

Research Paper



Performance evaluation of machine learning algorithm in various datasets

Md. Siraj-Ud-Douhah^{1*}, Md. Nazmul Islam²

^{1,2}Department of Statistics, Begum Rokeya University, Rangpur, Bangladesh.

Article Info

Article History:

Received: 18 November 2022

Revised: 30 January 2023

Accepted: 06 February 2023

Published: 25 March 2023

Keywords:

Machine Learning

Classification

Confusion Matrix

Performance Measures

Histograms



ABSTRACT

Machine learning is one of the fast-growing areas of computer science, with far-reaching applications. There are several applications for machine learning. The most significant of which is supervised learning. Supervised learning is common in classification problems. In this study, frequently used twelve machine learning algorithms are considered: NB, LDA, LR, ANN, SVM, K-NN, HT, DT, C4.5, CART, RF and BB. We apply these algorithms on seven datasets. The main goal of this study was to evaluate the performance of the machine learning algorithms on both binary and multiple classification problems using a variety of performance metrics: accuracy, kappa statistic, precision, recall, specificity, F-measure, MAE, RMSE and MCC. Here, we found that RF algorithm proved to have the best performance in three out of seven datasets. But the other four algorithms: NN, NB, BB and LR also performed well.

Corresponding Author:

Md. Siraj-Ud-Douhah

Department of Statistics, Begum Rokeya University, Rangpur, Bangladesh.

Email: sdouhah_brur@yahoo.com

Copyright © 2023 The Author(s). This is an open access article distributed under the Creative Commons Attribution License, (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Machine learning [1], [2] can be used in a variety of ways for algorithm selection. A few papers compare different machine learning algorithms in detail. J48, CART, and ADTree are compared by [3]. CART was found to be the most accurate algorithm, with a 98.5 percent accuracy rating. GRU-SVM, LR, MLP, NN, SR, and SVM for breast cancer dataset were used to discover Agarap [4]. MLP was determined to have the

highest classification accuracy of 99.04 percent. J48, Nave Bayes, LMT, REP Tree, DT, K-star, LR, ICO, IBK, and FC are used to analyze several machine learning methods for the breast cancer dataset in [3], [5], [6], [7], [8], [9]. Clean classifiers are good with a classification accuracy of 76 percent. For a breast cancer dataset, [9] examine LR, NB, and SVM. They discovered that SVM is the most effective. [10] Compared DT and SVM on a breast cancer dataset and discovered that the SVM had the highest accuracy of 96.99 percent. Doulah [11] compare three algorithms SVM, K-NN, DT and found that SVM grants the largest accuracy of (98.1%). [12] Examine five algorithms SVM; ANN; DT; NB; and K-NN. They located that SVM reap the best accuracy is ninety%. [13] Discover the use of lower BPNN and RBFN for breast tissue dataset and locate RBFN outperforms the BPNN in phrases of accuracy. Ravi [14] assessed NB, IBK and J48 to procedure classify the Breast Tissue dataset. They discovered that the J48 set of rules proficiency is higher than other algorithms. [15] compare 3 classification algorithms are SVM, KNN and DT and discovered that SVM performs. [16] examined algorithms NN, LR, SVM and determined that LR is the pleasant choice with 93.4%. Thomas [17] compares BB, DT, C4.5 and found that BB is outperform. [18] examined NN, BB, NB and found that NB, the highest accuracy is 93.87%. [19] Compares NB, NN, SMO, IBK, J48 and RIPPER, BB set of rules on various datasets. [20] Compare BPNN, RST-GA, RST-JA and discovered that BPNN performs higher than the opposite algorithms. [21] Evaluate the Quest, univariate splits, C4.5, Ind-Cart, and LR and discovered that C4.5 provides quality results.

2. METHODOLOGY

2.1 Naive Bayes (NB)

Naive Bayes classifier [22] considers that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors is given by-

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (1)$$

Where, $P(x|c) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$

Where, Equation (1) is the posterior probability $P(c)$, is the prior probability of class; $P(x|c)$, is the likelihood.

2.2 Logistic Regression (LR)

Let's assume a target variable Y and $X = x_1 + x_2 + \dots + x_j$ where j is the number of independent variables, then the conditional mean of Y is given x is given by:

$$\pi(x) = E\left(\frac{Y}{x}\right) \quad (2)$$

(2) Is the expected value of the target variable for a given X . The LR is given by the formula:

$$\pi(x) = \frac{e^{g(x)}}{1+e^{g(x)}} \quad (3)$$

Where, $g(x) = \ln \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$

Equation (3) is called the logit transformation, where $\beta_0, \beta_1, \dots, \beta_k$ are estimated using maximum likelihood. The log likelihood is given by: $L\left(\frac{\beta}{x} = \ln\left[\frac{\beta}{x}\right]\right) = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \{y_i \ln[1 - \pi(x_i)]\}$.

2.3 Linear Discriminant Analysis (LDA)

LDA was developed in 1936 by R.A. Fisher [23]. We assume the density function $P(X = x|Y = k)$ of X for an observation that comes from the k th class is known, then by Bayes Theorem we have.

$$P\left(\frac{Y = k}{X = x}\right) = \frac{P\left(\frac{X=x}{Y=k}\right) P(Y = k)}{\sum_{j=1}^N P\left(\frac{X=x}{Y=j}\right) P(Y = j)}$$

In general, $f_k(X) = P(X = x|Y = k)$ is the densities? For example, if we assume $X = (X_1, X_2, \dots, X_p)$ is drawn from a multivariate Gaussian distribution $N(\mu_k, \Sigma)$, then

$$f_k(x) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \mu_k)^T \Sigma^{-1}(x - \mu_k)\right\}$$

The LDA classifier assigns an observation $X = x$ to the class for which

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \ln \pi_k$$

is the largest. From here we can see that the decision boundary for LDA is linear.

2.4 Support Vector Machine (SVM)

SVM [24], [25] have been added by using Vapnik and collaborators in 1992. According to [26], the selection floor by using SVM for linearly separated space is a hyper-plane and is given below:

$$w \cdot x + b = 0$$

Where x is an arbitrary feature vector, b and w are learned from training set linearly separable data. The dot product $w \cdot x$ is defined by,

$$w \cdot x = \sum_{i=1}^n w_i x_i$$

This linear classifier is represented by the hyper-plane $H(w \cdot x + b = 0)$ and defines a region for class +1 patterns ($w \cdot x + b > 0$) and another region for class -1 pattern ($w \cdot x + b < 0$).

2.5 Classification and Regression Tree (Cart)

CART algorithm changed into advanced through Brieman, Friedman, Olshen, and Stone in 1984 [27]. For a binary magnificence the GINI degree of impurity is given by, $GINI(t) = 1 - \sum [p(\frac{t}{j})]^2$

Where $p(\frac{t}{j})$ is the relative frequency of class j at node t . When a node p is split into x partitions, the quality of split is given by

$$GINI_{split} = \sum_{i=1}^n \frac{n_i}{n} GINI$$

Where n_i = number of records at child i

n = number of records at node p C4.5

The first step in C4.5 set of rules [28] is to specify root of the tree. To determine the order of functions inside the choice tree, records benefit components is evaluated for each attribute as described in

$$Entropy(S) = \sum_{i=1}^n -p_i$$

Where i is a state, p_i is the possibility of outcome being in state i for the set S and n is the number of possible.

2.6 K-Nearest Neighbor (K-NN)

KNN [29] is distance-primarily based classifier wherein distance is used to classify information based on labels of its associates which can be decided on from training data.

$$\text{Euclidian Distance: } \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

If $k = 1$, the class where it's nearest neighbour belongs. Alternatively, if we provide a huge k input, it can bring about underneath-fitting.

2.6.1 Random Forest (RF)

Random Forests [29] similarly weighted is offered as inside the following system:

$$H(x) = \arg \left\{ \max \frac{1}{m} \sum_{i=1}^m (I(x; \theta_i) = y_j) \right\}$$

$$mg(X, Y) = \max_{j \neq Y_k} I(I_k(X) = Y) - j$$

The generalization error is given by,

$$PE^* = P_{X,Y}(mg(X, Y) < 0)$$

The strength of the set of classifiers $\{I(x, \theta)\}$ is

$$s = E_{X,Y} \left(P_{\theta}(I(x, \theta) = Y) - \max_{j \neq Y_{\theta}} (I(x, \theta) = j) \right) \quad (2)$$

An upper bound for the generalization error is given by

$$PE^* \leq \frac{\bar{\rho}(1 - s^2)}{s^2}$$

2.6.2 Bagging and Boosting (BB)

Let $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ be a set of m items and let $y_i \in Y = \{c_1, c_2, \dots, c_k\}$ be a set of k class labels; C denotes classification algorithm and n is number of learners.

1. Draw m items randomly with replacement from the dataset D , so generate bootstrap samples, D_1, D_2, \dots, D_n
2. Each dataset D_i is trained and multiple classification models are constructed, $M_i = C(D_i)$.
3. Consensus of classification models is tested to calculate out-of-bag error.
4. New sample x is given to classifiers as input and the outputs y_i are obtained from each model $y_i = M_i(x)$
5. The outputs of models $\{M_1, M_2, \dots, M_t\}$ are combined as in

$$M^*(x) = \arg \max_{y \in Y} \sum_{i: M_i(x)=y}^n 1$$

2.6.3 Hoeffding Tree (HT)

The Hoeffding Bound [30], [31], [32] is calculated the usage of the following equation: Hoeffding bound,

$$\epsilon = \sqrt{\frac{R^2 \ln(\frac{1}{\delta})}{2nl}}$$

As mentioned before, the Hoeffding Tree algorithm incrementally generates a decision tree from Data Streams.

2.6.4 Decision Table (DT)

Decision table is an accurate technique for making numerical predictions from decision trees. It is an ordered set of If-Then rules, which may be denser than decision trees and therefore easier to understand [33].

2.6.5 Evaluation Criteria

Table 1 shows a confusion matrix illustrating what might happen for positive and negative outcomes in a data mining model.

Table 1. Confusion Matrix

	Predicted	
Actual	TP	FN
	FP	TN

The performance criterions are shown in the following Table 2.

Table 2. Performance Criteria

Performance Criteria	Statistics
Accuracy	$A = \frac{TP+TN}{TP+TN+FP+FN}$
Cohen's kappa	$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$, Where, $P_o = \frac{TP+TN}{TP+TN+FP+FN}$ and $P_e = \left(\frac{TP + FN}{TP + TN + FP + FN} \cdot \frac{TP + TN + FP + FN}{TP + TN + FP + FN} \right) + \left(\frac{FP + TN}{TP + TN + FP + FN} \cdot \frac{FN + TN}{TP + TN + FP + FN} \right)$
Precision	$P = \frac{TP}{TP+FP} = \frac{C_{ii}}{\sum_{k=1}^{n+1} C_{ki}} = \frac{C_{ii}}{PTotal_i}$, Where PTotal _i is the Predicted total for the class and C _{ii} is the Predicted class for a multi class problem.
Recall or Sensitivity	$R/S = \frac{TP}{FN+TP} = \frac{C_{ii}}{\sum_{i=1}^{n+1} C_{ij}} = \frac{C_{ii}}{ATotal_i}$, Where ATotal _i is the total for the actual total and C _{ii} is the Predicted class for a multi class problem, and n is the sample size.
Specificity	$S = \frac{TN}{TN + FP} = \frac{T - ATotal_i - PTotal_i + C_{ii}}{T - ATotal_i}$
F-Measure	$F = \frac{2 * Recall * Precision}{Recall + Precision}$
Matthews's Correlation Coefficient	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
Mean Absolute Error	$MAE = \frac{1}{n} \sum_{j=1}^n y_j - \hat{y}_j $
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$

2.6.6 Software Used

The open-source programming language R is used to analyse statistical data, present graphics, and generate reports [34]. The experiments were conducted using open-source R software, version 3.5.2 (<https://www.r-project.org>).

2.6.7 Dataset Description

The summary of the datasets used in comparative studies in the following Table 3.

Table 3. Dataset Characteristics

Sl. No	Dataset	No. of Attributes (Including Class)	No. of Instances	No of Classes	Missing Values
1	Breast Cancer Wisconsin (Original)	11	699	2	Yes
2	Statlog (Vehicle Silhouettes)	19	846	4	No
3	Vertebral Column	7	310	3	No
4	Breast Tissue	10	106	6	No
5	Contraceptive method Choice	10	1473	3	No
6	Image segmentation	20	2310	7	No
7	Artificial Characters	8	10218	10	No

The datasets chosen for the article have been downloaded from UCI repository, <https://archive.ics.uci.edu/ml/datasets/> and <https://www.openml.org/search?type=data>.

2.6.7.1 Missing Value Estimation

There are numerous methods available for estimating missing values [35], to estimate missing values for our analysis, however, we use mode method.

2.6.7.2 Hold Out Approach

The hold out approach is a straightforward method for generating test data. This method divides the data set into two subsets, a training set and a test set, at random. In this study, the training set will contain 80% of the data, while the test set will contain the remaining 20%. We use the training data set to train the model and use the test data set to test the model based on the training model.

3. RESULTS AND DISCUSSION

A simple and sensible way to start is by looking at the data frame, using pairs to plot every variable against every other. There appears to be excellent data separation, and reasonable separation, but nothing obvious for the other variables. The commonest plots for a sample are histograms and box plots. Histograms are excellent for showing the mode, the spread, and the symmetry (skew) of a set of data whereas boxplots are outstanding for showing the spread [9], outliers [6], [7] and normality of a set of data [5]. From Figure 1, we found that the Wisconsin breast cancer dataset contains two classes. From Figure 2, we visualize that the breast tissue dataset contains six classes. From Figure 3, we envisage that the vehicle dataset contains four classes. From Figure 4, we found that the vertebral column dataset contains three classes. From Figure 5, we visualize that the contraceptive method choice dataset contains three classes. From Figure 6, we visualize that the image segmentation dataset contains seven classes. From Figure 7, we envisage that the artificial characters dataset contain ten classes and we also see the correlation coefficient between the attribute.

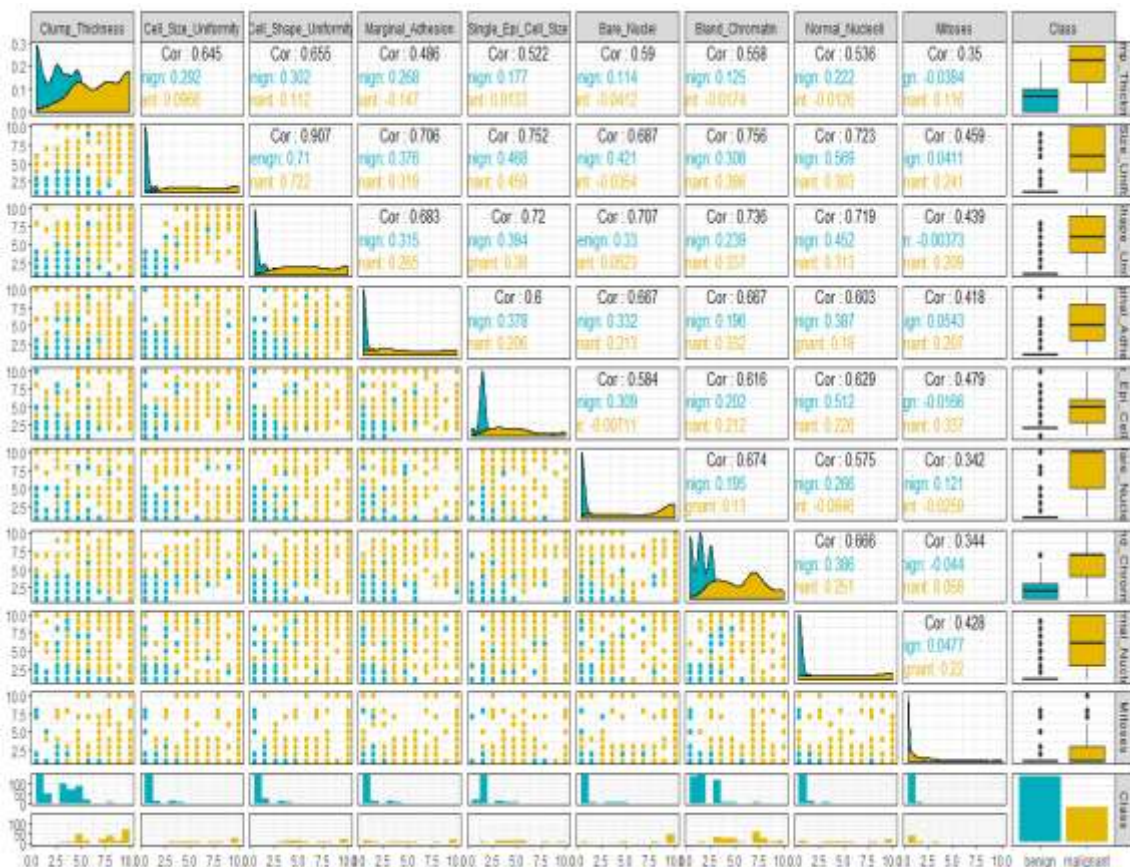


Figure 1. Scatterplot Matrix by Groups for Breast Cancer Dataset

Table 4. Performance Measures of Breast Cancer Dataset

CA	CCI (Accuracy)	ICI	Kappa	MAE	RMS E	Precision	Recall	Specificity	F-measure	MCC
NB	96.43%	3.57%	91.77%	0.035	0.187	0.966	0.964	0.971	0.965	0.919
LDA	95%	5%	88.02%	0.047	0.196	0.950	0.950	0.913	0.949	0.881
LR	97.14%	2.85%	93.37%	0.041	0.163	0.972	0.971	0.974	0.972	0.971
NN	96.42%	3.57%	91.77%	0.039	0.155	0.966	0.964	0.971	0.965	0.919
SVM	97.14%	2.85%	93.37%	0.063	0.169	0.972	0.971	0.974	0.972	0.934
KNN	95%	5%	88.33%	0.047	0.211	0.950	0.950	0.939	0.950	0.883
BG	95.71%	4.29%	89.93%	0.072	0.178	0.957	0.957	0.942	0.957	0.899
DT	95%	5%	88.33%	0.078	0.193	0.950	0.950	0.939	0.950	0.883
C4.5	93.57%	6.43%	84.99%	0.079	0.243	0.936	0.936	0.920	0.936	0.850
RF	97.85%	2.14%	95.03%	0.032	0.150	0.980	0.979	0.991	0.979	0.952
CART	93.57%	6.43%	84.99%	0.101	0.239	0.936	0.936	0.920	0.936	0.850
HT	96.43%	3.57%	91.77%	0.036	0.188	0.966	0.964	0.971	0.965	0.919



Figure 2. Scatterplot Matrix by Groups for Breast Tissue Dataset

Table 6. Performance Measures of Vehicle Dataset

CA	CCI	ICI	Kappa	MAE	RMS E	Precisio n	Recal l	Specificit y	F-measur e	MCC
NB	44.70 %	55.25 %	27.41 %	0.276	0.451	0.520	0.447	0.828	0.401	0.290
LDA	77.05 %	22.94 %	69.39 %	0.137	0.287	0.758	0.771	0.923	0.758	0.686
LR	81.17 %	18.82 %	74.87 %	0.116	0.272	0.804	0.812	0.934	0.805	0.744
NN	78.82 %	21.17 %	71.73 %	0.124	0.302	0.778	0.788	0.928	0.782	0.711
SVM	72.35 %	27.64 %	63.18 %	0.282	0.359	0.705	0.724	0.907	0.709	0.621
KNN	68.23 %	31.76 %	57.61 %	0.160	0.397	0.665	0.682	0.891	0.672	0.565
BG	71.17 %	28.83 %	61.56 %	0.167	0.288	0.687	0.712	0.902	0.695	0.602
DT	63.52 %	36.47 %	51.57 %	0.245	0.343	0.610	0.635	0.880	0.609	0.501
C4.5	68.82 %	31.18 %	58.44 %	0.162	0.363	0.692	0.688	0.895	0.688	0.583
RF	71.76 %	28.23 %	62.4%	0.158	0.282	0.705	0.718	0.905	0.705	0.615
CAR T	74.71 %	25.29 %	66.28 %	0.157	0.303	0.733	0.747	0.914	0.738	0.654
HT	45.29 %	54.70 %	28.16 %	0.274	0.448	0.530	0.453	0.830	0.411	0.300



Figure 4. Scatterplot Matrix by Groups for Vertebral Column Dataset

Table 7. Performance Measures of Vertebral Column Dataset

CA	CCI	ICI	Kappa	MAE	RMSE	Precision	Recall	Specificity	F-measure	MCC
NB	83.87 %	16.13 %	73.24 %	0.123	0.262	0.834	0.839	0.934	0.835	0.767
LDA	80.64 %	19.35 %	68.82 %	0.195	0.304	0.830	0.806	0.913	0.813	0.715
LR	88.71 %	11.29 %	81.38 %	0.102	0.224	0.896	0.887	0.951	0.886	0.843
NN	88.71 %	11.29 %	81.44 %	0.097	0.215	0.899	0.887	0.953	0.887	0.844
SVM	70.96 %	29.03 %	51.98 %	0.294	0.380	0.753	0.710	0.855	0.688	0.579
KNN	69.35 %	30.64 %	50.92 %	0.207	0.449	0.734	0.694	0.864	0.694	0.558
BG	87.10 %	12.90 %	78.71 %	0.128	0.245	0.885	0.871	0.946	0.870	0.824
DT	77.42 %	22.58 %	62.34 %	0.196	0.312	0.853	0.774	0.902	0.745	0.704
C4.5	87.10 %	12.90 %	78.74 %	0.110	0.265	0.871	0.871	0.947	0.871	0.818
RF	87.10 %	12.90 %	78.71 %	0.133	0.258	0.885	0.871	0.946	0.870	0.824
CART	87.10 %	12.90 %	78.71 %	0.124	0.266	0.885	0.871	0.946	0.870	0.824
HT	83.87 %	16.13 %	73.24 %	0.121	0.258	0.834	0.839	0.924	0.835	0.767

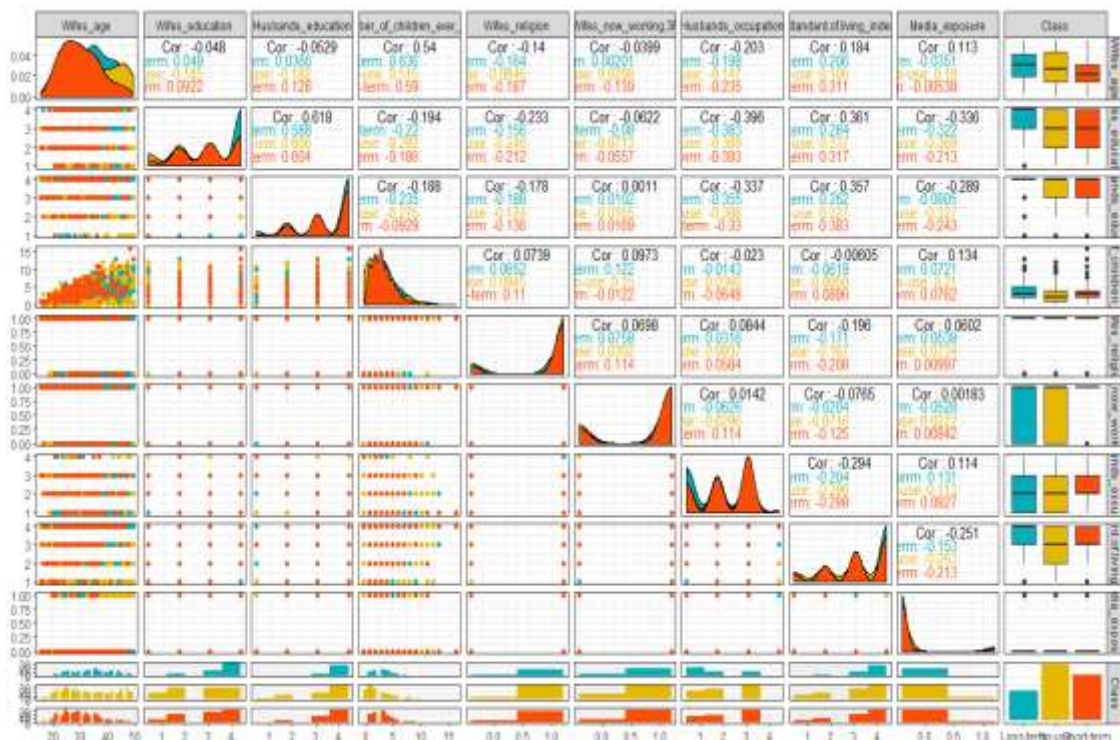


Figure 5. Scatterplot Matrix by Groups for CMC Dataset

Table 9. Performance Measures of Image Segmentation Dataset

CA	CCI	ICI	Kappa	MAE	RMS E	Precisio n	Recal l	Specificit y	F-measur e	MCC
NB	79.43 %	20.56 %	75.93 %	0.061	0.234	0.810	0.794	0.964	0.776	0.757
LDA	91.12 %	8.87%	89.64 %	0.03	0.142	0.912	0.911	0.985	0.910	0.896
LR	96.10 %	3.89%	95.45 %	0.017	0.094	0.963	0.961	0.993	0.961	0.955
NN	97.62 %	2.38%	97.22 %	0.04	0.082	0.977	0.976	0.996	0.976	0.972
SVM	93.07 %	6.92%	91.91 %	0.205	0.303	0.931	0.931	0.988	0.930	0.918
KNN	97.40 %	2.60%	96.97 %	0.017	0.086	0.974	0.974	0.995	0.974	0.969
BG	96.97 %	3.03%	96.46 %	0.020	0.085	0.970	0.970	0.995	0.970	0.965
DT	89.83 %	10.17 %	88.19 %	0.097	0.178	0.930	0.898	0.984	0.904	0.895
C4.5	96.32 %	3.67%	95.7%	0.122	0.099	0.963	0.963	0.994	0.963	0.957
RF	98.48 %	1.51%	98.23 %	0.014	0.069	0.985	0.985	0.997	0.985	0.982
CAR T	96.53 %	3.46%	95.95 %	0.017	0.091	0.966	0.965	0.994	0.965	0.960
HT	79%	21%	75.44 %	0.064	0.238	0.795	0.790	0.963	0.774	0.750

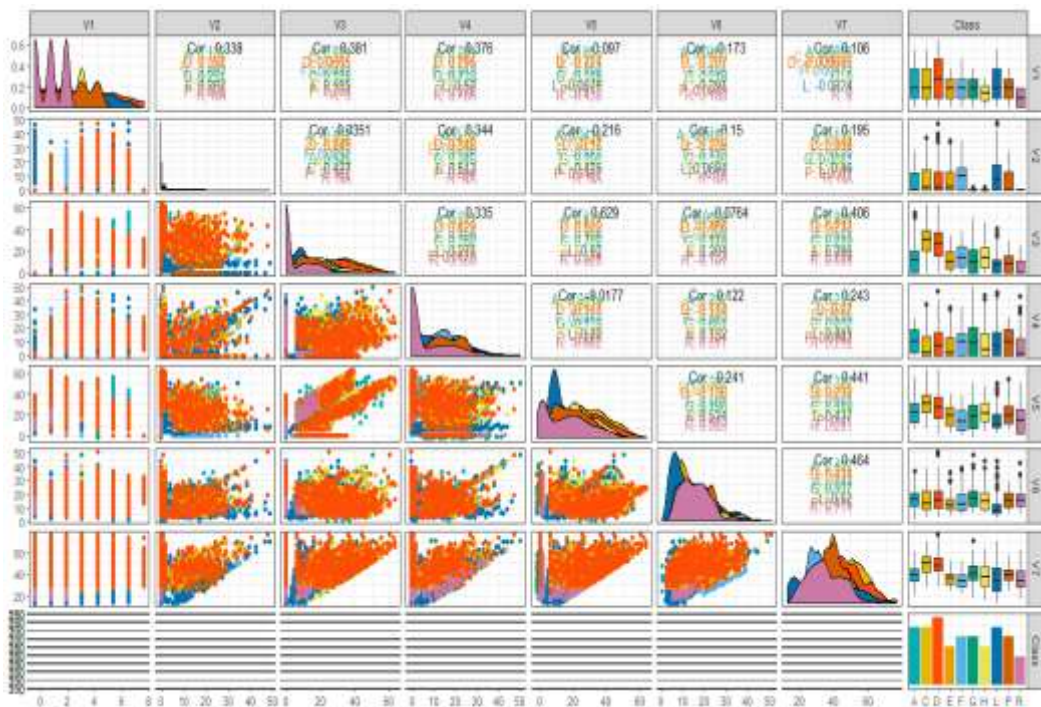


Figure 7. Scatterplot Matrix by Groups for Artificial Characters Dataset

Table 10 . Performance Measures of Artificial Characters Dataset

CA	CCI	ICI	Kappa	MAE	RMS E	Precision	Recall	Specificity	F-measure	MCC
NB	29.30 %	70.69 %	21.88 %	0.15	0.296	0.356	0.293	0.928	0.289	0.235
LDA	34.73 %	65.26 %	27.04 %	0.152	0.277	0.351	0.347	0.923	0.334	0.269
LR	34.83 %	65.16 %	27.19 %	0.149	0.273	0.334	0.348	0.924	0.333	0.263
NN	50.10 %	49.90 %	44.33 %	0.113	0.244	0.610	0.501	0.942	0.481	0.469
SVM	40.95 %	59.05 %	34.11 %	0.167	0.284	0.424	0.409	0.932	0.401	0.342
KNN	90.31 %	9.68%	89.18 %	0.019	0.136	0.904	0.903	0.989	0.903	0.892
BG	74.41 %	25.59 %	71.44 %	0.071	0.184	0.748	0.744	0.971	0.744	0.717
DT	65.02 %	34.98 %	60.81 %	0.120	0.223	0.669	0.650	0.958	0.642	0.612
C4.5	73.73 %	26.27 %	70.7%	0.053	0.192	0.744	0.737	0.971	0.737	0.709
RF	91.10 %	8.90%	90.06 %	0.019	0.132	0.912	0.911	0.990	0.911	0.901
CART	68.44 %	31.56 %	64.81 %	0.070	0.206	0.691	0.684	0.965	0.685	0.651
HT	35.81 %	64.18 %	28.91 %	0.135	0.306	0.481	0.358	0.933	0.340	0.316

Table 11. The Overall Performance

Dataset	CA	CCI	ICI	Kappa	MAE	RMS E	Precision	Recall	Specificity	F-measure	MCC
Breast cancer	RF	97.85 %	2.14 %	95.03 %	0.032	0.150	0.980	0.979	0.991	0.979	0.952
Breast Tissue	KN N	77.27 %	22.73 %	72.29 %	0.089	0.267	0.909	0.773	0.978	0.824	0.800
Vehicle	LR	81.17 %	18.82 %	74.87 %	0.116	0.272	0.804	0.812	0.934	0.805	0.744
Vertebral column	NN	88.71 %	11.29 %	81.44 %	0.097	0.215	0.899	0.887	0.953	0.887	0.844
CMC	BG	54.24 %	45.76 %	28.9 %	0.330	0.431	0.566	0.592	0.765	0.536	0.305
Image segmentat ion	RF	98.48 %	1.51 %	98.23 %	0.014	0.069	0.985	0.985	0.997	0.985	0.982
Artificial characters	RF	91.10 %	8.90 %	90.06 %	0.019	0.132	0.912	0.911	0.990	0.911	0.901

For **Table 4**, all algorithms provide relatively higher accuracy and random forest gives the highest accuracy (97.85%). It is worth noting that all algorithms provide relatively higher kappa value and random forest gives the highest value (95.03%). It is worth mentioned that all algorithms provide relatively higher precision and random forest gives the highest value (98%). It is notice that all algorithms provide relatively higher recall and random forest gives the highest value (97.9%). From the table, all algorithms provide relatively higher specificity and random forest gives the highest value (99.1%).

For **Table 5**, all the algorithm provides average performance and KNN has the highest accuracy (77.27%). all of the algorithm provides average performance and C4.5 has the highest kappa statistic (66.67%). It is worth noting that all the algorithm provides good performance and KNN has the highest precision value of (90.9%). It is noticed that all the algorithms provide good performance and KNN has the highest recall value of (77.3%). It is worth noting that all the algorithm provides good performance and KNN has the highest specificity value of (97.8%).

For **Table 6**, logistic regression provides the highest accuracy (81.17%) and other algorithms give relatively lower accuracy than logistic regression. It is worth noting that logistic regression provides the highest kappa value (74.87%) and other algorithms give relatively lower value than logistic regression. It is notice that logistic regression provides the highest precision value (80.40%) and other algorithms gives relatively lower value than logistic regression. it can be seen that logistic regression provides the highest recall value (81.2%) and other algorithms give relatively lower value than logistic regression. It is noticed that logistic regression provides the highest specificity value (93.4%) and other algorithms give relatively lower value than logistic regression.

For **Table 7**, NN and LR provide highest accuracy (88.71%) and Bagging, C4.5, RF, CART provide the same value (87.10%). It is notice that NN and LR provides highest accuracy (81.38%) and Bagging, C4.5, RF and CART has the kappa value (78.71%). In addition, NN and LR provide the highest precision value (89.60%) and Bagging, RF and CART have the precision of (88.50%). Moreover, NN and LR provide the highest recall value (88.70%) and Bagging, C4.5, RF and CART have the recall value of (87.1%). It is clear that NN provides the highest specificity value (95.3%).

For **Table 8**, the accuracy of all algorithms is less than 60%. It is noticed that the kappa value of all algorithms is less than 30%. It is worth noting that the precision of all algorithms is less than 60% and NN has the highest precision (56.6%). It is more notice that the recall of all algorithms is below 60% and NN has the highest precision (59.2%). It is worth mentioned that the specificity of NN (76.5%) algorithm are higher than the other algorithms.

In **Table 9**, all algorithms provide relatively higher accuracy and random forest gives the highest accuracy (98.48%). It is worth noting that all algorithms provide relatively higher kappa value and random forest gives the highest value (98.23%). It is notice that all algorithms provide relatively higher precision and random forest gives the highest value (98.5%). It can be clearly seen that all algorithms provide relatively higher recall value and random forest gives the highest value (98.5%). It is worth mentioned that all algorithms provide relatively higher specificity value and random forest gives the highest value (99.7%).

For **Table 10**, RF and KNN provide relatively higher accuracy and RF has the highest accuracy (91.10%). It is noticed that RF and KNN provide relatively higher kappa statistic and RF has the highest kappa value (90.06%). It is worth mentioning that RF and KNN provide relatively higher precision and RF has the highest value (91.20%). It can be clearly seen that RF and KNN provide relatively higher recall and RF has the highest value (91.10%). It is noticed that RF and KNN provide relatively higher specificity and RF has the highest value (99%).

From **Table 11**, it can be seen that all the algorithms are highly significant for breast cancer and image segmentation dataset. The Random Forest (RF) algorithm has the higher value of F-measure on 3 out of 7 datasets (breast cancer, image, and artificial characters). The LR also gives the higher value of F-measure for vehicle (80.5%) and vertebral column (88.70%). K-NN also provides the highest significant value of F-measure for breast tissue dataset. Bagging provides the highest value of F-measure (53.60%) for CMC dataset. For image and artificial dataset Random Forest (RF) provides highest value of F-measure (98.5% and 91.10%). It is seen that for image segmentation and artificial dataset, the RF provides the highest MCC values are (0.98 and 0.90). For breast cancer, vehicle, and vertebral column dataset, the LR

gives the highest MCC values are (0.97, 0.74 and 0.84). For breast tissue dataset, KNN provides the largest value of MCC (0.80) and for CMC dataset all algorithm shows the lowest value of MCC. We also see that the LR gives the best MCC value in 3 out of 7 datasets. It is worth mentioned that for breast cancer dataset, the SVM gives the lesser error rate (0.03) as it provides more perfect prediction and lesser variance in predictions. For vehicle dataset, the LR and NN algorithm provides smallest error rate (0.12). For vertebral column and breast tissue dataset, the LR provides smaller error rate (0.10 and 0.11). For CMC dataset, C4.5 and CART shows the lesser error rate are (0.35). For image dataset, the NN, KNN and CART provides the smaller error rate (0.01). For artificial dataset, the KNN and RF performs better as they provide lowest error rate (0.02). It is notice that the Random Forest (RF) algorithm provides lowest root mean square error of 5 (breast cancer (0.15), vehicle (0.28), breast tissue (0.27), image segmentation (0.07) and artificial characters (0.13)) out of 7 datasets. For vertebral column dataset, the LR provides lesser error rate (0.22) than the other algorithms. For CMC dataset, the KNN provides the lowest root mean square error (0.43).

4. CONCLUSION

Twelve different machine learning algorithms were considered: NB, LDA, LR, ANN, SVM, K-NN, HT, DT, C4.5, CART, RF and BB that have been applied on seven datasets. The results show that Random Forest (RF) was found to be the algorithms with most accuracy, precision and Matthew's correlation coefficient (MCC). Other four algorithms: NN, NB, BB and LR were found to be the next accurate after RF accordingly. While kappa statistic and RMSE is another factor. For future research, there is a plan to make hybridization of SVM and ANN to enhance and improve the performance of that type of successful famous machine learning algorithm.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Md. Siraj-Ud-Doulah	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	
Md. Nazmul Islam		✓		✓		✓		✓	✓			✓		✓

C: Conceptualization

M: Methodology

So: Software

Va: Validation

Fo: Formal analysis

I: Investigation

R: Resources

D: Data Curation

O: Writing- Original Draft

E: Writing- Review & Editing

Vi: Visualization

Su: Supervision

P: Project administration

Fu: Funding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study, and their voluntary consent was obtained prior to data collection.

Ethical Approval

The study was conducted in compliance with the ethical principles outlined in the Declaration of Helsinki and approved by the relevant institutional authorities.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.





REFERENCES

- [1] R. Caruana and A. Niculescu-Mizil, 'An Empirical Comparison of Supervised Learning Algorithms', in Proceedings of the 23rd International Conference on Machine Learning, 2006. doi.org/10.1145/1143844.1143865
- [2] 'Machine learning techniques for classification of breast tissue', 9th International Conference on Theory and Application of Soft Computing, vol. 120, pp. 402-410, 2017. doi.org/10.1016/j.procs.2017.11.256
- [3] B. Padmapriya and T. Velmurugan, 'Classification Algorithm Based Analysis of Breast Cancer Data', Breast Cancer Data. IJDMTA, vol. 5, pp. 43-49, 2016. doi.org/10.20894/IJDMTA.102.005.001.010
- [4] A. F. Agarap, 'On breast cancer detection: An application of machine learning algorithms on the Wisconsin diagnostic dataset', arXiv [cs.LG], 20-Nov-2017.
- [5] Tekur and J. Prerna, 'A Study on Classification Algorithms for Predicting Colon Cancer using Gene Tissue Parameters', Gene Tissue Parameters. IJPAM, vol. 119, pp. 2147-2166, 2018
- [6] An Introduction to Statistical Learning with Applications in R. Springer, 2014.
- [7] Handbook of Statistical Analysis and Data Mining Applications. 2009
- [8] 'Comparison of Classification Algorithms for Predicting Breast Cancer', IJSRD], vol. IJSRD], pp. 390-394, 2017.
- [9] M. Li and R. Sundararajan, 'Application of Machine Learning Algorithms on Breast Cancer Dataset, Electrostatics Joint Conference', pp. 1-7, 2018
- [10] 'Proportional learning of classifiers using breast cancer datasets', IJCSMC, vol. 3, no. 1, pp. 25-37, 2014.
- [11] M. S. U. Doulah, 'Time Series Forecasting: A Comparative Study of VAR ANN and SVM Models', Journal of Statistical and Econometric Methods, vol. 8, pp. 21-34, 2019
- [12] 'Diagnostic Accuracy of Different Machine Learning Algorithms for Breast Cancer Risk Calculation: a Meta-Analysis', Asian Pacific Journal of Cancer Prevention, vol. 19, no. 7, pp. 1747-1752, 2018
- [13] 'Machine learning techniques for classification of breast tissue', 9th International Conference on Theory and Application of Soft Computing, vol. 120, pp. 402-410, 2017. doi.org/10.1016/j.procs.2017.11.256
- [14] Ravi, 'Analysis of Classification Algorithms towards Breast Tissue Data Set', IJMTE, vol. 8, pp. 650-658, 2018
- [15] Vehicle Identification using Support Vector Machines, IOSRJEN', pp. 18-21, 2018
- [16] H. He and E. A. Garcia, 'Learning from imbalanced data', IEEE Trans. Knowledge Data Eng, vol. 21, pp. 1263-1284, 2009. doi.org/10.1109/TKDE.2008.239
- [17] T. G. Dietterich, 'An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: BBR', Kluwer-Academic-Publishers, vol. 40, pp. 139-157, 2000. doi.org/10.1023/A:1007607513941
- [18] 'Diagnosis of Vertebral Column Disorders Using Machine Learning Classifiers', in Conference: Information Science and Applications (ICISA), 2013, pp. 57-89
- [19] 'Guidelines to Select Machine Learning Scheme for Classification of Biomedical Datasets', pp. 128-139, 2009. doi.org/10.1007/978-3-642-01184-9_12
- [20] 'Diagnostic Prediction of Vertebral Column Using Rough Set Theory and Neural Network Technique', Information Technology Journal, vol. 13, no. 5, pp. 874-884, 2014. doi.org/10.3923/ijtj.2014.874.884
- [21] H. He and E. A. Garcia, 'Learning from imbalanced data', IEEE Trans. Knowledge Data Eng, vol. 21, pp. 1263-1284, 2009. doi.org/10.1109/TKDE.2008.239
- [22] 'Performance Analysis and Evaluation of Different Data Mining Algorithms used for Cancer Classification', IJARAI, vol. 2, no. 2, pp. 67-89, 2013. doi.org/10.14569/IJARAI.2013.020508

- [23] B.S. Everitt Cluster analysis. Edward Arnold, London, 1993.
- [24] V. Jakkula, Tutorial on Support Vector Machine (SVM). Retrieved August 15, 2011, from <http://eecs.wsu.edu/~vjakkula/SVMTutorial.doc>
- [25] 'Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease', Heart Disease. IJMLC, vol. 5, no. 5, pp. 414-419, 2015. doi.org/10.7763/IJMLC.2015.V5.544
- [26] M. S. U. Doulah, 'Application of Machine Learning Algorithms in Bioinformatics', Bioinformatics and Proteomics Open Access Journal, vol. 3, pp. 1-11, 2019. doi.org/10.23880/bpoj-16000127
- [27] Classification and Regression Trees. Wadsworth, Belmont, 1984
- [28] N. Dogan and Z. Tanrikulu, 'A Comparative Framework for Evaluating Classification Algorithms', in Proceedings of the World Congress of Engineering, 2010.
- [29] Duda et al., Pattern Classification, Wiley-New York, 2001
- [30] 'A principled Approach for Building and Evaluating Neural Network Classification Models', Decision Support Systems, pp. 233-246, 2004. [doi.org/10.1016/S0167-9236\(03\)00093-9](https://doi.org/10.1016/S0167-9236(03)00093-9)
- [31] 'Performance analysis of Hoeffding trees in data streams by using massive online analysis framework', Int. J. Data Mining, Modelling and Management, vol. 7, pp. 293-313, 2015. doi.org/10.1504/IJMMMM.2015.073865
- [32] 'An Algorithm Model for Incremental detection of Spam Reviews', International Journal of Modelling and Optimization, vol. 6, pp. 45-48, 2016. doi.org/10.7763/IJMO.2016.V6.501
- [33] Tharwat, 'Classification assessment methods', Applied Computing and Informatics, vol. 17, no. 1, pp. 39-56, 2021. doi.org/10.1016/j.aci.2018.08.003
- [34] D. Yu-Wei, 'Machine learning with R cookbook, Published by Packt Publishing Ltd', Birmingham, vol. 3, no. 2, 2015.
- [35] 'On the Comparison of Methods of Estimating Missing Values in Rectangular Lattice Designs', Journal of Mathematics and Statistics, vol. 14, pp. 201-208, 2018. doi.org/10.3844/jmssp.2018.201.208

How to Cite: Md. Siraj-Ud-Doulah, Md. Nazmul Islam. (2023). Performance evaluation of machine learning algorithm in various datasets. Journal of Artificial Intelligence, Machine Learning and Neural Network , 3(1), 68–84. <https://doi.org/10.55529/jaimlenn.32.14.32>

BIOGRAPHIES OF AUTHORS

	<p>Md. Siraj-Ud-Doulah , is associated with the Department of Statistics at Begum Rokeya University, Rangpur, Bangladesh. His academic interests focus on statistical modeling, data analysis, applied statistics, and quantitative research methodologies. He has contributed to academic and research activities related to statistical applications in social, economic, and scientific studies. He is dedicated to promoting evidence-based research and enhancing analytical approaches through statistical techniques and modern data-driven methods in higher education and research. Email: sdoulah_brur@yahoo.com</p>
	<p>Md. Nazmul Islam , is a faculty member/researcher in the Department of Statistics at Begum Rokeya University, Rangpur, Bangladesh. His research interests include statistical inference, biostatistics, data analytics, and applied quantitative methods. He actively participates in academic research and aims to apply statistical knowledge to solve real-world problems across various disciplines. His work emphasizes the importance of data interpretation, research accuracy, and innovative statistical applications in contemporary scientific and social research.</p>