Institute of Research and Publication Indonesia (IRPI)
Public Research Journal of Engineering, Data Technology and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/predatecs
Vol. 1 Iss. 1 July 2023, pp: 26-33
ISSN(P): 3024-921X | ISSN(E): 3024-8043

Application of the Supervised Learning Algorithm for Classification of Pregnancy Risk Levels

Zairy Cindy Dwinnie[1*], Luthfia Khairani[2], Margareta Amalia Miranti Putri[3], Jeni Adhiva[4], Muhammad Inas Farras Tsamarah[5]

[1,2,3]Department of Information Systems Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[4]Department of Computer Science, Faculty of Mathematics and Natural Science,
Institut Pertanian Bogor (IPB), Indonesia
[5]Department Computer Science and Information Engineering, College of Science and Engineering,
National Dong Hwa University, Hualien, Taiwan

E-Mail: [1]12050324081@students.uin-suska.ac.id, [2]12050323175@students.uin-suska.ac.id,
[3]12050320348@students.uin-suska.ac.id, [4]jeniadhiva@apps.ipb.ac.id, [5]inasfarras02@gmail.com

Received March 17th 2023; Revised Jun 4th 2023; Accepted Jul 2nd 2023
Corresponding Author: Zairy Cindy Dwinnie

Abstract

MMR is the number of women who die due to disorders during pregnancy or their treatment (excluding accidents, suicides, or incidental cases) during pregnancy, childbirth, and during the puerperium or 42 days after giving birth. This research aims to classify pregnancy risk datasets, namely to compare the performance of the NBC, K-NN, and SVM methods on the pregnancy risk status dataset and to find out the accuracy comparison of the algorithm results above. From the results of the analysis, it was found that of the three algorithms it resulted in a classification of pregnancy risk levels with the highest value occurring at a high level. To determine the accuracy of the data, a comparison was made between the three algorithms. Based on the confusion matrix namely Accuracy, Precision, and Recall. The results of the comparison can be concluded that the KNN algorithm provides the highest accuracy of 77.55%, NBC of 69.39%, and the lowest accuracy by SVM of 67.35%. These results state that the KNN algorithm classifies pregnancy risk level data better than the other two algorithms.

Keywords**:** Classification, K-Nearest Neighbor, Naïve Bayes Classifier, Pregnancy Risk Level, Support Vector Machine

1. INTRODUCTION

Pregnancy is a very important period for both a woman. In this period, the welfare and health of the mother and the fetus in her womb need attention. Maternal Mortality Rate (MMR) is one of several indicators that can assess the welfare of the community in an area/region. MMR, which means a woman who dies due to complications during pregnancy or its treatment (excluding accidents, suicide, or incidental cases) during pregnancy, childbirth, and during the puerperium or 42 days after giving birth [1]. The higher the MMR value in a country indicates the lower the health status of women in that country. This can also be one of the causes of the decline in the economy of both the family and the country.

In Indonesia, data from the Ministry of Health shows that there are 6,856 MMR in 2021, where this value shows an increase in MMR from the previous 4,197 maternal deaths in 2019. Meanwhile, according to WHO (2019), the global maternal mortality rate in 2019 reached 303,000 people. The high MMR can be caused by many factors. One of them is the author of Delays in processing cases. This late treatment can be prevented by early detection. Risk of harm during pregnancy [2].

At the sub-district and district/city levels, there are ways to reduce MMR and IMR involved in Making Pregnancy Safer (MPS), namely by providing directions for the management of 24-hour Comprehensive Emergency Obstetric Services (PONEK) with the most important steps: increasing early detection, pregnancy management high risk (staying) and strengthening city/regional level Program leadership skills in planning, performance management, monitoring and evaluation Admission of MMR and IMR [3].

In previous research, optimizing the data mining algorithm at the level of pregnancy risk, the accuracy of C4.5 is better than Naïve Bayes and the performance value of C4.5 increases with the addition of Particle Swarm Optimization (PSO) [1]. Furthermore, by using the M-KNN method in the risk classification of the pregnancy level, the results of the test accuracy were 85% which stated that the Modified K-Nearest Neighbor (MKNN) method was suitable for use in studies of disease risk levels of pregnant women.

The classification method applied is the SVM, K-NN, and NBC methods. This study aims to prove the best accuracy results in data mining classification on the risk level of pregnancy from the three algorithms. A previous study compared 7 Machine Learning Algorithms, namely the Logistic Regression algorithm, Decision Tree, MLPClassifier, SVM, Random Forest, Naive Bayes, and KKN for the Classification of Fetal Heart Rate with fetal heart rate. This research produces an accuracy value of 94% [4].

Other research in analyzing mathematics students' performance also uses seven methods, namely K-nearest neighbor, classification and regression trees, naïve Bayes, AdaBoost, extra tree, Bernoulli naïve Bayes, and random forest. The results obtained from the results of this study are the Random Forest G algorithm which is the algorithm with the best classification results of 89.79% [5]. Based on the above background, this research will be conducted to classify pregnancy risk datasets, namely to compare the performance of the NBC, K-NN, and SVM methods on the pregnancy risk status dataset and find out the accurate comparison of the algorithm results above.

2. MATERIALS AND METHODS

The data used in this study comes from UCI Machine Learning in the form of a pregnancy risk level dataset.

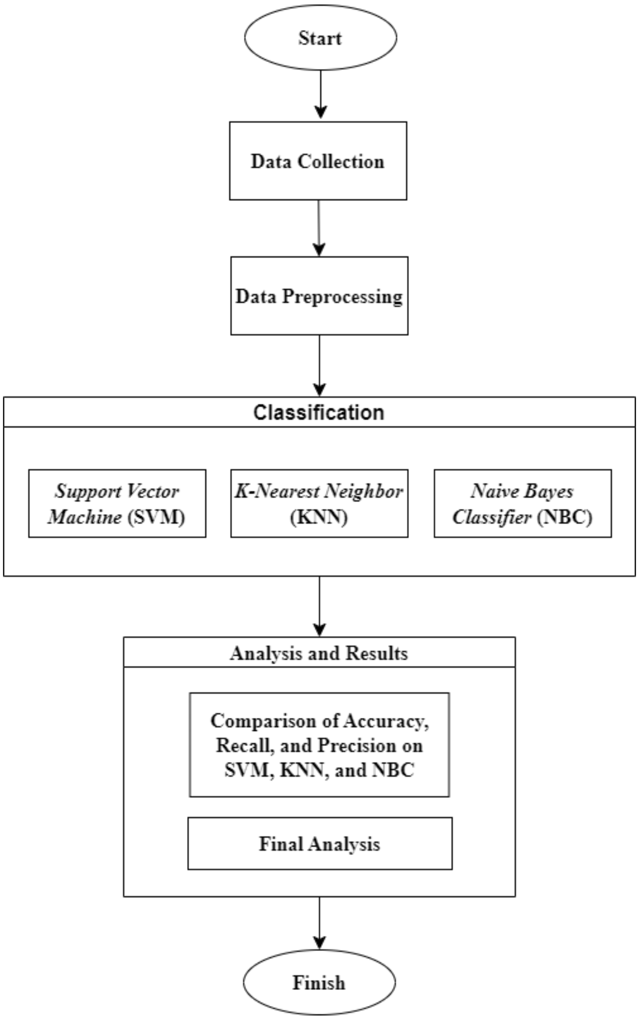
Start
Data Collection
Data Preprocessing
Classification
Support Vector Machine (SVM)
K-Nearest Neighbor (KNN)
Naive Bayes Classifier (NBC)
Analysis and Results
Comparison of Accuracy, Recall, and Precision on SVM, KNN, and NBC
Final Analysis
Finish

Figure 1. Research Methodology

2.1 Data collection

The data used in this study is a type of Maternal Risk Level classification data sourced from the UCI Machine Learning Repository website. Where the attributes in this data are indicators of pregnancy risk, namely Age, SystolicBP, DiastolisBP, blood glucose (BS), heart rate, and Risk Level.

Table 1. Data Attributes

Data Attributes	Information
Age	A person's age, namely the number of years since a person was born
Systolic Blood Pressure	The pressure of blood in the arteries when the heart is contracting and pumping blood into the bloodstream
Diastolic Blood Pressure	The pressure of blood in the arteries when the heart is resting between beats
Blood Glucose	The level of glucose (sugar) in a person's blood
Heart Rate	Number of heartbeats in one minute
Risk Level	The level of possibility of a person experiencing certain health problems is based on the factors above.

Analysis carried out based on reviewing a set of data so that it can be understood and useful which aims to find unpredictable relationships and summarize data in various ways that are different from before is one of the definitions of Data Mining[6]. Is a scientific field that utilizes techniques from machine learning, pattern recognition, statistics, databases, and visualization to be able to solve existing problems in the process of retrieving information from large databases[7].

2.2 Preprocessing Data

After collecting data, we need to prepare the data neatly before conducting analysis using machine learning models. There is a data preparation process that must be carried out to clean and modify the data so that it is more suitable for model training. This is called Data Preprocessing. This stage is cleaning missing value data, which removes data that is not used in calculations and is normalized[8]. This preprocessing is also the stage where the data is filled with empty data, duplicating data, checking data inconsistencies, cleaning data, and correcting errors in the data[9]. In this study, we use data that has been neat, so cleaning is not necessary. The transformation is to make all parts of the data attribute numeric (integer). The following table is before and after the transformation.

Table 2. Pregnancy Risk Level Dataset

Age	Systolic BP	Diastolic BP	BS	Body Temp	Heart Rate	Risk Levels
25	130	80	15	98	86	high risk
35	140	90	13	98	70	high risk
29	90	70	8	100	80	high risk
30	140	85	7	98	70	high risk
35	120	60	06.01	98	76	low risk
23	140	80	07.01	98	70	high risk
…	…	…	…	…	…	…
16	120	75	07.09	98	7	low risk

Table 3. Dataset After Transformation

Age	Systolic BP	Diastolic BP	BS	Body Temp	Heart Rate	Risk Levels
25	130	80	15	98	86	3
35	140	90	13	98	70	3
29	90	70	8	100	80	3
30	140	85	7	98	70	3
35	120	60	06.01	98	76	1
23	140	80	07.01	98	70	3
…	…	…	…	…	…	…
16	120	75	07.09	98	7	1

It can be seen in Table 3. The changes that occur in this transformation are the risk level attribute to be in the form of a numeric (integer) where this attribute is a label or target class in this classification. Where high risk = 3, mid risk = 2, and low risk = 1. Then select the features or data attributes to be used. Here we use all the attributes that support the classification of the target class, namely Age, Systolic Blood Pressure, Diastolic Blood Pressure, Blood Glucose, body temp, and heart rate. The amount of data used is 500 data.

2.3 Classification

The process of finding patterns (or functions) that describe and separate data classes or concepts to predict classes of objects with unknown class identifiers[10]. The classification algorithms that are widely used, namely Decision/classification trees, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Statistical Analysis, Genetic Algorithms, Rough sets, k-nearest neighbors, Rule-Based Methods, Memory based reasoning, and Support Vector Machines (SVM) [11].

2.4 Support Vector Machine (SVM)

Based on statistical learning theory, Support Vector Machine (SVM) is presented by Vapnik[12]. SVM is one of the algorithms that is often used for data classification analysis. The SVM method is based on the VC dimension of statistical learning and structural risk minimization (SRM) principles[13]. Classification of data in SVM is to get the optimal hyperplane separator between positive and negative[14]. For document classification, classifier selection is another major issue after dimension reduction. Found in statistical learning theory, the Support Vector Machine (SVM) classifier has attracted much attention because of its good performance in practical applications and its strong theoretical foundation[7]. The weakness of this model lies in the calculation process which is relatively long compared to other classification methods[15]. The SVM method divides the dataset into two classes. The first class separated by a hyperplane has a value of 1, while the class has a value of -1. The following is the equation of the SVM model.

$$\begin{gathered} Xi.W + b \geq 1 \text{ for } Yi = 1 \\ Xi.W + b \leq 1 \text{ for } Yi = -1 \end{gathered} \tag{1}$$

Information:

Xi	: data to -i
W	: weight value support Vector perpendicular to the hyperplane
b	: biased value
Yi	: data class to -i

2.5. K-Nearest Neighbor (KNN)

K-Nearest Neighbor is a classification algorithm that is often used in data classification. The working principle of the K-nearest neighbor (KNN) is to find the shortest distance between the data to be evaluated and the K-nearest neighbor of the training data. K is a positive integer that is determined before running the algorithm[16]. This model works classically to predict outcomes using a decrease in the value of k[17],[18]. Some researchers often use the Euclidean distance to calculate the distance between objects. The advantages possessed by KNN are data robustness and the effect on large amounts of training data and its performance is quite good. It's just that the computation time has passed very long if the training data is big and good sensitive to redundant or related features[19]. The equation of the KNN is as follows.

$$d_{\mathrm{Euclidian}} = \sqrt{\sum_{i=1}^{n} (x_{i2} - x_{i1})^2} \tag{2}$$

Information:

deEuclidian: Distance

xi1	: Sample Data
xi2	: Test Data
n	: Number of Attributes

2.6. Naïve Bayes Classifier (NBC)

NBC is one of the most widely used algorithms for making the most accurate predictions based on data collection because it is relatively easy to perform, understand, and very accurate.[16].Naïve Bayes has high speed and accuracy when applied to data owners with large enough data[20]. At the time of classification, the algorithm will look for the highest probability of all document categories tested[21]. The first step of the classifier is to calculate the average and standard deviation of the training data features of each class[22].
The NBC equation 3.

$$\mathrm{P}\ (\mathrm{Ci}|\mathrm{X}) = \frac{\mathrm{P}(\mathrm{X}|\mathrm{Ci})\mathrm{P}(\mathrm{Ci})}{\mathrm{P}(\mathrm{X})} \tag{3}$$

Information:

X	: Criteria for a case based on input
ci	: The i-th pattern solution class, where i is the number of class labels
P(Ci|X)	: Probability of appearance of class label Ci with the input criterion X
P(X|Ci)	: Probability of input criteria X with class label Ci
P(Ci)	: Class label probability Ci

2.7. Confusion Matrix

Confusion matrix is an evaluation method used to determine performance classification based on right and wrong. The confusion matrix has accuracy, precision, and recall. This formula performs calculations with four outputs, namely: recall, precision, accuracy, and error rate. The evaluation of the classification model is based on evaluating the correctness and falsity of the items in the test[23]. The confusion matrix has four important values, namely true positive (TP) and true negative (TN) which means the model gives correct prediction results, false positive (FP), and false negative (FN) which means the model gives wrong prediction results.[4]. The following is a diagram of the Confusion Matrix of figure 2.

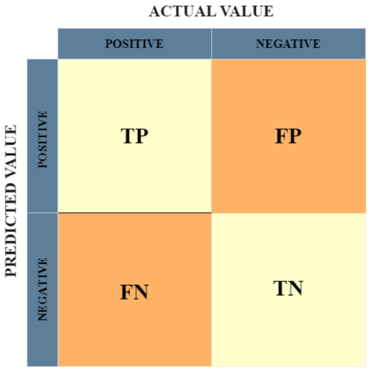

Figure 2. Confusion Matrix Diagrams

3. RESULTS AND ANALYSIS

Comparisons can be seen by comparing the highest Accuracy, Recall, and Precision among the three algorithm models tested.

3.1. Support Vector Machine (SVM)

In this study, the classification process was carried out using the Support Vector Machine (SVM) algorithm model.

Table 4. The result from Performance Model SVM

	True High Risk	True Low Risk	True Mid Risk	Class Precision
Pred. high risk	10	0	1	90.91%
Pred. low risk	1	20	8	68.87%
Pred. mid risk	4	2	3	33.33%
Class recall	66.67%	90.91%	25.00%	

From the analysis results table above, it shows the resulting SVM performance, namely the highest level of pregnancy risk occurs at a high level (high level) with a precision of 90.91%, then 90.91% of all positive predictions are correct, the remaining 9.09% are false positives. This model is also able to detect correctly with a recall result of 66.67% and the rest are said to be false negatives. The classification carried out by the model can be said to be quite good with an accuracy of 67.35%.

3.2. K-Nearest Neighbor (KNN)

The next classification process is carried out using the KNN algorithm. The following is one of the test results of the KNN algorithm.

Table 5. Result from KNN Performance Model

	True High Risk	True Low Risk	True Mid Risk	Class Precision
Pred. high risk	12	0	1	92.31%
Pred. low risk	1	17	2	85.00%
Pred. mid risk	2	5	9	56.25%
Class recall	80.00%	77.27%	75.00%	

From the results of testing using the K-NN model above, it was found that the risk level of pregnancy with the highest value occurred at a high level, with a precision of 92.31%, which means that 92.31% of all predictions of this positive model were correct at 7.69% and the rest are false positives. This model is also able to correctly detect all positive examples from the dataset with a recall obtained of 80.00% and the remaining

20.00% are said to be false negatives. The accuracy performance of the model is 77.55%. It can be said that the model is able to classify well.

3.3. Naïve Bayes Classifier (NBC)

After conducting two tests using the SVM and KNN models, the researcher conducted a third classification using the NBC model.

Table 6. Results from Performance Models NBC

	True High Risk	True Low Risk	True Mid Risk	Class Precision
Pred. high risk	10	0	1	90.91%
Pred. low risk	1	19	6	73.08%
Pred. mid risk	4	3	5	41.67%
Class recall	66.67%	86.36%	41.67%	

From the results of this third test, using the NBC model, it was found that the highest level of pregnancy risk occurs at a high level with a precision of 90.91%.then all positive predictions from this model are correct, while the remaining 9.09% are false positives. While the Recall generated was 66.67%, so this model was able to identify 66.67% of all positive examples in the dataset, while the remaining 33.33% were false negatives. And the truth of the classification results is said to be quite good with the accuracy obtained from the model testing of 69.39%.

3.5 Comparison of SVM, KNN and NBC Algorithms

What the algorithm needs to give the best value are standards and test equipment. The comparison of the 3 algorithms must be of the same standard to find out the best algorithm for comparison. This step tries to calculate the accuracy, memory, and precision values of the three algorithms[24]. A comparison of the Data Mining algorithm using the classification method between Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Naïve Bayes Classifier (NBC) is shown in Figure 2. As follows.

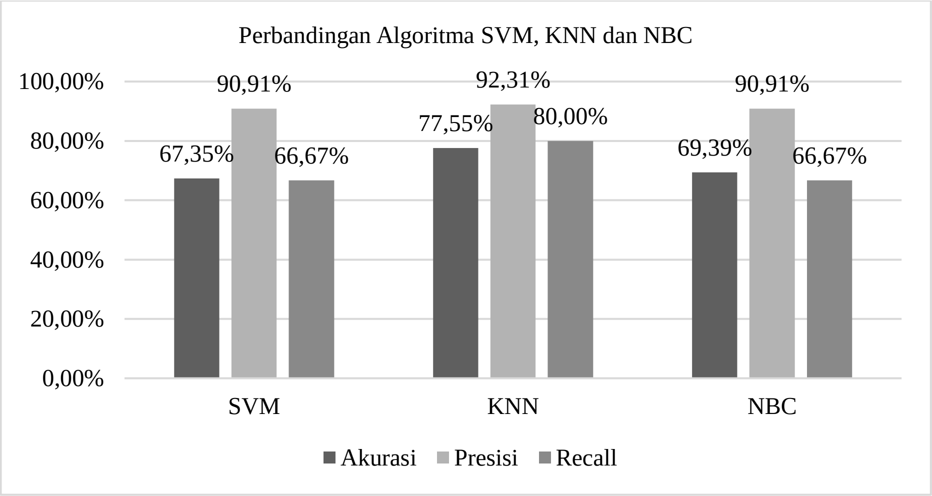

Figure 3. Bar Chart Comparison of SVM, KNN and NBC

The final results of the three classification algorithms show SVM with a precision of 90.91% and a recall of 66.67%. The accuracy of this model analysis is 67.35%. KNN with 92.31% precision and 80.00% recall. The performance accuracy of the model is 77.55%. And NBC with 90.91% precision and 66.67% recall. The accuracy of testing this model is 69.39%. There is a significant difference in accuracy for the three algorithms. From this comparison, the KNN algorithm has better accuracy than the other two algorithms. Several previous comparative studies of the KNN and NBC algorithms also produced KNN as the algorithm with the best accuracy compared to NBC with different types of datasets[25]. In previous research, namely a comparison of fingerprinting techniques. In the F-NBC dimension, 0.8627 is the harmonic mean, significantly superior to SVM (0.8085) and k-NN (0.6923). This is due to NBC's higher recovery rate (returns a higher number of correct location estimates as a fraction of all correct location estimates that must be returned) than SVM and k-NN[26].

4. CONCLUSIONS

From the results of the analysis that has been carried out, it is found that of the three algorithms, it produces a classification of pregnancy risk levels with the highest value occurring at a high level. To find out the accuracy of the data, researchers compared the three algorithms. Based on the confusion matrix namely Accuracy, Precision, and Recall, the first test using the SVM algorithm produced an accuracy of 67.35%, accuracy on KNN was 77.55% and NBC accuracy was 69.39%. It can be concluded from the results of the analysis that the KNN algorithm provides the highest accuracy of 77.55% and the lowest accuracy by SVM of 67.35%. These results state that the KNN algorithm classifies pregnancy risk level data better than the other two algorithms.

REFERENCES

[1] Wahyuni Windasari, “Wahyuni+Accept (2),” Journal of Data Science Theory and Application, vol. 01, no. 02, pp. 66–71, 2022.

[2] Y. Permatasari, U. Salamah, and R. Saptono, “Klasifikasi Risiko Bahaya Kehamilan dengan Metode Fuzzy C-Means,” Jurnal Teknologi & Informasi ITSmart, vol. 2, no. 1, p. 08, Mar. 2016, doi: 10.20961/its.v2i1.610.

[3] Q. Hasanah, A. Andrianto, and M. A. Hidayat, “Sistem Informasi Posyandu Ibu Hamil dengan Penerapan Klasifikasi Resiko Kehamilan Menggunakan Metode Naïve Bayes (Implementing Classification Risk in Posyandu System Information for Pregnant Using Naïve Bayes Method).”

[4] I. F. Nurahmadan, A. Agusta, P. A. Winarno, B. H. Sazali, Y. Thurfah, and A. Rosaliah, Perbandingan Algoritma Machine Learning Untuk Klasifikasi Denyut Jantung Janin. 2021.

[5] S. M. Dol, “Use of Classification Technique in Educational Data Mining,” in 2021 International Conference on Nascent Technologies in Engineering, ICNET 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/ICNTE51185.2021.9487739.

[6] A. Ilham Fatimah and S. Saepudin, “PENERAPAN DATA MINING DENGAN METODE APRIORI PADA PENJUALAN SEMBAKO (STUDI KASUS: GROSIR SEMBAKO LINA),” 2022. [Online]. Available: https://rekayasa.nusaputra.ac.id/index

[7] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 4, no. 2, p. 437, Apr. 2020, doi: 10.30865/mib.v4i2.2080.

[8] . Analisis Dan Penerapan, A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliyansyah, “Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi (Analysis and Application of Algorithm Support Vector Machine (SVM) in Data Mining to Support Promotional Strategies),” 2019.

[9] Sri Diantika, Windu Gata, Hiya Nalatissifa, and Mareanus Lase, “Komparasi Algoritma SVM Dan Naive Bayes Untuk Klasifikasi Kestabilan Jaringan Listrik,” JURNAL ILMIAH ELEKTRONIKA DAN KOMPUTER, vol. Vol.14, no. No.1, pp. 10–15, Oct. 2021.

[10] S. Ramadani, N. Zannah, S. Ayu, N. Nurhayati, F. Azzahra, and A. P. Windarto, “Analisis Data Mining Naive Bayes Klasifikasi Pada Kelayakan Penerima PKH,” KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer), vol. 4, no. 1, pp. 374–381, 2020, doi: 10.30865/komik.v4i1.2725.

[11] M. Gunawan, M. Zarlis, and R. Roslina, “Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu,” JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 2, p. 513, Apr. 2021, doi: 10.30865/mib.v5i2.2925.

[12] T. Prabhakar, “Classification of Signatures of Aircraft Prototypes using Support Vector Machine,” 2015. [Online]. Available: www.ijcrt.org

[13] R. Singla, B. Chambayil, A. Khosla, and J. Santosh, “Comparison of SVM and ANN for classification of eye events in EEG,” J Biomed Sci Eng, vol. 04, no. 01, pp. 62–69, 2011, doi: 10.4236/jbise.2011.41008.

[14] K. Li, H. Yu, Y. Xu, and X. Luo, “Detection of Marine Oil Spills Based on HOG Feature and SVM Classifier,” Journal of Sensors, vol. 2022, 2022, doi: 10.1155/2022/3296495.

[15] Y. Lukito and A. R. Chrismanto, “Perbandingan Metode-Metode Klasifikasi Untuk Indoor Positioning System,” 2015.

[16] T. M. Le, T. M. Vo, T. N. Pham, and S. V. T. Dao, “A Novel Wrapper-Based Feature Selection for Early Diabetes Prediction Enhanced with a Metaheuristic,” IEEE Access, vol. 9, pp. 7869–7884, 2021, doi: 10.1109/ACCESS.2020.3047942.

[17] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative performance analysis of K-nearest neighbor (KNN) algorithm and its different variants for disease prediction," Scientific Reports, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-10358-x.

[18] “A Practical Introduction to K-Nearest Neighbors Algorithm for Regression (with Python code).”

[19] S. Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, A. Nikmatul Kasanah, U. Pujianto, T. Elektro, F. Teknik, and U. Negeri Malang, “Terakreditasi SINTA Peringkat 2 Penerapan

Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," masa berlaku mulai, vol. 1, no. 3, pp. 196–201, 2017.

[20] J. Homepage, B. Delvika, S. Nurhidayarnis, P. D. Rinada, N. Abror, and A. Hidayat, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women," vol. 2, pp. 68–75, 2022.

[21] F. Handayani, D. Feddy, and S. Pribadi, "Implementasi Algoritma Naive Bayes Classifier dalam Pengklasifikasian Teks Otomatis Pengaduan dan Pelaporan Masyarakat melalui Layanan Call Center 110."

[22] J. Homepage, N. Semuel, and A. A. Pekuwali, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Pattern Recognition of Doctor's Prescription Handwriting Using the Naïve Bayes Classifier Method at Puskesmas Kambaniru," vol. 2, pp. 55–61, 2022.

[23] S. Thaufik Rizaldi, "Perbandingan Teknik Pembagian Data untuk Klasifikasi Sarana Akses Air pada Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier," 2020. [Online]. Available: https://sarpras.dikdasmen.kemdikbud.go.id

[24] Y. I. Kurniawan, "Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining," Jurnal Teknologi Informasi dan Ilmu Komputer, vol. 5, no. 4, p. 455, Oct. 2018, doi: 10.25126/jtiik.201854803.

[25] J. Homepage, P. Algoritma Klasifikasi untuk Analisis Sentimen, E. Ditendra, S. Romelah, M. Habil Arsyiddik Tanjung, and M. Sarah, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison of Classification Algorithms for Sentiment Analysis of Islam Nusantara in Indonesia," vol. 2, pp. 71–77, 2022.

[26] Ieee and Ieee, 2011 International Conference on ICT Convergence.