regression technique for prediction of cardiovascular disease, Glob. Transitions Proc., vol. 3, no. April, pp. 127–130, 2022, doi: 10.1016/j.gltp.2022.04.008.

[5] M. Muhathir, T.T. Sibarani, & A. Al-Khowarizmi. 2020. Analysis K-Nearest Neighbors (KNN) in Identifying Tuberculosis Disease (TB) By Utilizing Hog Feature Extraction, Int. Comput. Sci. Inf. Technol. Journal, vol. 1, no. 1, pp. 33–38, 2020.

[6] S. Widaningsih. 2019. Comparison of Data Mining Methods for Predicting Values and Graduation Time of Informatics Engineering Students with C4.5, Naïve Bayes, Knn and Svm Algorithms, (in Bahasa Indonesia), J. Tekno Insentif, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[7] R. A. Rizal, N. O. Purba, L. A. Siregar, K. Sinaga, and N. Azizah. 2020. Analysis of Tuberculosis (TB) on X-ray image using SURF feature extraction and the K-Nearest Neighbor (KNN) classification method, JAICT, vol. 5, no. 2, pp. 9–12, 2020, doi: 10.32497/jaict.v5i2.

[8] M. I. Gunawan, D. Sugiarto, and I. Mardianto. 2020. Improved Performance of Diabetes Mellitus Prediction Accuracy Using the Grid Search Method on the Logistic Regression Algorithm, (in Bahasa Indonesia) J. Edukasi dan Penelit. Inform., vol. 6, no. 3, p. 280, 2020, doi: 10.26418/jp.v6i3.40718.

[9] A. Wulandari. 2020. Comparison of classification of stock price movements PT. Astra Internasional Tbk uses stationary vector auto regressive (var) and logistic regression, (in Bahasa Indonesia) eProceedings Eng., vol. 7, no. 1, pp. 2614–2626, 2020.

[10] Utami et al. 2021. Integration face recognition and body temperature, JITE, vol. 5, no. July, pp. 53–64, 2021.

[11] K. Kristiawan and A. Widjaja. 2021. Comparison of Machine Learning Algorithms in Assessing a Retail Store Location, (in Bahasa Indonesia), J. Tek. Inform. dan Sist. Inf., vol. 7, no. 1, pp. 35–46, 2021, doi: 10.28932/jutisi.v7i1.3182.

[12] D. Stiawan, S. M. Daely, A. Heryanto, N. Afifah, M. Y. Idris, and R. Budiarto. 2021. Ransomware detection based on opcode behaviour using k-nearest neighbours algorithm, Inf. Technol. Control, vol. 50, no. 3, pp. 495–506, 2021, doi: 10.5755/j01.itc.50.3.25816.

[13] J. He and X. Fan. 2019. Evaluating the performance of the K-fold cross-validation approach for model selection in growth mixture modeling, Struct. Equ. Model., vol. 26, no. 1, pp. 66–79, 2019, doi: 10.1080/10705511.2018.1500140.

[14] S. Saud, B. Jamil, Y. Upadhyay, and K. Irshad. 2020. Performance improvement of empirical models for estimation of global solar radiation in India: A k-fold cross-validation approach, Sustain. Energy Technol. Assessments, vol. 40, no. June, p. 100768, 2020, doi: 10.1016/j.seta.2020.100768.

[15] T. & I. Hutapea. 2018. Application of Modified K-Nearest Neighbor Algorithm in Classification of Schizophrenic Mental Illness, (in Bahasa Indonesia), J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 10, pp. 3957–3961, 2018.

[16] F. D. Pramakrisna, F. D. Adhinata, N. Annisa, and F. Tanjung. 2022. Web-based sms classification application using logistic regression algorithm, (in Bahasa Indonesia), TEKNIKA,vol. 11, no. 2, pp. 90–97, 2022, doi: 10.34148/teknika.v11i2.466.

KERTAS KERJA

Ringkasan

Kertas kerja ini merupakan material kelengkapan artikel jurnal dengan judul "Analisis Tingkat Akurasi Penyebaran Penyakit TBC Menggunakan Algoritma Logistic Regression dan K-Nearest Neighbor di Kabupaten Muara Enim". Kertas kerja berisi semua material hasil penelitan Tugas Akhir yang tidak dimuat/atau disertakan di artikel jurnal. Di dalam kertas kerja ini disajikan: literature review, dataset yang digunakan, source code, dan hasil eksperimen secara keseluruhan.

Bagian I merupakan literature review dari hasil review jurnal yang terkait dalam penelitian ini. Bagian II merupakan source code yang digunakan dalam penelitian. Bagian III merupakan bagian yang menjelaskan dataset yang digunakan. Bagian IV merupakan tahapan eksperimen yang dilakukan dalam penelitian ini. Bagian V merupakan bagian yang menjelaskan hasil dari keseluruhan ekperimen.