JoCoSiR
Journal of Computer Science Research
ISSN 2986-2337 (Online)

Comparative Study of Machine Learning Approaches Based on Artificial Neural Network, Regression, and Clustering for Diabetes Prediction

Nauval Alfarizi[a], Adi Putra[b], Prima Lydia Yosophin Batubara[c], Satria Sinurat[d]

University Pembangunan of Panca Budi Medan, North Sumatera

email: [a] nauvalalfarizi026@gmail.com, [b] elangperak494@gmail.com, [c] primabatubara87@guru.smk.belajar.id, [b] satria.sinurat@yahoo.co.id

ARTICLE INFO

Keywords:
Artificial Neural Network
Logistic Regression
K-Means Clustering
Diabetes Prediction
Machine Learning

IEEE style in citing this article:
N. Alfarizi and A. Putra, P.L.Y. Batubara, S. Sinurat" Comparative Study of Machine Learning Approaches Based on Artificial Neural Network, Regression, and Clustering for Diabetes Prediction," JoCoSiR: Jurnal Ilmiah Teknologi Sistem Informasi, vol. 3, no. 3, pp. 73-80, 2025.

ABSTRACT

This study presents a comparative analysis of three machine learning model and algorithms Artificial Neural Network (ANN), Logistic Regression, and K-Means Clustering using the Pima Indians Diabetes dataset. The main objective is to evaluate the performance of supervised and unsupervised methods in predicting diabetes based on physiological and clinical features. he ANN model was developed using a feedforward and backpropagation approach, Logistic Regression applied the fundamental logit equation, and K-Means Clustering was employed as an unsupervised reference. Model performance was assessed using Accuracy, Precision, Recall, and F1-score for supervised models, and Adjusted Rand Index (ARI) for clustering. Experimental results indicate that Logistic Regression achieved the best accuracy of 0.7573, followed by ANN with 0.7078, while K-Means obtained an ARI of 0.1614. The heatmap comparison shows that supervised models outperform unsupervised approaches, with Logistic Regression offering better interpretability and stability, and ANN demonstrating the ability to model nonlinear relationships. K-Means, though less accurate, provided valuable insight into data structure and natural grouping. Overall, the findings confirm that supervised learning models, particularly Logistic Regression and ANN, are more effective for medical prediction tasks. Future research may explore hybrid or ensemble models that combine the interpretability of Logistic Regression, the adaptability of ANN, and the exploratory capability of clustering to enhance medical diagnostic performance.

1. Introduction

Diabetes mellitus is not a single, uniform disease; its definition varies depending on the viewpoint. Medically, it refers to a group of metabolic disorders marked by high blood sugar levels due to either partial or complete lack of insulin. Prolonged high blood sugar can damage small blood vessels in the eyes, kidneys, and nerves, but these complications take time to develop and thus aren't suitable for defining the disease. Larger blood vessel issues such as heart attacks, strokes, and peripheral artery disease—are more common and often appear even before diabetes, formally diagnosed. Some experts propose defining diabetes as “early-onset atherosclerosis accompanied by high blood sugar,” highlighting the serious health risks most patients face [1].

Gestational diabetes is defined as a form of carbohydrate intolerance first recognized during pregnancy, excluding individuals with clear evidence of pre-existing diabetes. It is among the most common metabolic disorders encountered during gestation and is associated with increased risks of adverse maternal and neonatal outcomes. Although glycemic levels typically normalize postpartum, affected individuals should undergo follow-up screening to identify persistent glucose intolerance and be counseled regarding their heightened risk of developing type 2 diabetes in the future. Less common forms of diabetes mellitus are characterized by identifiable etiological or pathologi

Diabetes Mellitus (DM) is a metabolic condition primarily triggered by unhealthy lifestyle habits. Non-Proliferative Diabetic Retinopathy (NPDR) marks the initial phase of DM, during which symptoms may be subtle or entirely absent. Without timely intervention, this condition can advance to diabetic retinopathy, potentially impairing vision. The objective of this study is to detect diabetes in its early stages, before ocular complications arise. Given that the human tongue displays measurable features—such as color, geometric patterns, and texture these characteristics can be effectively utilized for diagnosing both DM and NPDR.[2]

The World Health Organization (WHO) released its first Global Report on Diabetes on World Health Day,

ISSN 2502-3357 (Online)

April 7th, 2016, which was specifically dedicated to addressing diabetes. Although diabetes has been documented since ancient times and recognized as a severe metabolic disorder, historical records suggest it was relatively uncommon in clinical practice. In recent decades, however, the global burden of diabetes has increased significantly, posing a major challenge to public health and socioeconomic development. Diabetes, along with cardiovascular diseases (CVD), cancer, and chronic respiratory diseases, was identified as one of the four key noncommunicable diseases (NCDs) in the Political Declaration on the Prevention and Control of NCDs adopted during the United Nations High-Level Meeting in 2011. In 2013, WHO Member States approved a global monitoring framework for NCDs, establishing nine targets to be achieved by 2025 [3]

Type 1 Diabetes Mellitus (DM) is a metabolic disorder characterized by chronic hyperglycemia resulting from impaired glucose regulation. This condition arises due to the destruction of pancreatic β-cells, either through autoimmune mechanisms or idiopathic causes, leading to a reduction or complete cessation of insulin production. Experts project that the combined incidence of type 1 and type 2 diabetes will increase by approximately 64% by the year 2025, with an estimated 53.1 million individuals expected to be diagnosed with diabetes [4].

Diabetes places a considerable burden on society, as evidenced by rising direct healthcare costs, decreased workforce productivity, higher rates of premature mortality, and intangible consequences such as reduced social participation and diminished quality of life. In 2017, the overall economic impact of diabetes in the United States was estimated at approximately $327 billion, of which about $237 billion were attributed to direct medical expenditures [5].

State of the Art

Recent advancements have led to the development and publication of various methodologies for predicting diabetes. One prominent approach utilizes machine learning techniques, where researchers conducted empirical analysis employing supervised models—such as Artificial Neural Networks (ANN) and linear regression—alongside unsupervised clustering algorithms. The study aimed to enhance the accuracy of diabetes prediction using the Pima Indians Diabetes dataset, a widely accepted benchmark for evaluating diagnostic performance in medical machine learning applications. Within this framework, ANN and regression models are categorized as supervised learning methods, which rely on labeled datasets to establish relationships between input variables and target outputs. In contrast, clustering algorithms like K-Means fall under unsupervised learning, as they operate by detecting natural groupings within the data based on feature similarity, without the guidance of predefined class labels.Within this framework, ANN and regression models are categorized as supervised learning methods, which rely on labeled datasets to establish relationships between input variables and target outputs. In contrast, clustering algorithms like K-Means fall under unsupervised learning, as they operate by detecting natural groupings within the data based on feature similarity, without the guidance of predefined class labels [6].

Artificial Neural Networks (ANN)

One of the most widely adopted approaches in this field utilizes Artificial Neural Networks (ANNs), which are inspired by the structure and functioning of the human brain. An ANN consists of numerous simple processing elements, known as artificial neurons, that receive, interpret, and transmit information to other interconnected neurons. The connection strengths, referred to as weights, are optimized during the training phase, allowing the network to learn patterns and relationships from the given dataset [7].

Machine learning is capable of processing both structured data (well-organized datasets) and unstructured data such as images, videos, and text. In general, machine learning techniques are categorized into three primary types: Supervised learning, which involves predicting numerical values (regression) or categorical outcomes (classification) using labeled input–output data pairs; Unsupervised learning, which focuses on identifying patterns or clusters within unlabeled datasets. These methods can also be applied for dimensionality reduction to simplify datasets while retaining essential information, or for detecting anomalies (outliers) within the data.

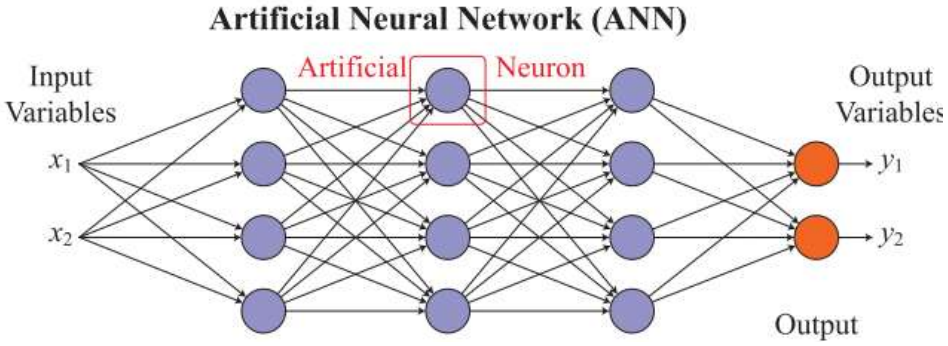


Figure 1. Workflow System of Artificial Neural Network

The output of each neuron in the network is computed through a forward propagation process, where the weighted sum of the inputs is calculated and passed through an activation function. The mathematical formulation of the neuron activation is expressed as:

DOI: https://doi.org/10.65126/jocosir.v3i3

ISSN 2502-3357 (Online)

$$z_j = \sum_{i=1}^{n} w_{ij}\, x_i + b_j$$

$$a_j = \frac{1}{1 + e^{-z_j}}$$

where w_{ij}denotes the connection weights, $b_j$is the bias, and the sigmoid function is applied to produce the neuron's output a_j. The process continues layer by layer until the final output is obtained for classification. During training, the model adjusts the connection weights using the backpropagation rule described by Kahramanli, but the forward propagation phase defines the main computation of the ANN structure [8]. In the learning phase of the Artificial Neural Network (ANN), the weights are updated iteratively to minimize the error between the predicted and target outputs. The adjustment of the connection weights follows the backpropagation learning rule as expressed in:

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \eta \cdot Err_j \cdot O_i$$

where ηrepresents the learning rate that controls the step size of the weight update, $Err_j$is the propagated error term of neuron j, and $O_i$denotes the output of neuron $i$in the previous layer. The update continues iteratively across all layers until the error is minimized, thereby improving the predictive accuracy of the model. This rule follows the standard formulation introduced by Kahramanli which has been widely adopted in medical data prediction, including diabetes classification. During the backpropagation process, the error from the output layer is propagated backward through the hidden layers to adjust the connection weights. The error term for a hidden neuron $j$is calculated as:

$$Err_j = O_j(1 - O_j) \sum_k E\, rr_k w_{jk}$$

where $O_j$represents the output of the hidden neuron j, $Err_k$is the error at the neuron $k$in the subsequent layer, and w_{jk}denotes the weight connecting neuron $j$to neuron k. The term $O_j(1 - O_j)$corresponds to the derivative of the sigmoid activation function. This step allows the network to distribute the total output error across all neurons in the hidden layers, thus improving the learning accuracy of the model The error term at the output neuron $j$quantifies the difference between the predicted output and the desired target. It is defined as:

$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

where $T_j$is the target (actual) value for neuron j, and $O_j$is the predicted output. The expression $O_j(1 - O_j)$is again the derivative of the sigmoid function, which facilitates the calculation of the gradient for weight.

Logistic regression

The simplest form of regression that can be constructed is linear regression. In gradient boosting, the final model is obtained by summing the outputs of all weak predictors. However, directly combining multiple linear regression predictors would simply yield another linear regression model. To address this, the proposed algorithm, RegBoost, partitions the training dataset into two branches based on the prediction outcomes of the current weak predictor. Linear regression is then recursively applied to each branch. During the testing phase, new data is routed to the appropriate branch to proceed with the subsequent weak predictor. The overall prediction is obtained by summing the contributions of all weak predictors along the traversal path. Since datasets often include redundant or irrelevant features that can be eliminated with minimal information loss, RegBoost employs a stepwise regression approach for feature selection [9].

Logistic regression is a regression method that links one or more independent variables to a categorical dependent variable—such as 0 and 1, true or false, or large and small. What sets logistic regression apart from multiple or linear regression is the categorical nature of its output [10].

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1 X$$

K-Mean Clustering

A comparative analysis of widely used clustering algorithms and evaluation techniques for clustering performance. An enhanced k-means text clustering approach is introduced, wherein the conventional centroid—typically derived via averaging—is substituted with the actual data point nearest to the computed centroid within the text dataset. Furthermore, a novel evaluation method utilizing the entropy of the silhouette coefficient is proposed to assess clustering quality. By varying the number of text samples, the optimal entropy value and corresponding ideal cluster count are determined. The proposed algorithm and evaluation strategy offer promising applications for fault data analysis in software-intensive systems [11].

$$ARI = \frac{\sum_{ij}\binom{a_{ij}}{2} - [\sum_i \binom{r_i}{2} \sum_j \binom{s_j}{2}]/\binom{\alpha}{2}}{\frac{1}{2}[\sum_i \binom{r_i}{2} + \sum_j \binom{s_j}{2}] - [\sum_i \binom{r_i}{2} \sum_j \binom{s_j}{2}]/\binom{\alpha}{2}}$$

DOI: https://doi.org/10.65126/jocosir.v3i3

ISSN 2502-3357 (Online)

Although several variants of the K-Means algorithm have been proposed to address its limitations, most of them are domain-specific and fail to generalize effectively. For instance, a K-Means method designed to handle categorical data may still perform poorly due to an unsuitable initialization procedure [12].

Research Framework

The review process was conducted based on a set of systematic guidelines and principles aimed at ensuring the theoretical rigor of the study. These guidelines highlight the importance of clearly defining the primary research domain, selecting appropriate samples, extracting and validating relevant data, and translating the findings into meaningful outcomes. In line with this approach, the present study focuses on identifying relevant literature, established standards, and evaluation frameworks. Furthermore, the research aims to construct a comprehensive framework that can serve as a reliable guide for tracking progress throughout the entire research process [13] .

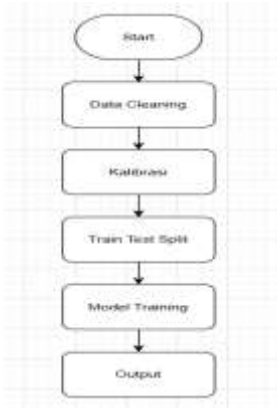


Figure 2. Research Framework and model training

2. Research Method

3.1 Visual Research

The research framework employed in this study involves a series of structured phases aimed at ensuring the consistency and validity of the experimental outcomes. The process commences with the data collection stage, in which relevant datasets are sourced from reliable and publicly available repositories. Next, data preprocessing is performed to enhance data quality through procedures such as cleaning, normalization, and the treatment of missing or inconsistent values. The subsequent feature selection phase focuses on identifying the most influential variables that contribute to diabetes prediction. Following this, model development and training are carried out using several machine learning approaches namely, Artificial Neural Networks (ANNs), Regression models, and Clustering algorithms to analyze their relative performance and predictive efficiency. In the final stage, performance evaluation is conducted to compare the accuracy, and generalization ability of each model. The complete workflow of the proposed methodology is depicted in Figure 2.

The Pima Indians Diabetes dataset, obtained from the UCI Machine Learning Repository, is used in this study. It contains 768 instances with 8 independent attributes, including glucose level, BMI, age, and insulin concentration, and one dependent attribute representing diabetes outcome (0 = negative, 1 = positive). Data preprocessing was performed to ensure the quality, consistency, and suitability of the Pima Indians Diabetes dataset before training the machine learning models. The preprocessing stage consisted of three primary steps: data cleaning, feature normalization, and data partitioning for model training and testing.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

3.2 Dataset Outlook

The dataset utilized in this study is primarily centered on diabetes prediction, encompassing medical and demographic data of individuals. As presented in Table 1, each record includes multiple clinical attributes—such as glucose concentration, blood pressure, body mass index (BMI), and insulin levels—that are essential for assessing the likelihood of diabetes occurrence. The dataset structure clearly supports the objective of evaluating machine learning models for early detection and classification of diabetes cases.

Table 1. Output Dataset Pima Diabetes

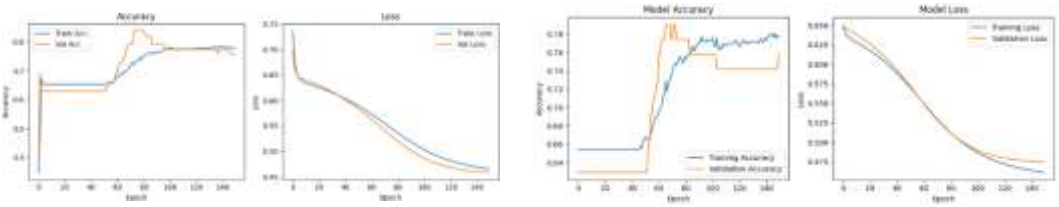
Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI
0	6	148	72	35	0 33.6
1	1	85	66	29	0 26.6
2	8	183	64	0	0 23.3
3	1	89	66	23	94 28.1
4	0	137	40	35	168 43.1

DOI: https://doi.org/10.65126/jocosir.v3i3

ISSN 2502-3357 (Online)

3.3 Model Training Artificial Neural Networks (ANN)

In this research, the authors adopted two established computational techniques to analyze the Pima Indians Diabetes dataset. The first technique employed a machine learning platform, TensorFlow/Keras, to construct an Artificial Neural Network (ANN) model that integrates both feedforward propagation and backpropagation algorithms. This combination enables the model to learn effectively from input features and optimize its weights during training with base manual calculation. The second technique involved utilizing a similar framework to perform comparative experiments, thereby assessing the accuracy and reliability of the model's predictive performance across different learning configurations.



(a) (b)

Figure 3. Comparison of ANN model testing results: (a) output derived from manual computation, and (b) output obtained from original data training using the implemented framework. Both ANN models achieved similar levels of accuracy (approximately 70%), with the framework-based implementation showing slightly better recall and F1-score. This indicates that while the manual ANN calculation correctly follows the theoretical formulation, the TensorFlow/Keras implementation provides improved weight optimization and generalization on the Pima Indians Diabetes dataset.

3.4 Logistic Regression

The training curve demonstrates a stable convergence pattern, where the loss function gradually decreases while the model accuracy increases and stabilizes after approximately 100 epochs. This behavior indicates that the logistic regression model successfully learns the underlying relationship between input variables and the target outcome without signs of overfitting or oscillation. The final accuracy of approximately 75% aligns with typical performance benchmarks for the Pima Indians Diabetes dataset, validating the model's reliability for binary classification tasks.

Table 2. training curve Logistic Regression

Metric	Nilai
Accuracy	0.757 (≈75.7%)
Precision	0.588
Recall	0.556
F1-score	0.571
Confusion Matrix	[[79, 21], [24, 30]]

The performance metrics obtained from the logistic regression model demonstrate a balanced and reliable classification ability for the Pima Indians Diabetes dataset.An accuracy of 75.7% indicates that the model is capable of correctly predicting the majority of diabetes and non-diabetes cases in the test set.

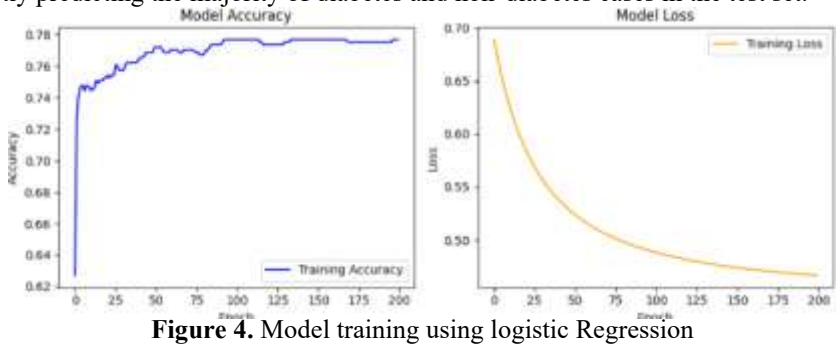


Figure 4. Model training using logistic Regression

Figure 4 presents the convergence pattern of the logistic regression model during training. The accuracy curve (left) shows a gradual improvement, stabilizing around 0.78 after approximately 100 epochs, indicating effective learning without overfitting. Meanwhile, the loss curve (right) decreases smoothly in an exponential manner, confirming that the model's learning rate and optimization process are properly tuned for efficient convergence.

3.4 K-mean Clustering

Base on figure 5 illustrates the K-Means clustering result for the Pima Indians Diabetes dataset projected into two principal components using PCA. The two color-coded clusters represent the algorithm's partitioning of the dataset without prior knowledge of the actual class labels. The visualization shows partial separation between clusters, with some overlap near the center, indicating that the features provide moderate discriminative power.

DOI: https://doi.org/10.65126/jocosir.v3i3

ISSN 2502-3357 (Online)

The obtained Adjusted Rand Index (ARI = 0.1614) confirms that while the clustering captures general structure, it does not closely align with the true diabetic and non-diabetic labels, reflecting the unsupervised nature of K-Means.

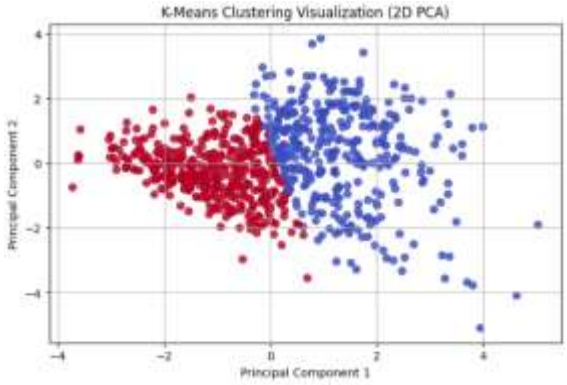


Figure 5. Model training using K-mean Clustering

3.5 Comparison model ANN with clustering and logical regression

Figure X illustrates the comparative performance of three machine learning algorithms—Artificial Neural Network (ANN), Logistic Regression, and K-Means Clustering—using multiple evaluation metrics.The color gradient in the heatmap represents the relative performance magnitude of each metric, where lighter shades indicate higher performance. As observed, Logistic Regression achieves the highest accuracy (0.757) and maintains a balanced precision (0.588), recall (0.556), and F1-score (0.571), indicating stable and consistent classification performance. The ANN model, while slightly lower in accuracy (0.708), still demonstrates reliable predictive capability with a comparable F1-score (0.563), suggesting that it successfully captures nonlinear relationships in the data.

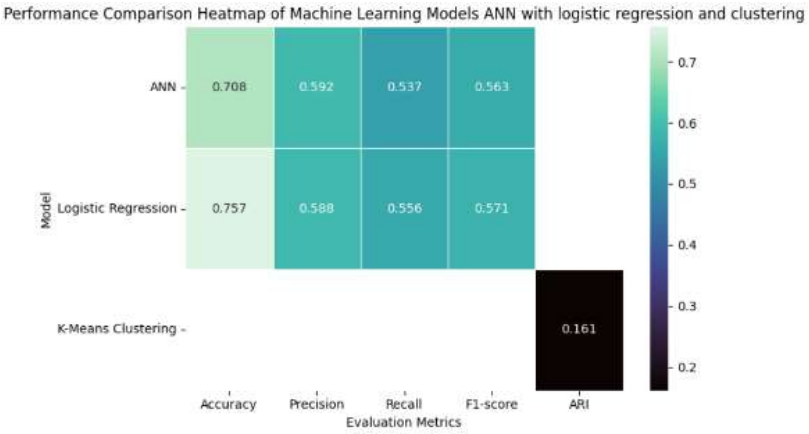


Figure 6. Comparison model training using Clustering & Logistic Regression

On the other hand, K-Means Clustering, being an unsupervised approach, produces a lower Adjusted Rand Index (ARI = 0.161), reflecting its limited alignment with the true class labels.

Overall, the heatmap confirms that supervised learning models outperform clustering-based methods in structured diagnostic datasets such as the Pima Indians Diabetes dataset.

Table 3. Intepreted model ANN testing

Model	Accuracy	Precision	Recall	F1-score	AdjustedRand Index (ARI)
ANN	0.7078	0.5918	0.5370	0.5631	–
Logistic Regression	0.7573	0.5882	0.5556	0.5714	–
K-Means Clustering	–	–	–	–	0.1614

3. Results and Discussion

This section presents the performance evaluation of three machine learning algorithms—Artificial Neural Network (ANN), Logistic Regression, and K-Means Clustering—implemented on the Pima Indians Diabetes dataset. Each model was trained and evaluated using relevant performance metrics, including Accuracy, Precision, Recall, and F1-score for supervised learning models, and Adjusted Rand Index (ARI) for the unsupervised clustering approach.

The ANN model was trained using a feedforward and backpropagation architecture with two hidden layers and a sigmoid activation function. The training process, illustrated in Figure 3, demonstrates a consistent increase in accuracy and a gradual decrease in loss across 200 epochs, indicating stable convergence and effective learning. The model achieved an accuracy of 0.7078, with precision = 0.5918, recall = 0.5370, and F1-score = 0.5631.

DOI: https://doi.org/10.65126/jocosir.v3i3

ISSN 2502-3357 (Online)

These results show that the ANN was able to generalize moderately well, capturing nonlinear relationships between the input features (such as glucose, BMI, and insulin levels) and diabetes outcomes. However, due to the relatively small dataset size and potential data imbalance, the performance did not surpass that of the simpler Logistic Regression model.

The Logistic Regression model achieved the highest overall accuracy of 0.7573 with precision = 0.5882, recall = 0.5556, and F1-score = 0.5714.The convergence curve, shown in Figure 4, depicts smooth and stable learning behavior, with loss decreasing exponentially and accuracy plateauing around epoch 100. This demonstrates that Logistic Regression effectively captured linear separability in the dataset and produced well-balanced predictive outcomes. As a result, it serves as the most stable and interpretable model among the tested algorithms. The comparative analysis, visualized in Figure 6 and summarized in Table 3, highlights that Logistic Regression outperforms ANN and K-Means Clustering in terms of classification accuracy and stability.

4. Conclusions

This study conducted a comparative analysis of three machine learning algorithms—Artificial Neural Network (ANN), Logistic Regression, and K-Means Clustering—for diabetes prediction using the Pima Indians Diabetes dataset. The experimental results reveal that Logistic Regression achieved the best overall performance, with an accuracy of 0.7573, outperforming the ANN model (0.7078) and the unsupervised K-Means clustering approach (ARI = 0.1614).Logistic Regression demonstrated stable convergence and strong interpretability, making it the most efficient model for structured medical data classification.

The ANN model, although slightly less accurate, showed reliable predictive capability and the ability to capture nonlinear relationships between physiological variables. This suggests its potential for improvement through hyperparameter tuning and architectural optimization.Meanwhile, K-Means clustering, despite its relatively low ARI value, provided insight into natural data patterns and patient grouping, highlighting its exploratory potential in unsupervised learning contexts.

Overall, the findings confirm that supervised learning methods, particularly Logistic Regression and ANN, are more effective for medical diagnosis prediction tasks, whereas clustering techniques are better suited for data exploration. In the future research may focus on hybrid or ensemble models combining the interpretability of Logistic Regression, the flexibility of ANN, and the unsupervised capability of clustering to enhance predictive performance in healthcare analytics.

5. Acknowledgment

The authors gratefully acknowledge the UCI Machine Learning Repository for granting access to the Pima Indians Diabetes Dataset, which formed the core basis of this study. They also extend their appreciation to the individuals and research teams responsible for compiling and maintaining the dataset, whose contributions have played a vital role in advancing the fields of machine learning and medical data analytics.

6. References

[1]A. M. Egan and S. F. Dinneen, “What is diabetes?,” Med. (United Kingdom), vol. 47, no. 1, pp. 1–4, 2019, doi: 10.1016/j.mpmed.2018.10.002.

[2]J. K. Mathew and S. S. Lakshmi, “A Study on Diagnosis of Diabetes Mellitus Based on Tongue Images with Various Methods,” Proc. Int. Conf. Comput. Commun. Secur. Intell. Syst. IC3SIS 2022, 2022, doi: 10.1109/IC3SIS54991.2022.9885616.

[3]G. Roglic, “WHO Global report on diabetes: A summary,” Int. J. Noncommunicable Dis., vol. 1, no. 1, pp. 3–8, 2016, doi: 10.4103/2468-8827.184853.

[4]I. Kusumastuty, D. M. Halimatussa’diah, C. S. Wilujeng, and F. A. Nugroho, “Gambaran Pola Asuh terhadap Kepatuhan Diet Anak dan Remaja dengan Diabetes Mellitus: Studi Kasus,” Indones. J. Hum. Nutr., vol. 7, no. 2, pp. 139–152, 2020, [Online]. Available: https://www.researchgate.net/profile/Fajar_Ari_Nugroho/publication/314713055_Kadar_NF-_Kb_Pankreas_Tikus_Model_Type_2_Diabetes_Mellitus_dengan_Pemberian_Tepung_Susu_Sapi/links/5b4dbf09aca27217ff9b6fcb/Kadar-NF-Kb-Pankreas-Tikus-Model-Type-2-Diabetes-Melli

[5]E. D. Parker et al., “Economic costs of diabetes in the u.S. in 2022,” Diabetes Care, vol. 47, no. 1, pp. 26–43, 2024, doi: 10.2337/dci23-0085.

[6]M. K. Hasan, M. A. Alam, D. Das, E. Hossain, and M. Hasan, “Diabetes prediction using ensembling of different machine learning classifiers,” IEEE Access, vol. 8, pp. 76516–76531, 2020, doi: 10.1109/ACCESS.2020.2989857.

[7]T. Guillod, P. Papamanolis, and J. W. Kolar, “Artificial neural network (ann) based fast and accurate inductor modeling and design,” IEEE Open J. Power Electron., vol. 1, pp. 284–299, 2020, doi: 10.1109/OJPEL.2020.3012777.

[8]H. Kahramanli and N. Allahverdi, “Design of a hybrid system for the diabetes and heart diseases,” Expert Syst. Appl., vol. 35, no. 1–2, pp. 82–89, 2008, doi: 10.1016/j.eswa.2007.06.004.

[9]W. Li, W. Wang, and W. Huo, “RegBoost: a gradient boosted multivariate regression algorithm,” Int. J. Crowd Sci., vol. 4, no. 1, pp. 60–72, 2020, doi: 10.1108/IJCS-10-2019-0029.

DOI: https://doi.org/10.65126/jocosir.v3i3

ISSN 2502-3357 (Online)
[10]M. R. Romadhon and F. Kurniawan, "A Comparison of Naive Bayes Methods, Logistic Regression and KNN for Predicting Healing of Covid-19 Patients in Indonesia," 3rd 2021 East Indones. Conf. Comput. Inf. Technol. EIConCIT 2021, pp. 41–44, 2021, doi: 10.1109/EIConCIT50028.2021.9431845.

[11]J. Li, W. Liu, M. Liu, and M. Huang, "Study on Chinese Text Clustering Algorithm Based on K-mean and Evaluation Method on Effect of Clustering for Software-intensive System," Proc. - 2020 Int. Conf. Comput. Eng. Appl. ICCEA 2020, pp. 513–519, 2020, doi: 10.1109/ICCEA50009.2020.00114.

[12]M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," Electron., vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.

[13]N. M. Karie, N. M. Sahri, W. Yang, C. Valli, and V. R. Kebande, "A Review of Security Standards and Frameworks for IoT-Based Smart Environments," IEEE Access, vol. 9, pp. 121975–121995, 2021, doi: 10.1109/ACCESS.2021.3109886.