STM JOURNALS

Journal of
Artificial Intelligence Research & Advances

ISSN: 2395-6720
Volume 9, Issue 1, 2022
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Review JoAIRA

Diabetes Mellitus Prediction and Classification with a Randomizable filtered Classifier Ensemble Algorithm of the K-Nearest Neighbor

Baku Agyo Raphael[1,*], Usman Zailani[2], Fred Moveh[3], John Abiodun Oladunjoye[4], Useni Adi Aji[5]

Abstract

Diabetes is a chronic and life-threatening condition caused by high levels of sugar glucose in the blood, which primarily affects the elderly in all societies. This has become a huge global health issue. Patients, diseases, causes, and medical services are all collected in vast amounts by medical care systems and hospitals. Diabetes has been the subject of numerous studies aimed at detecting and preventing the disease. This study presents an Ensemble of Randomizable Filtered Classifier K-Nearest Neighbor Algorithms as a data mining tool for predicting patients who are prone to develop diabetes mellitus. The Pima Indian Diabetes Dataset was used, which contains information on patients with diabetes and the possibility of acquiring diabetes. For design and implementation, the study used an ensemble of randomizable filtered classifiers K-Nearest Neighbor Algorithm and WEKA Software. The proposed approach predicts the likelihood of developing diabetes mellitus, according to the findings. It is suggested that the plan be implemented by the health-care system in order to discover diabetes mellitus patients early.

Keywords: Diabetes dataset, diabetes mellitus, K-nearest neighbor algorithm, machine learning, Random Forest

***Author for Correspondence**
Baku Agyo Raphael
E-mail: bakuralph@fuwukari.edu.ng

[1]Lecturer II, Department of Computer Science, Federal University Wukari, Km 10 Katsina-ala Road, Wukari, PMB 2010, Taraba State, Nigeria
[2]Principal Data processing officer, Department of Information Technology, Federal University Wukari, Km 10 Katsina-ala Road, Wukari, PMB 2010, Taraba State, Nigeria
[3]Assistant Lecturer, Department of Information Technology, Modibbo Adama University, Yola, Adamawa State, Nigeria
[4] Senior Lecturer, Department of Computer Science, Federal University Wukari, Km 10 Katsina-ala Road, Wukari, PMB 2010, Taraba State, Nigeria
[5]Graduate Assistant, Department of Computer Science, Federal University Wukari, Km 10 Katsina-ala Road, Wukari, PMB2010,Taraba State, Nigeria

Received Date: April 20, 2022
Accepted Date: April 26, 2022
Published Date: April 30, 2022

Citation: Baku Agyo Raphael, Usman Zailani, Fred Moveh, John Abiodun Oladunjoye, Useni Adi Aji. Diabetes Mellitus Prediction and Classification with a Randomizable filtered Classifier Ensemble Algorithm of the K-Nearest Neighbor. Journal of Artificial Intelligence Research & Advances. 2022; 9(1): 43–58p.

INTRODUCTION

Diabetes Mellitus is a common human disease caused by a group of metabolic disorders wherein blood sugar levels are abnormally high for an extended length of time. It affects various organs in the human body, causing damage to a vast variety of physiological systems, including the blood vessels and neurons [1]. Diabetes Mellitus has now become a frequent worldwide health concern that can lead to a variety of global health complications such as heart disease, kidney failure, vision impairment, and so on. According to the World Health Organization (WHO), in 2013, 300 million people around the world will be affected. By 2025, diabetes will be the leading cause of death. Currently, 20 million people in Sub-Saharan Africa have diabetes; roughly 62% remain undiagnosed, and the figure is anticipated to rise to 41.4 million by 2035, representing a 109.1% increase. Nigeria has the greatest number of diabetics in Sub-Saharan Africa, with an estimated 3.9 million adults aged 20 to 77 suffering from the disease [2].

Type I and Type II diabetes are the two forms of

diabetes. Type I diabetes, commonly known as insulin-dependent diabetes, is a chronic illness in which the pancreas produces little or no insulin. The human body cannot utilize insulin properly in type II diabetes. Non-insulin-dependent diabetes is yet another name for it. Several approaches are being used by many studies to predict diabetes at an early stage. Diabetes is diagnosed using a variety of readily available conventional procedures that rely on physical and substance tests. To address these concerns, a number of data mining techniques have been developed. Classification techniques such as Decision Tree were used to classify patients with Diabetes Mellitus [3-9].

PROBLEM STATEMENT

According to the 6th Edition of the International Diabetes Federation (IDF) Atlas, Nigeria is the leading country in Africa in terms of diabetes prevalence; 3.9 million people have diabetes, with 105,091 diabetes-related deaths in 2013, and this number is expected to rise by 125,000 per year between 2010 and 2030.

According to studies, diabetes prevalence ranged from a low of 0.8% among adults in rural highland residents to over 7% in urban Lagos, with a national average of 2.2% [10, 11]. Diabetes leads to the development of heart disease, renal disease, pneumonia, bacteremia, and tuberculosis in terms of morbidity [12–14]. People with diabetes mellitus are three times more likely to get tuberculosis, and diabetes is estimated to be a predisposing factor in around 15% of tuberculosis cases worldwide. Many studies and researches have been conducted to predict and classify the level of diabetes mellitus in patients, but the most commonly used classification techniques are supervised machine learning algorithms, which include a target/outcome variable or dependent variable that must be predicted from a set of predictors or independent variables [15–20]. Unlike unsupervised learning, which has no aim or outcome variable, supervised learning has a target or outcome variable.

This work highlighted this as a problem and developed a supervised deep learning, an ensemble of randomizable filtered classifier KNN machine learning method for diabetes mellitus prediction and classification [21–23].

In light of the foregoing, this study provides an improved technique that uses an ensemble of randomizable filtered classifier K-Nearest Neighbor Algorithm to predict and classify diabetes mellitus.

RESULTS

Experimental Results

Data preparation is a crucial step in the data mining process. The WEKA tool is freeware application software that includes built-in data mining and knowledge analysis tools [24–28]. Data preparation, grouping, classification, regression, and association are among the technologies available. As illustrated in Figures 1–7, the results at each level can be evaluated through visualization and plain text. The data to be studied is fed into WEKA, and a prediction can be generated. As shown in Figure 4, there are two approaches for rescaling data: standardization and normalizing [29].

As illustrated in Figure 5, the Pima Indian Diabetes dataset comprises two cases that were tested positive and two cases that were tested negative.

There are no missing values in the database, but there are zeros in several fields. There were 392 occurrences with no missing values after removing the zero values (130 tested positive and 292 tested negative). The values in the range [0,1] are scaled during normalisation.

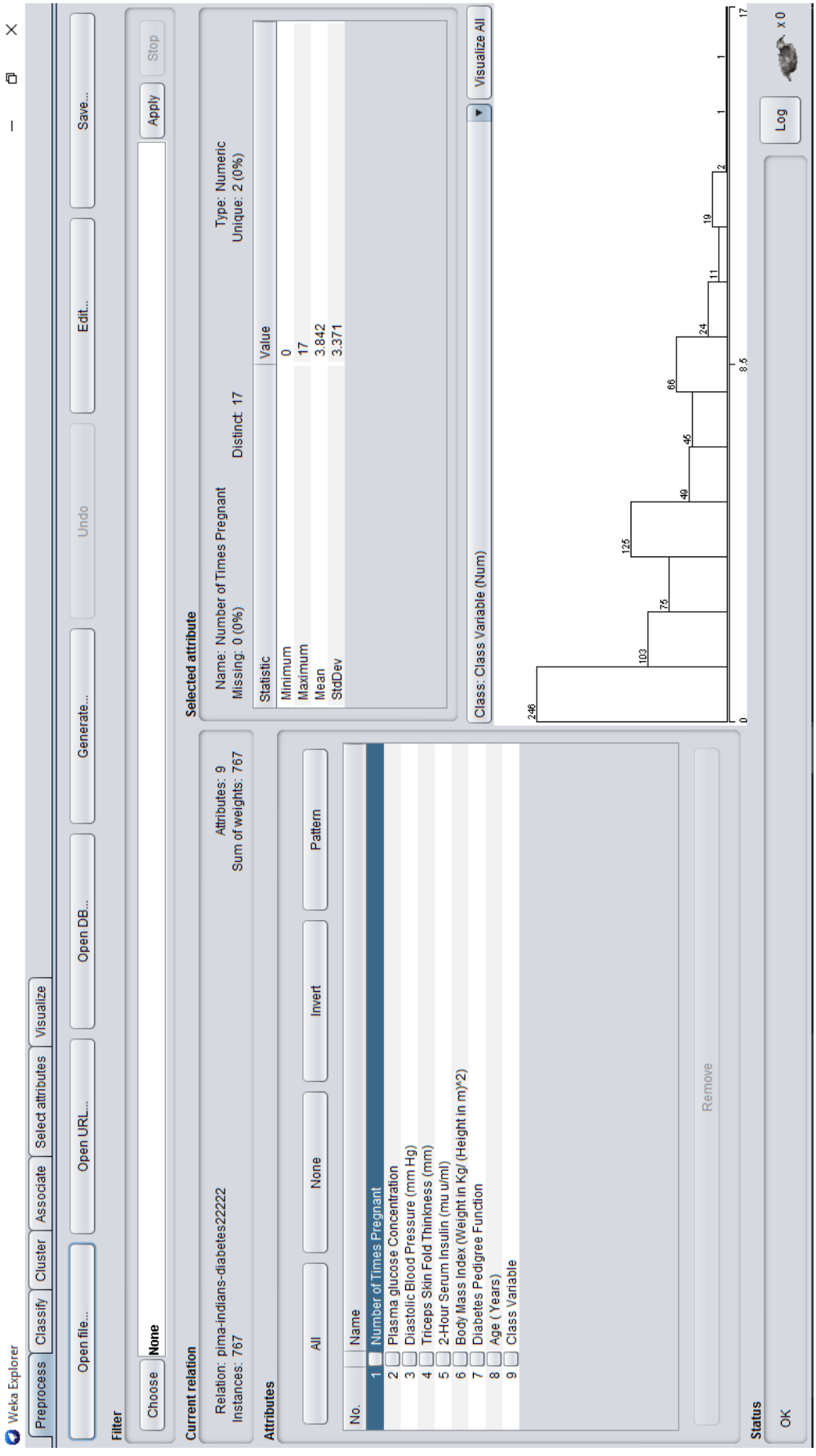

Figure 1. Dataset Preprocessing.

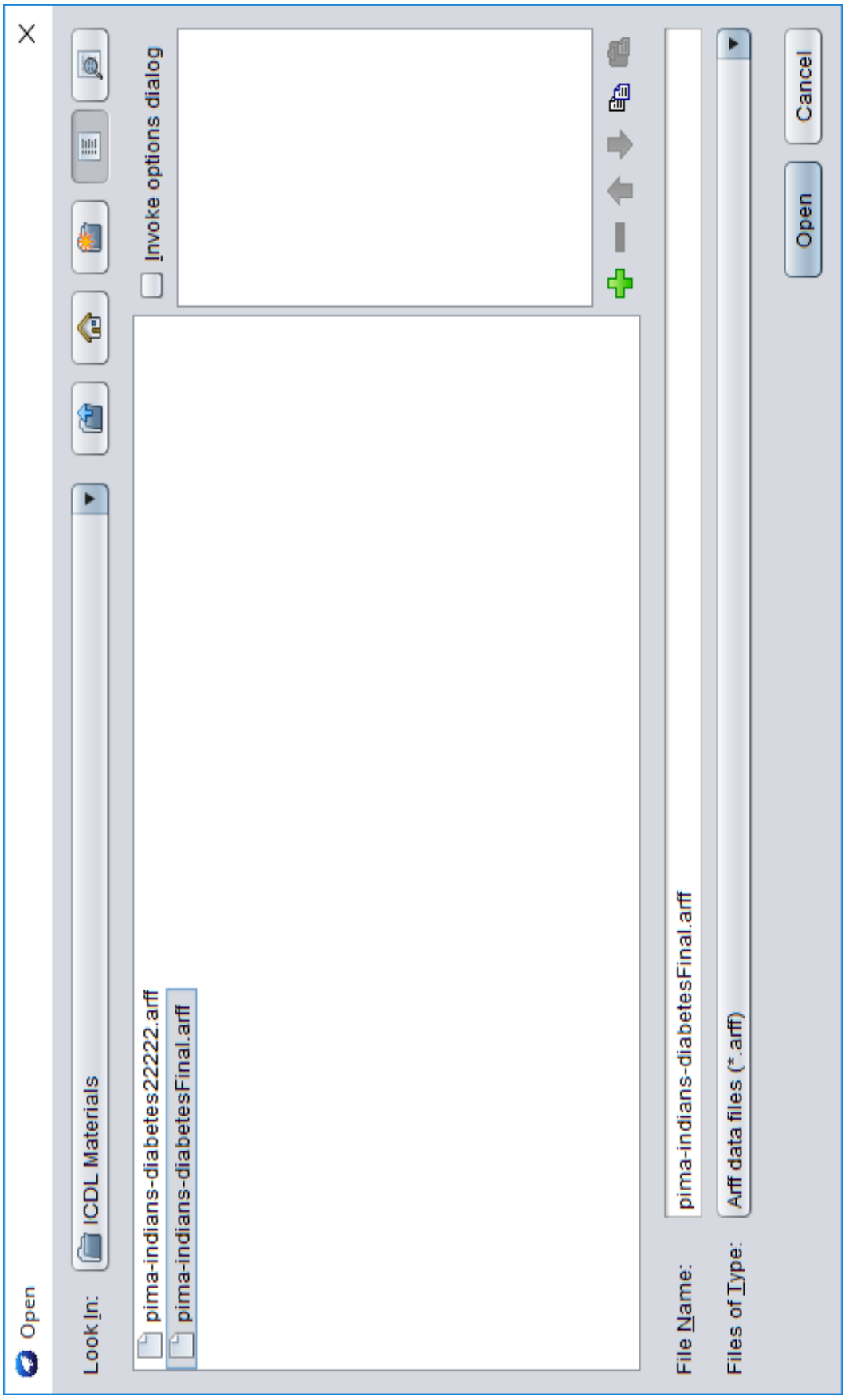

Figure 2. The Pima Indian Dataset was put into Weka for preprocessing.

ARFF-Viewer - C:\Users\compvil\Desktop\ICDL Materials\pima-indians-diabetesFinal.arff

File Edit View

pima-indians-diabetesFinal.arff

Relation: pima-indians-diabetes22222

No.	1: Number of Times Pregnant Numeric	2: Plasma glucose Concentration Numeric	3: Diastolic Blood Pressure (mm Hg) Numeric	4: Triceps Skin Fold Thinkness (mm) Numeric	5: 2-Hour Serum Insulin (mu u/ml) Numeric	6: Body Mass Index (Weight in Kg/ (Height in m)^2) Numeric	7: Diabetes Pe Nur
1	0.0	60.0	57.0	0.0	0.0	21.7	
2	0.0	76.0	67.0	0.0	0.0	45.3	
3	0.0	0.0	73.0	0.0	0.0	21.1	
4	0.0	52.0	74.0	10.0	36.0	27.8	
5	0.0	88.0	78.0	29.0	40.0	36.9	
6	0.0	64.0	84.0	22.0	66.0	35.8	
7	0.0	82.0	84.0	31.0	125.0	38.2	
8	0.0	68.0	86.0	32.0	0.0	35.8	
9	0.0	68.0	91.0	32.0	210.0	39.9	
10	0.0	80.0	91.0	0.0	0.0	32.4	
11	0.0	60.0	93.0	25.0	92.0	28.7	
12	0.0	100.0	93.0	39.0	72.0	43.4	
13	0.0	60.0	93.0	0.0	0.0	35.3	
14	0.0	0.0	94.0	0.0	0.0	0.0	
15	0.0	70.0	94.0	27.0	115.0	43.5	
16	0.0	85.0	95.0	25.0	36.0	37.4	
17	0.0	80.0	95.0	45.0	92.0	36.5	
18	0.0	64.0	95.0	39.0	105.0	44.6	
19	0.0	64.0	97.0	36.0	100.0	36.8	
20	0.0	82.0	98.0	15.0	84.0	25.2	
21	0.0	0.0	99.0	0.0	0.0	25.0	
22	0.0	88.0	100.0	60.0	110.0	46.8	
23	0.0	70.0	100.0	26.0	50.0	30.8	
24	0.0	65.0	101.0	28.0	0.0	24.6	
25	0.0	76.0	101.0	0.0	0.0	35.7	
26	0.0	64.0	101.0	17.0	0.0	21.0	
27	0.0	62.0	101.0	0.0	0.0	21.9	
28	0.0	75.0	102.0	23.0	0.0	0.0	
29	0.0	52.0	102.0	0.0	0.0	25.1	
30	0.0	64.0	102.0	46.0	78.0	40.6	
31	0.0	78.0	102.0	40.0	90.0	34.5	
32	0.0	86.0	102.0	17.0	105.0	29.3	
33	0.0	76.0	104.0	0.0	0.0	18.4	
34	0.0	64.0	104.0	23.0	116.0	27.8	
35	0.0	64.0	104.0	37.0	64.0	33.6	
36	0.0	64.0	105.0	41.0	142.0	41.5	
37	0.0	84.0	105.0	0.0	0.0	27.9	
38	0.0	68.0	105.0	22.0	0.0	20.0	
39	0.0	90.0	105.0	0.0	0.0	29.6	

Figure 3. In Weka, a CSV file is converted to an Attributes Relative File Format.

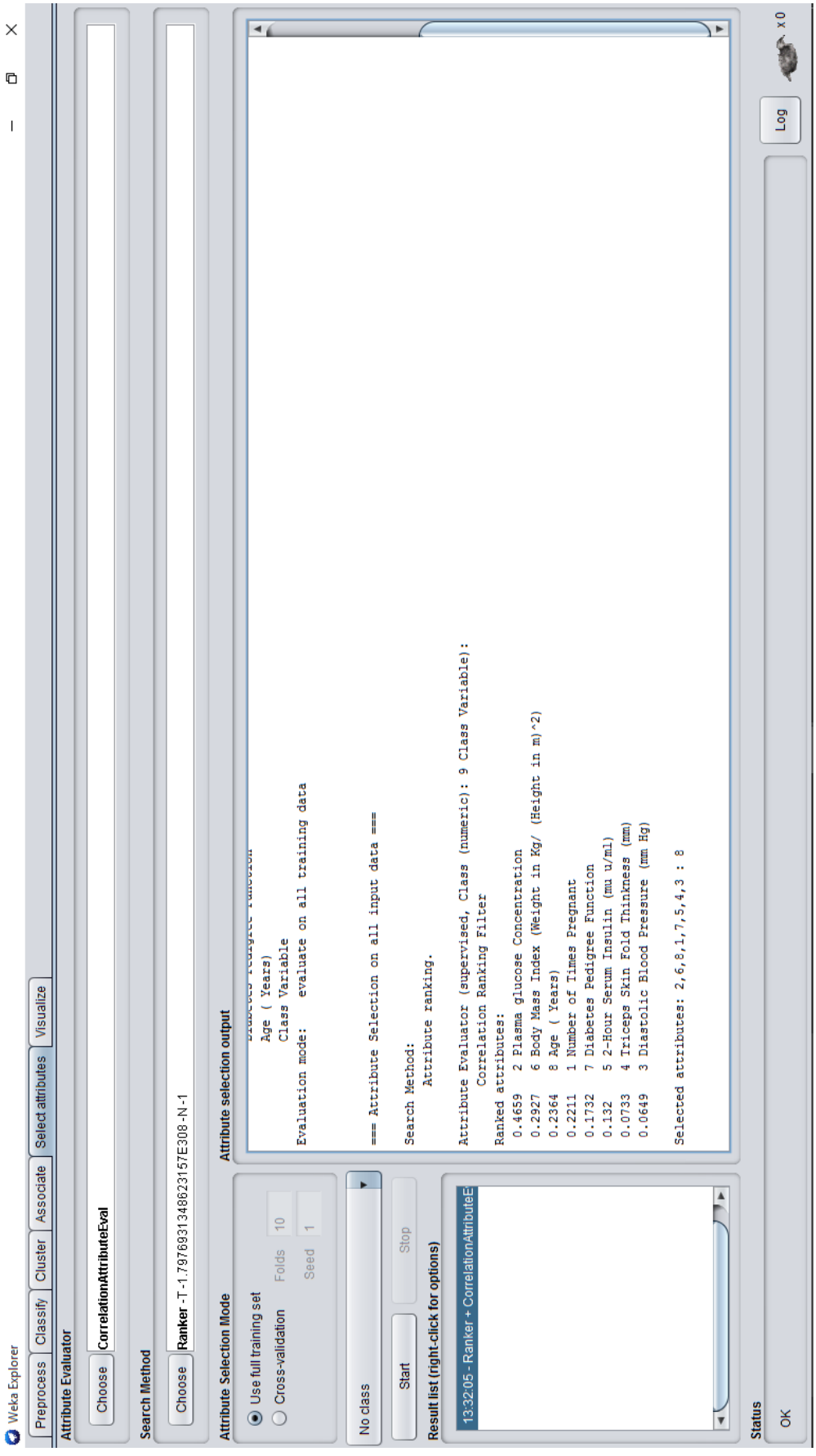

Figure 4. Normalization data rescaling.

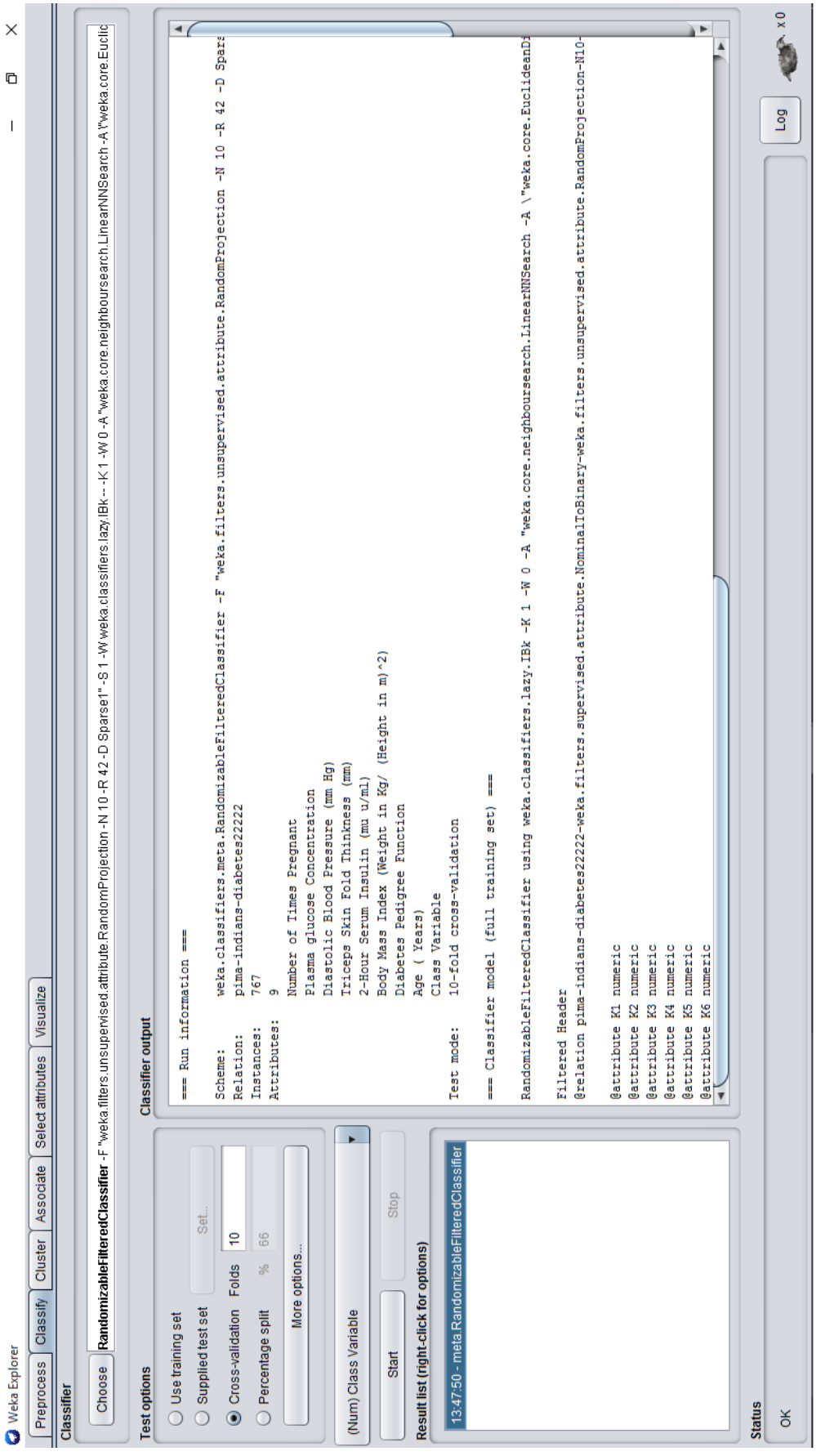

Figure 5. Pima dataset.

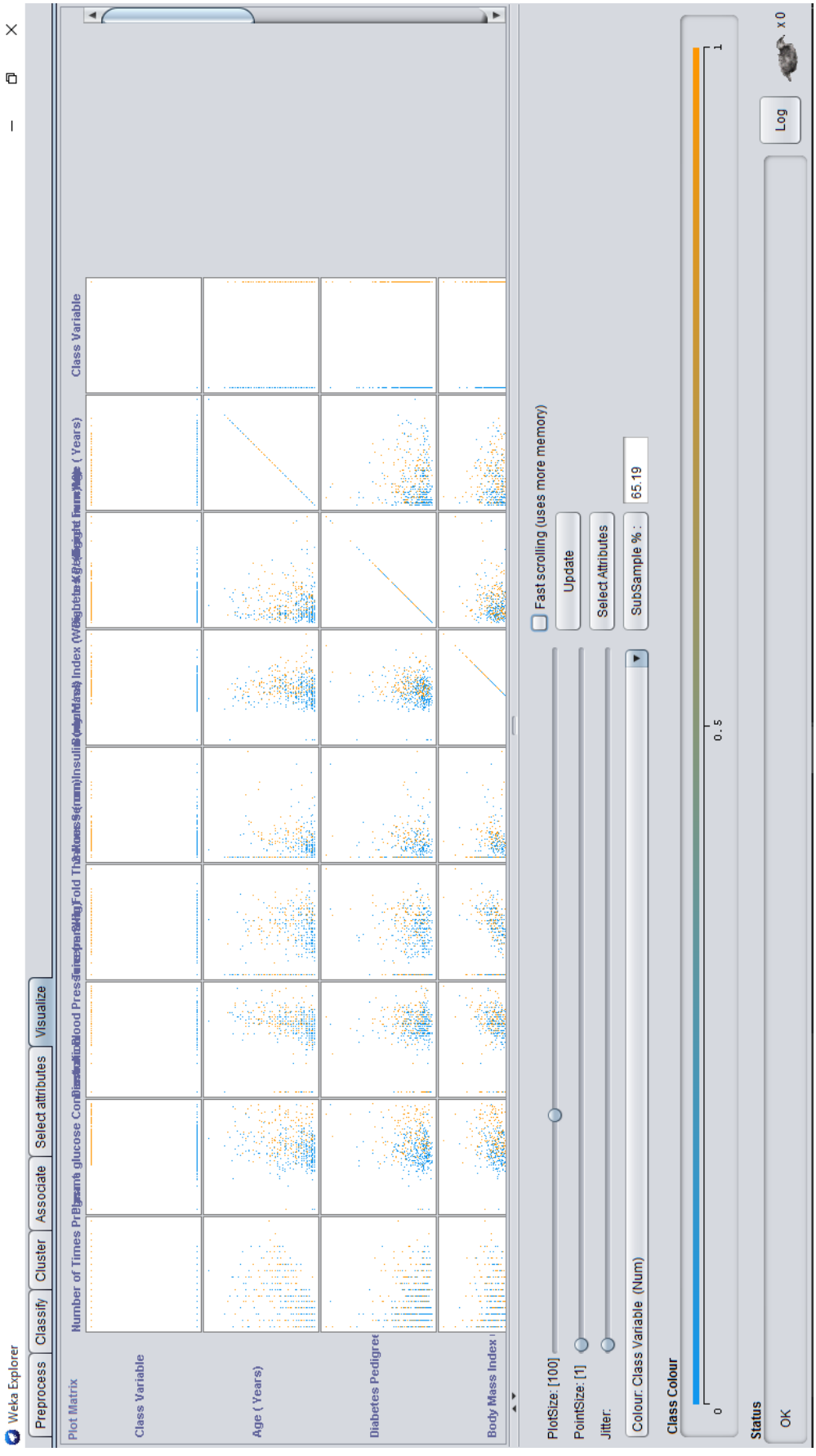

Figure 6. Visualization of a Randomizable filtered classier ensemble that has been categorized.

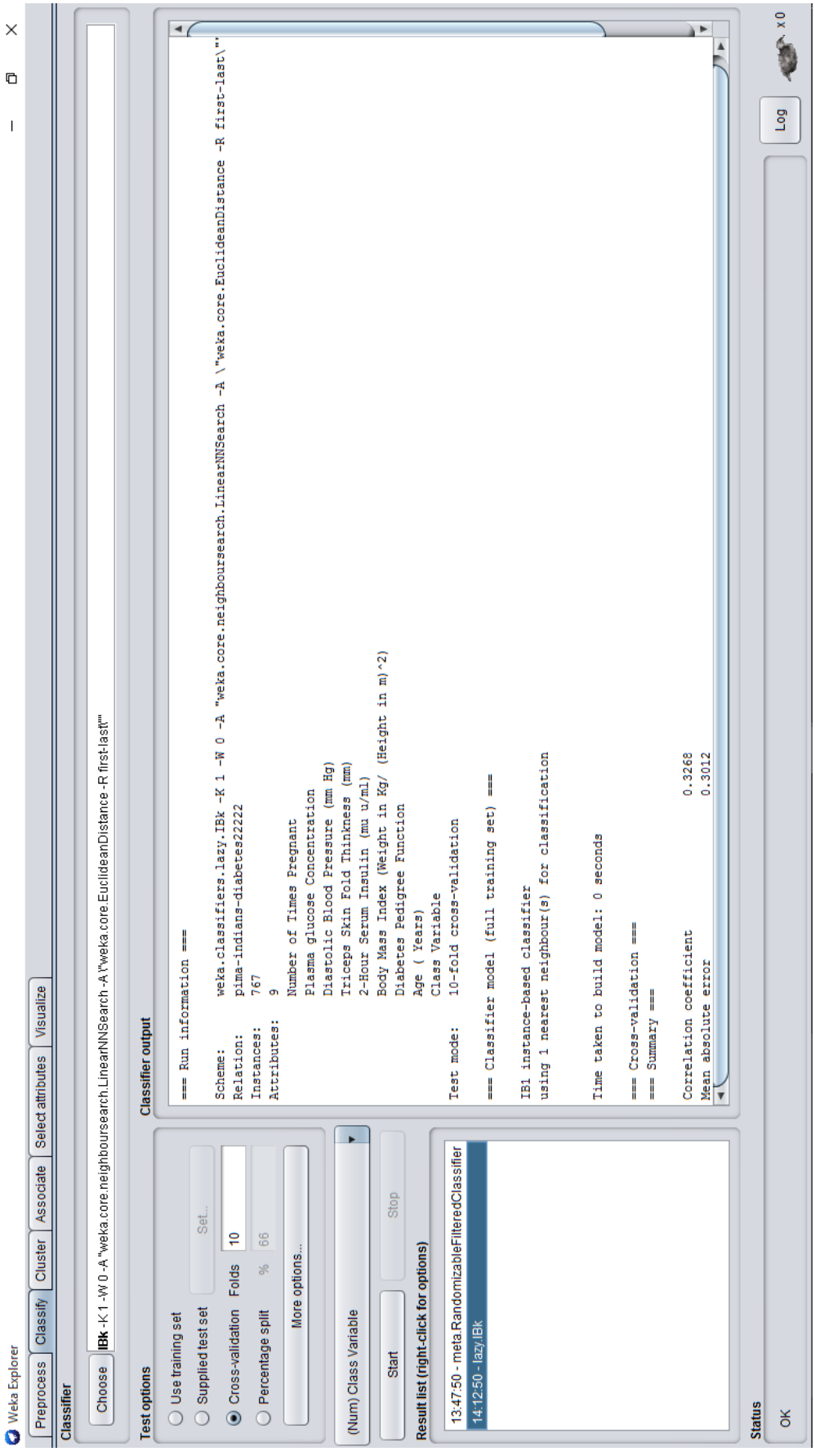

Figure 7. Shows a comparison of the algorithms' performance.

TABLES

Datasets

The study's data came from PIMA Indians with diabetes at the National Institute of Diabetes, Digestive and Kidney Diseases. Many constraints were imposed on the selection of these examples from a larger database [30]. The number of instances is 768, and the number of attributes is 8, with two positive and negative examples checked in each phase, as indicated in Table 1.

Table 1. PIMA Indian database attributes.

Attribute No.	Attribute	Description	Type
A1	Pregnant	Number of times pregnant	Numeric
A2	GTT	Plasma glucose concentration a 2-hour in an oral glucose tolerance test	Numeric
A3	BP	Diastolic blood pressure (mmHg)	Numeric
A4	Skin	Triceps skin fold thickness	Numeric
A5	Insulin	2-hour serum insulin (mm u/m)	Numeric
A6	BMI	Body Mass Index (kg/m)	Numeric
A7	DPF	Diabetes Pedigree function	Numeric
A8	Age	Age of patient (years)	Numeric
Class	Diabetes	Diabetes onset within 5 years (0,1)	Numeric

There are no missing values in the database, yet there are zeros in several areas. The values in the range [0, 1] are scaled after normalization.

Table 2. Before Normalization.

Attribute No.	Means	Standard Deviation
A-1	3.8	3.4
A-2	120.9	32.0
A-3	69.1	19.4
A-4	20.5	16.0
A-5	79.8	115.2
A-6	32.0	7.9
A-7	0.5	0.3
A-8	33.2	11.

The data in Tables 1 and 2 are irregular; normalization is used to scale them between the stated ranges.

Table 3. The values after normalization.

Attribute No.	Mean	Standard Deviation
Atri-1	0.226	0.19
Atri-2	0.608	0.16
Atri-3	0.566	0.15
Atri-4	0.207	0.16
Atri-5	0.094	0.13
Atri-6	0.477	0.11
Atri-7	0.168	0.14
Atri-8	0.204	0.19

The values are now in the range [0, 1] after normalization. WEKA was employed in the research for filtering and normalization as shown in Table 3.

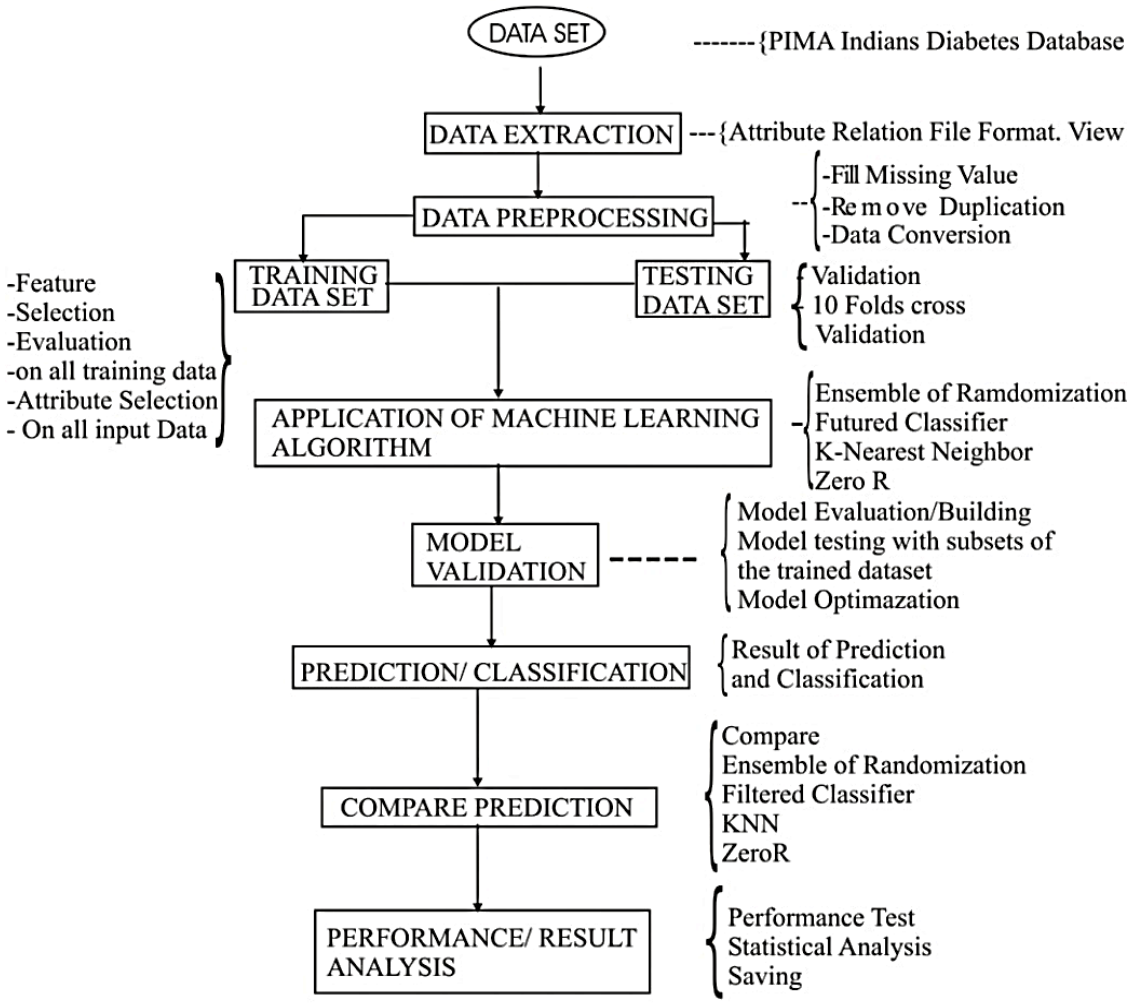

Figure 8. System flow architecture.

The study examines some of the existing approaches for diabetes mellitus classification and prediction. K-Nearest Neighbor (KNN), Decision Tree, J48, Support Vector Machine (SVM), Random Forest, Gradient Boosting, and light GBM are among the approaches [31].

The proposed methodology's design and system flow architecture are shown in Figures 8 and 9. The suggested classification and prediction approach is a combination of randomizable filtered Classifier KNN algorithms for Diabetes Mellitus classification and prediction [32].

Kindest Neighbor (KNN)

K-Nearest Neighbor is an algorithm strategy for storing all different examples and classifying new situations using comparable measures or procedures. It is also known as the slow learning algorithm since it considers all acceptable qualities before classifying new ones based on their similarity measure. For example, K is the number of dataset objects examined for classification. A case may be classified by a majority vote of its neighbors, with the case being assigned to the most common class among its KNN as determined by the A distance function. Euclidean _(i=i)K(Xi-Yi)^2 is the Euclidean distance between two points x and y in ICT. Inspection of the data set yields the value of K. If the numbers are K=1, 3, and 5, the result is K=5, which means the k values are layered, and the result is more accurate. Both regression and classification issues can be solved with KNN. KNN is simple and intuitive, but it is computationally expensive and requires a high number of samples for accuracy because it is based on k values [33].

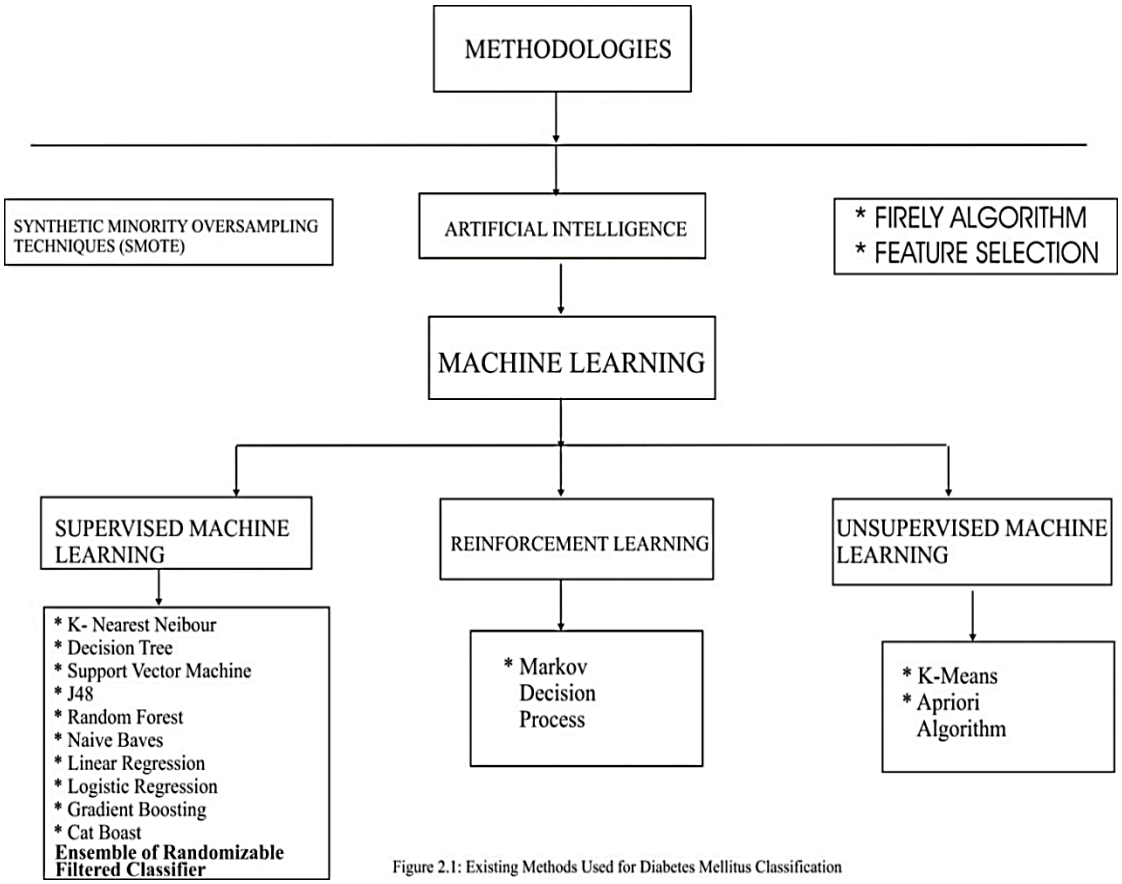

Figure 9. Existing methodologies.

Tree of Decisions

J48 A decision tree is a categorization approach that can be used to predict and categorize data. Paths, branches, and leave nodes make up a tree. Each route in a decision tree represents a rule that is used for classification or prediction. The simple algorithm used by the J48 Decision Tree Classifier is shown below. J48 is a pseudo code. Make sure there are not any base cases.

- For each “a” attribute:

It looks at the value of "a" in terms of normalized information gain.

- Amount of the attribute a's best information gain.
- Best: It chooses the attribute with the greatest amount of knowledge gain.
- It adds that characteristic to a decision node.
- This approach is performed for each node's sub list before being applied to the node's child node.

To categorize new items, it first creates a decision tree based on the attribute values of the existing training data. J48 is an expansion of ID3, with the added features of accounting for missing values, tree pruning continuous attribute value ranges, and rule derivation. Because the trees formed can be used for categorization, it is known as a statistical classifier.

Support Vector Machine (SVM)

One of the most famous and commonly used machine learning approaches is the Support Vector Machine algorithm. SVM is a collection of related supervised learning methods that are always used in medical diagnostics for classification and regression; it may minimize the empirical classification error while also increasing the geometric margin. This algorithm's main purpose is to forecast class

membership for categorical target tasks by generating hyper planes in a multidimensional space that partition examples with various class labels.

- Step 1: We locate the appropriate hyperplane.
- Step 2: Maximizing the distances between neighboring data points is the second step.
- Step 3: Add a feature Z=X2+Y2 that denotes that SVM solves this problem.
- Step 4: To categorize the class, use the SVM Classifier. It is a binary class.

The SVM can handle a variety of continuous and categorical variables and can be used for regression as well as classification.

Random Forest

Random Forest is a statistical and machine learning technique that employs numerous learning algorithms to achieve superior predicting performance. Bagging techniques were used to construct the unsupervised machine learning Random Forest algorithm, which was a blend of classification and regression methodologies.

There are two components to this algorithm.

Tree bagging: Each tree is grown as follows from tree bagging to Random Forest:

1. If the training set contains N cases, randomly sample N of them, replacing the original data as needed. This sample will be used to train the tree.
2. If there are M input variables, the best splits the node by selecting a random number of characteristics at random. The values of M are kept constant during the forest's growth.
3. Each tree is allowed to reach its full potential. Pruning is out of the question.
4. The subspace approach is used by Random Forest to select fixed features for decision trees.

Gradient Boosting GBM

Gradient Boosting GBM is a boosting algorithm that is utilized when a large amount of data has to be forecasted with great accuracy. Boosting is a collection of learning techniques that combine the predictions of several different base estimators to increase resilience over a single estimator.

XCbind (_train, y train) Rcode Library (caret) Model that is appropriate.

Fit control and train control are two terms that are often used interchangeably. (number=4, repeats=), method="representedcv"

Train to get in shape. (yw, x, method="gbm," trcontrol="fitcontrol", verbose="FALSE") (fit, x test, type=“prob”) Predicted=predict [0, 2]

GBM Light

Light GBM is a tree-based learning algorithm-based gradient boosting framework. Light GBM is a high-performance gradient boosting algorithm that is used for ranking, classification, and a variety of other machine learning applications.

Python Source Code

np.random.rand=data (500, 10) # 500 entities, each with ten characteristics

np.random.randint(2, size=500) label=np.random.randint(2, size=500) train data=lgb.Dataset(data, label=label) # binary target

train data.create valid('test.svm') test data=train data.create valid('test.svm')

'num leaves':31, 'num trees':100, 'objective':'binary' param='num leaves':31, 'num trees':100, 'objective':'binary'

num round=10 param['metric']='auc'

bst=lgb.train(param, train data, num round, valid sets=[test data]) bst=lgb.train(param, train data, num round, valid sets=[test data])

bst.save model('model.txt')

7 entities, each has its own set of rules.

data=np.random.rand np.random.rand np.random.rand np.random (7, 10)

ypred=bst.predict ypred=bst.predict ypred=b(data)

DISCUSSIONS

The PIMA database was classified using an ensemble of Randomizable Filtered Classifier K-Nearest Neighbor Algorithms. There are 768 patients in the dataset, each with eight attributes and a class attribute. This includes the training and testing sets. The system is trained using the training dataset, and it is tested using the testing dataset. The missing values are first deleted, and the data is then scaled using data normalization. Following normalization, 392 instances were discovered. The data is organized into a number of classes, each of which has a representative sample and associated samples. The training and testing portions of the dataset are split 60 and 40.

Experiments were conducted using k=10 clusters, and the distance between them was determined using Euclidean distance. With cross fold validation =10 and a kappa statistics value of 0.3625, the proposed an ensemble of Randomizable Filtered Classifier K-Nearest Neighbor model achieves a percentage of accuracy of 78%. Figure 7 depicts the results of the experiment conducted in WEKA.

CONCLUSION

Data mining is the process of extracting interesting knowledge from huge databases, such as associations, patterns, and anomalies. For the categorization of the PIMA database, an ensemble of randomizable filtered classifiers was utilized. There are 768 patients in the dataset, each with eight attributes and a class attribute. The information is organized into a number of classes, each of which has a super sample and related samples. The distance between the clusters was determined using Euclidean distance in an experiment with k=10.

In conclusion, an ensemble of randomizable filtered classifiers was used to analyze, preprocess, and forecast diabetes mellitus, and the performance of other methods was compared to the model described in this study. The categorization is performed on the PIMA diabetes data set, and the results demonstrate that the study performed with 78% accuracy. Other algorithms, such as smart Average KNN and partial Average KNN, are advised for achieving improved accuracy.

REFERENCES

1. Talha Mahboob Alam, Muhammad AtifIqbal, Yasir Ali, Abdul Wahab, Safdar Ijaz, Talha Imtiaz Baig, Ayaz Hussain, Muhammad Awais Malik, Muhammad Mehdi Raza, Salman Ibrar, Zunish Abbas. A Model for Diabetes Prediction in the Early Stages. Inform Med Unlocked. 2019; 16: 100204. 2352-9148 https://www.elsevier.com/locate/imu
2. Tukur Dahiru**,** Aliyu Alhaji A**,** Shehu AU. A Review of Nigerian Population-Based Diabetes Mellitus Studies. Sub-Saharan African Journal of Medicine (SSAJM). 2016; 3(2): 59–64.
3. Hasan MK, Alam MA, Das D, Hossain E, Hasan M. Prediction of Diabetes Using a Combination of Machine Learning Classifiers. IEEE Access. 2020; 8: 76516–76531.
4. Ankit Narendrakumar Soni. (2020). Diabetes Mellitus Prediction Using Ensemble Machine

Learning Techniques. [Online]. https://www.ssrn.com/abstract=3642877.
5. Alghamdi M, Al-Mallah M, Keteyian S, Brawner C, Ehrman J, Sakr S. Predicting diabetes mellitus using SMOTE and an ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) Project. PLoS One. 2017 Jul 24; 12(7): e0179805. https://www.doi.org/10.1371/journal.pone.0179805.s001.
6. Faruque S.S., 2019. Performance Analysis of Machine Learning Techniques to Predict Diabetes Mellitus. Volume 2 Issue 4 of the International Journal of Computer Science Trends and Technology (IJCST).
7. Minyechil Alehegn, Rahul Raghvendra Joshi, Preeti Mulay. Diabetes Analysis and Prediction Using Random Forest, KNN, Nave Bayes, and J48: An Ensemble Approach. Int J Sci Technol Res. 2019; 8(09): 1346–1354. www.ijstr.org
8. Ala'raj M, Majdalawieh M, Abbod MF. Improving binary classification using Filtering Based on K-NN Proximity Graphs. J Big Data. 2020; 7: 15. www.doi.org/10.1186/s40537-020-00207-7
9. Uswa Ali Zia, Naeem Khan. Predicting Diabetes in Medical Datasets Using Machine Learning Techniques. Int J Sci Eng Res. 2017; 8(5): 1538–1551. https://www.ijser.org
10. Safial Islam Ayon, Md. Milon Islam. Diabetes Prediction: A Deep Learning Approach. Int J Inf Eng Electron Bus. 2019; 2(2): 21–27. https://www.mecs-press.org/)
11. Balamuradi S, Pradeep Kandhasamy. Performance Analysis of Classifier Models to Predict Diabetes Mellitus. Procedia Comput Sci. 2015; 47: 45–51.
12. Naveen Kishore G, Rajesh V, Vamsi Akki Reddy A, Sumedh K, Rajesh Sai Reddy T. Prediction of Diabetes Using Machine Learning Classification Algorithms. Int J Sci Technol Res. 2020; 9(01): 1805–1808. ISSN 2277-8616,
13. BalkauB, Lange C, Fezeu L, Tichet J, de Lauzon-Guillain B, Czernichow S, Fumeron F, Froguel P, Vaxillaire M, Cauchi S, Ducimetière P, Eschwège E. (2008). Predicting diabetes: Clinical, Biological, and Genetic Approaches: Data from the Epidemiological Study on the Insulin Resistance Syndrome (DESIR). Diabetes Care. 2008 Oct; 31(10): 2056–2061. https://doi.org/10.2337/dc08-0368.
14. Gopi Battineni, Getu Gamo Sagaro, Chintalapudi Nalini, Francesco Amenta, Seyed Khosrow Tayebati. Comparative Machine-Learning Approach: A Follow-Up Study on Type 2 Diabetes Predictions by Cross-Validation Methods. Machines. 2019; 7: 74. https://www.mdpi.com/journal/machines
15. Bharathi Kavitha S, Dhavamani M, Srihariakash K. A Hybrid approach for an analysis of diabetes and prediction using machine learning techniques. Int J Psychosoc Rehabilitation. 2020; 24(8): 10485–10491. IJPR/V24I8/PR281043 DOI: 10.37200/IJPR/V24I8/PR281043. ISSN:1475-7192
16. Bhavana N, Chadaga Meghana S, Pradeep KR. A Review of Ensemble Machine Learning Approach in Prediction of Diabetes Disease. International Journal on Future Revolutions in Computer Science and Communication Engineering. 2018; 4(3): 463–466. https://www.ijfrcsce.org
17. Birjails Roshan, Ashish Kumar Mourya, Ritu Chauhan, Harleen Kaur. Prediction and Diagnosis of Future Diabetes Risk: A Machine Learning Approach. SN Appl Sci. 2019; 1(9): 1–8.
18. Desmond Bala Bisandu, Dorcas Dachollom Datiri, Eva Onokpasa, Godwin Thomas, Musa Maaji Haruna, Aminu Aliyu, Jerry Zachariah Yakubu. Diabetes Prediction using Data Mining Techniques. International Journal of Research and Innovation in Applied Science (IJRIAS). 2019; IV(VI): 103–111. ISSN 2454-6194
19. Challa M, Chinnaiyan R. (2019). Optimized Machine Learning Approach for Prediction of Diabetes Mellitus. In: In: Smys S, Tavares J, Balas V, Iliyasu A, editors. Computational Vision and Bio-Inspired Computing. International Conference on Computational Vision and Bio-Inspired Computing, ICCVBIC 2019: 2019; 321–328. https://link.springer.com/chapter/10.1007/978-3-030-37218-7 37. (http://www.rsisinternational.org/)
20. Choudhary Ishan. (2020). PIMA Indian Diabetes Prediction. Diabetic onset can be predicted. [Online]. Contreras and colleagues (2017). A hybrid technique combines grammatical evolution and physiological data to predict personalized blood glucose levels. https://doi.org/10.1371/journal.pone.0187754

21. Dagliati A, Marini S, Sacchi L, Cogni G, Teliti M, Tibollo V, De Cata P, Chiovato L, Bellazzi R. Machine Learning Methods to Predict Diabetes Complications. J Diabetes Sci Technol. 2018 Mar; 12(2): 295–302. https://doi.org/10.1177/1932296817706375
22. Mahmoudinejad Dezfuli SA, Mahmoudinejad Dezfuli SR, Mahmoudinejad Dezfuli SV, Kiani Y. Early Diagnosis of Diabetes Mellitus using Data Mining and Classification Techniques. Jundishapur J Chronic Dis Care. 2019; 8(3): e94173. https://www.Doi:10.5812/jjcdc.94173.
23. Dey SK, Hossain A, Rahman MM. Implementation of a Web Application to Predict Diabetes Disease: An Approach Using Machine Learning Algorithm. 2018 21st International Conference of Computer and Information Technology (ICCIT). 2018; 1–5. DOI: 10.1109/ICCITECHN.2018.8631968
24. Dinh A, Miertschin S, Young A, et al. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak. 2019; 19: 211. https://www.doi.org/s12911.019-0918-5
25. Guo and colleagues (2012). The Bayes Network is being used to predict type 2 diabetes. The 7th International Conference on Internet Technology and Secured Transactions is a gathering of experts in the field of internet technology and secure transactions (ICITST).
26. Minyechil Alehegn, Rahul Joshi, Preeti Mulay. Analysis and Prediction of Diabetes Mellitus using Machine Learning Algorithm. Int J Pure Appl Math. 2018; 118(9): 871–878.
27. J. M. Alehegn (2018). Machine Learning Algorithm for Diabetes Mellitus Analysis and Prediction Volume 118, Number 9, 871-878, International Journal of Pure and Applied Mathematics, 2018. ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (electronic version) (on-line version) http://www.ijpam.eu
28. KodamaS, Fujihara K, Shiozaki H, Horikawa C, Yamada MH, Sato T, Yaguchi Y, Yamamoto M, Kitazawa M, Iwanaga M, Matsubayashi Y, Sone H. (2021). Ability of current machine learning algorithms to predict and detect hypoglycemia in patients With Diabetes Mellitus: Meta-analysis. JMIR Diabetes. 2021 Jan 29; 6(1): e22458.
29. Sung Hye Kong, Daehwan Ahn, Buomsoo (Raymond) Kim, Karthik Srinivasan, Sudha Ram, Hana Kim, A Ram Hong, Jung Hee Kim, Nam H Cho, Chan Soo Shin. A novel fracture prediction model using machine learning in a community-based cohort. JBMR Plus. 2020; 4(3): e10337.
30. Saxena K, Zubair Khan, Shefali Singh. Diagnosis of Diabetes Mellitus Using K Nearest Neighbor Algorithm. International Journal of Computer Science Trends and Technology (IJCST). 2014; 2(4): 36–43. www.standford.edu/hastie/papers/LARS/diabetes.data
31. Kumar Senthil B, Gunavathi R. An Enhanced Model for Diabetes Prediction using Improved Firefly Feature Selection and Hybrid Random Forest Algorithm. 2019; 9(1): 3765–3769. ISSN: 2249-8958.
32. Kushner T, Breton MD, Sankaranarayanan S. Multi-Hour Blood Glucose Prediction in Type 1 Diabetes: A Patient-Specific Approach Using Shallow Neural Network Models. Diabetes Technol Ther. 2020 Dec; 22(12): 883–891. https://doi.org/10.1089/dia.2020.0061
33. Malik S, Harous S, El-Sayed H. Comparative Analysis of Machine Learning Algorithms for Early Prediction of Diabetes Mellitus in Women. In: Chikhi S, Amine A, Chaoui A, Saidouni D, Kholladi M, editors. Modelling and Implementation of Complex Systems. MISC 2020. Lecture Notes in Networks and Systems, vol 156. Springer, Cham. Souad Larabi-Marie-Sainte, Linah Aburahmah, Rana Almohaini, Tanzila Saba. Current Techniques of Diabetes Prediction: Review and Case Study. Appl Sci. 2019; 9(21): 4604. https://www.mdpi.com/journal/applsci

