Journal of Information Systems and Informatics
Vol. 5, No. 4, December 2023 e-ISSN: **2656-4882** p-ISSN: **2656-5935**
DOI: 10.51519/journalisi.v5i4.568 Published By DRPM-UBD

Empowering Pregnancy Risk Assessment: A Web-Based Classification Framework with K-Means Clustering Enhanced Models

Bernard Pramata Wongso[1], Monika Evelin Johan[2], Melissa Indah Fianty[3]

[1,2,3]Information System Department, Universitas Multimedia Nusantara, Tangerang, Indonesia
Email: [1]bernard.pratama@student.umn.ac.id, [2]monika.evelin@umn.ac.id, [3]melissa.indah@umn.ac.id

Abstract

This study aims to determine whether there is an increase in accuracy results for predicting pregnancy risk with a classification algorithm that goes through and without going through the clustering stage. After that, compare which classification algorithm gets the best improvement. This study uses the K-Means clustering approach, as well as the SVM, Naive Bayes, and K-Nearest Neighbor (KNN) classification algorithms. The pregnancy risk dataset used comes from the UCI Machine Learning Repository. Evaluation metrics used include accuracy, precision, recall, and F1-score. The results of the study revealed that the K-Means model with KNN provided the highest performance compared to the other two, with an accuracy of 79.53% and an average F1-score of 0.8. The implementation of K-Means resulted in an increase in accuracy of 0.4%, 1.57%, and 2.76% on KNN, SVM, and Naive Bayes respectively, which confirms the impact of clustering in improving classification performance. The resulting model can be used in real-time via a website built using the Flask API, and offers tools that can help health practitioners to plan treatments effectively and minimize the risk of pregnancy.

Keywords: Classification, Clustering, Pregnancy Risk, Web-based

1. INTRODUCTION

Pregnancy is a process that involves significant physical and emotional changes for a mother [1]. During pregnancy, there are several factors that can increase the risk of complications for both the mother and the fetus [2]. These factors include the mother's age being too young or too old, obesity, history of diseases such as diabetes or hypertension, and history of abnormal pregnancies [3]. Older maternal age (above 35 years) is associated with an increased risk of chromosomal abnormalities in the fetus, such as Down syndrome [4]. In addition, research has found that obesity during pregnancy can increase the risk of pregnancy complications, including gestational diabetes and high blood pressure [5]. According to data from 2018, about 76% of maternal deaths occurred during

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

childbirth and postpartum, with 24% occurring during pregnancy, 36% during childbirth, and 40% postpartum [6]. These percentages indicate the importance of developing an early detection system that is related to the level of pregnancy risk. To assess the level of pregnancy risk for a mother, data on the physical and emotional condition of the mother during pregnancy are usually collected [7]. This data is then processed using clustering and classification methods that help identify the risk level for each individual. The dataset used to assess the pregnancy risk level in this study was obtained from the UCI Machine Learning repository.

Previous studies on identifying the level of pregnancy risk for a mother using classification methods have been commonly done, such as in the journal article "Model for Predicting Risk Levels in Maternal Healthcare" [8]. This study focuses on developing a predictive model to assess the level of risk in maternal health care. The main objective is to identify risk factors that affect maternal health during pregnancy and childbirth, to provide better care, and reduce maternal mortality. The available data is then analyzed using statistical and mathematical modeling techniques to identify factors related to maternal health risk. The data includes vital parameters such as blood pressure, heart rate, body temperature, and so on [8]. This study only involved the use of classification algorithms, while the ongoing research involves the use of clustering algorithms (clustering) whose results are used in classification algorithms.

The research in the journal article used KNN, Neural Networks, Random Forest, and AdaBoost algorithms to build a predictive model in assessing the level of pregnancy risk. This research produced a predictive model to assess the level of pregnancy risk using classification techniques. This model can help doctors predict pregnancy risk and provide appropriate care recommendations. The research findings show that the model developed can predict the level of risk in maternal health care with relatively good accuracy, namely 86% for the best model, namely the AdaBoost algorithm, while the algorithm with the lowest performance is the KNN algorithm with an accuracy of 64%. Some of the main risk factors identified include maternal age, previous birth history, nutritional status, and the presence of medical complications history [8].

The K-Means algorithm is one of the clustering algorithms used to group data based on the similarity of their features [9]. After the clustering process is completed, the next step is to perform classification on each group formed by the K-Means algorithm [10]. Classification algorithms are needed to interpret the clustering results because clustering only groups data into clusters based on their similarity, without providing labels or category information for each cluster [11]. Classification algorithms help identify patterns or characteristics of each cluster and provide labels or categories to the clusters based on the learned patterns [12]. By using classification algorithms, we can classify new data into existing clusters or identify new clusters that were previously unknown [13].

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

To perform this classification, there are several classification algorithms that can be used, such as Naive Bayes, SVM (Support Vector Machine), and KNN (K-Nearest Neighbors). Each of these algorithms has its own advantages and disadvantages. Naive Bayes, SVM, and KNN are three classification algorithms that are commonly used in data analysis. These algorithms can also be used to classify the results of clustering with high accuracy and effectiveness [14]. Therefore, these three algorithms are used in the comparison of classification algorithms to determine the most suitable algorithm in determining the level of pregnancy risk for a mother based on the clustering results. The comparison analysis is done by measuring the accuracy, precision, recall, and f1-score of each algorithm [15]. Thus, it will be determined which classification algorithm is the best in determining the level of pregnancy risk for a mother based on the clustering results. Accuracy measures how well an algorithm can classify data correctly overall, while precision measures how well an algorithm can identify positive classes correctly [16]. Recall measures how well an algorithm can find all relevant positive classes, and f1-score is the harmonic mean of precision and recall, providing a more comprehensive understanding of the algorithm's performance [17].

Some studies have used clustering techniques to improve classification accuracy in various fields [18]. One of these studies focuses on using semi-supervised clustering techniques to improve the accuracy of DDoS attack classification in networks. The research results show that semi-supervised clustering techniques can improve classification accuracy compared to traditional supervised learning methods [19]. Research on pregnancy risk has been done using risk classification modified by the World Health Organization (mWHO) on 2742 pregnant women in developed and developing countries. The research results show that the mWHO classification only has moderate performance in distinguishing women who experience risk and those who do not experience risk during pregnancy with an accuracy level of 71.1% [20]. Therefore, further research is needed to develop and improve the performance of the model created.

In this study, a predictive model for pregnancy risk level will be created using sequential clustering and classification techniques, which are useful in determining whether a mother is at low, medium, or high risk during pregnancy. The difference from previous research [8] lies in the use of clustering methods, especially the K-Means algorithm, before performing classification using Naive Bayes, SVM (Support Vector Machine), and KNN (K-Nearest Neighbors) algorithms in determining the level of pregnancy risk for a mother based on the clustering results.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2. METHODS

The research framework used is based on previous research titled "Model for Predicting Risk Levels in Maternal Healthcare" [8]. The stages of research process as shown in Figure 1.

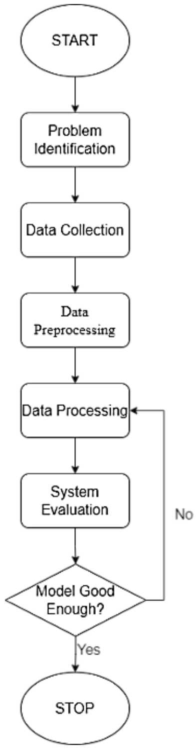


Figure 1. The Research Framework

The following is the detail of the research methodology used.

1. Problem Identification, the main objective of this research is to develop an accurate risk level prediction system using the K-means clustering approach and the proposed classification approach. This is expected to help doctors in determining the level of pregnancy risk for a mother. The object of this research is the Maternal Health Risk Dataset from the UCI Machine Learning repository, consisting of 1,015 medical histories of randomly selected pregnant mothers. This dataset includes information such as maternal age, upper and lower blood pressure values, blood glucose levels, body temperature, normal heart rate, and pregnancy risk level. By analyzing the data using machine learning techniques, researchers can identify patterns and provide recommendations to improve pregnancy monitoring and intervention for pregnant mothers with high risk, which can ultimately increase the safety level of pregnant mothers and their babies.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

2. Data Collection, the data used for analysis comes from the UCI Machine Learning Repository website and consists of medical records from a hospital in Bangladesh. This dataset consists of 1,015 randomly selected pregnant women who have undergone pregnancy examinations at the same hospital. The data includes various parameters such as age, blood pressure, blood glucose levels, body temperature, heart rate, and pregnancy risk level, and was collected between 2000 and 2010, with access and download done on January 9, 2023. The data was accessed through the website https://archive.ics.uci.edu/ by searching for maternal health dataset. The data in Excel format was downloaded from the website and stored on the storage device used. After that, the data is ready to be used for the required analysis purposes. The study involves two variables, specifically the dependent variable and the independent variable.
   a) In this study, the independent variables include the mother's age, represented as 'Age' in the dataset, the upper value of blood pressure in mmHg, represented as 'SystolicBP' in the dataset, the lower value of blood pressure in mmHg, represented as 'DiastolicBP' in the dataset, the glucose level in molar concentration mmol/L, represented as 'BS' in the dataset, body temperature, represented as 'BodyTemp' in the dataset, and the normal heart rate in beats per minute, represented as 'HeartRate' in the dataset.
   b) In this study, the dependent variable is the pregnancy risk level, named 'RiskLevel', in the dataset used.
3. Data Preprocessing, the collected data was then processed using data preprocessing techniques such as handling missing values, selecting relevant features, and transforming data into a suitable format for further processing. In this research, the preprocessing stage includes data exploration and data cleaning processes.
4. Data Processing, the processed data was then analyzed using the K-means clustering method to group patients into similar clusters. Next, the proposed classification approach was used to predict the level of pregnancy risk for a mother based on other variables or attributes that are used in determining the level of pregnancy risk. In this research, the data processing stage involves dividing the data into train and test sets, creating a classification model without clustering, and creating a classification model with clustering. For a detailed modeling workflow, see Figure 2. The modeling workflow used is based on previous research titled “Clustering based semi-supervised machine learning for DDoS attack classification” [19].

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

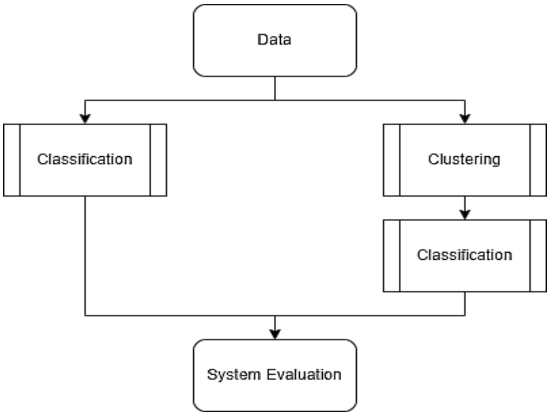


Figure 2. Flowchart of the Used Modeling Process

Two types of model development were done. The first type involved direct classification of data using SVM, Naïve Bayes, and KNN algorithms. The second type involved clustering data using the K-Means algorithm, followed by classification using SVM, Naïve Bayes, and KNN algorithms. This modeling process resulted in six models, three from the classification stage only, and three from the clustering stage followed by the classification stage.

5. System Evaluation, the system developed was evaluated using several metrics such as accuracy, precision, recall, and F1 score. This system evaluation is important to ensure that the system developed is accurate and reliable in predicting the level of pregnancy risk for patients. If the model formed is not satisfactory, adjustments will be made to the data processing section. Overall, the development of a pregnancy risk level prediction system using the K-means clustering approach and the proposed classification approach can help doctors determine the level of pregnancy risk more accurately and reliably.

Google Colab is a better choice than Jupyter Notebook, RStudio, and Microsoft Azure Machine Learning Studio for several reasons as data analysis tools. Google Colab is free and easy to access, does not require strong local resources, provides free access to GPU/TPU, integrates well with Google Drive and Docs, and supports interactive and collaborative data exploration [21].

3. RESULTS AND DISCUSSION

3.1. Problem Identification

Pregnancy risk determination involves factors such as maternal age, health history, nutritional status, and lifestyle habits, which can contribute to complications such as preeclampsia, gestational diabetes, and premature birth. The application of machine learning models, through clustering and classification techniques, offers

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

good potential in helping health practitioners identify and minimize potential risks accurately, contribute to public health policy research, and improve the overall quality of health care by providing more complex and accurate prediction models based on diverse data sources.

3.2. Data Collection

To collect the Maternal Health Risk Dataset from secondary data, users must visit the UCI Machine Learning Repository website and use the search function to find datasets related to maternal health and evaluate the description and attributes of each data to find the most suitable one. After the most relevant dataset is identified, it must be downloaded in a suitable format such as CSV or Excel, followed by a manual inspection to understand the characteristics and features of the data.

3.3. Data Preprocessing

The results obtained from creating the Heatmap can be seen in Figure 3, which is the Heatmap of Correlation Between Variables in the Dataframe. The results show that the highest correlation between the RiskLevel variable and other variables is found in the correlation between RiskLevel and the BS variable, with a value of 0.57. On the other hand, the lowest correlation between the RiskLevel variable and other variables is found in the correlation between RiskLevel and the BodyTemp variable, with a value of 0.16.

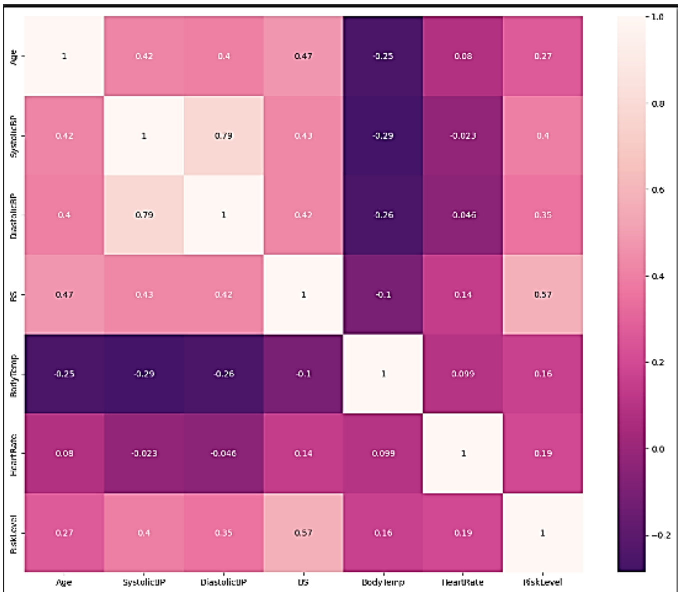


Figure 3. Correlation Heatmap Among Variables in a Dataframe

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

Checking for empty data is done to determine whether there is any empty data or if there are still missing values in the data used. Because it is known that there is no more empty data, it continues to the data processing stage.

3.4. Data Processing

In this stage, the dataset is divided into train and test data using the 'train_test_split' function from the sklearn library, with a test size of 0.25 determined to produce the best model accuracy, and the results are stored in 'X_train', 'X_test', 'y_train', and 'y_test'. After splitting the data into train and test sets, it is important to use standard scaler to adjust these variables, which eliminates scale differences between different variables, which are important for algorithms such as SVM that will be used in this study. GridSearchCV is used to find the best combination of parameters given in the 'parameters' variable. This process is followed by the Classification Stage without Clustering beforehand. The classification results obtained from the SVM algorithm can be seen in Figure 4, with an accuracy of 74% achieved by the SVM model.

Accuracy: 74.01574803149606 %

	precision	recall	f1-score	support
high risk	0.83	0.88	0.85	57
low risk	0.74	0.69	0.71	102
mid risk	0.68	0.72	0.70	95
accuracy			0.74	254
macro avg	0.75	0.76	0.76	254
weighted avg	0.74	0.74	0.74	254

Figure 4. The Results of SVM Model Classification

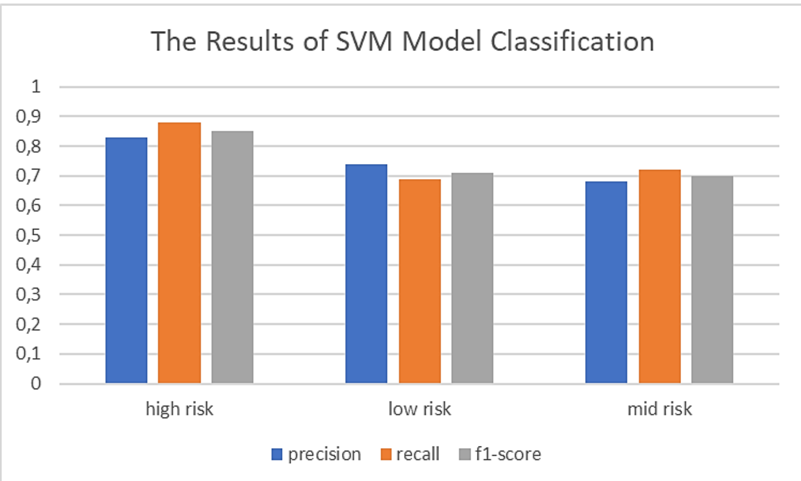


Figure 5. Chart SVM Model Classification Result

The classification results obtained from the Naïve Bayes algorithm can be seen in Figure 6, with an accuracy of 58% achieved by the Naïve Bayes model.

Accuracy: 57.874015748031496 %

	precision	recall	f1-score	support
high risk	0.76	0.61	0.68	57
low risk	0.53	0.92	0.67	102
mid risk	0.58	0.19	0.29	95
accuracy			0.58	254
macro avg	0.62	0.58	0.55	254
weighted avg	0.60	0.58	0.53	254

Figure 6. The Results of Naïve Bayes Model Classification

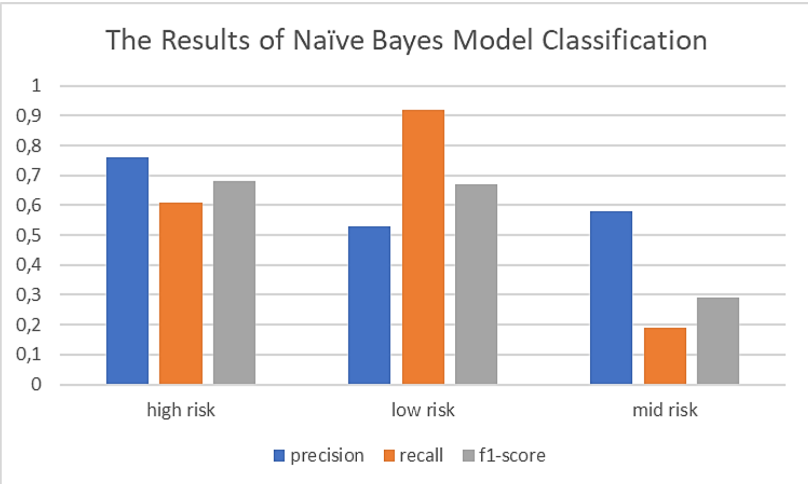


Figure 7. Chart Naïve Bayes Model Classification Result

The classification results obtained from the KNN algorithm can be seen in Figure 8, with an accuracy of 79% achieved by the KNN model.

Accuracy: 79.13385826771653 %

	precision	recall	f1-score	support
high risk	0.83	0.86	0.84	57
low risk	0.81	0.75	0.78	102
mid risk	0.75	0.79	0.77	95
accuracy			0.79	254
macro avg	0.80	0.80	0.80	254
weighted avg	0.79	0.79	0.79	254

Figure 8. The Results of KNN Model Classification

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

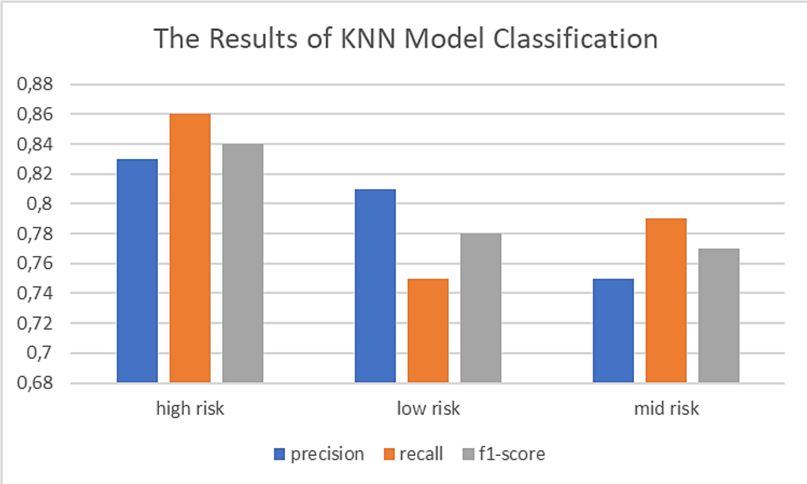


Figure 9. Chart KNN Model Classification Result

The process continues with the clustering stage, followed by classification. Based on the obtained scores from silhouette method in Figure 10, the optimal number of clusters is 2, with the highest silhouette score. Based on these results, the best number of clusters chosen is only 2.

```
For n_clusters=2, the silhouette score is 0.43300932071030607
For n_clusters=3, the silhouette score is 0.374703285906805
For n_clusters=4, the silhouette score is 0.39976707588496746
For n_clusters=5, the silhouette score is 0.3671232054031672
```

Figure 10. The Result Score of Silhouette Method for Determining the Optimal Number of Clusters

The classification results obtained from the SVM algorithm after clustering implementation can be seen in Figure 11, with an accuracy of 76% achieved by the SVM model.

```
Accuracy: 75.59055118110236 %
              precision    recall  f1-score   support

   high risk       0.83      0.88      0.85        57
    low risk       0.77      0.71      0.73       102
    mid risk       0.70      0.74      0.72        95

    accuracy                           0.76       254
   macro avg       0.77      0.77      0.77       254
weighted avg       0.76      0.76      0.76       254
```

Figure 11. The Classification Results of the SVM Model after Clustering Implementation

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

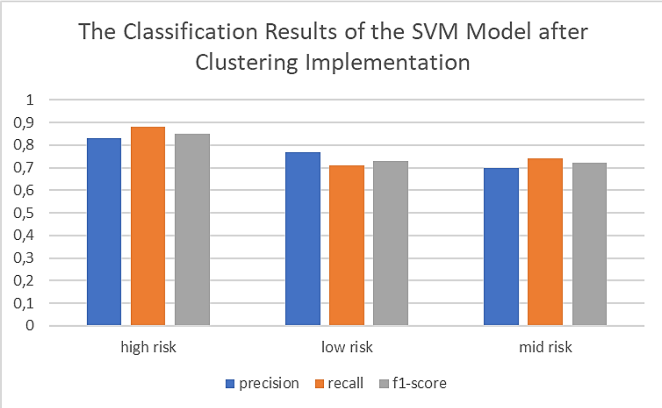
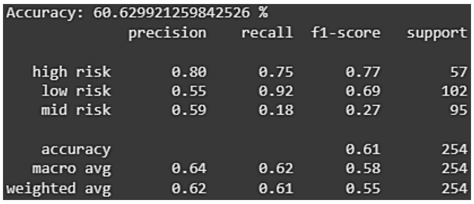


Figure 12. Chart SVM Model Classification Result after Clustering Implementation

The classification results obtained from the Naïve Bayes algorithm after clustering implementation can be seen in Figure 13, with an accuracy of 61% achieved by the Naïve Bayes model.



Accuracy: 60.629921259842526 %

	precision	recall	f1-score	support
high risk	0.80	0.75	0.77	57
low risk	0.55	0.92	0.69	102
mid risk	0.59	0.18	0.27	95
accuracy			0.61	254
macro avg	0.64	0.62	0.58	254
weighted avg	0.62	0.61	0.55	254

Figure 13. The Classification Results of the Naïve Bayes Model after Implementing Clustering

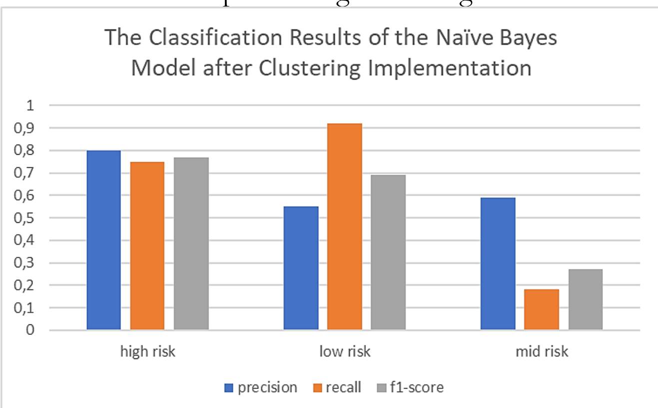


Figure 14. Naïve Bayes Model Classification Result after Clustering Implementation

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

The classification results obtained from the KNN algorithm after clustering implementation can be seen in Figure 15, with an accuracy of 80% achieved by the KNN model.

Accuracy: 79.52755905511812 %

	precision	recall	f1-score	support
high risk	0.83	0.88	0.85	57
low risk	0.81	0.75	0.78	102
mid risk	0.76	0.79	0.77	95
accuracy			0.80	254
macro avg	0.80	0.81	0.80	254
weighted avg	0.80	0.80	0.79	254

Figure 15. The Classification Results of the KNN Model after Implementing Clustering

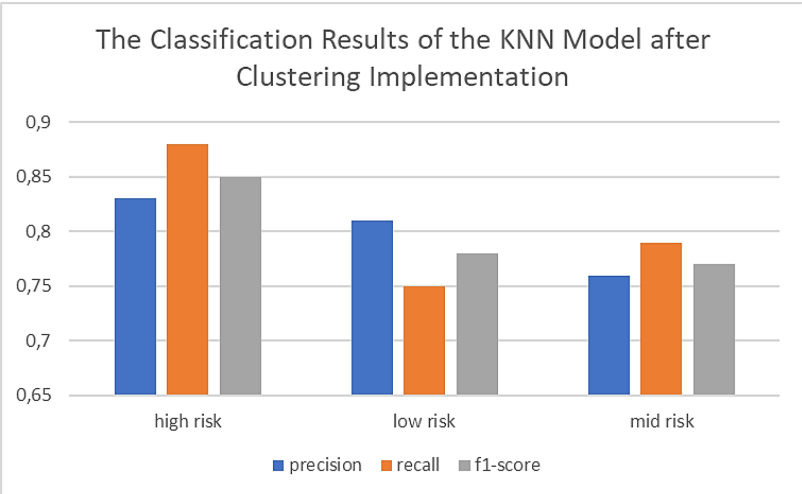


Figure 16. Chart KNN Model Classification Result after Clustering Implementation

3.5. System Evaluation

To develop a system that can predict the level of pregnancy risk for a mother, six machine learning models have been developed and tested. The first three models use SVM, Naive Bayes, and KNN classification techniques directly, while the other three models integrate the K-means clustering approach before applying the same classification algorithms (SVM, Naive Bayes, and KNN). The evaluation of these six models focuses on the performance differences between the methods used.

The KNN model with clustering proved to be the best choice, achieving an accuracy of 79.53% and an average F1 score of 0.79. This model outperformed

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

the KNN model without clustering and the SVM model with clustering, both of which also showed good performance. The KNN model without clustering achieved an accuracy of 79.13% and an average F1 score of 0.79, while the SVM model with clustering achieved an accuracy of 75.59% and an average F1 score of 0.76.

At the same time, the Naive Bayes model, both with and without clustering, showed lower performance compared to other models. The Naive Bayes model without clustering only achieved an accuracy of 57.87% and an average F1 score of 0.53, while the Naive Bayes model with clustering achieved an accuracy of 60.63% and an average F1 score of 0.55. The overall performance of the prediction models can be seen in Table 1 and Table 2.

Table 1. Performance of The Prediction Model without K-Means Clustering

	Support Vector Machine	Naïve Bayes	K-Nearest Neighbors
Accuracy	0.7402	0.5787	0.7913
Precision	0.75	0.62	0.8
Recall	0.76	0.58	0.8
F1-score	0.76	0.55	0.8

Table 2. Performance of The Prediction Model with K-Means Clustering

	Support Vector Machine	Naïve Bayes	K-Nearest Neighbors
Accuracy	0.7559	0.6063	0.7953
Precision	0.77	0.64	0.8
Recall	0.77	0.62	0.81
F1-score	0.77	0.58	0.8

```
Nilai F: 0.03350597330921975
Nilai p: 0.8636651735997966
Tidak terdapat perbedaan signifikan antara kedua kelompok model
```

Figure 17. The ANOVA Test Result

The results of the ANOVA test can be seen in Figure 17, which shows that there is no significant difference between the two groups of models, as indicated by the p-value of 0.8637, which is larger than 0.05. This means that we fail to reject the null hypothesis that there is no difference between the groups. The small F value, which is 0.0335, also indicates that the variation between the groups is not larger than the variation within the groups. Therefore, statistically, the two groups of models are not significantly different, and both have almost similar performance.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

Based on this evaluation, it is suggested to use the KNN model with clustering to predict the level of pregnancy risk for a mother. This decision is based on the superior performance and the ability of the model to provide more accurate prediction results. In addition, the integration of the K-means clustering method before classification helps improve the performance of the model, especially in the case of KNN and SVM. The performance results of the KNN algorithm obtained with or without K-means clustering in this study can exceed the performance of previous research, which only achieved 64%. This increase is caused by the use of GridSearchCV to find the best parameters for the algorithm used, thus allowing the performance results obtained for the same algorithm to exceed the results of previous research using KNN. For a comparison of the performance of the best research model with previous research, see Table 3.

Table 3. Comparison of Performance Between the Best Research Model and Previous Research

Penelitian	Accuracy	F1-score	Precision	Recall
Application of Classification Algorithm Comparison Results in Implementation of Web-based K-Means Algorithm Clustering Results to Assess Pregnancy Risk Levels	0.7953	0.7900	0.8000	0.8100
Model for Predicting Risk Levels in Maternal Healthcare [8]	0.6400	0.6400	0.6400	0.6400

3.6. Deployment

To implement the previously trained machine learning model into a web-based HTML application using Flask API in Python, several steps need to be followed. The trained model is saved as a .pkl file using libraries such as 'pickle' or 'joblib', and Flask is used to import the model into the application. The API is built with Flask to serve as a connection between the web application and the machine learning model, allowing the application to access and use the model. On the web-

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

based application, Flask creates API routes to receive user data, process it using the model, and return predictions or results. The API is connected to the HTML web application using technologies such as AJAX or Fetch API, allowing smooth communication without page reload. The components that are exported include the k-means clustering model, the scaler used for preprocessing, and the best classification model (such as k-nearest neighbors). Figure 18 shows the resulting HTML interface.

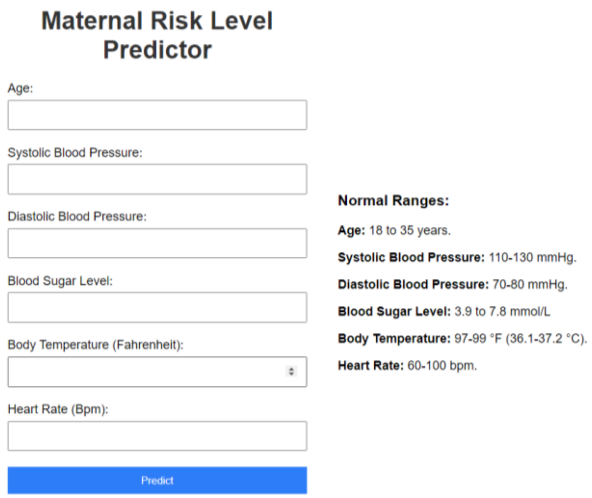


Figure 18. HTML Initial Display

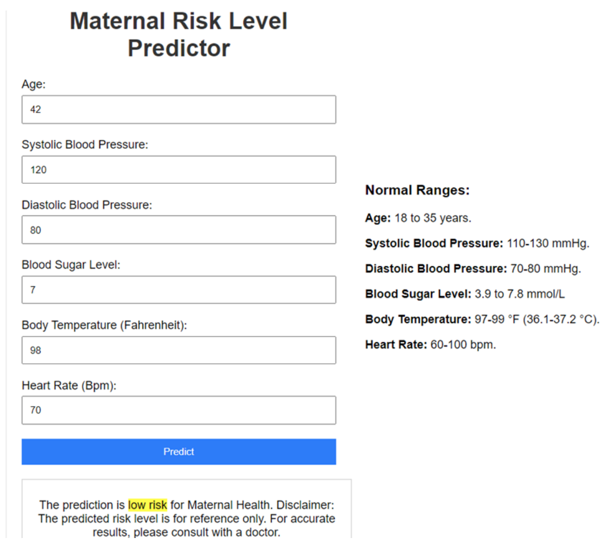


Figure 19. The Prediction Result of Low Risk in Pregnancy Risk

Journal of Information Systems and Informatics
Vol. 5, No. 4, December 2023

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

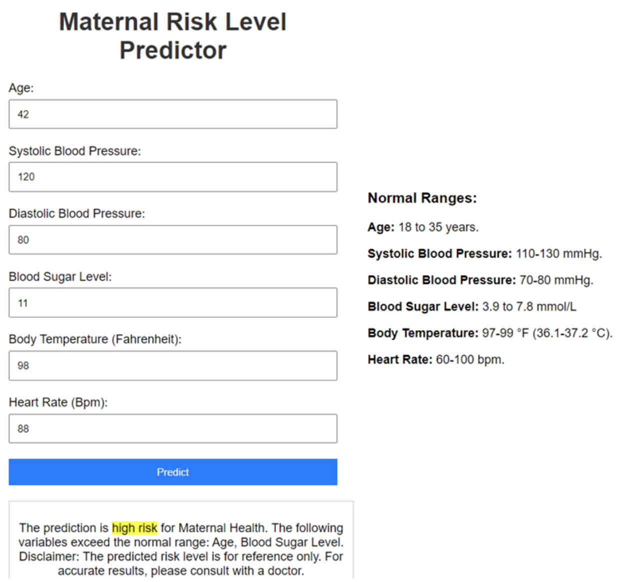


Figure 20. The Prediction Result of High Risk in Pregnancy Risk

To make the desired prediction, users can enter data in the available columns by filling in the appropriate numbers for the metrics provided. After all columns are successfully filled, users can press the prediction button, and on the same page, the level of pregnancy risk based on the previously entered input will be displayed. Examples of predicted results can be seen in Figure 19 and Figure 20.

This study developed and integrated a machine learning model for pregnancy risk assessment using KNN, SVM, and Naïve Bayes, and enhanced it with K-Means clustering. The results showed that KNN was the best classifier, and that K-Means clustering improved the accuracy and F1 score of all the classifiers. The web application of the model provided a practical and user-friendly tool for health practitioners to obtain real-time predictions and monitor the pregnancy status of their patients. The study discussed the implications, limitations, and recommendations of the model for the field of health care and machine learning.

4. CONCLUSION

A predictive model has been developed to assess the level of pregnancy risk using three main classification algorithms, Support Vector Machine (SVM), Naïve Bayes, and K-Nearest Neighbors (KNN). This model incorporates the K-Means clustering algorithm, which improves its performance. The KNN model achieved an accuracy of 79.53% and an average F1 score of 0.8, surpassing other models. By integrating this model into a Flask web application, health practitioners can utilize pregnancy risk predictions in real-time, facilitate pregnancy monitoring, and

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

improve the quality of patient care provided to pregnant mothers. This implementation offers a simple yet functional HTML interface, ensuring ease of use and adaptability for users from diverse backgrounds. The use of K-Means clustering technique before classification resulted in an accuracy increase of 0.4% for the KNN model, while SVM and Naïve Bayes achieved higher increases, namely 1.57% and 2.76%.

REFERENCES

[1] H. B. Shulman, D. V. D'Angelo, L. Harrison, R. A. Smith and L. Warner, "The Pregnancy Risk Assessment Monitoring System (PRAMS): Overview of Design and Methodology," *American Journal of Public Health,* vol. CVIII, no. 10, pp. 1305-1313, 2018.

[2] G. Stephen, M. Mgongo, T. H. Hashim, J. Katanga, B. Stray-Pedersen and S. E. Msuya, "Anaemia in Pregnancy: Prevalence, Risk Factors, and Adverse Perinatal Outcomes in Northern Tanzania," *Anemia,* pp. 1-9, 2018.

[3] S. Sangin and C. Phonkusol, "Perception of Pregnancy Risk and Related Obstetric Factors among Women of Advanced Maternal Age," *PRIJNR,* vol. XXV, no. 3, pp. 494-500, 2021.

[4] S. Lou, K. Carstensen, O. B. Petersen, C. P. Nielsen, L. Hvidman, M. R. Lanther and I. Vogel, "Termination of Pregnancy following a Prenatal Diagnosis of Down Syndrome: A Qualitative Study of the Decision-Making Process of Pregnant Couples.," *Acta Obstetricia Et Gynecologica Scandinavica,* vol. XCVII, no. 10, pp. 1228-1236, 2018.

[5] J. A. Grieger, M. J. Hutchesson, S. D. Cooray, M. B. Khomami, S. Zaman, L. Segan, H. Teede and L. J. Moran, "A review of maternal overweight and obesity and its impact on cardiometabolic outcomes during pregnancy and postpartum," *Therapeutic Advances in Reproductive Health,* vol. XV, pp. 1-16, 2021.

[6] Rokom, "Kemenkes Perkuat Upaya Penyelamatan Ibu dan Bayi," Kemenkes, 15 September 2021. [Online]. Available: https://sehatnegeriku.kemkes.go.id/baca/umum/20210914/3738491/kemenkes-perkuat-upaya-penyelamatan-ibu-dan-bayi/. [Accessed 25 May 2023].

[7] S. Rajbanshi, M. N. Norhayati and N. H. Nik Hazlina, "High-risk pregnancies and their association with severe maternal morbidity in Nepal: A prospective cohort study," *PLoS ONE,* vol. XV, no. 12, pp. 1-14, 2020.

[8] R. S. Kambli and Nirmala, "Model for Predicting Risk Levels in Maternal Healthcare," *IJARIIE,* vol. VIII, no. 6, pp. 1633-1637, 2022.

[9] A. Singh, J. C. Mehta, D. Anand, P. Nath, B. Pandey and A. Khamparia, "An intelligent hybrid approach for hepatitis disease diagnosis: Combining

p-ISSN: 2656-5935 http://journal-isi.org/index.php/isi e-ISSN: 2656-4882

enhanced k-means clustering and improved ensemble learning," *Expert Systems,* vol. XXXVIII, no. 1, pp. 1-13, 2021.

[10] D. Zhao, X. Hu, S. Xiong, J. Tian, J. Xiang, J. Zhou and H. Li, "K-Means Clustering and KNN Classification Based on Negative Databases," *Applied Soft Computing,* vol. CX, pp. 1-15, 2021.

[11] L. M. Abualigah, A. T. Khader and E. S. Hanandeh, "A Combination of Objective Functions and Hybrid Krill Herd Algorithm for Text Document Clustering Analysis," *Engineering Applications of Artificial Intelligence,* vol. LXXIII, pp. 111-125, 2018.

[12] A. W. Syaputri, E. Irwandi and M. Mustakim, "NAÏVE BAYES ALGORITHM FOR CLASSIFICATION OF STUDENT MAJOR'S SPECIALIZATION," *Journal of Intelligent Computing and Health Informatics,* vol. I, no. 1, pp. 1-19, 2020.

[13] L. Wang, Y. Zhang and Y. Rong, "Research on the Classification Algorithm for Cluster Analysis," *In 2019 5th International Conference on Control, Automation and Robotics (ICCAR),* pp. 469-474, 2019.

[14] A. C. Khotimah and E. Utami, "Comparison Naïve Bayes Classifier, K-Nearest Neighbor And Support Vector Machine in The Classification of Individual on Twitter Account," *Jurnal Teknik Informatika (JUTIF),* vol. III, pp. 673-680, 2022.

[15] K. L. Kohsasih and Z. Situmorang, "Analisis Perbandingan Algoritma C4.5 Dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular," *JURNAL INFORMATIKA,* vol. IX, no. 1, pp. 13-17, 2022.

[16] F. Jiang, Y. Jiang and H. Zhi, "Evaluation metrics for classification models," *Computational and mathematical methods in medicine,* pp. 1-7, 2019.

[17] S. Orozco-Arias, J. S. Piña, R. Tabares-Soto, L. F. Castillo-Ossa, R. Guyot and G. Isaza, "Measuring Performance Metrics of Machine Learning Algorithms for Detecting and Classifying Transposable Elements," *Processes,* vol. VIII, no. 6, pp. 10-18, 2020.

[18] C. K. P. Rukma and Alamsyah, "Increase Accuracy of Naïve Bayes Classifier Algorithm with K-Means Clustering for Prediction of Potential Blood Donors," *Journal of Advances in Information Systems and Technology,* vol. IV, no. 1, pp. 42-49, 2022.

[19] M. &. Z. S. M. A. Aamir, "Clustering based semi-supervised machine learning for DDoS attack classification," *Journal of King Saud University - Computer and Information Sciences,* vol. XXXI, no. 3, pp. 377-384, 2019.

[20] I. M. v. Hagen, E. Boersma, M. R. Johnson, S. A. Thorne, W. A. Parsonage, P. E. Subías, A. Leśniak-Sobelga, O. Irtyuga, K. A. Sorour, N. Taha, A. P. Maggioni, R. Hall and J. W. Roos-Hesselink, "Global cardiac risk assessment in the Registry Of Pregnancy And Cardiac disease: results of a registry from

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

the European Society of Cardiology," *European Journal of Heart Failure,* vol. XVIII, pp. 523-533, 2016.

[21] W. Vallejo, C. Díaz-Uribe and C. Fajardo, "Google Colab and Virtual Simulations: Practical e-Learning Tools to Support thes Teaching of Thermodynamics and to Introduce Coding to Students," *ACS OMEGA,* vol. VII, no. 8, pp. 7421-7429, 2022.