Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 15, Issue. 5, May 2026, pg.1 – 6

Evaluating Predictive Accuracy: A Systematic Comparison of CART and M5P

Abhijeet Mankar[1]; Sudhakar Bhoite[2]; Kirti Raskar[3]

[1]Department of Computer Science, T. C. College, Baramati (Affiliated to SPPU, Pune), India
[2]School of Computer Science and Applications, CSIBER, Kolhapur (Affiliated to Shivaji University), India
[3]Department of Science and Computer Science, MIT College, Alandi, Pune (Affiliated to SPPU, Pune), India
[1] abhijeetmankar@gmail.com; [2] sdbhoite05@gmail.com; [3] kirtiraskar1711@gmail.com

DOI: https://doi.org/10.47760/ijcsmc.2026.v15i05.001

Abstract: Predictive models following tree representation are favored because of resilience and interpretability. Still their relative predictive behavior across heterogeneous datasets remains insufficiently understood. This study presents a statistically rigorous comparison of CART and M5P models across five regression datasets. The rpart package is used in RStudio to build CART regression tree while RWeka package is used to build M5P tree in RStudio. A repeated 70-30 cross-validation framework is used, and model performance is estimated using paired statistical hypothesis testing. Besides performance comparison, the study discusses the tree structure details of M5P and rpart. Experimental results display that M5P demonstrates statistically considerable improvements, while CART remains competitive on low-dimensional datasets.

Keywords: Tree-based models, CART, M5P, Statistical model comparison, Cross-validation, Predictive modeling

I. INTRODUCTION

Decision tree algorithms give us an idea about intermediate decision nodes. They are known as white box approach. On the contrary side, artificial neural networks are known as black box approach, because they are indeed complicated to interpret. CART is among the popular algorithms used in machine learning. In machine learning, the original dataset is divided into training and test set.

While various research papers report predictive accuracy of machine learning algorithms, several studies show absence of statistical rigor and unfit to assess whether observed performance differences are significant.

Particularly, relative evaluations of classical tree-based models rely on single train-test splits, confining the reliability of conclusions. The research paper deal with this gap by performing a relative analysis of CART and M5P models across multiple datasets having statistical base. The research question addressed in this research paper is: "Is there a considerable distinctness in predictive accomplishment between CART and M5P models across datasets?".

II. LITERATURE REVIEW

CART algorithm is operationalized in RStudio through rpart package. There are diverse application areas of decision trees in which CART algorithm is prominent. The M5P algorithm in RStudio is a part of RWeka package. The basic difference between CART (i.e. rpart in RStudio) and M5P is that, CART regression tree has leaf nodes that represent mean(average) value of all the observations in the leaf node for the response variable column. On the other hand, M5P algorithm applies linear regression to the leaf nodes obtained in regression tree [1][2].

Behnood *et al.* [3] performed a study to create a predictive model for estimating the binding power of usual concrete and high-performance concrete. In their research, they developed a model based on M5P algorithm for forecasting the binding power of usual concrete and high-performance concrete. Their findings indicated effectiveness of M5P algorithm in predicting the binding power of usual concrete and high-performance concrete using recognized figures of composition proportions and age of testing.

Bienvenido-Huertas *et al.* [4] conducted a study to investigate about capability of six algorithms related to regression in estimating the threat of energy shortage precisely. Their study simulated 38800 cases and found that M5P, multilayer perceptron and SVR provided the highest accuracy. Regarding computing time, M5P performs better than the rest. The aim of this exploratory study was to sort out the most suitable algorithm for forecasting economically marginalized households in Chile. The findings of this study provide the starting point towards environment conscious development of social housing by considering the financial condition of economically marginalized households.

E. Chen and X. J. He [5] conducted a study to forecast crude oil prices. They compared the accomplishment of five distinct models including CART as well as M5P. They discovered that CART model was the worst performer among all models, while M5P outperformed CART.

Elangovan *et al.* [6] conducted a study to highlight the incorporation of clustering, regression, and classification techniques to strengthen our understanding of climate dynamics. The researchers utilized 10 different machine learning approaches including M5P. The researchers found that, mean absolute error of random forest model was lowest and mean absolute error of M5P model was second lowest among all the 10 models.

Jayasinghe *et al.* [7] explained the use of regression trees in place of other machine learning and statistical modelling methods for predicting the energy generation at 5 unceasing energy facilities in Sri Lanka. The researchers used CART algorithm and implemented it in R 4.2.3 with "rpart" and "rpart.plot" packages. The researchers found that highly precise predictive models might be evolved via raising the regression tree depth.

Kebonye *et al.* [8] showed how to enhance the performance of CART model for digital soil mapping. The study was conducted in Czech Republic. Their study demonstrated that using various data splitting techniques and feature selection methods improved the performance of the CART decision tree for digital soil mapping. The researchers found soil organic carbon level were higher in high-altitude forested regions in this study. They also highlighted the interpretability achieved using CART model.

Nhu *et al.* [9] applied M5P, RF, random tree (RT) and reduced error pruning tree (REPT). The researchers predicted the day-to-day level of water of a lake in Iran through the four decision tree techniques. The model attainment was assessed using six different quantitative measures including mainly RMSE, MAE, coefficient of determination (R^2). The result shown excellent prediction competence by each of the developed models. But the M5P model was better in performance as compared to others, then RF and RT and then REPT.

Nikoo *et al.* [10] developed rules towards optimum reservoir operation and water withdrawal from water bodies such as river by considering supply of water and pollution regulation targets. The Zayandehrood Dam downstream sub-basin within Iran is utilized as the information by the researchers. They utilized M5P and SVR model for comparison. As per the researchers, M5P model is white-box means it shows the rules and relationships within input and output, while SVR maintains it concealed (black-box). M5P model assists the decision maker about the parameters influencing the decision and its influence. As per the researchers, M5P model forecasted the results comparatively superior to SVR model.

Pande *et al.* [11] evaluated the effectiveness of various machine learning models, such as ANN and M5P Tree for predicting the SPI (standardized precipitation index), which is a measure of drought. The study area selected was Godavari basin in Maharashtra. The rainfall data from two stations were collected for 20 years (2000-2019) of which beginning 14-year data were used to train the model and extra 6-year data were exercised to validate the model. The researchers found out of the qualitative and quantitative examination that, M5P model is the most precise model for forecasting the aridity index at two distinct measures (SPI-3, SPI-6) at both the stations under study.

Shah *et al.* [12] carried out research for building models for predicting the binding power and tensile stress of concrete incorporating silica fume. They well trained M5P model and expressed the association between predictor and response variables, which predicted the tensile stress of silica fume concrete more accurately. The M5P model exhibited enhanced estimative performance in comparison to linear regression analysis for the duo binding power and tensile stress.

P. Skurowski and M. Pawlyta [13] discovered that traditional tree-based regression methods are inapplicable to the domain of their study, which aimed at reconstructing gaps in motion capture sequences. They observed that the M5P algorithm was appropriate for this task and outperformed other methods, including neural networks, particularly when dealing with longer gaps.

Thai *et al.* [14] analyzed the winter wheat grain yield under dissimilar fertilizer applications in Germany. They used four models including M5P regression tree model for evaluating the response of grain yield. The results of ANOVA show the principal components influencing the winter wheat yield. The M5P demonstrated predictive performance under LMM, but M5P revealed principal yield independent variables which were not represented by the LMM. The researchers further concluded to co-use the ANOVA, M5P, and LMM model for the conclusion and predictions in long-term yield studies.

Yukseler *et al.* [15] conducted a study to assess the flood hazard areas along the Black Sea coast of Trabzon province in Turkiye. They utilized M5P Rule Tree, logistic regression, M5P Regression Tree models. As per the flood inventory, in the study area sixteen flood events took place in five distinct places. The areas under investigation were transformed into point data constituting into point data having 1600 points, of which 800 were identified as waterlogged and 800 as non-waterlogged, selected through arbitrary sampling. The assessment metrics utilized by the researchers in this study were AUC, accuracy, precision, recall, F-score indices. For the training dataset, M5P Rule Tree achieved the highest AUC of 0.989 followed by M5P Regression Tree with 0.927. For the test dataset, M5P Rule Tree attained the highest AUC with 0.983 followed by M5P Regression Tree with 0.935 and logistic regression with 0.851. According to researchers, principal factors affecting flood hazards were precipitation and elevation. For the logistic regression model, distance from roads was found to be the third principal factor. By considering the fine performance of M5P Regression Tree Model and M5P Rule Tree Model, the researchers stated that these models will be utilized in the research hereafter to decrease damages related with flood.

III. METHODOLOGY

A. *Datasets*

The datasets chosen for this study are taken from UC Irvine Machine Learning Repository. The datasets used are of different size in terms of number of observations. The smallest dataset used is 'Liver Disorders' with 345 observations while the largest dataset used is 'Combined Cycle Power Plant' dataset with 9568 observations. Table I gives the details about the datasets used in the study.

The fish toxicity dataset has the response variable 'LC50' which means the acute aquatic toxicity towards the fish. The response variable of concrete compressive strength dataset is evaluated in Megapascal. The combined cycle power plant dataset has the target variable 'PE' which means net hourly electrical energy output. It is measured in Megawatts. The white wine quality dataset has the target variable 'Quality' whose score is between 0 and 10. The liver disorders dataset has the target variable 'Drinks'.

TABLE I. DATASETS DESCRIPTION

Name of the dataset	**Target variable considered**	**Subject Area (as per UCI ML repository)**	**Total no. of variables (including target)**	**Total no. of observations**
QSAR Fish toxicity[16]	LC50	Physics and Chemistry	7	908
Concrete Compressive Strength[17]	Concrete Compressive Strength	Physics and Chemistry	9	1030
Combined Cycle Power Plant[18]	PE	Computer Science	5	9568
White Wine Quality[19]	Quality	Business	12	4898
Liver Disorders[20]	Drinks	Heath and Medicine	6 (excluding "selector" variable)	345

The prediction at the leaf nodes in CART regression tree is mean of the values of response variable. On the other side, M5P assimilate linear regression models at leaf nodes. This likely offers benefits for datasets with linear trends. While rpart package in RStudio is used for both classification and regression models, the regression model uses variance reduction. M5P is generally much more accurate than CART on numeric targets.

B. Experimental Design

The datasets are divided into 70% training and 30% testing set. Then 10-split model evaluation indices were created on the dataset of training. For both the CART and M5P, hyperparameter grid was designed. Table II gives the details about the hyperparameter grid. In case of M5P algorithm, there is no 'maxdepth' hyperparameter available. Then grid search accompanied by 10-fold cross-validation was performed to choose the optimal hyperparameters. Table III and IV mentions the optimal hyperparameters for both algorithms. Finally, the optimal hyperparameters selected for training final model over whole dataset of training. The eventual trained model was assessed on the held-out test dataset to produce test RMSE. It is discussed in TABLE V.

TABLE II. HYPERPARAMETERS SEARCH SPACE

<table>
<tr><th>Algorithm</th><th>Hyperparameter</th><th>Information</th><th>Hyperparameter Values considered</th></tr>
<tr><td rowspan="2">M5P</td><td>M</td><td>Minimum number of observations per leaf</td><td>{7, 8, 10, 15, 20}</td></tr>
<tr><td>U</td><td>Unpruned tree (TRUE=unpruned, FALSE=pruned)</td><td>{TRUE, FALSE}</td></tr>
<tr><td rowspan="4">RPART</td><td>Cp</td><td>Complexity parameter</td><td>{0.001, 0.005, 0.01}</td></tr>
<tr><td>Minsplit</td><td>Minimum no. of observations required to split a node</td><td>{20, 30}</td></tr>
<tr><td>Minbucket</td><td>Minimum no. of observations in a leaf node</td><td>{7, 8, 10, 15, 20}</td></tr>
<tr><td>Maxdepth</td><td>Maximum tree depth</td><td>{5, 10}</td></tr>
</table>

TABLE III. OPTIMAL HYPERPARAMETERS FOR M5P

<table>
<tr><th>Dataset</th><th>M</th><th>U</th><th>RMSE(CV)</th></tr>
<tr><td>QSAR Fish toxicity</td><td rowspan="2">20</td><td rowspan="5">TRUE</td><td>0.9495095</td></tr>
<tr><td>Concrete Compressive Strength</td><td>7.071132</td></tr>
<tr><td>Combined Cycle Power Plant</td><td>15</td><td>3.933137</td></tr>
<tr><td>White Wine Quality</td><td>20</td><td>0.7223753</td></tr>
<tr><td>Liver Disorders</td><td>10</td><td>3.367252</td></tr>
</table>

TABLE IV. OPTIMAL HYPERPARAMETERS FOR RPART

<table>
<tr><th>Dataset</th><th>Minsplit</th><th>Cp</th><th>Min bucket</th><th>Max depth</th><th>RMSE(CV)</th></tr>
<tr><td>QSAR Fish toxicity</td><td rowspan="4">20</td><td rowspan="4">0.001</td><td>20</td><td rowspan="4">10</td><td>1.017939</td></tr>
<tr><td>Concrete Compressive Strength</td><td>10</td><td>7.825106</td></tr>
<tr><td>Combined Cycle Power Plant</td><td>7</td><td>4.612911</td></tr>
<tr><td>White Wine Quality</td><td>20</td><td>0.7332439</td></tr>
<tr><td>Liver Disorders</td><td>30</td><td>0.005</td><td>7</td><td>5</td><td>3.040126</td></tr>
</table>

TABLE V. TEST RMSE FOR M5P AND RPART

Dataset	*M5P (RMSE)*	*RPART (RMSE)*
QSAR Fish toxicity	0.8925693	0.9727028
Concrete Compressive Strength	7.210149	8.022088
Combined Cycle Power Plant	4.189853	4.64459

White Wine Quality	0.7489722	0.7634385
Liver Disorders	3.535404	3.541029

C. *Statistical Testing*

Statistical test like 'Paired t-test' was applied between rpart and M5P to determine whether observed performance differences were significant.

TABLE VI. PAIRED STATISTICAL TEST RESULTS

Dataset	Paired t-test (p-value)	Mean difference	Significant Difference
QSAR Fish toxicity	4.922e-05	-0.1349274	Yes
Concrete Compressive Strength	2.107e-06	-2.560883	Yes
Combined Cycle Power Plant	5.187e-09	-1.866014	Yes
White Wine Quality	0.004712	-0.0340745	Yes
Liver Disorders	0.1321	0.2957289	No

IV. CONCLUSIONS

Table VI describes the paired t-test performed between RPART and M5P.

For QSAR Fish toxicity dataset, the paired t-test gives $p<0.05$ indicating significant difference. Also, mean difference is less than 0 which means M5P significantly outperforms RPART.

For Concrete compressive strength dataset, the paired t-test gives $p<0.05$ indicating significant difference. Also, mean difference is less than 0 which means M5P significantly outperforms RPART. For combined cycle power plant dataset, the paired t-test gives $p<0.05$ indicating significant difference. Also, mean difference is less than 0 which means M5P significantly outperforms RPART. For White wine quality dataset, the paired t-test gives $p<0.05$ indicating significant difference. Also, mean difference is less than 0 which means M5P significantly outperforms RPART. For the Liver disorders dataset, the paired t-test gives $p>0.05$ indicating no significant difference. Also, mean difference is not less than 0, hence RPART outperforms M5P.

The TABLE VII discusses the tree structure of both M5P model and RPART trees.

TABLE VII. DECISION TREE DETAILS

	M5P			RPART		
Dataset	**Total Nodes**	**Leaf Nodes**	**Depth**	**Total Nodes**	**Leaf Nodes**	**Depth**
QSAR Fish toxicity	14	8	4	39	20	7
Concrete Compressive Strength	12	7	5	77	39	9
Combined Cycle Power Plant	424	213	12	35	18	6
White Wine Quality	55	28	5	171	86	9
Liver Disorders	10	6	4	19	10	5

The depth of tree produced by rpart model for QSAR fish toxicity dataset is 9 whereas the depth of tree produced by M5P model is 7. Similarly, the depth of the tree produced by rpart model for Concrete compressive strength dataset is 9 and for M5P model the depth is 5. Also, the depth of the tree produced by rpart model for White wine quality dataset is 9 and that of M5P model is 5. These three datasets are medium sized in terms of number of observations. The liver disorders dataset is comparatively small having 345 observations. For this dataset, rpart model and M5P model trees have depth values 5 and 4 respectively. The combined cycle power plant dataset, having 9568 observations, is the largest in terms of number of observations. The depth of the rpart tree model for combined cycle power plant dataset is 6 and for the depth of M5P tree model for combined cycle power plant is 12. From the TABLE VII, for the small size datasets, both the models are nearly equal in depth. For the medium size datasets, the depth of M5P is nearly half of the depth of tree produced by rpart model. For the large size dataset, the rpart tree model depth is half the depth of M5P model. From the TABLE VI, we observe that the predictive performance of M5P is significantly better than rpart for the four datasets. Only for the liver disorders dataset which is small size dataset, the predictive performance of rpart is better than M5P.

ACKNOWLEDGEMENT

We are thankful to our family members, higher authorities of our institutions and our colleagues.

REFERENCES

[1]. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.

[2]. Y. Wang and I. H. Witten, "Inducing model trees for continuous classes", *Proceedings of the ninth European conference on machine learning*, vol. 9(1), pp. 128-137, 1997.

[3]. A. Behnood, V. Behnood, M. M. Gharehveran, and K. E. Alyamac, "Prediction of the compressive strength of normal and high-performance concretes using M5P model tree algorithm", *Construction and Building Materials*, vol. 142, pp.199-207, 2017.

[4]. D. Bienvenido-Huertas, J. A. Pulido-Arcas, C. Rubio-Bellido, and A. Pérez-Fargallo, "Prediction of fuel poverty potential risk index using six regression algorithms: a case-study of Chilean social dwellings", *Sustainability*, vol. 13, no. 5, p. 2426, 2021.

[5]. E. Chen and X. J. He, "Crude oil price prediction with decision tree based regression approach", *J. Int. Technol. Inf. Manag*., vol. 27, no. 4, pp. 2-16, May 2019.

[6]. R. Elangovan, P. Rajesh, R. Arulpavai, and M. Govindhan, "Advanced computational approaches to climate change with global temperature trends: a data mining and machine learning perspective", in *2025 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE)*, IEEE, pp. 1-6, 2025.

[7]. J. Jayasinghe, P. Ekanayake, O. Panahatipola, C. I. Madhushani, and U. Rathnayake, "Forecasting the power generation at renewable power plants in Sri Lanka using regression trees", *Results in Engineering*, vol. 22, p.102111, 2024.

[8]. N. M. Kebonye, P. C. Agyeman, and J. K. Biney, "Optimized modelling of countrywide soil organic carbon levels via an interpretable decision tree", *Smart Agricultural Technology*, vol. 3, p.100106, 2023.

[9]. V. H. Nhu, H. Shahabi, E. Nohani, A. Shirzadi, N. Al-Ansari, S. Bahrami, S. Miraki, M. Geertsema, and H. Nguyen, "Daily water level prediction of Zrebar Lake (Iran): a comparison between M5P, random forest, random tree and reduced error pruning trees algorithms", *ISPRS International Journal of Geo-Information*, vol. 9, no. 8, p.479, 2020.

[10]. M. R. Nikoo, A. Karimi, R. Kerachian, H. Poorsepahy-Samian, and F. Daneshmand, "Rules for optimal operation of reservoir-river-groundwater systems considering water quality targets: application of M5P model", *Water Resources Management*, vol. 27, no. 8, pp.2771-2784, 2013.

[11]. C. B. Pande, N. Al-Ansari, N. L. Kushwaha, A. Srivastava, R. Noor, M. Kumar, K. N. Moharir, and A. Elbeltagi, "Forecasting of SPI and meteorological drought based on the artificial neural network and M5P model tree", *Land*, vol. 11, no. 11, p.2040, 2022.

[12]. H. A. Shah, M. L. Nehdi, M. I. Khan, U. Akmal, H. Alabduljabbar, A. Mohamed, and M. Sheraz, "Predicting Compressive and Splitting Tensile Strengths of Silica Fume Concrete Using M5P Model Tree Algorithm", *Materials*, vol. 15, no. 15, p. 5436, 2022.

[13]. P. Skurowski and M. Pawlyta, "Tree based regression methods for gap reconstruction of motion capture sequences", *Biomedical Signal Processing and Control*, vol. 88, p.105641, 2024.

[14]. T. H. Thai, R. A Omari, D. Barkusky, and S. D. Bellingrath-Kimura, "Statistical analysis versus the m5p machine learning algorithm to analyze the yield of winter wheat in a long-term fertilizer experiment", *Agronomy*, vol. 10, no. 11, p.1779, 2020.

[15]. U. Yukseler, A. Toprak, E. Gul, and O. F. Dursun, "Flood hazard mapping using M5 tree algorithms and logistic regression: a case study in East Black Sea Region", *Earth Science Informatics*, vol. 16, no. 3, pp.2033-2047, 2023.

[16]. D. Ballabio, M. Cassotti, V. Consonni, and R. Todeschini, "QSAR fish toxicity", UCI Machine Learning Repository, 2015. [Online]. Available: https://doi.org/10.24432/C5JG7B

[17]. I. Yeh, "Concrete Compressive Strength", UCI Machine Learning Repository, 1998. [Online]. Available:https://doi.org/10.24432/C5PK67

[18]. P. Tfekci, and H. Kaya, "Combined Cycle Power Plant", UCI Machine Learning Repository, 2014. [Online]. Available: https://doi.org/10.24432/C5002N

[19]. P. Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis, "Wine Quality", UCI Machine Learning Repository, 2009. [Online]. Available: https://doi.org/10.24432/C56S3T

[20]. "Liver Disorders", UCI Machine Learning Repository, 2016. [Online]. Available: https://doi.org/10.24432/C54G67

