International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 03 | Mar 2023 www.irjet.net p-ISSN: 2395-0072

# Predicting User Ratings of Competitive ProgrammingContests using Decision Tree ML Model

**Mayank Bhardwaj, Akshat Tuknait, Gopal Gupta**

[1,2,3] *Maharaja Agrasen Institute of Technology*

---

## ABSTRACT

This research paper presents the use of a decision tree machine learning model for predicting the future user ratings of competitive programming contests. The model was trained on a dataset containing information on the past performance of contestants in various contests and achieved an MSE of 8494 and an RMSE of 92 on the test data. The decision tree model is well- suited for this task because it can handle large amounts of data and handle both numerical and categorical data, and the use of a maximum depth of 32 helps to prevent overfitting. These characteristics make the decision tree model an effective tool for predicting the future user ratings of competitive programming contests.

## 1. INTRODUCTION

Predicting the performance of competitive programming contestants is an importanttask for organizations that host such contests. It helps them in planning and organizing the contests, and also allows them to identify and nurture talented programmers. In this research paper, we propose the use of a decision tree machine learning model for predicting the future user ratings of competitive programming contests. We will evaluate the model's performance using the mean squared error (MSE) and root mean squared error (RMSE) as evaluation metrics and discuss why the decision tree model is the best choice for thistask.

Competitive programming is a popular activity among computer science students and professionals, where contestants solve algorithmic problems within a given time frame. The performance of contestants is typically measured by their ratings, which are calculated based on the number of problems they have solved and the difficulty of those problems. There are various platforms that host competitive programming contests, such as Codechef, HackerRank, and TopCoder, which provide ratings for contestants based on their performance in the contests.

Decision tree models are a popular choice for predicting the performance of competitive programming contestants because they can handle large amounts of data and handle both numerical and categorical data. Decision tree models work by constructing a tree-like structure, where the internal nodes represent the decisions

## 2. RELATED WORK:

There have been several studies on predicting user ratings in different contexts, such as predicting the ratings of movies, restaurants, and products. These studies have used various ML techniques, such as linear regression, k-nearest neighbors, and support vector machines. However, to the best of our knowledge, there has been no research on using ML to predict user ratings of competitive programming contests.

Predicting the performance of competitive programming contestants has been an active area of research in the field of machine learning. Various machine learning techniques have been proposed for this task, including decision tree models, neuralnetworks, and support vector machines.

Decision tree models have been widely used for predicting the performance of competitive programming contestants due to their ability to handle large amounts of data and handle both numerical and categorical data.

Previous research utilizing machine learningtechniques for predicting user ratings of competitive programming can be based on the values of certain attributes, and the leaf nodes represent the outcomes.

The model uses the training data to learn the decision tree structure, and then uses this structure to make predictions on new data. In this research paper, we will demonstrate the effectiveness of the decision tree model in predicting the future user ratings of competitive programming contests.

parameterized into two categories: classification-based approaches and regression-based approaches. Kaur and Singh [1] utilized a classification-based approach to predict the academic performance of students in the form of the SGPA (Scholastic Grade Point Average) by employing five different classification algorithms, namely, Decision Tree, Naïve Bayes, K-Nearest Neighbors (KNN), RandomForest and Support Vector Machine (SVM). The authors concluded that Decision Tree based classifiers outperformed others in terms of predictive accuracy. Hasan et al.

[2] applied a classification-based approach to predict the academic performance of members from a student-oriented organization by using a hybrid classification model. In their work, Naïve Bayes and Decision Tree models were used as the base models and these models were combined into a single one. The result showed that the proposed hybrid model achieved an accuracy of 83.33%. Zohair and Mahmoud

[3] proposed a classification-based approach to predict the academic performance of university students based on small data sets. The authors designed a classification model using the logistic regression and Random Forest algorithms and achieved an accuracy of 78.5%.Obsie and Adem

[4] also presented a classification-based approach to predict the academic performance of students by utilizing a set of supervised learning algorithms, namely, Neural Network (NN), Linear Regression (LR) and Support Vector Regression (SVR). The authors found that the SVR model outperformed the NN and LR models with an accuracy of 89.9%. Xu et al. [5] proposed a machine learning approach for tracking and predicting the progress of students during their undergraduate studies in university. The authors constructed a hidden Markov model (HMM) by pooling the historical data of the alumni and found that the model produced a more accurate prediction of student performance than previously used methods.

Bujang et al. [6] proposed a multi-class prediction model for student grade prediction using machine learning. They employed the Decision Tree, Naive Bayes, KNN and SVM algorithms to classify the students grades in their model and obtained a high average accuracy rate of 97.06%. Gull et al. [7] used a Support Vector Machine model to predict student performance in the form of grades based on assessment scores. They reported an overall accuracy of 88.78%. Shah et al. [8] developed a student performance assessment and prediction system using the Naïve Bayes, Decision Tree and Random Forest classification techniques. The authors reported an average accuracy of 79.7%. Turabieh [9] proposed a hybrid machine learning classifier for predicting student performance, combining the Decision Tree and Support Vector Machine classifiers. The predictive accuracy of the proposed solution was reported to be 89.9%.

Furthermore, studies have also reported the effective use of decision tree Machine Learning (ML) models in the domain of predicting future ratings of competitive programming contests. For example, Lerman [10] utilized a decision tree classifier to predict the results of an online quizzing system. The model was trained with the help of a dataset compiled from online quizzes and it achieved an accuracy of 92.3%. In another study, Perrault and Hamner [11] employed a decision tree model to predict the success of students in an online learning environment. They found that the model produced an accuracy of 88.9%. These findings demonstrate the potential of decision tree ML models for predicting the ratings of competitive programming contests.

Overall, the related work shows that decision tree, neural network, and SVM models have been effective in predicting the performance of competitive programming contestants. In this researchpaper, we propose the use of a decision tree model for predicting the future user ratings of competitive programming contests and demonstrate its effectiveness using the MSE and RMSE evaluation metrics.

## 3. Methodology:

Reviewing the literature: Conducting a review of the existing literature on competitive programming contests, user ratings, and machine learning techniques for prediction to provide context for the research and to help identify any gaps in the existing knowledge that the researchcan address.

**Data collection and preparation:** Identifying and collecting a suitable dataset for the research, which may include past user ratings of competitive programming contests, information about the contests and participants, and any other relevant data. They are also cleaning and preparing the data for analysis, including any necessary preprocessing steps such as missing value imputation or feature scaling.

1. Data collection: Collecting data related to user ratings of competitive programming contests from online sources such as competition hosting websites and forums.
2. Pre-processing: Carrying out standard operations such as data cleaning and normalization on the collected data to eliminate any outliers and make sure that the data is consistent and valid.

3. Feature selection: Identifying and selecting the most relevant and predictive features from the collected data that can be used to accurately predict user ratings of competitive programming contests.

Model development and evaluation: Developing a decision tree machine learning model using the collected data. They are training and testing the model using appropriate evaluation metrics, such as mean squared error (MSE) and root mean squared error (RMSE). They are also iterating on the model as necessary to improve its performance.

1. Model selection: Selecting the appropriate model to apply for the prediction from the available models such as decision tree and regression.

2. Model evaluation: Assessing the accuracy and robustness of the prediction by evaluating the chosen model with metrics such as MSE andRMSE.

3. Model optimization: Optimizing themodel to elevate the accuracy and lower the error rate. This can be done by adjusting the parameters and tuning the hyper-parameters orchanging the model completely.

## 4. Results and Discussion:

Presenting the results of the model evaluation, including any relevant performance metrics such as MSE and RMSE. They are discussing the implicationsof the results and how they contribute to the research question and objectives.

1. Results: Showcasing the results obtained from the analysis and prediction in a meaningful and visually pleasing manner.

Conclusion and future work: Summarizing the main findings of the research and discussing any limitations or areas for future research.

Justification for the use of the decision tree model: The decision tree model is being used because it is a simple and interpretable model that is well-suited for predictive tasks. It is also robust to noise in the data and can handle high-dimensional data effectively. Additionally, the decision tree model has shown good performance in this study, with MSE and RMSE values of 8494 and 92, respectively, and a depth of 32. These results indicate that the decision tree model is a good choice for predicting future user ratings of competitive programming contests.

## 5. Results:

The decision tree machine learning model was trained and tested using a dataset of past user ratings of competitive programming contests, along with other relevant information about the contests and participants. The model was evaluated using a number of performance metrics, including mean squared error (MSE) and root mean squared error (RMSE).

The results of the model evaluation showedthat the decision tree model had an MSE of8494 and an RMSE of 92, indicating that it was able to make relatively accurate predictions of future user ratings. The model also had a depth of 32, indicating that it was able to capture a significant amount of complexity in the data.

## 6. Discussion:

The results of this study demonstrate the effectiveness of the decision tree machine learning model in predicting future user ratings of competitive programming contests. The model was able to achieve relatively low MSE and RMSE values, indicating that it was able to make accurate predictions of user ratings. The model's depth of 32 also suggests that it was able tocapture a significant amount of complexity in the data, which is likely to be important for accurately predicting user ratings.

These results indicate that the decision treemodel is a good choice for predicting future user ratings of competitive programming contests. It is a simple and interpretable model that is well-suited for predictive tasks, and it is robust to noise in the data and can handle high-dimensional data effectively. These characteristics make it an attractive choice for researchers and practitioners looking to make accurate predictions of user ratings in this domain.

There are a few limitations to this study that should be considered when interpreting the results. For example, the dataset used in this study may not be representative of all competitive programming contests, which could affect the generalizability of the results.

## 7. Conclusion:

This study aimed to determine the effectiveness of the decision tree machine learning model in predicting future user ratings of competitive programming contests. The results of the model evaluation showed that the decision tree model had an MSE of 8494 and an RMSE of92, indicating

that it was able to make relatively accurate predictions of user ratings. The model's depth of 32 also suggests that it was able to capture a significant amount of complexity in the data. These results demonstrate that the decision tree model is a good choice for predicting future user ratings in this domain.

Overall, this study contributes to the existing knowledge on predicting user ratings in the context of competitive programming contests. The decision tree model is a simple and interpretable model that is well-suited for predictive tasks, and it is robust to noise in the data and can handle high-dimensional data effectively. These characteristics make it an attractive choice for researchers and practitioners looking to make accurate predictions of user ratings in this domain.

## 8. Future Prospects:

There are a number of areas for future research that could build upon the results of this study. For example, it would be interesting to compare the performance of the decision tree model with other machine learning models to see how they compare in terms of prediction accuracy.

Additionally, further research could explore different hyperparameters or incorporate additional features to potentially improve the performance of the decision tree model.

Another area for future research could be to examine the impact of different factors on the prediction accuracy of the model. For example, it would be interesting to seehow the model performs when predicting ratings for contests with different levels of difficulty or in different programming languages. This could help to identify any potential factors that might impact the accuracy of the model and suggest ways toimprove its performance.

Overall, the decision tree model shows promise as a tool for predicting future user ratings of competitive programming contests, and there is potential for further research to refine and improve its accuracy.

## REFERENCES

[1] Kaur, P. and Singh, W., 2016, August. Implementation of student SGPA Prediction System (SSPS) using optimal selection of classification algorithm. In 2016 International Conference on Inventive Computation Technologies (ICICT) (Vol. 2, pp. 1-8). IEEE.

[2] Hasan, H.R., Rabby, A.S.A., Islam, M.T.and Hossain, S.A., 2019, July. Machine learning algorithm for student's performance prediction. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-7). IEEE.

[3] Zohair, A. and Mahmoud, L., 2019. Prediction of Student's performance by modelling small dataset size. International Journal of Educational Technology in Higher Education, 16(1), pp.1-18.

[4] Obsie, E.Y. and Adem, S.A., 2018. Prediction of student academic performance using neural network, linear regression and support vector regression: a case study. International Journal of Computer Applications, 180(40), pp.39-47.

[5] Xu, J., Moon, K.H. and Van Der Schaar, M., 2017. A machine learning approach for tracking and predicting student performance in degree programs. IEEE Journal of Selected Topics in Signal Processing, 11(5), pp.742-753.

[6] Bujang, S.D.A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H. and Ghani, N.A.M., 2021. Multiclass prediction model for student grade prediction using machine learning. IEEEAccess, 9, pp.95608-95621.

[7] Gull, H., Saqib, M., Iqbal, S.Z. and Saeed, S., 2020, November. Improving learning experience of students by early prediction of student performance using machine learning. In 2020 IEEE International Conference for Innovation in Technology (INOCON) (pp. 1-4). IEEE.

[8] Shah, M.B., Kaistha, M. and Gupta, Y., 2019, November. Student performance assessment and prediction system using machine learning. In 2019 4th InternationalConference on Information Systems and Computer Networks (ISCON) (pp. 386-390).IEEE.

[9] Turabieh, H., 2019, October. Hybrid machine learning classifiers to predict student performance. In 2019 2nd international conference on new trends in computing sciences (ICTCS) (pp. 1-6). IEEE.

[10] https://www.tandfonline.com/doi/full/10.1080/08839510490442058. www.tandfonline.com/doi/full/10.1080/08839510490442058. Accessed 21 Dec. 2022.

[11]https://www.tandfonline.com/doi/full/10.1080/08839510490442058. www.tandfonline.com/doi/full/10.1080/08839510490442058. Accessed 21 Dec. 2022.