IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol. 9, Issue.8, August 2021 E-ISSN: 2347-2693

Classification of Pulsar Candidates Using an Ensemble Model

Sanat Kumar Sahu

Department of Computer Science, Govt. Kaktiya P.G. College Jagdalpur, C.G., India

**Corresponding Author:sanat.kosa1@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v9i8.8183 | Available online at: www.ijcseonline.org

Received: 16/Aug/2021, Accepted: 20/Aug/2021, Published: 31/Aug/2021

Abstract— In the past, researchers study candidate filters used to solve the problem for the last years. Pulsar is a type of star, which is interested in the great scientific topic. Through which we discover this celestial pulsar. Here we have used the decision tree under the new machine learning in this research. We use two classification techniques C4.5 Tree and classification and regression tree CART to classify the HTRU2 dataset and we set a model C4.5 Tree and CART from the ensemble of the classification and regression tree. The Model Ensemble C4.5 Tree and CART provides the best performance compared to the individual models of each classifier. Ensemble Model is useful for classifying candidates in pulsar and non-pulsar.

Keywords— Classification, C4.5, CART, Ensemble Model, HTRU2

I. Introduction

Highly magnetic high-speed neutron stars are known by the name of pulsar. Whose linear transmitted polarized electromagnetic radiation spreads along their magnetic poles? While Pulsar keeps variance. It's radiating travels from time to time to the observer's vision, such as the rotating beacon, which is the result of a periodic train of narrow broadband radiation pulses, which was detected, using a radio telescope, can go [1], [2].Machine learning has become one of the cornerstones of information technology and, through this, a rather central, though normally hidden, part of our life. With the increasing amount of data available, there is good reason to believe that intelligent data analysis will be even more widespread as a basic factor for technological progress [3].Data mining and Machine learning is useful for searching or finding new pulsars. We have to use classification techniques to determine pulsar and non-pulsar candidates in our research. For which we have used two classification techniques C4.5 Tree and CART and both of them have been made a new technique. Which is called ensemble model, we get increased accuracy from the combination of both of them ensemble model. These machine-learning techniques are very important for pulsar candidate's selection.

Many researchers have worked in search of pulsar made in the past. The brief description of his contribution is as follows.

Each candidate must be inspected both by an automatic method like machine learning techniques and by a human expert to determine its authenticity [4]. The process for deciding which candidates are worth investigating is known as selection of candidates is known as "candidates" to the press, a possible detection of a new pulsar [5] they also presented in new model it selecting promising candidate using a purpose built in tree –based machine learning classifiers. With the help new approaches they have discovered 20 new pulsars. The authors [6] have explained the discovered of a new pulsar survey by using the Parkes Radio Telescope. The high time and frequency resolution of our digital backend system leads to increased sensitivity to short period, high-DM pulsars compared to previous surveys.

The rest of the study structured as follows. Section II defines the methodology as have used C4.5 Tree, CART and their ensemble model, also describes dataset and process flow of work. Section III explores the experimental results and discussion. Finally, Section III concludes the findings of the research work and future.

II. METHODS AND MATERIAL

CLASSIFICATION: classification is a specific form of data analysis that divides important data into class [7]. With which we divide the different dataset into the class. The model used for this task. That model called a classifier, with the help of which the class estimates.

C4.5 Tree: C 4.5 is an algorithm that was produced by a decision tree developed by Ross Queenville. The C 4.5 Queenville is considered to be an extended form of the ID3 algorithm. C 4.5 has to be used for classification. C4.5 is known as statistical classifier [8].

CART: CART adopt a greedy (i.e., non back tracking) approach in which decision trees are constructed in a top-down recursive divide-and-conquer manner. Most algorithms for decision tree induction also follow a top-down approach, which starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is being built [9]. CART is classification and regression tree uses recursive partitioning to split the training records into subdivision with similar target field ideals using Gini index.

Ensemble Models: When two classification techniques like C4.5 Tree and CART combined it is called hybrid or ensemble model [10].Stacking is similar to the boosting: it also applies different models to its original data. The difference here is, however, that you do not have a single experimental formula for your weight function, but you enter a meta level and use one more model to calculate approximately the input together with the results of each model to estimate weights or, in other words, to decide which models work well and what's wrong with this input data.

HTRU2 DATASET: Each candidate is described by 8 continuous variables, and a single class variable. The first four are simple statistics obtained from the integrated pulse profile (folded profile). This is an array of continuous variables that describe a longitude-resolved version of the signal that has been averaged in both time and frequency. The remaining four variables are similarly obtained from the DM-SNR curve [11].

The HTRU (High Time Resolution Universe Survey) 2 dataset have total number of instance is 17898 with 1639 are positive instances and 16259 are negative instances. The total number of attributes (features) is 8 with class label.

Table 1: Descriptions HTRU 2 Data Set

Sl. No.	Attributes	Details
1	Profile_mean	Mean of the integrated profile
2	Profile_stdev	Standard deviation of the integrated profile
3	Profile_skewness	Skewness of the integrated profile
4	Profile_kurtosis	Excess kurtosis of the integrated profile
5	DM_mean	Mean of the DM-SNR curve
6	DM_stdev	Standard deviation of the DM-SNR curve
7	DM_skewness	Skewness of the DM-SNR curve
8	DM_kurtosis	Excess kurtosis of the DM-SNR curve
9	Class	Negative and Positive

III. CLASSIFICATION OUTLINE PROCEDURE FLOW

Relevant details should be given including experimental design and the technique (s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory). Figure 1 shows how classifiers used by us like tree like C4.5 tree and CART dataset HTRU2 classify. To classify the HTRU 2 dataset, we are using 10 cross validation techniques. We are using dataset for classification of individual model C4.5 tree and CART and their ensemble model. Eventually, got the performance of the classifier, which are the accuracy, sensitivity and specificity.

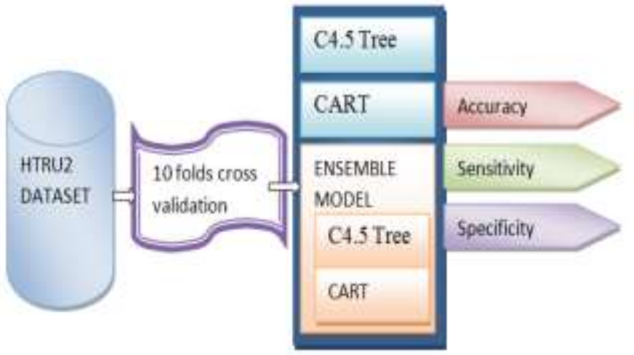

Figure 1: Classification Outline Procedure Flow

Dataset is classified with 10 folds cross validation after loading the HTRU 2 (R. Lyon, 2016) dataset. When Classifiers do, dataset classifies. Therefore, we get three performance measurements like accuracy, sensitivity and specificity. Which we thought, because in these three measures the efficiency of the classification model clearly shown.

IV. RESULTS AND DISCUSSION

In the table 2 below we have shown that the Algorithm and their parameters of classifiers that reflects its efficiency. The model of classification C4.5 and CART and their Ensemble Model result shows which is a comparative study among them.

Table 2: Comparison of classification performances

Sl. no.	Name of Algorithm	Accuracy	Sensitivity	Specificity
1	C4.5 Tree	97.93	88.26	84.74
2	CART	92.96	87.48	84.25
3	Ensemble model C4.5, CART	97.96	88.48	85.35

In Table 2, we have shown remarkable results in the classification model C4.5 tree and CART and their ensemble model C4.5 tree, CART. It is clear that the ensemble model C4.5 tree, CART has achieved the highest accuracy compared to the Model C4.5 tree and CART individual model. Similarly, ensemble model C4.5 tree, CART has received the highest specificity compared to the model C4.5 tree and CART individual model. Ensemble models of classification, the C4.5 tree, CART has received the highest sensitivity compared to the C4.5 tree and CART individual model.

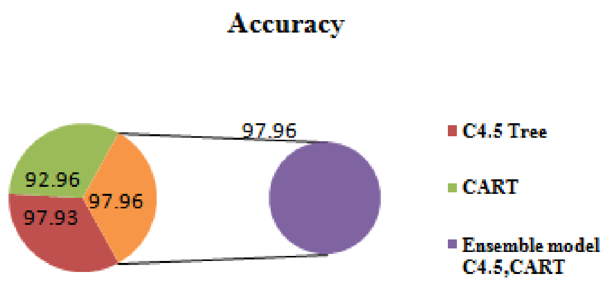

Figure 2: Comparative Accuracy graphs of the different Classifier

The figure 2 shows maximum accuracy achieved by ensemble model (C4.5 Tree and CART) compared to the individual models C4.5 Tree and CART.

V. Conclusion and Future Scope

To properly detect Pulsar's candidate, we use many types of data mining and machine-learning classifiers. The purpose of this study was to analyze the application of data mining algorithms and machine learning in HTRU2 dataset and had to predict Pulsar and non-Pulsar. In this paper, Pulsar candidate has been predicted using three types of classification techniques. In which we have used C4.5 and CART and their ensemble model. The result is that the classifier C4.5 accuracy 97.93% and CART accuracy 92.96% and the 97.96% accuracy in the ensemble model. We conclude that the accuracy of the ensemble model is 0.03% higher than C4.5. Similarly, the accuracy of the ensemble model is 5.00% higher than the CART. Therefore, we can say that the ensemble model is better than the C4.5 and CART in prediction of pulsar and non-pulsar.

Use feature selection techniques in the future. With the help of, which we can reduce its feature and identify important features. We can use other classification techniques, which will get us enhanced accuracy. Similarly, we can use both feature selection techniques and classification techniques.

References

[1] S. K. Saha, S. Sarkar, and P. Mitra, “Feature selection techniques for maximum entropy based biomedical named entity recognition,” J. Biomed. Inform., **vol. 42, no. 5, pp. 905–911, 2009.**

[2] P. S. Ramkumar and a. a. Deshpande, “Real-time signal processor for pulsar studies,” J. Astrophys. Astron., **vol. 22, no. 4, pp. 321–342, 2001.**

[3] E. Alpaydın, “Introduction to machine learning,” Methods Mol. Biol., **vol. 1107, pp. 105–128, 2014.**

[4] R. J. Lyon, “WHY ARE PULSARS HARD TO FIND ?,” **2016.**

[5] R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, and J. D. Knowles, “Fifty Years of Pulsar Candidate Selection : From simple filters to a new principled real-time classification approach,” **vol. 22, no. March, pp. 1–22, 2016.**

[6] M. J. Keith et al., “The High Time Resolution Universe Pulsar Survey – I . System configuration and initial discoveries Introduction simulation and survey strategy,” **vol. 627, pp. 619–627, 2010.**

[7] Sivanandam and Deepa, Principles of Soft Computing, Second. wiley, **2014.**

[8] A. Pujari, Data mining techniques, Third. University press, **2013.**

[9] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques, Third. Elsevier, **2012.**

[10] S. Haykin, Neural Networks and Learning Machines, vol. 3. 2008.

[11] R. Lyon, “HTRU2,” 2016. [Online]. Available: https://figshare.com/articles/HTRU2/3080389/1. [Accessed: 01-Dec-2017].

AUTHORS PROFILE

Dr. Sanat Kumar Sahu is working as an Assistant Professor in the Department of Computer Science, Govt. Kaktiya PG College, Jagdalpur (Bastar) Chhattisgarh, India. He has more than 11 years teaching Experience. His area of interest includes soft computing, machine learning, and data mining.. He has more than 20 research paper in national and international journals.