Analysis of the prediction performance of decision tree-based algorithms

Fadwa AABOUB
Department of Mathematics and Informatics
Fundamental and Applied Mathematics Laboratory
Faculty of Sciences Ain Chock, Hassan II University
Casablanca, Morocco
f.aaboub@gmail.com

Hasna CHAMLAL
Department of Mathematics and Informatics
Fundamental and Applied Mathematics Laboratory
Faculty of Sciences Ain Chock, Hassan II University
Casablanca, Morocco
hasna.chamlal@univh2c.ma

Tayeb OUADERHMAN
Department of Mathematics and Informatics
Fundamental and Applied Mathematics Laboratory
Faculty of Sciences Ain Chock, Hassan II University
Casablanca, Morocco
t.ouaderhman@gmail.com

2023 International Conference on Decision Aid Sciences and Applications (DASA) | 979-8-3503-4205-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/DASA59624.2023.10286809

***Abstract*—Due to their many benefits, decision trees are extensively utilized to solve a variety of classification problems in the real world. Consequently, creating efficient and effective decision trees remains a common challenge for researchers in the fields of machine learning and data mining. In the literature, various decision tree algorithms with different attribute selection criteria are introduced. Although the currently available decision tree algorithms achieve different performances, none of them can produce optimal decision trees for a range of data sets. In this study, the efficacy of two traditional decision tree techniques is contrasted with the efficacy of two more recent decision tree approaches. On eleven real-world data sets, the four methods are compared in terms of four evaluation metrics: classification accuracy, tree depth, leaf nodes, and tree construction time.**

***Index Terms*—data set, decision tree, performance**

I. Introduction

Decision trees, non-parametric supervised classifiers that were first developed by Quinlan in 1986 [1], are one of the most effective and widely used approaches for tackling the classification challenge in the fields of machine learning and data mining. Decision trees are frequently employed due to their flexibility, efficiency, simple structure, strong robustness, low number of parameters, and other advantages.

Decision trees are often created using the top-down partitioning process. At each level of the tree construction, the most important attribute is selected from the attribute set using an attribute selection method. The training data set will subsequently be split into smaller, disjoint subsets based on the chosen splitting attribute. Recursively repeating the steps of choosing a splitting attribute and partitioning the data set for each non-empty subset until certain stopping conditions are satisfied.

Choosing an attribute selection technique that will be utilized to pick the splitting attributes presents the biggest obstacle in the development of decision trees. Consequently, a wide variety of decision tree methodologies are suggested in the literature, each of which is based on a different attribute evaluation metric. Even though there are numerous decision tree techniques, none of them can create optimal decision trees that perform effectively for several data sets.

In this study, the performances of four decision tree approaches are compared in terms of four evaluation metrics, including classification accuracy, tree depth (total number of nodes), leaf nodes, and tree construction time. These four strategies consist of the Iterative Dichotomizer 3 (ID3) [1], Classification And Regression Trees (CART) [2], Pearson's Correlation Coefficient based decision Tree (PCC-Tree) [3], and Dispersion Ratio based Decision Tree (DRDT) [4].

These are the main contributions of this work:

- Related work in the area of decision tree induction is first provided.
- An overview of the four decision tree processes that were compared is also provided, including ID3, CART, PCC-Tree, and DRDT.
- On eleven data sets from various fields, the performances of the four techniques are examined using four different evaluation metrics: classification accuracy, tree depth, leaf nodes, and tree construction time.
- Further analysis of the obtained results is conducted using the Friedman and post-hoc Nemenyi tests.

The remaining portions of the paper are organized as follows: Section II provides an overview of related work in the field of decision tree induction. Then, the ID3, CART, PCC-Tree, and DRDT decision tree algorithms are briefly described in Section III. The experimental results of the evaluation of the efficacy of the four approaches, along with a discussion, are then presented in Section IV. Finally, Section V gives a succinct conclusion.

II. Related work

In many fields, classification models have been constructed using decision tree approaches that are based on various splitting criteria. The Iterative Dichotomizer 3 (ID3) strategy [1], the C4.5 technique [5], and the Classification And Regression Trees (CART) method [2], are some of the popular traditional algorithms that implement the top-down greedy process to build decision trees. The ID3 methodology uses the Information Gain (IG) metric to evaluate the relevance of attributes. Whereas the Gain Ratio (GR) criterion, a kind of IG measure modification, is used by the C4.5 procedure to determine the splitting attributes. In contrast, the Gini Index (GI) is the node splitting criterion for the CART approach.
Moreover, decision tree development research has increased in recent years. Sun and Hu [6] select splitting attributes using the heuristic class constraint uncertainty for building decision trees. In addition, the Pearson's Correlation Coefficient based decision Tree (PCC-Tree) technique [3] uses the Pearson's Correlation Coefficient (PCC) as an impurity measure to define the splitting rule (splitting attribute as well as splitting point). Furthermore, Karabadji et al. [7] introduced a potent method for creating decision trees called PSO-DT, which creates the best decision tree by selecting the best pairing of training samples and attribute subsets. The Particle Swarm Optimization (PSO) methodology is used to combine the processes of data sampling and attribute selection. Additionally, in 2019, the Dispersion Ratio (DR) concept [4], which is a variant of the Correlation Ratio (CR) metric [8], is employed as an attribute evaluation measure for selecting significant attributes while creating decision trees. Moreover, Zhou et al. in 2021 [9] describe the Feature Weight-based Decision Tree (FWDT), which is an approach based on the feature weight principle, where the weights of features are determined by employing the ReliefF algorithm.

III. Background of the ID3, CART, PCC-Tree, and DRDT decision tree strategies

The four decision tree approaches that will be compared in this framework are briefly described in this section. These techniques consist of two classic decision tree strategies (ID3 and CART) as well as two newly developed procedures (PCC-Tree and DRDT). All these techniques follow the top-down partitioning strategy to generate decision trees; the only distinction is in the criteria used to select the attributes. An overview of the various attribute selection methods utilized by each of these four decision tree approaches is presented in the following sections:

Let $\mid D \mid$ represents the number of observations x_i in a training data set D. Assume that these observations are divided into N distinct classes denoted $C_1, C_2, \ldots, C_N$, and that the sizes of these classes are recorded as $\mid C_1 \mid, \mid C_2 \mid, \ldots, \mid C_N \mid$, respectively.

A. Iterative Dichotomizer 3 (ID3) method

The Information Gain (IG) metric serves as the foundation for the attribute selection methodology in the ID3 technique. The attribute that maximizes the IG measure is selected as the splitting attribute at each step of the tree's development. Before measuring the information gain of attributes, the entropy of information of the training data set as well as the entropy of information associated with each attribute must be determined.

The entropy of information of the training data set D is given by the following equation [1]:

$$\textit{Entropy}(D) = -\sum_{i=1}^{N} \frac{\mid C_i \mid}{\mid D \mid} \log_2 \frac{\mid C_i \mid}{\mid D \mid}. \tag{1}$$

Later, (2) [1] can be used to determine the entropy of information associated with a categorical attribute A:

$$\textit{Entropy}(D, A) = \sum_{v \in V} \frac{\mid D_v \mid}{\mid D \mid} \textit{Entropy}(D_v), \tag{2}$$

where v is a value of the attribute A, V refers to the set of all distinct values of A, $D_v = \{x_i \in D / A(x_i) = v\}$, and $\mid D_v \mid$ denotes the cardinal of D_v.
Finally, the IG of the attribute A in the data set D can be calculated as follows [1]:

$$\textit{IG}(D, A) = \textit{Entropy}(D) - \textit{Entropy}(D, A). \tag{3}$$

According to (3), the greater the IG of an attribute, the more significant it is.
The main drawbacks of the ID3 strategy are that it can only handle categorical attributes and is predisposed towards multi-valued attributes.

B. Classification And Regression Trees (CART) strategy

In contrast to the ID3 method, the CART technique produces binary decision trees by splitting each node into two child nodes. Additionally, it can handle attributes that are either continuous or discrete, as well as problems involving classification and regression. The CART technique employs the GI as a splitting criterion to select the best splitting attributes.

The GI of the whole data set D is first determined as follows [2]:

$$\textit{GI}(D) = 1 - \sum_{i=1}^{N} \left(\frac{\mid C_i \mid}{\mid D \mid} \right)^2. \tag{4}$$

Hence, for an attribute A, the GI metric can be calculated by (5) [2].

$$\textit{GI}(D, A) = \frac{\mid D_1 \mid}{\mid D \mid} \textit{GI}(D_1) + \frac{\mid D_2 \mid}{\mid D \mid} \textit{GI}(D_2), \tag{5}$$

where the attribute A divides the training data set D into two subsets, D_1 and D_2.
Finally, the reduction in impurity of the attribute A is computed as [2]:

$$\textit{GI}_{red}(A) = \textit{GI}(D) - \textit{GI}(D, A). \tag{6}$$

Therefore, the attribute that maximizes the measure $\textit{GI}_{red}$ will be the most efficient dividing attribute.

C. Pearson's Correlation Coefficient based decision Tree (PCC-Tree) algorithm

The PCC-Tree technique utilizes the PCC as an impurity measure to identify the ideal splitting rule (best splitting attribute and best splitting point). The main benefit of this strategy is its ability to manage several types of attributes.

The PCC-Tree technique begins by creating an intermediate vector for each attribute A in the data set (either a condition attribute or a class attribute Y) [3], so that:

$$V(A) = \begin{cases} A & if\ A \text{ is numerical} \\ \in \{0, 1, \ldots, F-1\} & if\ A \text{ is categorical,} \end{cases} \quad (7)$$

where F represents the number of distinct values for the attribute A.
Based on the first vector, a second vector is defined as follows [3]:

$$V(A,c) = \{v(A(x_i), c)\}_{i=1}^{|D|} = \begin{array}{cc} 1 & if \quad V(A(x_i)) \leq c \\ 2 & if \quad V(A(x_i)) > c, \end{array} \quad (8)$$

where c is a numerical value to split the value domain of the attribute A.
Finally, the splitting rule is formed by choosing the splitting attribute A^* and the splitting point c that satisfy the following condition [3]:

$$\begin{aligned} A^* &= \underset{A}{argmax} \mid P(V(A,c), V(Y)) \mid \\ &= \underset{A}{argmax} \frac{\mid cov(V(A,c), V(Y)) \mid}{\sqrt{var(V(A,c)) \times var(V(Y))}}, \end{aligned} \quad (9)$$

where P denotes the Pearson's correlation coefficient, cov represents the covariance, and var is the variance.

D. Dispersion Ratio based Decision Tree (DRDT) approach

The full procedure of the DRDT technique is divided into two stages. The numerical attributes in the data set are first discretized using the k-means methodology, and then the DR measure, which is a variation of the CR metric, is employed to evaluate attributes and choose the splitting attributes.

The formula for calculating the DR of an attribute A is given as follows [4]:

$$DR(A) = \sqrt{\frac{\sum_{i=1}^{N} \mid C_i \mid \left(\overline{m}_{C_i}^{(A)} - \overline{m}^{(A)}\right)^2}{\sum_{i=1}^{N} \sum_{j=1}^{C_i} \left(v_{jC_i}^{(A)} - \overline{m}^{(A)}\right)^2}}, \quad (10)$$

where $\overline{m}_{C_i}^{(A)}$ is the relative importance of the attribute A to the class label C_i, $\overline{m}^{(A)}$ reflects the overall importance of the attribute A, and $v_{jC_i}^{(A)}$ signifies the relative importance of the j-th value of the attribute A in class C_i.
The attribute with the highest value of DR will be utilized as the splitting attribute since it is the most significant attribute according to the DR measure.

Table I: Detailed information on the data sets used in the experiment.

N^o.	Data sets	Instances	Attributes	Classes
1	Bankruptcy	250	6	2
2	Breast Cancer	286	9	2
3	Glass	214	9	2
4	Hepatitis	155	19	2
5	Iris	150	4	3
6	New Thyroid	215	5	3
7	Somerville	143	6	2
8	Sonar	208	60	2
9	Statlog	270	13	2
10	TAE	151	5	3
11	Wine	178	13	3

Table II: Classification accuracies (%) achieved by the ID3, CART, PCC-Tree, and DRDT approaches on the eleven data sets.

Data sets	ID3	CART	PCC-Tree	DRDT
Bankruptcy	**98.34**	98.30	91.12	97.52
Breast Cancer	69.94	**69.97**	68.67	67.37
Glass	74.10	**74.51**	67.29	66.35
Hepatitis	**80.13**	78.47	77.50	77.74
Iris	93.49	**93.54**	93.49	88.17
New Thyroid	**91.68**	89.89	89.71	90.09
Somerville	62.03	**62.18**	58.68	60.43
Sonar	**73.13**	71.12	71.39	70.20
Statlog	**79.09**	78.08	74.70	76.83
TAE	46.65	45.85	49.19	**57.64**
Wine	90.28	88.39	**93.19**	82.29
Average	**78.08**	77.30	75.90	75.88

IV. Experimental study

For the purpose of experimentally comparing the effectiveness of the ID3, CART, PCC-Tree, and DRDT methodologies, several real-world data sets were collected from various fields. Table I describes the data sets employed to conduct this comparison.

All of these data sets are treated using the five-fold cross-validation approach [10]. Consequently, each estimated result presented in this study is based on an average of 100 times five-fold cross-validation.

The ID3, CART, PCC-Tree, and DRDT decision tree techniques are applied to the eleven data sets, and the performances are compared in terms of four evaluation measurements: classification accuracy [11], tree depth, leaf node count, and tree construction time. The attained results for each of these four metrics are depicted in Tables II, III, IV, and V, respectively.

The subsequent conclusions are formed from the examination of the obtained results:

- It is clear from the examination of Table II that both the PCC-Tree and DRDT algorithms can only reach the highest classification accuracy on a single data set. In contrast to the ID3 and CART techniques, which produce the highest classification accuracy for five and four data sets, respectively. Furthermore, the ID3 method yields the best average classification accuracy, which is 0.78%,

Table III: Tree depths produced by the ID3, CART, PCC-Tree, and DRDT techniques.

Data sets	ID3	CART	PCC-Tree	DRDT
Bankruptcy	**3.02**	**3.02**	30.07	50.83
Breast Cancer	12.55	**11.70**	115.03	409.66
Glass	13.87	**13.20**	61.46	120.14
Hepatitis	8.01	**6.54**	36.90	140.81
Iris	5.20	**5.10**	12.44	81.96
New Thyroid	6.58	**6.03**	17.98	30.24
Somerville	11.92	**11.33**	49.41	205.35
Sonar	13.65	**12.31**	35.34	161.57
Statlog	13.38	**13.20**	81.51	275.88
TAE	14.62	**13.33**	32.39	412.85
Wine	7.84	**7.70**	13.69	157.99
Average	10.06	**9.41**	44.20	186.12

Table IV: Leaf nodes generated by the ID3, CART, PCC-Tree, and DRDT algorithms.

Data sets	ID3	CART	PCC-Tree	DRDT
Bankruptcy	**2.01**	**2.01**	15.54	34.22
Breast Cancer	6.77	**6.35**	58.01	318.80
Glass	7.43	**7.10**	31.23	83.38
Hepatitis	4.50	**3.77**	18.95	98.42
Iris	3.10	**3.05**	6.72	67.94
New Thyroid	3.79	**3.51**	9.49	22.05
Somerville	6.48	**6.18**	25.21	164.06
Sonar	7.32	**6.66**	18.17	91.22
Statlog	7.19	**7.10**	41.26	243.72
TAE	7.81	**7.17**	16.70	384.74
Wine	4.42	**4.35**	7.35	158.35
Average	5.53	**5.20**	22.60	151.54

Table V: Comparison among the tree construction times (s) of the ID3, CART, PCC-Tree, and DRDT procedures.

Data sets	ID3	CART	PCC-Tree	DRDT
Bankruptcy	**0.002**	**0.002**	0.018	0.039
Breast Cancer	**0.003**	**0.003**	0.081	0.372
Glass	0.006	**0.003**	0.149	0.101
Hepatitis	0.004	**0.003**	0.073	0.177
Iris	**0.002**	**0.002**	0.014	0.067
New Thyroid	0.003	**0.002**	0.034	0.022
Somerville	**0.002**	**0.002**	0.032	0.170
Sonar	0.033	**0.012**	1.298	0.680
Statlog	0.005	**0.004**	0.133	0.361
TAE	31.976	2.843	**0.029**	0.300
Wine	0.006	**0.003**	0.105	0.240
Average	2.913	0.262	**0.179**	0.230

2.18%, and 2.20% higher than that reached by the CART, PCC-Tree, and DRDT techniques, respectively.

- The analysis of Table III reveals that both the PCC-Tree and DRDT methods are unable to produce the best result for any given data set in terms of the tree depth measure. However, the ID3 method produces the smallest tree depth on the Bankruptcy data set. On the other hand, the CART methodology can generate decision trees with the smallest tree depths for all eleven data sets.
- When comparing the CART technique to the ID3, PCC-Tree, and DRDT algorithms according to the number of leaf nodes, Table IV analysis demonstrates that it produces the fewest leaf nodes on all the tested data sets.

Table VI: Average ranks for the four measurements obtained by the ID3, CART, PCC-Tree, and DRDT techniques.

Methods	Testing accuracy	Tree depth	Leaf nodes	Construction time
ID3	1.68	1.95	1.95	2.00
CART	2.09	1.05	1.05	1.36
PCC-Tree	3.05	3.00	3.00	3.09
DRDT	3.18	4.00	4.00	3.55

- With the exception of the TAE data set, it is clear from the examination of Table V that the CART decision tree strategy achieves the shortest construction time on all data sets. Nevertheless, the best average tree-building time is provided by the PCC-Tree approach.

Furthermore, in order to determine whether there is any significant difference in the effectiveness of the ID3, CART, PCC-Tree, and DRDT procedures, they are put through the Friedman test [12].

The Friedman test is a statistical non-parametric test that enables the comparison of a set of classifier models using various data sets. In general, the null hypothesis to be tested states that the performances of the compared classifiers are equivalent at a significance level α. For k classifiers and N data sets, the Friedman statistic χ_F^2 is calculated as follows [12]:

$$\chi_F^2 = \frac{12N}{k(k+1)} \left[\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4} \right], \qquad (11)$$

where R_j is the average rank of the j-th classifier according to a specific metric.

Then, the Iman's F statistic can be determined as follows [12]:

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1) - \chi_F^2}. \qquad (12)$$

The statistic F_F, which is distributed according to the F-distribution with $(k-1)$ and $(k-1)(N-1)$ degrees of freedom, has a critical value $F_\alpha((k-1), (k-1)(N-1))$. Finally, based on the Friedman test, if the statistic F_F exceeds the critical value, the null hypothesis will be rejected. Consequently, there is a significant difference among the performances of the classifiers.

In this study, four independent Friedman tests are conducted, with the null hypotheses stating, respectively, that the testing accuracies, tree depths, leaf nodes, and tree building times reached by the ID3, CART, PCC-Tree, and DRDT decision tree algorithms are identical at the significance level $\alpha = 0.05$.

Table VI depicts the average ranks for classification accuracies, tree depths, leaf nodes, and tree construction times of the ID3, CART, PCC-Tree, and DRDT techniques.

Given that $k = 4$ decision tree techniques and $N = 11$ data sets are available, the Friedman test results for the four metrics are presented in Table VII.

Since Iman's F statistics for the four measures are higher than

Table VII: Friedman test results for the four metrics.

Measures	Accuracy	Tree depth	Leaf nodes	Building time
χ_F^2	10.55	32.43	32.43	19.69
F_F	4.70	566.19	566.19	14.80

Table VIII: Pairwise differences of average ranks for the four metrics.

Strategies	ID3	CART	PCC-Tree	DRDT
ID3	–			
	–			
	–			
	–			
CART	0.41	–		
	0.91	–		
	0.91	–		
	0.64	–		
PCC-Tree	1.36	0.95	–	
	1.05	**1.95**	–	
	1.05	**1.95**	–	
	1.09	**1.73**	–	
DRDT	**1.50**	1.09	0.14	–
	2.05	**2.95**	1.00	–
	2.05	**2.95**	1.00	–
	1.55	**2.18**	0.45	–

the critical value $F_{0.05}(3, 30) = 2.92$, as shown in Table VII, the four null hypotheses are consistently rejected. Finally, the classification accuracy, tree depth, leaf node count, and tree development time of the ID3, CART, PCC-Tree, and DRDT approaches differ from one another.

In the case of rejecting the null hypothesis, a post-hoc Nemenyi test [12] is further used to assess which classifier performs better than the others at the significance level α. According to the post-hoc Nemenyi test, the performances of two classifiers are significantly different if their corresponding average ranks differ by at least the Critical Difference (CD), which is determined by the following equation [12]:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \tag{13}$$

here q_α is the critical value for post-hoc tests.

The pairwise differences (in absolute value) of average ranks for classification accuracies (in black), tree depths (in orange), leaf nodes (in purple), and tree construction times (in blue) obtained by the ID3, CART, PCC-Tree, and DRDT methods are provided in Table VIII. Pairwise differences greater than $CD = 1.41$ are indicated in bold font.

Through analysis of Table VIII, the ID3 methodology significantly outperforms the DRDT method in terms of each of the four metrics. Additionally, although the difference is not particularly significant, the ID3 strategy can perform better than the PCC-Tree approach. Whereas all of the PCC-Tree and DRDT techniques are outperformed by the CART procedure in terms of tree depth, leaf nodes, and tree construction time. In contrast, the ID3 and CART approaches, as well as the PCC-Tree and DRDT methodologies, do not differ significantly in terms of the four tested measures.

Overall, the post-hoc Nemenyi test shows that the ID3 and CART procedures are both better and more effective than the PCC-Tree and DRDT decision tree methods in terms of classification accuracy, tree depth, leaf node count, and tree building time.

V. Conclusion

In this study, the performances of the traditional decision tree methods ID3 and CART are contrasted with those of the recently created decision tree approaches PCC-Tree and DRDT. In order to compare the performances according to classification accuracy, tree depth, leaf nodes, and time needed to build the tree using the four strategies, several data sets with mixed-type attributes were gathered. The obtained results clearly demonstrate that the efficacy of the ID3 approach is comparable to that of the CART algorithm, and that both techniques outperform the recently proposed procedures PCC-Tree and DRDT in terms of efficacity. The results are also supported by the Friedman and post-hoc Nemenyi tests.

To create efficient decision trees, a novel splitting measure based on preordonance theory will be introduced in future work. The effectiveness of the new decision tree approach will be contrasted with that of the currently used techniques.

Acknowledgements

This work was supported by the National Center for Scientific and Technical Research of Morocco (CNRST).

References

[1] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.
[2] L. Brieman, J. Friedman, C. J. Stone, and R. Olshen, "Classification and regression tree analysis," 1984.
[3] Y. Mu, X. Liu, and L. Wang, "A pearson's correlation coefficient based decision tree and its parallel implementation," *Information Sciences*, vol. 435, pp. 40–58, 2018.
[4] S. Roy, S. Mondal, A. Ekbal, and M. S. Desarkar, "Dispersion ratio based decision tree model for classification," *Expert Systems with Applications*, vol. 116, pp. 1–9, 2019.
[5] J. R. Quinlan, "C4.5: Program for machine learning," *Morgan Kaufmann Pub*, 1993.
[6] H. Sun and X. Hu, "Attribute selection for decision tree learning with class constraint," *Chemometrics and Intelligent Laboratory Systems*, vol. 163, pp. 16–23, 2017.
[7] N. E. I. Karabadji, I. Khelf, H. Seridi, S. Aridhi, D. Remond, and W. Dhifli, "A data sampling and attribute selection strategy for improving decision tree construction," *Expert Systems with Applications*, vol. 129, pp. 84–96, 2019.
[8] S. Roy, S. Mondal, A. Ekbal, and M. S. Desarkar, "Crdt: correlation ratio based decision tree model for healthcare data mining," in *2016 IEEE 16th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 36–43, IEEE, 2016.
[9] H. Zhou, J. Zhang, Y. Zhou, X. Guo, and Y. Ma, "A feature selection algorithm of decision tree based on feature weight," *Expert Systems with Applications*, vol. 164, p. 113842, 2021.
[10] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, pp. 1137–1145, Montreal, Canada, 1995.
[11] A. Ghasempour and M. Martinez-Ramon, "Short-term electric load prediction in smart grid using multi-output gaussian processes regression," *IEEE Kansas Power and Energy Conference (IEEE KPEC)*, 2023.
[12] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine learning research*, vol. 7, pp. 1–30, 2006.