JOURNAL OF INFORMATICS DEVELOPMENT

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

# Detection of Diabetes in Pregnant Women Using Machine Learning as an Effort Towards Golden Indonesia 2045

Agung Muliawan[1*], Muhamat Abdul Rohim[2], Difari Afreyna Fauziah[3], Hamzah Fansuri Yusuf[4]

Software Engineering, Faculty Science Techonology and Industry, Institute Technology and Sciene Mandala, Indonesia[1]
System and Technology Information, Faculty Science Techonology and Industry, Institute Technology and Sciene Mandala, Indonesia [2,3]
Management, Faculty Economics and Business, Institute Technology and Sciene Mandala, Indonesia[4]

Corresponding Author: Agung Muliawan (agung.muliawan@itsm.ac.id)

ARTICLE INFO

Date of entry:
*24 April 2024*
Revision Date:
*01 October 2024*
Date Received:
*31 October 2024*

ABSTRACT

One of the goals of the Golden Indonesia 2045 program is to utilize health technology to enhance public health, with diabetes being a prominent concern. This research aims to employ ensemble classifier optimization techniques in machine learning for the early detection of diabetes among pregnant women. The study uses physiological data, including variables such as glucose levels, number of pregnancies, skin thickness, blood pressure, insulin levels, body weight, family history, and age. By combining multiple models, ensemble classifiers can enhance prediction accuracy, stability, and overall model performance. This research utilizes an open Kaggle dataset on pregnant women to train and test machine learning models, specifically Support Vector Machine (SVM) and Deep Learning, incorporating ensemble techniques such as bagging and boosting. Experimental results indicate that the ensemble classifier approach significantly enhances diabetes classification, with SVM using bagging achieving the highest accuracy at 76.95%. These findings suggest that ensemble classifier methods could be a valuable tool for early diabetes detection, providing timely intervention and improved risk management during pregnancy, which supports the objectives of improving public health under the Golden Indonesia 2045 initiative.

Keywords: Diabetes, Machine learning, Ensamble Classifer, SVM, Deep Learning



Cite this as: Muliawan, A., Rohim, M. A., Fauziah, D. A., & Yusuf, H. F. (2024). Detection of Diabetes in Pregnant Women Using Machine Learning as an Effort Towards Golden Indonesia 2045. *Journal of Informatics Development*, *3*(1), 11–17. https://doi.org/10.30741/jid.v3i1.1418

## INTRODUCTION

Diabetes is a chronic illness characterized by elevated levels of blood glucose, accompanied by disturbed metabolism of fats and proteins (Roglic, 2016). According to the International Diabetes Organization (IDF) in 2019 at least 463 million people aged around 20-79 years in the world had diabetes, equivalent to a prevalence rate of 9.3% of the world's population of the same age. In 2021, there will be a 16% increase in people with diabetes to 537 million (Syamsudin et al., 2024). Apart from causing premature death, diabetes also causes several other diseases such as blindness, heart

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

disease, and kidney failure. According to the IDF, people with diabetes around the age of 20-79 years, some of these countries are China, India and the United States with the highest number of cases in the world. Indonesia is in the top 10 category of countries with the most sufferers. Diabetics in Indonesia reached 10.7 million which made Indonesia number 7 in the world and became the Southeast Asian country with the most diabetics.

The development of technology today can help humans to obtain information and predict the disease (Wiranto et al., 2023). According to the World Health of Organization (WHO), Digital Health is the use of digital, mobile and wireless technologies to support the achievement of health goals (Fauziah et al., 2023). One form of Digital Health is a digital-based application in the health sector that provides interaction space through media with an internet connection. This research is one of the implementations of Digital Health, which uses a machine learning approach through classification techniques to predict diabetes (Syamsudin et al., 2024).

With the rapid development of technology, diabetes can be detected early using Data Mining approach classification. Data mining is the process of collecting and processing data that aims to extract important information or a pattern in the data. The extraction process can be done with machine learning algorithms, besides that the learning system is able to make human work more efficient (Sabilirrasyad et al., 2024). Machine learning (ML) is used to teach machines how to handle the data more efficiently (Mahesh, 2020). The role of machine learning has penetrated in various fields in human activities such as speech recognition (Tahseen Ali et al., 2021). In addition, ML also helps human roles in the medical field such as early detection of cancer (Mazlan et al., 2021) and diabetes (Chauhan et al., 2021).

Previous studies have explored classification and regression methods for predicting bank bankruptcy, with a focus on the use of machine learning algorithms to detect early signs of financial distress. These studies have shown that models such as Decision Tree and Random Forest are effective in predicting bankruptcy risk with a high degree of accuracy, especially when applied to bank financial ratio data (Urrochman et al, 2024). A study by (Sisodia & Sisodia, 2018) analyzed diabetes classification results from three machine learning algorithms, namely Naive Bayes, Support Vector Machine (SVM), and Decision Tree. The dataset used is Pima Indian from the National Institute of Diabetes and Digestive and Kidney Diseases. Naïve Bayes algorithm has the highest accuracy compared to other algorithms. Previous research has not been able to present accuracy above 80%. In other words, the research that has been done has not achieved maximum performance. In addition, the normalization method for the unbalanced range of values for the variables that determine diabetes is suspected to be one of the factors for the low performance of the classification model (Henderi et al., 2021). In this study using machine learning classification, including Support Vector Machine and Deep learning using ensemble classifiers bagging and bosting as parameter optimization. The results of the research on the accuracy of diabetes classification in pregnant women to support the Indonesia 2045 health program.

## METHOD

This research phase is divided into three parts, namely: collecting data, selecting and applying methods to machine learning and comparing accuracy information on each method used. At the stage of collecting data is done by searching for public datasets in Kaggle related to diabetes. At the stage of using the Pima Indians Diabetes Dataset data obtained from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Furthermore, the process of selecting and applying methods to machine learning is carried out with the help of rapid minner tools to process data using predetermined algorithms. The results of rapid minner processing will compare the accuracy of each method to determine the highest accuracy obtained. The following are the stages of the method in this study:

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

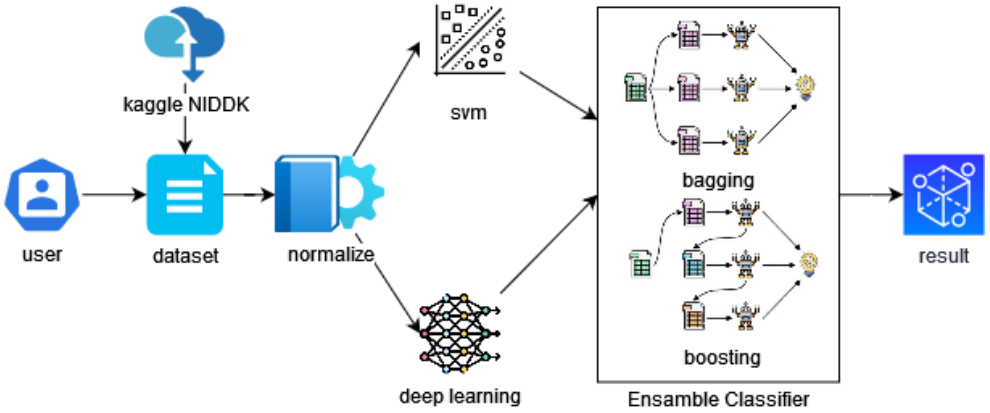


Fig 1. Methodology
Source : Researcher, 2024

1. Dataset

This research uses an open dataset in the form of Pima Indians Diabates (PID) (Pima Indians Diabetes Database Dataset, 2024). The Pima Indians Diabetes dataset is obtained from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) which is data on the results of diagnoses of women with a minimum age of 21 years, and have lineage or blood from the Pima Indian tribe, United States. Dataset 1 contains 768 records with 268 positive data and 500 negative data. In Table 1 is the attribute data in the dataset:

Table 1. Attribute variables

| No | Informant | Information |
|---|---|---|
| 1. | Pregnancies | Number of pregnancies |
| 2. | Glucose | 2-hour plasma glucose concentration |
| 3. | BloodPressure | Diastolic blood pressure (mm Hg) |
| 4 | SkinThickness | Triceps skinfold thickness (mm) |
| 5 | Insulin | 2-hour serum insulin (mu U/ml) |
| 6 | BMI | Body Mass Index |
| 7 | DiabetesPedigree | Family history |
| 8 | Age | Age |

Source : Researcher, 2024

2. Machine Learning – SVM

The computer will be taught using techniques derived from statistical learning theory and optimization theory using SVM, a learning system used to hypothesize linear functions in a high-dimensional feature space. By applying a kernel technique to the data set's initial characteristics, SVM may operate on non-linear data. Lower dimensions are mapped to higher dimensions using kernel functions (Hermansyah et al., 2023). In order to determine the new features created, the kernel function must be utilized to dot product substitution in the feature space. The classification technique in this study uses the Radial Basic Function (RBF) kernel idea to improve accuracy using the following formula:

$$K(Xi, Xj) = \exp\left(\frac{|Xi - Xj|^2}{2\,\alpha 2}\right) \quad (1)$$

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

Keterangan :
Xi, Xj = pair of two traning data
α = constant

3. Machine Learning – Deep Learning

Deep learning is suited for analyzing and extracting useful knowledge from both large huge amounts of data and data collected from different sources (Zhang et al, 2017). Deep learning is a subfield of machine learning whose algorithms are inspired by the structure of the human brain. Deep learning technology has been applied in various high-tech products such as self-driving cars and everyday services. Here is a formula for using deep learning:

$$ht = \alpha(Whht\text{-}1 + WxXt + b) \qquad (2)$$

Keterangan :
Wh = weight matrix of the previous state
Wx = weight matrix of the inputs at run time
b = bias value
α = sigmoid activation function / tanh

4. Ensamble Classifier

Ensemble classifier is a data level approach aimed at improving class balance. The ensemble algorithm approach has the same goal of improving the algorithm without changing the data so that there can be 2 approaches, namely the data level approach and the algorithm level approach (Cendani & Wibowo, 2022). Popular ensemble algorithms use boosting and bagging. ThereBoost is an algorithm that shows improved classification performance while bagging is a simple but effective ensemble method that has been applied to many real-world applications used in classification algorithms with the aim of increasing accuracy and the results are better than random sampling (Muliawan et al., 2023).

## RESULTS AND DISCUSSION

In this research, RapidMiner tools are utilized to develop data mining models aimed at classifying diabetes among pregnant women. The primary process is divided into two main approaches: one that incorporates ensemble classifiers and one that does not. The inclusion of ensemble classifiers is intended to enhance model accuracy and mitigate the risk of overfitting. Ensemble classifiers generally provide greater stability and resilience to noise compared to individual models. This robustness is attributed to the diversity of predictions generated by multiple classifiers, which collectively balance out potential errors.

Ensemble classifiers can integrate various types of base models, such as Decision Trees, SVM, Logistic Regression, and others, thereby leveraging the strengths of different algorithms to address complex data patterns effectively. This flexibility allows for the combination of diverse modeling techniques, which can enhance overall predictive performance and adaptability. The following section presents the data mining models developed for diabetes classification in pregnant women, detailing how ensemble approaches contribute to improved diagnostic accuracy and reliability in early detection efforts. The following is a data mining model that has been created in the classification of diabetes in pregnant women:

JOURNAL OF INFORMATICS DEVELOPMENT
E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

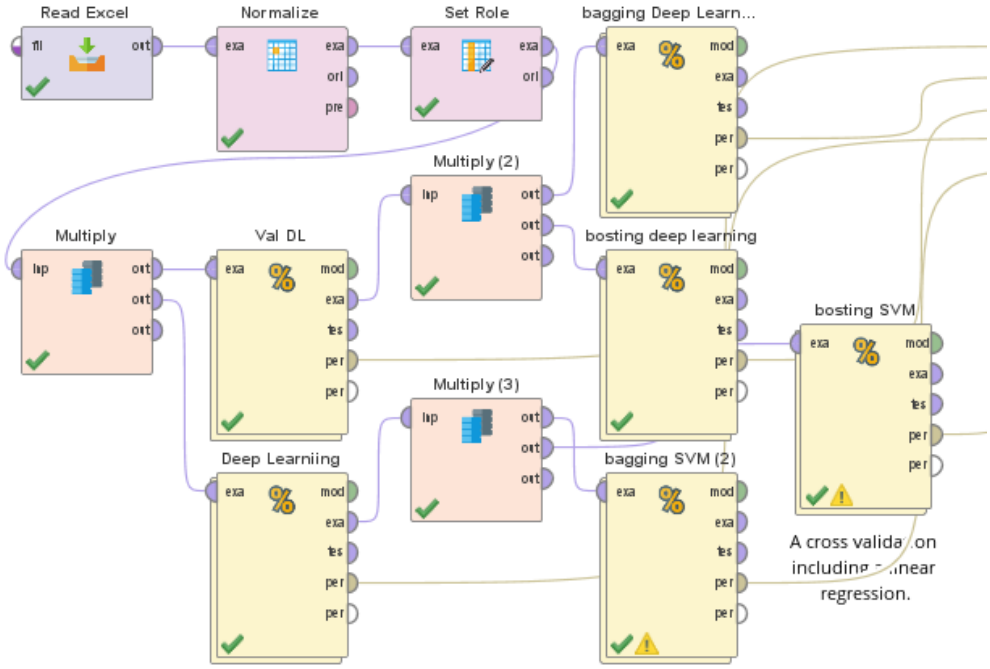


Fig 2. Modeling
Source : Researcher, 2024

From making data mining models, the accuracy results of using machine learning with ensamble classifiers and without using ensamble classifiers are as follows:

Table 2. Accuracy Result

| No | Method | Ensamble Classifier | Accucary |
|---|---|---|---|
| 1. | SVM | Bagging | 76.95 |
| 2. | SVM | Boosting | 76.91 |
| 3. | SVM | - | 76.56 |
| 4 | DL | Bagging | 76.31 |
| 5 | DL | Boosting | 75.26 |
| 6 | DL | - | 74.62 |

Source : Data Processing, 2024

The results in Table 2 highlight the effectiveness of ensemble classifier techniques, specifically bagging and boosting, in enhancing the accuracy of machine learning models for diabetes detection. The SVM with Bagging achieved the highest accuracy of 76.95%, surpassing the standalone SVM model, which reached an accuracy of 76.56%. This indicates that Bagging can provide a notable improvement in predictive performance by reducing variance and enhancing stability. Similarly, the Deep Learning (DL) model shows a substantial improvement when using Bagging, achieving an accuracy of 76.31% compared to 74.62% without the ensemble classifier. Boosting also contributes to accuracy enhancement; however, its effect is slightly less pronounced compared to Bagging for both SVM and DL models. For example, the SVM with Boosting achieved an accuracy of 76.91%, slightly lower than SVM with Bagging but still above the standalone SVM model. These results suggest that the use of ensemble techniques, particularly Bagging, is effective in improving model accuracy. Consequently, adopting these optimized machine learning models could be a significant and relevant step in supporting Indonesia's Golden Vision 2045 by contributing to early and accurate diabetes detection, which can lead to timely interventions and better health outcomes.

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

## CONCLUSIONS

This study reveals that the application of Machine Learning technology in diabetes detection in pregnant women is a progressive and relevant step in the context of the vision of Indonesia Emas 2045. It was found that Machine Learning methods, especially classification algorithms such as Support Vector Machine (SVM) with a combination of ensemble classifiers using bagging were able to provide predictions and even improve the accuracy quite high based on the medical data collected by 76.95%. This finding shows the great potential of Machine Learning technology in improving early diagnosis and management of diabetes, which can significantly minimize the health risks associated with pregnant women and fetuses. Thus, the implementation of Machine Learning for diabetes detection in pregnant women is not only a technological innovation, but also a representation of the commitment to improve the quality of life and welfare of the Indonesian people towards the vision of a Golden Indonesia 2045.

## REFERENCES

Chauhan, T., Rawat, S., Malik, S., & Singh, P. (2021). Supervised and Unsupervised Machine Learning based Review on Diabetes Care. *2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, *1*, 581–585. https://doi.org/10.1109/ICACCS51430.2021.9442021

Fauziah, D. A., Muliawan, A., & Rohim, M. A. (2023). Stock Price Prediction of PT. Jasa Marga (Persero) Tbk Using Linear Regression Algorithm. *PROCEEDING INTERNATIONAL CONFERENCE ON ECONOMICS, BUSINESS AND INFORMATION TECHNOLOGY (ICEBIT)*, *4*, 810–815.

Henderi, H., Wahyuningsih, T., & Rahwanto, E. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *International Journal of Informatics and Information Systems*, *4*(1), Article 1. https://doi.org/10.47738/ijiis.v4i1.73

Hermansyah, M., Fauziah, D. A., & Muliawan, A. (2023). Decision Support System for Identifying Best Students in the Class with Simple Additive Weighting (SAW) Method. *INSIDE - Jurnal Sistem Informatika Cerdas*, *1*(1), Article 1. https://doi.org/10.31967/inside.v1i1.866

Mahesh, B. (2020). Machine learning algorithms-a review. International Journal of Science and Research (IJSR).[Internet], 9(1), 381-386.

Mazlan, A. U., Sahabudin, N. A. binti, Remli, M. A., Ismail, N. S. N., Mohamad, M. S., & Warif, N. B. A. (2021). Supervised and Unsupervised Machine Learning for Cancer Classification: Recent Development. *2021 IEEE International Conference on Automatic Control & Intelligent Systems (I2CACIS)*, 392–395. https://doi.org/10.1109/I2CACIS52118.2021.9495888

Muliawan, A., Rizal, A., & Hadiyoso, S. (2023). Heart Disease Prediction based on Physiological Parameters Using Ensemble Classifier and Parameter Optimization. *Journal of Applied Engineering and Technological Science (JAETS)*, *5*(1), 258–267.

Roglic, G. (2016). WHO Global report on diabetes: A summary. International Journal of Noncommunicable Diseases, 1(1), 3-8.

Sabilirrasyad, I., Hermansyah, M., Prasetyo, N. A., Muliawan, A., & Wahid, A. (2024). Unveiling X/Twitter's Sentiment Landscape: A Python Crawler That Maps Opinion Using Advanced Search. *LOREM: Computational Engineering and Computer Information Systems*, *1*(1), Article 1.

Sisodia, D., & Sisodia, D. S. (2018). Prediction of Diabetes using Classification Algorithms. *Procedia Computer Science*, *132*, 1578–1585. https://doi.org/10.1016/j.procs.2018.05.122

Syamsudin, T., Handhayani, T., & Syaifudin, M. I. (2024). PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE MACHINE LEARNING. *Jurnal*

E-ISSN : 2963-0568, P-ISSN : 2963-055X
Available online at:
https://ejournal.itbwigalumajang.ac.id/index.php/jid

*Ilmu Komputer Dan Sistem Informasi*, *12*(1), Article 1. https://doi.org/10.24912/jiksi.v12i1.28197

Tahseen Ali, A., Abdullah, H. S., & Fadhil, M. N. (2021). WITHDRAWN: Voice recognition system using machine learning techniques. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.04.075

Urrochman, M. Y., Asy'ari, H., & Ro'uf, A. (2024). Using Machine Learning Techniques to Predict Financial Distress in Rural Banks in Indonesia . Journal of Informatics Development, 2(2), 36–44. https://doi.org/10.30741/jid.v2i2.1341

Wiranto, F., Rohim, M. A., Firdausi, M. F., Muliawan, A., & Afrianto, E. (2023). *Optimizing Coffee Crop Selection for Plant-Ready Coffee Fields: A Study Using the "Predict.in" Decision Support System at the Coffee and Cocoa Research Center in Jember*. *6*(1).

Zhang, L., Wang, S., & Liu, B. (2017). Deep learning for sentiment analysis: A survey. National Science Foundation (NSF), and Huawei Technologies Co. Ltd.