JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.8 No.1 January 2026: 65-71
e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v8i1.15332

A Comparative Study of Machine Learning and Stacking Ensemble Models for Diabetes Prediction

Bakti Amirul Jabar[1*], Albertus Januario[2], Davin Miguel Sanjaya[3], James Tanuwijaya[4]

[1-4]Computer Science Department, School of Computer Science
Bina Nusantara University,
Jakarta, Indonesia 11480
bakti.jabar@binus.ac.id; albertus.januario@binus.ac.id; davin.sanjaya@binus.ac.id; james.tanuwijaya@binus.ac.id

*Correspondence: bakti.jabar@binus.ac.id

Abstract – *Diabetes is a chronic metabolic disease and an increasingly widespread disease around the world, and early diagnosis is crucial. Methodology In this study, the performance of three machine learning models (Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes) is reviewed under the task of diabetes classification using the Pima Indians Diabetes Dataset. To tackle the class imbalance, we applied imputation, SMOTE for the data preprocessing, and min-max scaling to enhance the prediction performance. Further, we have applied ensemble learning and stacking, where all three models have been used as meta-classifiers. The results indicate that KNN had the best individual model performance (accuracy 77.27%, AUC 0.8444), but the stacking ensemble with a meta-model being Logistic Regression is superior to any model (accuracy 80.52%, AUC 0.8604). This suggests that ensemble learning can also improve the accuracy of diabetes diagnosis. These findings demonstrate that combining multiple classification approaches may provide more stable predictions across different patient conditions and clinical attributes. In addition, the preprocessing stages contributed to reducing noise and improving data consistency before model training. The study also highlights the potential use of ensemble-based systems in supporting healthcare professionals during preliminary diabetes screening, particularly in environments with limited medical resources and increasing numbers of diabetes cases requiring rapid assessment.*

Keywords: *Diabetes prediction; Machine learning; Pima Indians Dataset; Ensamble Model*

I. INTRODUCTION

Diabetes is a long-term metabolic condition that is now a global health issue. Different factors, including unhealthy diet, hereditary predisposition, and sedentary lifestyle, are the culprits for the growing disease burden of diabetes (Ghosh et al., 2021). Blindness, kidney failure, cardiovascular disease, and neuropathy are just some of the serious complications that can develop as a result of uncontrolled diabetes (Daghistani & Alshammari, 2020). To reduce these risks and make patient outcomes more effective, early detection and proper care are essential (Ooka et al., 2021). Computerized diagnostic model systems for illness diagnosis and health guidance have become increasingly popular owing to advancements in artificial intelligence (AI) and machine learning (ML). This paper will compare some Machine Learning algorithms such as Logistic Regression, KNN, Naive Bayes, stacking ensemble with meta model logistic, and their performance and accuracy in diabetes prediction based on the literature.

Logistic Regression is the most frequently used classification algorithm for binary outcomes, which is heavily utilized for medical prediction issues. Prediabetes, estimated (Ooka et al., 2021), also defines a group at high risk of diabetes (but not sufficient criteria for the disease) with several studies showing its use in diabetes prediction (Khanam & Foo, 2021) (Samet et al., 2021). Logistic Regression is an acceptable choice given its interpretability, something which is especially important in clinical decision-making (Rajendra & Latifi, 2021) (Hassan et al., 2017) (Joshi & Chawan, 2018).

K-Nearest Neighbors (KNN) is another simple and easy to implement instance based learning algorithm, which categorizes instances based on the nearest training instances. Ali and others (Ali et al.,

2020) also applied KNN for diabetes prediction on clinical datasets with success. Similarly, Rikatsih and others (Rikatsih et al., 2024) used symptom-based data for the algorithm and validated that it is also helpful for diagnostics at an early stage. Furthermore, Pertiwi and others (Pertiwi et al., 2020) proposed an improved KNN model by implementing the SMOTE approach to mitigate the class imbalanced data, and thus sharpen diagnostic accuracy.

The Naive Bayes is a probabilistic classifier algorithm and commonly used to predict tasks in medical diagnosis, such as diabetes prediction. Because of its simplicity and effectiveness, this algorithm has the ability to handle high-dimensional data and noisy inputs. Okikiola et al. (Okikiola et al., 2023) showed its capability in classifying the presence of diabetes using clinical attributes with high accuracy and low computational cost. Arrayyan et al. (Arrayyan et al., 2024) built a classification model for specific populations, confirming that Naive Bayes can be adapted to various demographic contexts. Beyond diabetes, Sopharak et al. (Sopharak et al., 2023) applied the Naive Bayes classifier for detection of diabetic retinopathy automatically, demonstrating its wider applicability to diabetes-related diagnostic problems.

Ensemble learning techniques like stacking is the most used in medical diagnostics due to their ability to improve predictive accuracy by combining multiple base models. Stacking models employs a meta-model trained on the outputs of base learners, allowing it to learn how best to combine these predictions for more accurate results. Rajendra and Latifi (Rajendra & Latifi, 2021) showed that both methods improved diabetes prediction over single models. Kibria et al. (Kibria et al., 2022).

Early studies have used single machine learning classifiers such as Logistic Regression, KNN, and Naive Bayes for diabetes prediction (Ali et al., 2020). Although these methods have produced moderate accuracy, they are susceptible to performance degradation when applied to imbalanced datasets or when confronted with noisy clinical data. Recent works have indicated that ensemble methods such as stacking can be more powerful in terms of prediction if ensemble power can be made by merging the power of the individual models (Kibria et al., 2022) (Rousyati et al., 2021). However, few studies have evaluated these approaches in a systematic manner on the Pima Indians Diabetes Dataset with standardized preprocessing techniques (UCI Machine Learning Repository, n.d.). So, this study is an attempt to fill the gap by comparing individual models to stacked ensemble approach for enhanced prediction of diabetes.

II. METHODS

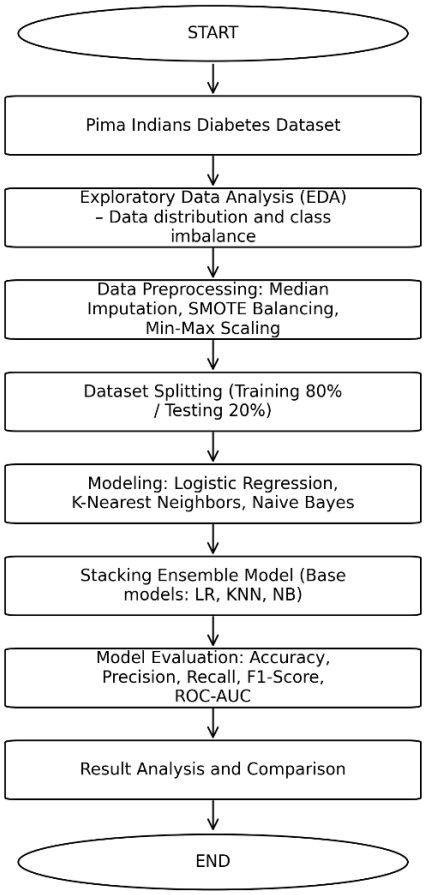

Figure 1. Flowchart Method

The methodology used in this study consists of several stages, including data preprocessing, model development, and performance evaluation. The overall workflow of the proposed approach is illustrated in Figure 1.

2.1 Selecting a DataSet

This study uses the publicly accessible Pima Indian dataset using the Kaggle platform to detect diabetes in people (UCI Machine Learning Repository, 2016). In machine learning and deep learning studies, the Pima Indian Dataset has been one of the most often used datasets to detect diabetes because of its large sample size and many features, including demographics and medical history (Rousyati et al., 2021). The sample's 768 individual data reflect ages ranging from 21 to 81. In total, there are 500 data records, which reflect a negative class, or individuals who have not been diagnosed with diabetes. People who have been diagnosed with diabetes provide the remaining amount of data, or 268 total.

2.2 Exploratory Data Analysis

The first examination of the dataset based on the medical features such as Glucose, Blood Pressure, Skin Thickness, Insulin and BMI had been performed. The target variable Outcome was explored, indicating an imbalanced class of the target variable, one in which the number of non-diabetics outweigh the number of diabetic individuals. The class imbalance was also intuitively visualized in a count plot, which provided an initial sense of the dataset structure, while highlighting a consideration for the later modelling phases.

In addition, boxplot visualizations were created for the numerical features to examine the distribution and spread of the data. The analysis of boxplots showed separate patterns across each of the features; Glucose, BMI, and Age almost appearing normally distributed whereas Insulin and Skin Thickness had a predominance of lower values. These visualizations provide some insights relating to the overall properties of the entire dataset prior to the preprocessing and model training phase.

2.3 Data PreProcessing

The data preprocessing stage was completed as part of the effort to ensure the dataset consisting of medical features is clean and contains reliable values when training the model. One of the first steps taken was to address zero values for certain medical features such as Glucose, Blood Pressure, Skin Thickness, Insulin, and BMI which are considered biologically impossible values. Instead of excluding these values or treating the zero as a missing value, the zero values were treated as missing values and were subsequently imputed with the median value. In addition to using the median value, the imputation of load zero values was done on a class-wise basis, that is, the no diabetes class was one imputation and the diabetes class group were in a separate imputation group. This is different from a traditional method where the imputation would occur for this features combined and could be introduced with bias towards one of the imputed groups in its variance. Imputing based on each original class helps maintain the uniqueness of each class and to limit any added bias.

The problem of class imbalance in the target variable was handled using the Synthetic Minority Over-sampling Technique (SMOTE) in the target classes to oversample the minority class. SMOTE oversamples the minority class by producing synthetic data based on the positive class feature distribution. SMOTE allows for better representation of that class without duplicating existing records, nor will the new synthetic records alter the class distribution of the retrieved dataset. In addition, data normalization was accomplished through the use of Min-Max Scaling to place all data features on the same scale. After pre-processing, the dataset was stratified sampled into a train and test set with an 80:20 ratio to maintain overall class distribution in both classes.

2.4 Modeling

During the first phase of the modeling, three machine learning algorithms were used to classify diabetes risk: Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes. Each of the models was trained on training data which had been balanced through SMOTE, resolving the class imbalance in the positive (diabetic) and negative (non-diabetic) class. Each model was conducted with the default parameters set to the libraries utilized, with no changes to the settings. After the training process was finished, a performance assessment was completed on the test dataset, with accuracy, ROC AUC and classification reports used as evaluation metrics. The performance evaluation provided a first pass indication of each models ability to differentiate between individuals with diabetes and with no diabetes.

2.5 Stacking Classifier

In order to increase the accuracy of diabetes diagnosis, this study uses an ensemble learning technique using a stacking classifier. The basis estimators were determined to be three machine learning algorithms: Naive Bayes, K-Nearest Neighbors (KNN), and Logistic Regression. These models were chosen because they use different categorization techniques: Naive Bayes uses a probability model, KNN determines how close the data points are to one another, and Logistic Regression uses linear relationships. In the ensemble setting, the variety of these models is probably going to provide more benefits.

The same base estimators were used on three different stacking setups. The most prominent distinction among these setups is that the meta-model for blending the predictions of the base models differs. All three stacking variants utilized a different meta-model: Logistic Regression, KNN, and Naive Bayes respectively. That is, while the base estimators remained unchanged, each variant employs a different meta-model to estimate the impact of meta-model choice on the overall accuracy of classification.

All the stacking classifiers were trained with pre-processed training data that had been class balanced using SMOTE and feature normalized using Min-Max Scaling. To enhance the model robustness during training, cross-validation was employed in the training process. This enables an exhaustive analysis of the degree to which predictions from the base models can be optimally utilized by various meta-models, as well as the

influence of the selected meta-model on classification performance in diabetes diagnosis.

2.6 Model Evaluation

We evaluated our model in this study using the confusion matrix to evaluate the predictive success of the model in terms of the actual classification of the data. The confusion matrix consists of four elements: True Positive (TP) positive data labeled correctly as positive; True Negative (TN) negative data labeled correctly as negative; False Positive (FP) negative data labeled incorrectly as positive; and False Negative (FN) positive data labeled incorrectly as negative. Using these values we calculated four of the primary evaluation metrics comprised of accuracy (overall correct classifications), precision (the correctness of the positive predictions), recall (to determine the model's ability to detect all positive occurrences) and F1-score (the harmonic mean of precision and recall). The metrics of ROC AUC (Receiver Operating Characteristic - Area Under the Curve) was also calculated to determine the model ability to come close to distinguish between positive and negative classes. A high ROC AUC score indicates strong classification performance.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$Precesion = \frac{TP}{(TP+FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

$$F1-Score = \frac{2 \times (Precision \times Recall)}{(Precision+Recall)} \quad (4)$$

2.7 Majority Baseline and McNemar's Test

To compare baseline performance, we also included a majority class classifier that predicts the most common class or the dominant class in the dataset, providing a reference for the minimum expected performance. In addition, to evaluate whether the performance differences between models were statistically significant, we applied McNemar's test on the predictions of the best-performing individual model (KNN) and the stacking ensemble (Meta Logistic Regression). This allowed us to assess not only raw performance metrics themselves but also the robustness of improvements values performed by the ensemble approach.

IV. RESULTS AND DISCUSSION

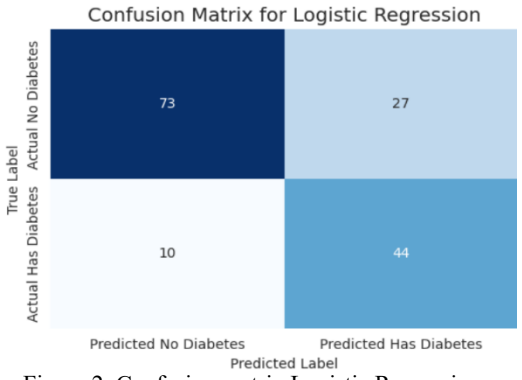

Figure 2. Confusion matrix Logistic Regression

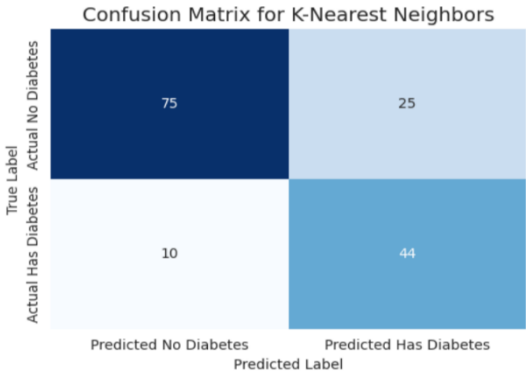

Figure 3. Confusion matrix K-Nearest Neighbors (KNN)

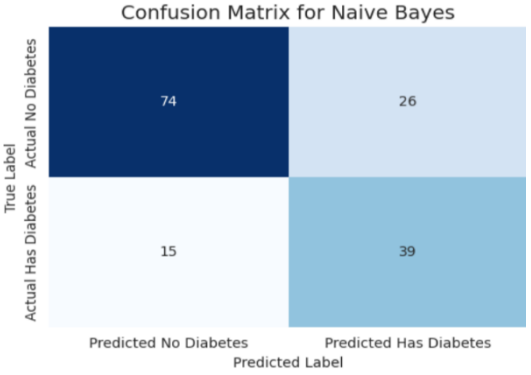

Figure 4. Confusion matrix Naive Bayes

Figure 2 , Figure 3 and Figure 4 shows the confusion matrix of each methods. The prediction results obtained from three machine learning techniques, such as Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes, are all appropriate in the diabetes risk detection task. It can be seen by the measure of accuracy score and AUC score that the KNN model is the most effective having an accuracy of 77.27% and an AUC of 0.8444 followed by the Logistic Regression which predicts 75.97% accuracy with an AUC of 0.8261. The poor performance is shown for Naive Bayes, and it is 73.38% (AUC=0.8133) in accuracy. The classification report also highlights such an overall good performance of KNN, regarding also as

substantially good identification of the positive diabetes cases (81% recall and 0.72 F1-score for the positive class). This implies that the KNN model is the best among the three in diagnosing diabetic patients correctly with lower false negatives.

The pattern of confusion matrices is same in all the models as non-diabetic patients are being predicted as diabetic and are considered as false positives. In particular, 27 out of 100 non-diabetics were misclassified by Logit Regression, vs. 26 by Naive Bayes. The precision of the positive class with Naive Bayes was also lower (0.60) and a higher number of false positive predictions.

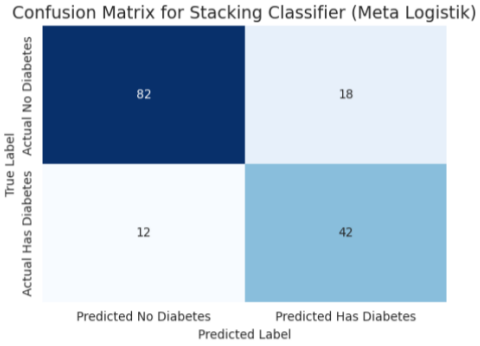

Figure 5. StackingClassifier (Meta model Logistic Regression)

In Figure 5, All individual models were outperformed by Stacking Classifier base models, the logistic regression, K-nearest neighbor (KNN) and naive Bayes, and its meta-learner, logistic regression. As a result, this voting model, with accuracy of 80.52% and AUC = 0.8604 predict decent enough with the report card reflecting a good value of precision and recall around the positive class ("Has Diabetes"). Precision: 0.70 Recall: 0.78 F1-score: 0.74 As for the confusion matrix, the model was able to identify 42 out of 54 patients with diabetes and 82 out of the 100 non-diabetic patients correctly. This suggests that stacking is an effective method to combine sensitivity and specificity for various base models and holds huge potential for diabetes prediction.

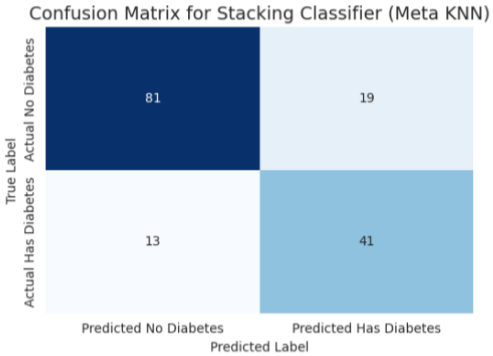

Figure 6. StackingClassifier (Meta model KNN)

In Figure 6, the Stacked Model using KNN-based meta model and stacked base learners, including Logistic Regression, KNN and Naive Bayes, was promising in predicting diabetes attendance. It achieved 79.22% prediction accuracy in a test set and an AUC score of 0.8417. The positive class (the diabetic patients) had precision, recall and F1-score equal to 0.68, 0.76 and 0.72, according to the classification report. As per the confusion matrix, out of all diabetes cases (54), the model has correctly captured 41 diabetes cases. Furthermore, from the total of 100 non-diabetic cases, only 19 were diagnosed as cases of diabetics. These results suggest that KNN as meta-classifier strikes a good trade-off between sensitivity and specificity for the detection of individuals with diabetes, false positive rate of detection is kept low. The performance of the new model is a bit lower than that of the stacking model with a logistic regre ssion meta-learner, but it's actually a reliable prediction method for diabetes.meta-learned, but it's indeed a credible predictive approach for diabetes.

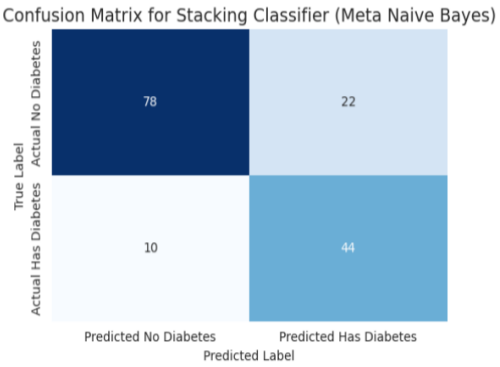

Figure 7. StackingClassifier (Meta model Naive Bayes)

In Figure 7, the Stacking Ensemble with Naive Bayes as the meta-model performs competitively with Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes base models to predict diabetes cases. The model achieves an accuracy of 79.22% and AUC value of 0.8517. It achieves 0.67, 0.81, and 0.73 for precision, recall, and F1-score for the positive class (people who have diabetes). This suggests that the model can predict diabetic patients more accurately and with less false negative prediction. It can be inferred from the confusion matrix that 44 of 54 of the diabetic patients were classified correctly and 78 of 100 of the non-diabetic patients were tested correctly. These results suggest that using Naive Bayes as a meta-classifier provides a powerful strength in prediction of suicidal groups meanwhile decreasing misclassifications, and can be an competitive option to its well-known meta-models in ensemble methods.

In this study, we tested three individual classifiers: Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes. Logistic

Regression yielded an accuracy of 75.97% and an AUC of 0.826, with sensitivity of 0.81 and specificity of 0.73, showing balanced performance in diagnostic both diabetic and non-diabetic cases. KNN showed better performance than Logistic Regression with an accuracy of 77.27% and AUC of 0.844, yielded the highest sensitivity (0.81) but with moderately lower specificity (0.75). Naive Bayes achieved 73.38% accuracy and an AUC of 0.813, with sensitivity of 0.72 and specificity of 0.74, thus having reasonable ability to predict diabetic patients but lower precision compared to the other models.

We also experimented stacking ensembles with the three models as base learners and three different meta-models. Of these, the stacking classifier with Logistic Regression as meta-learner achieved the best performance, with an accuracy of 80.52% and an AUC of 0.860, with good trade-off between sensitivity (0.78) and specificity (0.82). The stacking variant, Naive Bayes and KNN as meta-learners also improved their base models and each model achieving an accuracy of 79.22% with slightly lower AUC values. These results ensure that stacking provides a consistent enhanced performance over individual models.

Table 1. Accuracy and F1- Score All Model

Model	Accuracy	AUC-Score	Precision	Recall	F1-Score
Logistic Regression	0.7597	0.8261	0.6197	0.8141	0.7040
K-Nearest Neighbors	0.7727	0.8444	0.6376	0.8141	0.7154
Naive Bayes	0.7338	0.8133	0.6000	0.7222	0.6554
Stacking Logistic	0.8052	0.8604	0.7000	0.7777	0.7368
Stacking KNN	0.7922	0.8417	0.8833	0.7592	0.7192
Stacking Naive	0.7922	0.8517	0.6666	0.8148	0.7333

Table I shows the performance of all models. The enhancement reflects the complementary strengths of the base learners: where Naive Bayes for modeling the probabilistic relationships, KNN captures local instance-based patterns, and Logistic Regression detecting linear trends. The stacking model combines these classifiers through a meta-model. With this method predictions are achieving more accurate, stable, and generalizable predictions than any single classifier.

All models performed above the baseline of the majority (65% accuracy), indicating their ability to capturing relevant patterns. Among individual classifiers, KNN was the best performing one by capturing non-linear relationships, while Logistic Regression resulted in more balanced sensitivity and specificity, and Naive Bayes was weaker due to its independence assumption. Stacking enhanced the results by integrating the complementary strengths of these models, yielding a higher accuracy, AUC and a more stable sensitivity–specificity balance. McNemar's test ($p = 0.179$) confirmed that the difference between Stacking and KNN was not statistically significant, the consistent performance result indicate that the stacking method is a more robust method for this task.

V. CONCLUSION

According to the performance of multiple machine learning models in diabetes prediction, the Stacking Classifier with the meta leaner as Logistic Regression achieved the highest accuracy in diabetes prediction (80.52%) and AUC 0.8604. It demonstrates better predictive power than popular one-model methods, i.e. Logistic regression, K-Nearest Neighbors, and Naïve Bayes. All models achieved recall higher than 70% for positive diabetes cases detection, nevertheless Stacking Classifier had the best trade-off between precision and recall giving it the highest F1-scores for both classes. This indicates that there is a great deal of room for improvement in terms of how ensemble method combined with a suitable meta learning algorithm can increase prediction accuracy and robustness over individual models in this dataset.

AUTHOR'S CONTRIBUTION

Bakti Amirul Jabar: Conceptualization, Methodology, Validation, Supervision, Project administration, Writing – Review & Editing. Albertus Januario: Methodology, Software, Validation, Formal analysis, Resources, Data curation, Writing – Original draft, Writing – Review & Editing, Visualization. Davin Miguel Sanjaya: Software, Conceptualization, Methodology, Formal analysis, Investigation, Writing – Original draft, Writing – Review & Editing. James Tanuwijaya: Validation, Formal analysis, Conceptualization, Writing – Original draft, Writing – Review & Editing.

AVAILABILITY DATA AND MATERIALS

The dataset used can be accessed via https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database

REFERENCES

Ali, A., Alrubei, M., Hassan, L. F. M., Al-Ja'afari, M., & Abdulwahed, S. (2020). Diabetes classification based on KNN. IIUM Engineering Journal, 21(1), 175–181. https://doi.org/10.31436/iiumej.v21i1.1206

Arrayyan, A. Z., Setiawan, H., & Putra, K. T. (2024). Naive Bayes for diabetes prediction: Developing a classification model for risk identification in specific populations. Semesta Teknika, 27(1), 28–36. https://doi.org/10.18196/st.v27i1.21008

Daghistani, T., & Alshammari, R. (2020). Comparison of statistical logistic regression and RandomForest machine learning techniques in predicting diabetes. Journal of Advances in Information Technology, 11(2), 78–83. https://doi.org/10.12720/jait.11.2.78-83

Ghosh, P., Azam, S., Karim, A., Hassan, M., Roy, K., & Jonkman, M. (2021). A comparative study of different machine learning tools in detecting diabetes. Procedia Computer Science, 192, 467–477. https://doi.org/10.1016/j.procs.2021.08.048

Hassan, M., Butt, M. A., & Baba, M. Z. (2017). Logistic regression versus neural networks: The best accuracy in prediction of diabetes disease. Asian Journal of Computer Science and Technology, 6(2), 33–42. https://doi.org/10.51983/ajcst-2017.6.2.1782

Joshi, T. N., & Chawan, P. M. (2018). Logistic regression and SVM based diabetes prediction system. International Journal For Technological Research In Engineering, 5(11), 4347–4350.

Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. ICT Express, 7(4), 432–439. https://doi.org/10.1016/j.icte.2021.02.004

Kibria, H. B., Nahiduzzaman, M., Goni, M. O. F., Ahsan, M., & Haider, J. (2022). An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI. Sensors, 22(19), 7268. https://doi.org/10.3390/s22197268

Okikiola, F. M., Adewale, O. S., & Obe, O. O. (2023). A diabetes prediction classifier model using Naive Bayes algorithm. FUDMA Journal of Sciences, 7(1), 253–260. https://doi.org/10.33003/fjs-2023-0701-1301

Ooka, T., et al. (2021). Random forest approach for determining risk prediction and predictive factors of type 2 diabetes: Large-scale health check-up data in Japan. BMJ Nutrition, Prevention & Health, 4(1). https://doi.org/10.1136/bmjnph-2020-000200

Pertiwi, A. G., Bachtiar, N., Kusumaningrum, R., Waspada, I., & Wibowo, A. (2020). Comparison of performance of k-nearest neighbor algorithm using SMOTE and without SMOTE in diagnosis of diabetes disease in balanced data. Journal of Physics: Conference Series, 1524, 012048. https://doi.org/10.1088/1742-6596/1524/1/012048

Rajendra, P., & Latifi, S. (2021). Prediction of diabetes using logistic regression and ensemble techniques. Computer Methods and Programs in Biomedicine Update, 1, 100032. https://doi.org/10.1016/j.cmpbup.2021.100032

Rikatsih, N., Anshori, M., Pradini, R. S., & Faurika. (2024). K-Nearest Neighbor method for early detection of diabetes patients based on symptoms and clinical data. Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, 9(2), 187–192. https://doi.org/10.25139/inform.v9i2.8582

Rousyati, R., Rais, A. N., Rahmawati, E., & Amir, R. F. (2021). Prediksi Pima Indians Diabetes Database dengan ensemble Adaboost dan Bagging. EVOLUSI: Jurnal Sains dan Manajemen, 9(2), 36–42. https://doi.org/10.31294/evolusi.v9i2.11159

Samet, S., Laouar, M. R., & Bendib, I. (2021). Use of machine learning techniques to predict diabetes at an early stage. Proceedings of the International Conference on Networking and Advanced Systems (ICNAS). https://doi.org/10.1109/ICNAS53565.2021.9628903

Sopharak, A., Nwe, K. T., Moe, Y. A., Dailey, M. N., & Uyyanonvara, B. (2023). Automatic exudate detection with a Naive Bayes classifier. 13th International Conference on Engineering, Science and Information Technology (ICESIT).

UCI Machine Learning Repository. (2016). Pima Indians diabetes database [Dataset]. Kaggle. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database