# Carrier Dome: Prediction of Student Career Interest

**Prof. R. R. Shewale[1] Shidore Varsha B.[2] Pardeshi Gayatri N.[3] Pingale Ruchira A.[4] Wagh Kamini N.[5]**
[1]Professor [2,3,4,5]U. G. Student
[1,2,3,4,5]Department of Computer Engineering
[1,2,3,4,5]NDMVP Samaj's, KBTCOE, Nashik (India)

*Abstract*— Attempting to deepen the understanding of factors that explain student career interest, this tries to identify and characterize profiles of students based on personal details, academic performance, student and family demographic (background details), family educational details, personality traits, activity aspects. Existing systems like paper work process and web-based services for determining career interest. In existing systems, CHAID (CHI-square Adjusted Interaction Detection), ID3 (Iterative Dichotomiser 3) and C4.5 (Classification Algorithm) algorithms were used but are only for specific factor so, to overcome existing system drawbacks and work on more factors at a time 'Carrier Dome' Android-based Application is developed by using RF (Random Forest) and CART (Classification And Regression Tree) algorithms. These two algorithms are very powerful and more accurate. These techniques are more advantageous, as they run efficiently on large databases, generate accurate predictive models, suitable for high prediction accuracy of new data, supports high speed deployment, estimates which variables are important in classification, it is an effective method for estimating missing data and improve accuracy even if large proportion of data are missing, no need for prior feature selection and data pre-processing, works on large dataset. In this system, student should have to fill only mandatory data and it is cost effective. Carrier Dome aimed to develop a more accurate and powerful application than existing system for prediction of career interest.

***Key words:*** CART (Classification and Regression Tree) algorithm, Machine learning, RF (Random Forest) algorithm, Training data set etc

## I. INTRODUCTION

As we know that organizations cannot promise their employees lifetime job and the employer is forced to focus on the test interests of the company where which may, or may not, be the same as the best interests of each employee. It is a real fact that individuals of our current working population will undertake five to seven occupational chances throughout their lifetime in many fields. This means that we have to think about that where we are, what we have done, and where we want to go. So, that having a career plans with us, provides the steps and possible outcomes required to reach those goals. This in turn teaches us a vision and motivation to achieve that goal which is critically important. For choosing a career in one best suitable field the main point is the consideration of interest, according to interest one can have successful career without any trouble. Prediction of career interest can help students to know their career in which field likes Engineering, Medicine and Health Professions, B.Com, B.Pharm., Agriculture, Law, Accountant, Actor, Politic field and so on.

Our idea is to develop an application to predict student career interest using Data Mining and Knowledge Processing. We are developing the application because earlier methodologies are not so accurate and techniques used are not efficient. Also, the way of giving result of earlier techniques is not well and trusty. So, we are overcoming all these disadvantages in our application. Now-a-days, there are many education facilities and options are available for students so, students are confused to select a particular career option. Therefore, we developed an application which helps student to select career. In Carrier Dome application there are forms like Registration, Data entry etc.

## II. RELATED WORK

There are many existing systems are available for prediction of student career interest and also to predict student performance based on, Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses [3], Academic success\ performance [10] [14], business domain like Enterprises, Socio-demographic factors and subject like Mathematics [1], Effects of home environment and center-based child care quality on children's language, communication, and literacy outcomes [7], Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program [8], Parents' math–gender stereotypes, children's self-perception of ability [11], Single-parent households and children's educational achievement [12], The interaction between social goals and self-construal on achievement motivation [13], etc. Also, existing approaches includes paper based and web based services. Now-a-days we are using those approaches but are not efficient to use. Sometimes, result will be wrong as paper based process is manual, it is very lengthy process and is not cost efficient. And also in web based approach ID3, C4.5 and CHAID algorithms are used. Those algorithms are having less accuracy, works on very small dataset etc. [5]

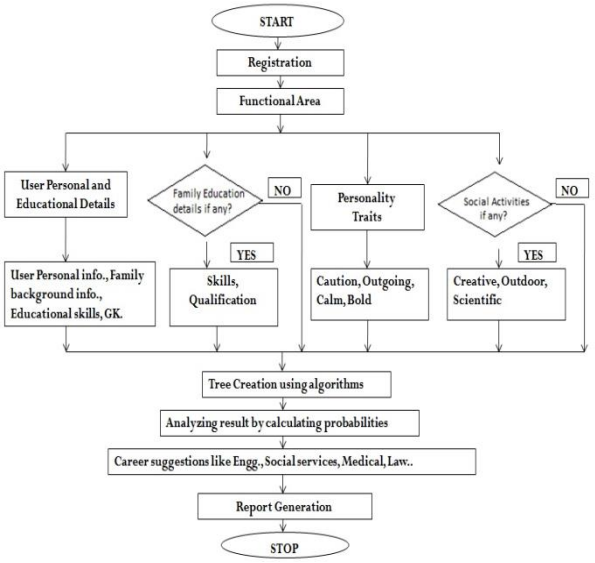


Fig. 1: Flowchart of Carrier Dome

 

To overcome existing system drawbacks 'Carrier Dome' Android-based Application is developed by using RF (Random Forest) and CART (Classification And Regression Tree) algorithms. And also in previous studies career is predicted using only the specific one or two factors and so that we are considering all the factors which are essential and/or important for predicting the career. Most probably, student's demographic and educational details play an important role in choosing the career. And also here no internet connection is required while performing test and generating result but only to download an application and to sync the report to store in database internet is required.

## III. WORKING

First user has to download the app from the play store. Our application is an Android application and so that it is freely available to use in play store.

To use this app user has to register by entering Contact Number, Name and Class etc. In Class field there are different eight educations are given like for SSC, Diploma, HSC_Science_PCM, HSC_Science_PCMB, HSC_Arts, HSC_Science_PCB, HSC_Commerce, Graduation etc and user has to select appropriate qualified/applied education. After registration login window Carrier Dome actual application window will display. User Interfaces like Import Questions, Import Careers, Start Test, See Report, Close etc. are there. Then by click on Start Test data entry form (questions) will open in that form various types of questions will included. The questions will be based on the factors like User Personal and Educational Details, Family Educational Details, Personality Traits and Social Activities. Considering these factors tree is generated with the help of algorithms like CART and RF based on training dataset.

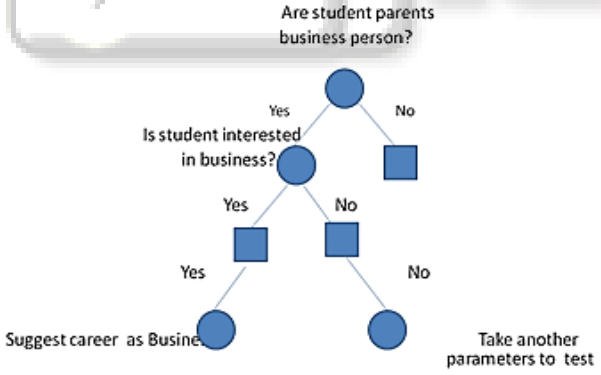


Fig. 2: CART algorithm example

Then it analyze the result by calculating the probabilities and suggest the career like Engineering, I.T.I., Diploma, Architecture, B.Com, B.Sc., Agriculture, Social Services, Medical, Law and many more, to the user according to his/her entered information while testing. Fig. (a) Shows flowchart of system.

## IV. TOOLS AND TECHNIQUES

### *A. Requirements*

*1) Software Resources Required*

- Operating System: Android OS(specifically Android 4.0 and above)
- Server: Windows 2x
- Database: SQLite
- Programming Language: Mono, C#, .net (Framework 4.0)

*2) Hardware Resources Required*

| SR NO. | PARAMETER | MINIMUM REQUIREMENT |
|---|---|---|
| 1. | CPU Speed | 1.8 GHz |
| 2. | Processor Type | |
| 3. | RAM | 2 GB |
| 4. | Hard Disk | 500 GB |

Table 1. Hardware Resources

### *B. Techniques*

[I] CART Algorithm: CART allows analyzing associations between individual and family features with the user's career interest. Using CART unstructured tree is generated according to user information. [2] [6]. It classifies objects or predicts outcomes by selecting from a large number of variables the most important ones in determining the outcome variable. CART analysis is a form of binary recursive partitioning.

There are three steps in CART algorithm as follows:

Tree Building : There Are Some Steps Of Tree Building.

1) Start the splitting with a variable at all split points. Here, sample splits into two binary nodes at each split point.
2) Select the best split in the variable in terms of the reduction in impurity (heterogeneity)
3) Repeat steps 1,2 for all variables at the root node.
4) Then rank all possible best splits and select the variable which achieves the highest purity placed at root.
5) Assign classes to the nodes according to a rule that minimizes misclassification costs.
6) Repeat 1-5 for each non-terminal node.
7) Grow a very large tree Tmax until all terminal nodes are either small or pure or contain identical measurement vectors.
8) Prune and choose final tree using the cross validation.

Pruning : In pruning, instead of finding out an appropriate stopping rules, grow a Tmax and again prune it to the root of the tree. Before pruning, for growing a sufficiently large initial tree Tmax specifies Nmin and split until each terminal node either is pure or N(t)c Nmin. Generally Nmin has been set at 5.

Optimal Tree Selection : For optimal tree selection there is use of cross-validation (CV) after, it built into the CART algorithm. Basic idea behind this is "grow the tree" out as far as you can then "prune back". Here CV tells you when to stop pruning.

Fig. (b) Shows CART algorithm example

[II] RF Algorithm: RF helps to select the best subset from available variables to build the tree model. A RF is a group of un-pruned classification or regression trees made from different with-replacement bootstrap sample from the data. RF also generate unstructured tree and by combining CART and RF algorithms final career suggestion is calculated. [2] [4] [9]

Fig. (c) Shows RF algorithm process. The RF algorithm is applied on the number of trees generated from CART algorithm. It is having following four processes:

1) Tree growing

 

2) Tree combination
3) Self-testing
4) Post processing

Tree Growing: Tree growing is the process of growing tree randomly using a binary partitioning. Here, is the use of splitter. Splitter is the square root of total number of predicators available. Tree is split across splitter node.

Tree Combination: In tree combination all the sub-trees obtained from previous stage are combined to give next results.

Self Testing: In this stage tree is tested for cross validation. For cross validation there is about 63% original training data and about 37% data available for testing a single tree.

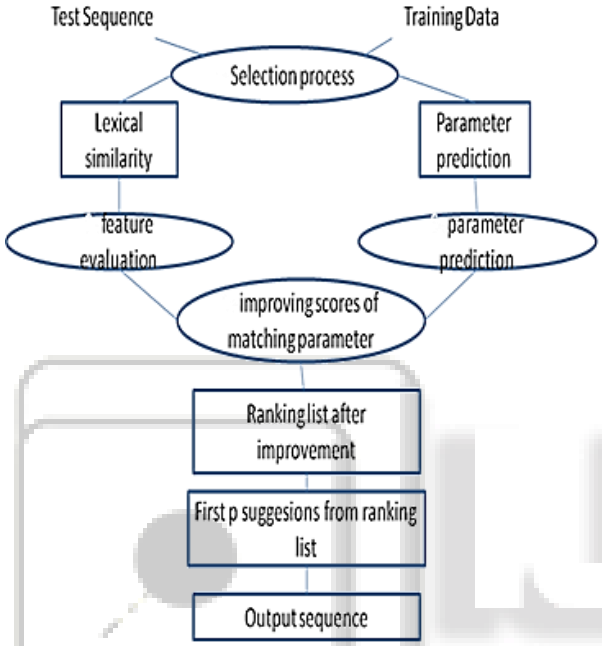


Fig. 3: RF Algorithm

Post Processing: It improves score of matching parameter.

Test sequence and Training data are inputs to the selection process. Selection process has two sub process lexical similarity and parameter prediction. From Lexical similarity features are evaluated then predicted parameters and evaluated features are matched and scores of matching parameter are improved. After that, improved scores ranking is given to that scores and first P suggestions are considered for further feature evaluation process. After all these processes the actual result as a suggested career is obtained for that particular user.

RF algorithm is totally depends on random selection of parameters. In this application RF use to select questions randomly. When the user runs the application then there is option to choose class for login where user has to select particular class. According to that particular class questions are displayed on screen of user's mobile not all class's questions are displayed and user has to give answers of that much questions only. It will save time and avoid complexity by choosing random questions.

## V. System Description

- Input: Functional Area- Personal Details, Family Educational Details, Personality Traits, Social Activities etc.
- Output: Suggestion of Career interest to user.
- Identify data structures, classes, divide and conquer strategies to exploit distributed/parallel/concurrent processing, constraints.
- Success Conditions: User should have to enter valid data, valid information about his/her family, personal details etc. to get more accurate result.

## VI. Tasks for Carrier Dome

- Creation of training dataset: Large dataset so that result is more accurate and efficient.
- Creation of Question set: Questions functional area as User's Personal details, Family Education and Background details, Personality traits, Social activity aspects etc.
- Framework: Framework design using Mono software compatible with Android version 4.0 and later.
- Application Programming Interface: Designed API for our application is an emulator and as well as it runs on Android handset.
- Algorithms Coding: Apply CART and RF algorithms on functional area.
- Analysis with respect to training dataset.
- Feature classification: Classify Career based on coding of algorithms.
- Testing of all implemented modules and corrections if any.
- Result as career interest to user.

## VII. Results

To predict the student career interest we had conducted test of 200 students and according to that we got the minimum accuracy as 90-95%. A total of 75% accuracy got when the students did not give accurate information about their skills, interest etc.

The RF algorithm procedure was done by taking around P number of sets where P the total number of parameters which are fixed with another different 100 parameters and also, the given number of variables per level taken as 15. Out of all variables five were considered to predict the career interest: "Personal education details", "Parents expectation", "Demographic details", "Social aspects" and "Personality traits".

The CART algorithm generates a tree as a final decision tree. The CART procedure was applied on 15 numbers of groups to give final resultant decision tree as terminal nodes of tree which was the career suggested to user. Result was obtained by prediction of all tress.

## VIII. Conclusion

As Carrier Dome application implement fast, accurate and powerful application for prediction of student career interest which efficiently works on large dataset based on factors like student personal details, educational details, family background and educational details, personality traits, social activities etc. We used two algorithms, CART and RF which

works as reverse engineering so, calculated career interest result is more accurate. In our application internet required is only to sync reports to store on database and to export the reports if user want to email it and free application memory space. This application can be helpful for students, faculty members, institute members and anyone who has educational background.

## ACKNOWLEDGEMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help of our project guide and organizations.

We would like to thanks to all of them. We are really thanks to Prof. R. R. Shewale for their guidance and constant supervision as well as for providing necessary information and their support in completing the project. We would like to express our gratitude towards our institute N.D.M.V.P.'s K.B.T.C.O.E

## REFERENCES

[1] Monica Bravo Sanzana , Sonia Salvo Garrido & Carlos Munoz Poblete (2015). "Profiles of Chilean students according to academic performance in mathematics: An exploratory study using classification trees and random forests". Studies in Educational Evaluation 44 (2015) 50–59.

[2] Irimia-Dieguez, A.I.a, Blanco-Oliver, A.a, Vazquez-Cueto & M.J.(2015). "A Comparison of Classification/Regression Trees and Logistic Regression in Failure Models". 2nd Global Conference on Business, Economics, Management and Tourism, 30-31 October 2014, Prague, Czech Republic. Procedia Economics and Finance 23 ( 2015 ) 9 – 14.

[3] Yu-Chun Kuo, Andrew E. Walker b, Kerstin E.E. Schroder c, Brian R. Belland, "Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses", Internet and Higher Education 20 (2014) 35–50.

[4] Kennedy Were, Dieu Tien Bui, Øystein B. Dick & Bal Ram Singh (2015). "A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape". Ecological Indicators 52 (2015) 394–403.

[5] M. Ramaswami, and R. Bhaskaran, "A CHAID Based Performance Prediction Model in Educational Data Mining", International Journal of Computer Science, Vol. 7, Issue 1, No. 1.of 2010.

[6] Jing Li, Xinpu Ji, Yuhan Jia, Bingpeng Zhu, Gang Wang, Zhongwei Li & Xiaoguang Liu, "Hard Drive Failure Prediction Using Classification and Regression Trees" , 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks.

[7] Ana Isabel Pinto, Manuela Pessanha, & Cecília Aguiar (2013)."Effects of home environment and center-based child care quality on children's language, communication, and literacy outcomes". Early Childhood Research Quarterly 28 (2013) 94– 101.

[8] Pimpa Cheewaprakobkit, "Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program", Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I, IMECS 2013, March 13 - 15, 2013, Hong Kong.

[9] A. Hapfelmeier & K. Ulm (2014), "Variable selection by Random Forests using data with missing values", Computational Statistics and Data Analysis 80 (2014) 129–139.

[10] Zlatko J. Kovacic, Open Polytechnic, Wellington, New Zealand (2010). "Early Prediction of Student Success: Mining Students Enrolment Data", Proceedings of Informing Science & IT Education Conference (InSITE) 2010.

[11] Carlo Tomasetto, Alberto Mirisola, Silvia Galdi, Mara Cadinu (2015)," Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds", Contemporary Educational Psychology 42 (2015) 186–198.

[12] Paul R. Amato, Sarah Patterson, Brett Beattie (2015), "Single-parent households and children's educational achievement: A state-level analysis", Social Science Research 53 (2015) 191–202.

[13] Rebecca Wing-yi Cheng, Shui-fong Lam (2013), "The interaction between social goals and self-construal on achievement motivation", Contemporary Educational Psychology 38 (2013) 136–148.

[14] Nguyen Thai-Nghe, Andre Busche, and Lars Schmidt-Thieme (2009), "Improving Academic Performance Prediction by Dealing with Class Imbalance", 2009 Ninth International Conference on Intelligent Systems Design and Applications.

 