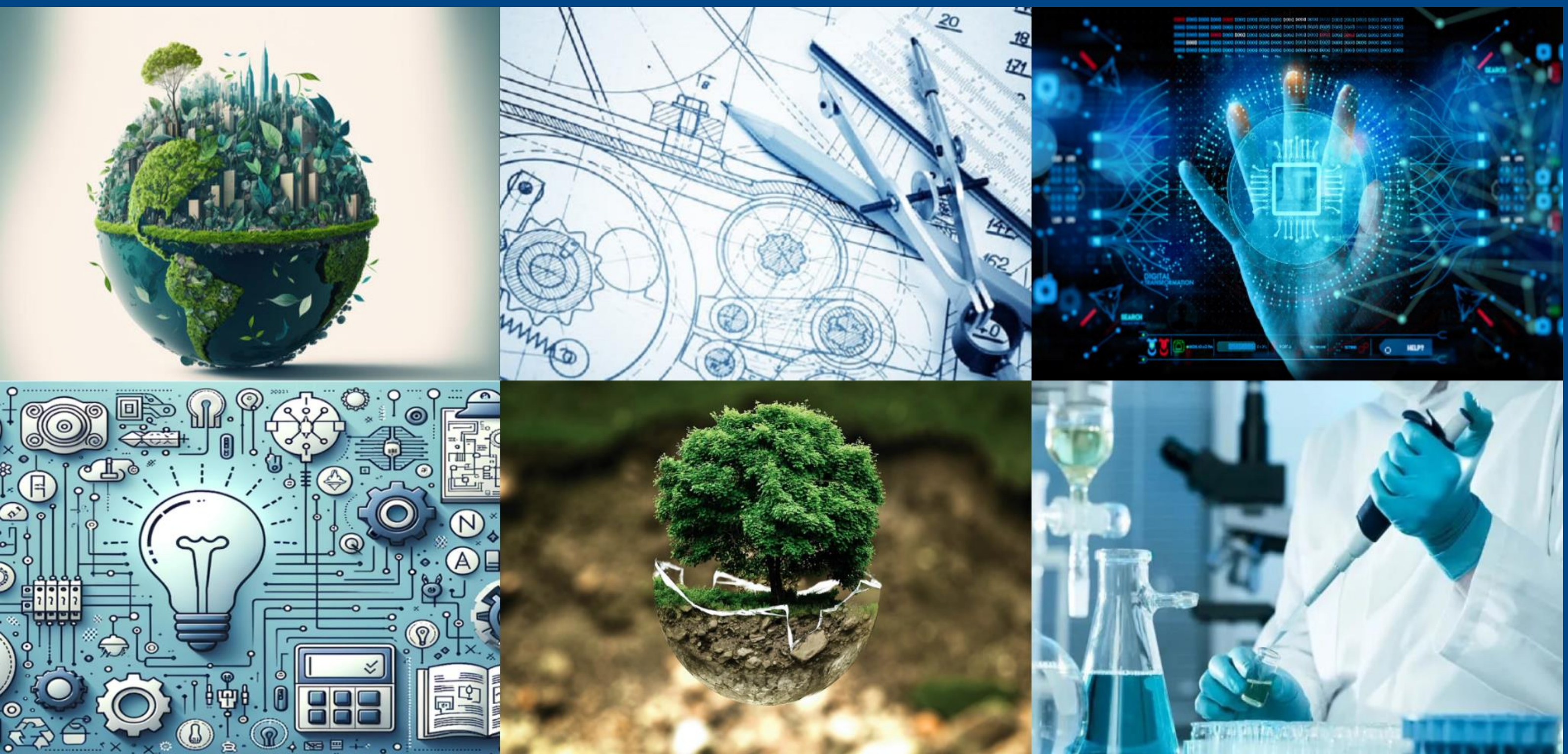
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 6, June 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Precision Monitoring: Unveiling the Science Behind Blood Sugar Monitoring using Machine Learning

Sowjanya M, Prakriti Bhandary

PG Student, St Josheph Engineering College, Vamanjoor, Manglore, India

Assistance Professor, St Josheph Engineering College, Vamanjoor, Manglore, India

ABSTRACT: Blood sugar monitoring is crucial to managing diabetes, as it allows individuals to maintain optimal glycemic control and prevent complications. People with diabetes should monitor their blood sugar levels because it allows them to control their illness more effectively and avoid problems. Traditional procedures, such as finger pricks, can be unpleasant and time-consuming for patients. New devices that employ machine learning algorithms are being developed to make blood sugar monitoring more convenient and less intrusive. In addition to doses of insulin or other specific to the patient's characteristics, these approaches estimate blood sugar levels using data gathered from sources such as continuous glucose monitoring (CGM), blood pressure measurements, or food information. This article discusses how machine learning algorithms may reliably predict blood sugar levels based on these parameters. It also goes into the steps involved in developing these systems, such as feature engineering, model selection, and assessment approaches. The suitability of several machine learning methods, including Support Vector Classification (SVC), Random forests, Naive Bayes Classifier, and K-Nearest Neighbour (KNN) Classifier, for blood sugar prediction tasks is examined. The potential for personalized diabetes care, early detection of hypo- and hyperglycaemic episodes, and integration with digital health platforms is emphasized. Overall, highlights the transformative potential of machine learning in revolutionizing blood sugar monitoring and improving the lives of people with diabetes.

KEYWORDS: Blood sugar monitoring, machine learning, diabetes management, glycaemic control, noninvasive, automated methods, continuous glucose monitoring (CGM), support vector classification, random forests, and KNN classifier.

I. INTRODUCTION

Diabetes, a chronic metabolic disorder, affects millions worldwide, necessitating continuous monitoring of blood glucose levels to prevent complications. Traditional methods of blood sugar monitoring often pose challenges in terms of accuracy, convenience, and real-time feedback. Enter Precision Monitoring, a revolutionary approach that harnesses the power of machine learning algorithms to transform the landscape of diabetes management. Advancements in healthcare technology have paved the way for a new era of personalized and precise monitoring of medical conditions. One such groundbreaking development is the integration of machine learning into blood sugar monitoring, ushering in a paradigm shift in diabetes management. This introduction aims to delve into the realm of Precision Monitoring, shedding light on the intersection of cutting-edge technology and the critical task of blood sugar regulation.

At its core, Precision Monitoring combines the principles of machine learning with the vast datasets generated by continuous blood glucose monitoring devices. By leveraging intricate patterns, correlations, and predictive analytics, these algorithms empower healthcare professionals and individuals with diabetes to make informed decisions about insulin dosages, dietary choices, and lifestyle modifications. This interdisciplinary fusion of medicine and machine learning not only enhances the accuracy of blood sugar predictions but also allows for the early detection of trends and anomalies. The ability to anticipate glucose fluctuations enables proactive interventions, ultimately preventing severe episodes and promoting long-term health outcomes for individuals managing diabetes.

Machine learning algorithms such as SVM, RF, KNN, and NB Classifier are currently utilized for predicting diabetes using age, BMI, blood pressure, and glucose levels. SVM performs the best with the highest accuracy when tested against the PIMA Indian dataset, which contains data on 768 women with diabetes. These algorithms are compared based on parameters such as precision, accuracy, sensitivity, and specificity. This indicates that SVM can reliably determine if an individual has diabetes using this dataset; other algorithms, such as RF, KNN, among others, and NB Classifier, function well but may not be as accurate as SVM in more complicated datasets. These discoveries have practical implications for diabetes treatment and clinical decision-making.

II. LITERATURE REVIEW

Chou et al. (2023) [1] Diabetes is a serious chronic condition. Using cutting-edge machine intelligence techniques, the study aimed to improve knowledge of the elements that contribute to diabetes onset. The study carried very extensive data processing using eight essential characteristics: age, BMI, sebum thickness, insulin level, number of pregnancies, blood glucose level, systolic blood pressure, and diabetes genetic functionality. The next processes included training, testing, cross-validation, and comparison to provide data for model performance analysis.

Ahmad et al. (2021) [2] Gestational diabetes mellitus (GDM) is a kind of diabetes that occurs during pregnancy as a result of hormonal or metabolic abnormalities. This disease is characterized by insulin resistance, which can result in increased blood sugar levels.

M.K.Hasan et al. (2020) [3] Diabetes may be reliably predicted using an ensemble model using a specific dataset known as PID. They stress that accurate findings may be obtained from low-quality data, hence the quality of the dataset is essential for making trustworthy predictions. They propose a preprocessing method that emphasizes the removal of outliers and the filling in of missing values as critical components to improve the quality of the dataset. This strategy considerably enhances prediction reliability, as seen by their findings, which demonstrate that using this preprocessing method improves dataset quality, resulting in more accurate diabetes prediction results.

Vieira S et al. (2020) [4] A study found that machine learning (ML) approaches may effectively predict diabetic complications. These approaches, such as random forest, which performed the best overall, are trained on labeled datasets using algorithms that detect patterns in the data and generate predictions. ML approaches employ existing data to create predictions without being specifically trained to do so, classifying them as a type of artificial intelligence.

Abhari S, Handelman GS (2018, 2019) [5] Machine learning (ML) technologies have grown in popularity in healthcare over the last several years due to their capacity to enhance diagnosis, treatment planning, and overall healthcare decision-making. These techniques employ algorithms to quickly and precisely analyze vast volumes of medical information and produce insight that may assist healthcare professionals in making defensible judgments.

Khanam et al. (2021) [6] Five unique models were evaluated using the PIMA Indian Diabetes dataset, with an accuracy of 7-97%. The algorithms KNN, RF, NB, AB, LR, and SVM were used. With two hidden layers and 400 epochs, the NN model achieved an accuracy rate of around 86% over all epochs (200, 400, 800). These models all have accuracy ratings of 70% or above. This shows that the neural network model's capacity to anticipate diabetes is quite accurate.

Soni et al. (2020) [7] The study's goal was to design and develop a machine-learning system capable of predicting diabetes. The technique employed a variety of algorithms, including Support Vector Machine, K-Nearest Neighbours, Random Forest, Decision Tree, Logistic Regression, and Gradient Boosting classifiers. The method predicted diabetes with a success rate of around 77%. This implies that physicians may use this technique to anticipate diabetes sooner, potentially saving lives by preventing consequences such as kidney damage or blindness.

Mujumdar et al. (2019) [8] The machine learning techniques applied to the dataset for this study yielded the best accuracy, 96%, for Logistic Regression. When using a pipeline strategy, the AdaBoost classifier achieved 98.8% accuracy, making it the most effective model. The study employed two separate datasets to test the accuracies of machine learning algorithms, and the findings revealed that diabetes prediction using the present set of data was more accurate and precise than with the prior one. The study advises more investigation into the likelihood of non-diabetics getting diabetes in the future.

© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806097

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. METHODOLOGY

Data Collection: 'Diabetes.csv' is a CSV file from which the data is gathered using the Pandas package. There are 9 columns and 768 rows in this data collection. The column contains the following attributes: age, results, BMI, insulin, glucose, pregnancies, skin thickness, blood pressure, and insulin.

Data Preprocessing: Data preprocessing involves preparing the data for modelling. The dataset is explored using methods like head(), info(), and describe() to understand its structure and characteristics. Features and labels are separated: X contains the features, and y contains the target variable ('Outcome'). Feature scaling is applied using StandardScaler from sklearn. preprocessing to standardize the feature values.

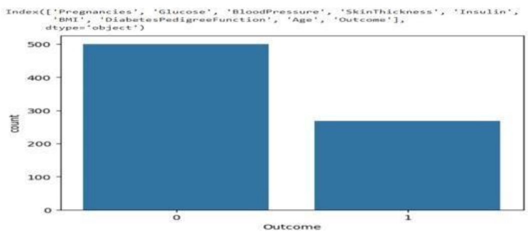
Data Splitting: Utilizing the train_test_split function from sklearn.model_selection, the data is divided into training and testing sets. The training phase uses 70% of the data, while the testing phase uses 30%.

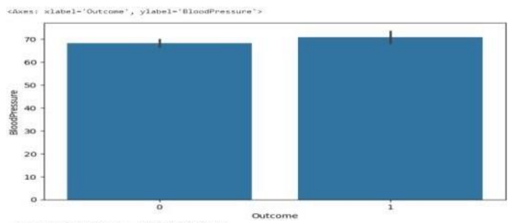
Data Visualization: Data visualization is performed to gain insights into the dataset and understand relationships between features and the target variable. Seaborn and Matplotlib are used to create count plots and bar plots, visualizing the distribution of classes and the relationship between specific features and the diabetes outcome.

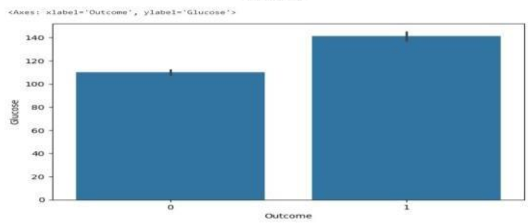
Machine Learning Classification Algorithms:
Several machine learning classification algorithms are implemented to predict diabetes outcomes: Support Vector Classifier (SVC), Naive Bayes Classifier, K-Nearest Neighbors (KNN) Classifier, Random Forest Classifier (RFC). Each algorithm is trained, and its performance is evaluated using metrics such as confusion matrix, classification report, and accuracy score. The Random Forest Classifier's results are also saved in a CSV file named 'diabetes_submission.csv'.

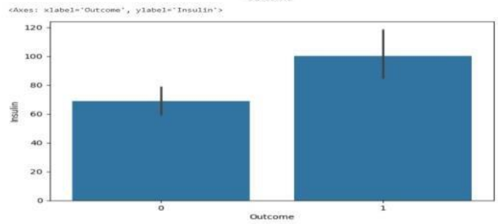
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 768 entries, 0 to 767
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               768 non-null    int64
 1   Glucose                   768 non-null    int64
 2   BloodPressure             768 non-null    int64
 3   SkinThickness             768 non-null    int64
 4   Insulin                   768 non-null    int64
 5   BMI                       768 non-null    float64
 6   DiabetesPedigreeFunction  768 non-null    float64
 7   Age                       768 non-null    int64
 8   Outcome                   768 non-null    int64
dtypes: float64(2), int64(7)
memory usage: 54.1 KB
```

	count	mean	std	min	25%	50%	75%	max
Pregnancies	768.0	3.845052	3.369578	0.000	1.00000	3.0000	6.00000	17.00
Glucose	768.0	120.894531	31.972618	0.000	99.00000	117.0000	140.25000	199.00
BloodPressure	768.0	69.105469	19.355807	0.000	62.00000	72.0000	80.00000	122.00
SkinThickness	768.0	20.536458	15.952218	0.000	0.00000	23.0000	32.00000	99.00
Insulin	768.0	79.799479	115.244002	0.000	0.00000	30.5000	127.25000	846.00
BMI	768.0	31.992578	7.884160	0.000	27.30000	32.0000	36.60000	67.10
DiabetesPedigreeFunction	768.0	0.471876	0.331329	0.078	0.24375	0.3725	0.62625	2.42
Age	768.0	33.240885	11.760232	21.000	24.00000	29.0000	41.00000	81.00
Outcome	768.0	0.348958	0.476951	0.000	0.00000	0.0000	1.00000	1.00









© 2025 IJMRSET | Volume 8, Issue 6, June 2025| DOI:10.15680/IJMRSET.2025.0806097

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Grouping the data according to the 'Outcome' variable—where '1' denotes the presence of diabetes and '0' denotes its absence—was the first step in the analysis of the diabetes dataset. A count plot was then used to display the number of people in each class, giving a clear picture of the distribution of results in the dataset. Following that, particular characteristics like ‘Glucose’, ‘Blood Pressure’, and ‘Insulin’ were visually examined in connection to the diabetes result. The average amounts of these characteristics for people with and without diabetes were compared using bar graphs. The visuals show possible variations in blood pressure, insulin, and glucose levels between the two result groups, laying the groundwork for additional research and feature analysis to determine the variables influencing the existence of diabetes in the dataset.

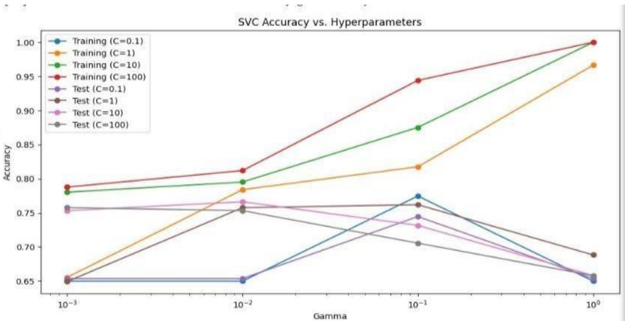
IV. IMPLEMENTATION

Support Vector Classification (SVC):

Support Vector Classification (SVC) is a machine learning technique classified as Support Vector Machines (SVM). Its primary purpose is to divide data points into two or more categories. It is very effective for binary classification tasks. In layman's words, SVC aids in distinguishing between two groups by determining the optimal line or plane that divides them.

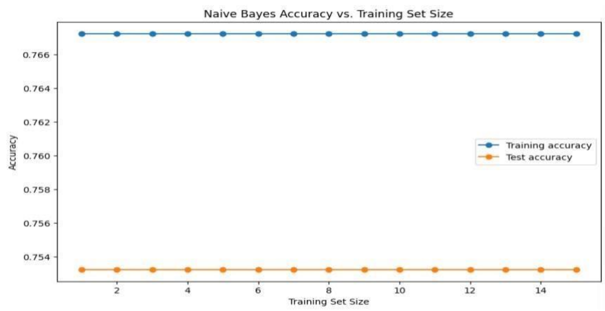
To train a Support Vector Classifier (SVC), we employ an accuracy versus. hyperparameter plot. This graphic depicts how the model's performance varies as we modify two critical parameters, gamma and C (regularisation parameter We can identify the optimal SVC model that has the most effective hyper parameters by experimenting with different values. After determining the best model, we train it using the scaled training set and then test it on a distinct collection of data known as the test set. The confusion matrix and classification report show how well the model performs in each class. This model's precision and recall vary per class, and its overall accuracy is 75.32%. The model was 78% accurate on the training set and 75% on the test set.

```
Confusion Matrix:
[[128  23]
 [ 34  46]]

Classification Report:
              precision    recall  f1-score   support

           0       0.79      0.85      0.82       151
           1       0.67      0.57      0.62        80

    accuracy                           0.75       231
   macro avg       0.73      0.71      0.72       231
weighted avg       0.75      0.75      0.75       231

Accuracy Score: 0.7532467532467533
Accuracy of SVC classifier on training set: 0.78
Accuracy of SVC classifier on test set: 0.75
```



Naive Bayes Classifier:

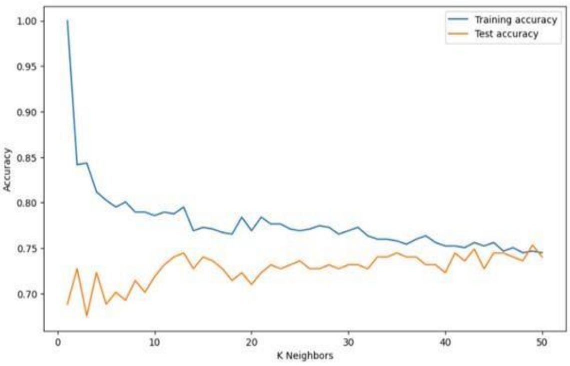
When predicting the category or class of an observation is the aim of the job, Naive Bayes is more frequently employed. How well the Naive Bayes classifier generalizes to fresh, untested data is shown by the final accuracy scores. Both the Support Vector Classifier (SVC) and the Naive Bayes classifier have an overall accuracy score of about 75%. The Naive Bayes classifier performs somewhat worse than the SVC classifier, but it still does well, with an accuracy of 77% on the training set and 75% on the test set. The confusion matrix demonstrates that both models perform better for class 0 (accuracy of about 81%) than for class 1 (accuracy of approximately 64% for the SVC classifier and 65% for the Naive Bayes classifier).



K-Nearest Neighbour (KNN) Classifier:

The K-Nearest Neighbours (KNN) algorithm is an easy-to-use categorization approach in machine learning that allows us to predict the category of a new data point based on the categories of its nearest neighbors. To determine the optimal value of K(number of neighbors), we ran the algorithm with different values of K and compared the accuracy of the training and test sets. The findings indicated that K=8 performed well on both sets of data (78% training accuracy and 71% test accuracy). The confusion matrix and classification report offered further information about the model's performance in each class. Class 0 had better recall and accuracy than class 1, suggesting superior prediction of the majority class. Overall, the KNN classifier with K=8 obtained a competitive accuracy of 71% on the test set but fell somewhat short of the Support Vector Classifier (SVC) and Naive Bayes classifiers stated previously in this discussion.

```
              precision    recall  f1-score   support

           0       0.76      0.82      0.79       151
           1       0.60      0.51      0.55        80

    accuracy                           0.71       231
   macro avg       0.68      0.67      0.67       231
weighted avg       0.71      0.71      0.71       231


Accuracy Score : 0.7142857142857143
Accuracy of KNN classifier on training set: 0.79
Accuracy of KNN classifier on test set: 0.71
```

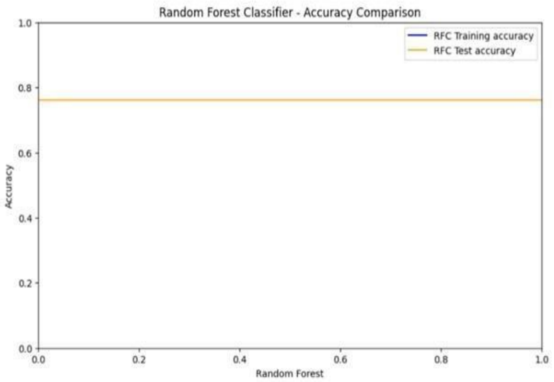


Random Forest:

Random Forest is a prominent machine learning approach that makes predictions using multiple decision trees. The Random Forest classifier outperformed the prior models, with an overall accuracy score of around 76%. The model achieved an accuracy score of exactly 100% on the training set, indicating that it correctly predicted every sample during training. The confusion matrix and classification report revealed that the model had greater accuracy and recall for class 0 (approximately 81% and 83%, respectively) than for class 1. This suggests that the Random Forest classifier performs better at predicting the majority class than at properly categorizing class 1 samples. It is critical to understand that Random Forest is an ensemble learning approach that integrates numerous decision trees during training, which can produce better results than a single decision tree.

When we run a machine learning model on given data, it generates labeled predictions for each piece of data. In this scenario, the labels are either 0 or 1, which might indicate two distinct groups or classes. Each row in the data represents a unique piece of data, and the model guesses which class it belongs to based on its characteristics. To illustrate, the model predicts that the row containing "Id 31" will belong to class 1 if the "Outcome" column has a value of 1, which could indicate either positive or true for that class. By comparing these projected labels to the actual labels (which may be from another dataset), we can assess the model's ability to make accurate predictions for new, unknown data.

```
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.83      0.82       151
           1       0.67      0.62      0.65        80

    accuracy                           0.76       231
   macro avg       0.74      0.73      0.73       231
weighted avg       0.76      0.76      0.76       231


Accuracy Score: 0.76
Training Accuracy Score: 1.00
Testing Accuracy Score: 0.76
```



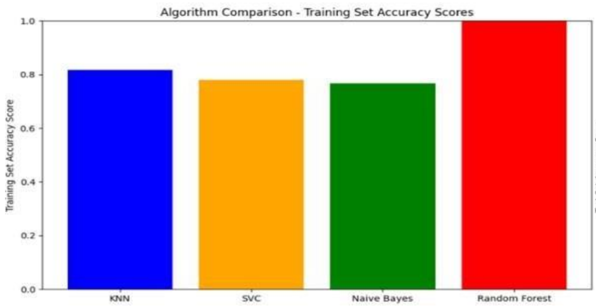
ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

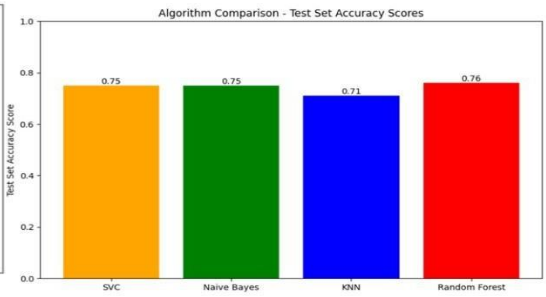
International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. RESULTS

S.no:	Methodologies:	Accuracy:
01	Support Vector Classifier	75.32%
02	Naïve Bayes Classifier	75%
03	KNN	71%
04	Random Forest	76%





Here Random Forest has more accuracy with 76%.

VI. CONCLUSION

Blood sugar monitoring model using machine learning algorithms to enhance diabetes management. The methodologies employed included Support Vector Classifier (SVC), Naive Bayes Classifier, K-Nearest Neighbors (KNN) Classifier, and Random Forest Classifier. Our analysis revealed that the Random Forest Classifier achieved the highest accuracy at 76%, outperforming the other models.

This comparative analysis indicates that while all the algorithms demonstrated commendable performance in predicting diabetes outcomes, Random Forest proved to be the most effective, followed closely by SVC and Naive Bayes. KNN exhibited competitive accuracy but fell slightly short of the top-performing models. These findings underscore the potential of machine learning in revolutionizing blood sugar monitoring by providing more accurate and non-invasive methods for diabetes management.

REFERENCES

[1] Chou, C. Y., Hsu, D. Y., & Chou, C. H. (2023). Predicting the onset of diabetes with machine learning methods. Journal of Personalized Medicine, 13(3), 406.
[2] Ahmad, H.F.; Mukhtar, H.; Alaqail, H.; Seliaman, M.; Alhumam, A. Investigating Health-Related Features and Their Impact on the Prediction of Diabetes Using Machine Learning. Appl. Sci. 2021, 11, 1173.
[3] M. K. Hasan, M. A. Alam, D. Das, E. Hossain and M. Hasan, "Diabetes Prediction Using Ensembling of Different Machine Learning Classifiers," in IEEE Access, vol. 8, pp. 76516-76531, 2020, doi: 10.1109/ACCESS.2020.2989857.
[4] Vieira S, Lopez Pinaya WH, Mechelli A. Introduction to machine learning. In: Mechelli A, Vieira S eds. Machine Learning. London: Academic Press; 2020:1-20.
[5] Abhari S, Niakan Kalhori SR, Ebrahimi M, Hasannejadasl H, Garavand A. Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods. Healthc Inform Res. 2019;25(4):248-261.
[6] Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

medicine. J Intern Med. 2018;284(6):603-619.

[7] Khanam, J. J., & Foo, S. Y. (2021). A comparison of machine learning algorithms for diabetes prediction. Ict Express, 7(4), 432-439.

[8] Soni, M., & Varma, S. (2020). Diabetes prediction using machine learning techniques. International Journal of Engineering Research & Technology (Ijert) Volume, 9.

[9] Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. Procedia Computer Science, 165, 292-299.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY