

Android Application for finding Tutors using Data Mining techniques

Dolly Panchal, Mili Sanghavi, ShikhaDevi Pandey, Elizabeth George

Abstract— This paper explains about our application i.e., Android application for finding tutors using a data mining technique. The increasing trend of private or group tuitions has made it more difficult for a parent to find a perfect tutor for his/her child. The process of finding a tutor is time consuming. The rise in educational standards has led to the need for additional coaching along with basic schooling. An Android application for finding tutors can help a parent find tutors without consulting any third party. The application encourages parents to find tutors and tutors to find prospective students in their nearby vicinities. The tutor application also considers the fact that students need personal attention and guidance in studies. Naïve Bayes algorithm is implemented in order to compare tutor profiles. Along with the Android platform used for Graphical User Interface (GUI), SQLite is used to store tutor data. Naive Bayes algorithm classifies the tutor records and thus helps in achieving maximum accuracy. WEKA is the platform used for implementing data mining techniques. The overall aim of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. When tutor registers or signs up his/her data is stored in SQLite. Parent can find tutors based on three parameters, namely area, board and preference. Based on stipulated requirements, search results are retrieved from SQLite and shall be displayed to the user.

Index Terms Android, Data Mining, Classification, Weka, Naïve Bayes, Sql Lite, Tutoring.

1 INTRODUCTION

With the rising competition emerging between students parent find the task of finding a qualified and experienced tutor tedious The existing system based on finding tutors for parents or finding prospective students for tutors is a very time consuming and tiring process for the parents, students and tutors. Even after spending hours on the task, the websites currently available do not give relevant results. Our project overcomes these flaws by reducing the time consumed for searching and displaying the results within shortest time. Here is where "Tutor Hunt" application comes into picture.

Android application for finding tutors is to be used in the education center. "Tutor Hunt" provides information regarding tuitions or coaching for students. It involves management of records, schedules, updates and search. This application is easy to use and understand. The main objective behind making this project is to enhance student and teacher relationship. Android application for finding tutors has been designed to make it convenient for students and teachers find each other. Students or parents can find tutors for online private tutoring. One can find tutors residing in the same vicinity as that of students. Students can find tutors for home tutoring or group tutoring. The intention behind creating the application is completely non-profit. Tutors can list there profiles so that students can contact them. The application not only updates teacher information in terms of preferences of location, but also has options to choose specializations in their respective domain of education. The application makes an effort to ensure registration of skilled professionals and also encourages and attracts talented, highly enthusiastic and extremely passionate tutors to join the platform in order to contribute to the needs of students. This means the students are not only able to get best learning/training experience but also attain the support required to achieve their potential/goals. "Tutor Hunt"

Dolly panchal is pursuing bachelor program in information technology engineering in Mumbai University, India, ph: 9870581311. Email: dollypanchal181@gmail.com

Mili Sanghavi is pursuing bachelor program in information technology engineering in Mumbai University, India.PH: 9702321299. Email: fransis2015mili@gmail.com

Shikha Pandey is pursuing bachelor program in information technology engineering in Mumbai University, India, ph:8655608916. Email: shikha.pamdey123@gmail.com

allows individual teachers and institutes to create their profiles, add skills or experience etc. The application provides a demography with which parents can choose the best tutor by comparing the attributes and skillsets of the tutor.

Using this application a parent can easily find tutors for their child instead of asking any relative or friends. Parents would get their result instantly instead of waiting for days or waiting for someone else to respond. For finding a good tutor, there are many parameters like gender of the tutor, education qualification, expertise in a particular subject/area, command over English language; based on which the parent will finalize the tutor. Also this application will help the tutors to coordinate and save their time by managing their schedule. The application considers location preference as the prime attribute in order to assign teachers and students in the same vicinity. The significant features of the application are maintainability, where data regarding tutor and parent will be maintained by database, profiles uploaded can be modified based on requirement, responsive, reliable and user-friendly.

The primary tasks involved in this project are: collection of data, database creation, information retrieval and implementation of data mining techniques in order to find the preferred tutor. The further sections describe the project on finding tutors using data mining techniques. Section II depicts the literature survey, Section III gives a detailed gist about the proposed application. Section IV describes the results obtained from the implementation of the application with the help of screenshots taken during project implementation. Section V refers to the conclusions and future scope drawn from the "Tutor Hunt" application.

2 LITERATURE REVIEW

Since the evolution of data mining technology, surveys and experiments have been carried out and data mining has been a major topic of research for many researchers. Data Mining is the process of converting data to knowledge.[4] It is the synonym for knowledge discovery from database. This means it collects data from large datasets and analyses the result to predict behavior and future trends allowing the organization to make

decisions. It scours the database for hidden patterns, finding predictive information which experts may miss while manually trying to identify patterns.[10]. Data Mining is widely used for market analysis, business management, and decision support as the need of transforming data to knowledge increases. Retail, manufacturing, health, finance, transportation based companies use data mining tools to take full advantage of historical data Data Mining works by building models of known data and applying this model to unknown data. Classification, Clustering, Association, Frequent patterns, etc. falls under Data Mining

Prediction of a certain outcome based on a given input is the main task under Classification. Prediction of the outcome is done when the algorithm processes a training set containing a set of attributes and the respective prediction attribute. The relationship between the attributes is discovered by the algorithm, and the discovery of result supports prediction of outcome. The prediction set without the predicted result is passed to the algorithm. The prediction is then produced by the algorithm by analyzing the input.[13].

2.1 Naïve Bayes

The Bayes rule of conditional probability forms the basis of the Naïve Bayes algorithm. The attributes in the data are analyzed individually considering each one to be important for the prediction and independent of the other attributes [11]. The various datasets used for 26 experiments shows that the performance of Naïve Bayes algorithm drops only in 3-4 cases in comparison with J48 and Support Vector Machines which proves that along with being simple, it classifies most of the problems well. Main features like simplicity and good performance makes Naïve Bayes is what it is used for. Naïve Bayes algorithm assigns probabilities as,

$$p(C_k | x_1, \dots, x_n)$$

The conditional probability is decomposed to get:

$$p(C_k | \mathbf{x}) = \frac{p(C_k) p(\mathbf{x} | C_k)}{p(\mathbf{x})}$$

Sumit Garg[4] defines data mining to be used as method at the interaction for Artificial Intelligence (AI), Machine Learning, etc. The research lists and explains the techniques like classification, clustering, association, prediction, sequential patterns, etc. which fall under data mining. The merits of Naïve Bayes include: High accuracy and speed on large database, minimum error rate, easy to understand, handles streaming of data well.

Xindong Wu [] has conducted a survey based on the top 10 algorithms identified by the IEEE International Conference. This survey consisted of those 10 algorithms, with their evolution, concept, limitations and how the concepts were gradually overcome. K-means, C4.5, Support Vector Machine, Apriori, EM, PageRank, AdaBoost, k- Nearest Neighbor, Naïve Bayes, CART algorithms.

Raj Kumar and Dr. Rajesh Verma [3], considered Naïve Bayes, KKN and Decision tree type. They executed four datasets and concluded that Naïve Bayes is better than the other two.

3 PROPOSED WORK

Considering the rising needs of private tuitions, "Tutor Hunt" application is introduced. This application runs on the Android platform. As described before, the tasks comprising this application development are data collection, database creation, information retrieval and implementation of data mining techniques in order to find the preferred tutor.

3.1 Data Collection:

For implementing the application, sample data from Borivali (West), Mumbai area has been taken. The teacher profiles participating in the tutor hunt dataset have been collected and the valid attributes like age, qualification, experience, whether state, ICSE or CBSE board, preferred medium to teach, any specializations, etc. These attributes are the key fields that have been considered for obtaining preferred tutor results.

3.2 Database Creation and Information Retrieval

The data collected from survey is stored in SQLite database with .db extension and when parent enters his/her requirement, according to their requirement list of tutors are displayed from database.

3.3 Implementation of Naïve Bayes algorithm

The working of the application is further described in the flow diagram given in Figure 1. When user starts "Tutor Hunt" application, two option are displayed; one for tutor and second for parent. When the user selects "tutor", the tutor has to register and give his/her details according to the requirement fields in application, also he/she can manage the schedule or upload videos and save their profile. When user selects "parent", he/she can search tutor with or without registration. When user wants to comment or give rating to a particular tutor then he/she needs to complete registration process.

A training dataset is already generated and processed in WEKA. When a tutor's record has to be tested, the record is supplied as a test set to the WEKA tool. The result of which is retrieved back in the application to the parent. The result is in the form of whether the tutor is preferable according to the parent's requirements.

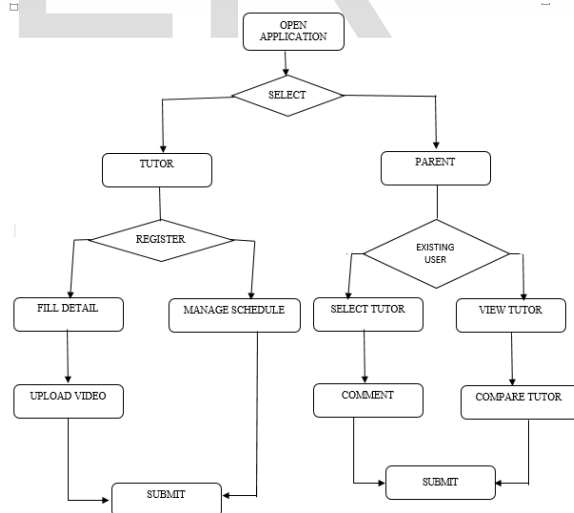


Figure 1: Flowchart of "Tutor Hunt" application

Searching operation is applicable to both tutor as well as student. From parent's side, searching will be done based on various fields such as experience, board, area, etc. From tutor's side, tutor has to give their personal information based on which application can search on parent's requirement.

4 RESULTS

“Tutor Hunt” is an application where there are three sections: parent, tutor and home. “Home” includes list of teachers with their names, experience and specialization with a button wherein one can click and view the profile of a particular tutor with his/her full record.

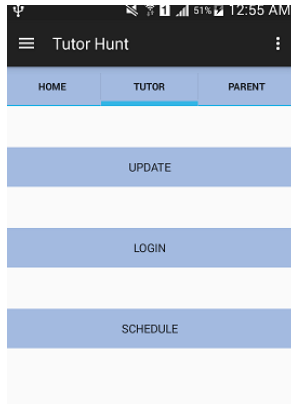


Figure 2: Tutor tab

Figure 2 shows the “tutor” tab along with a list of buttons namely update, login, schedule. When the user clicks on “update”, user has to login and update his/her data based on his/her preference.

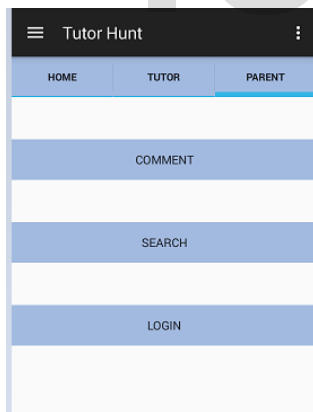


Figure3: Parent tab

Figure(3) shows parent tab that comprises of list of buttons namely search, login and comment. Parents can perform search operation with or without registration and can search for tutors near their homes, according to their preferences whether tutors prefer private, group tuitions or both and boards they would prefer to teach whether it is SSC, ICSE, CBSE.

WEKA IMPLEMENTATION

1. Create a training dataset.

The training data set is created and the data is preprocessed in WEKA. All the attributes (name, subject specialized, etc.) which are not required are removed and rests (age, experience, private/ public, ratings) are converted in nominal form for better discretization. Further, the Training data set is classified and the classified model is saved. The figure below depicts the preprocessed file. The @ relation specifies the table whose attributes are to be tested. The @ attribute denotes the attribute and the values it may contain. The data is contained at the last.

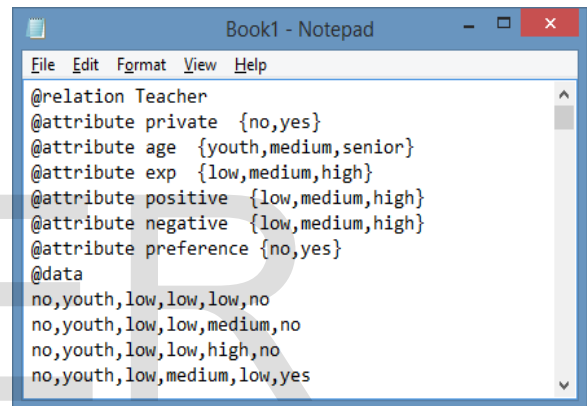


Figure 4: Training Dataset in arff format

2.Create the test set

The next step is selecting the records that are to be tested. The records are preprocessed like the training data set and here the last column is left blank as further it will be filled with the result classified by the WEKA .

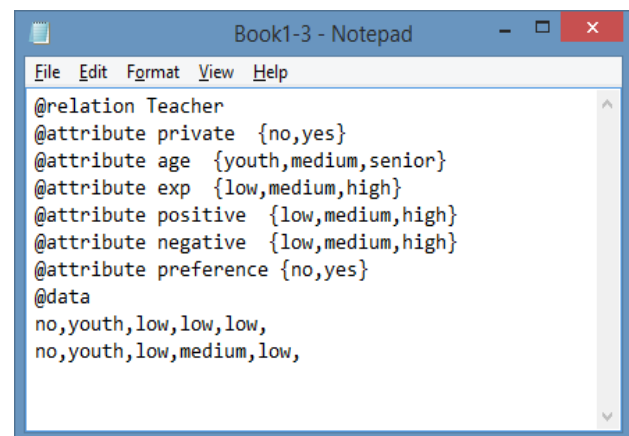


Figure 5: Preprocessed Test Dataset in arff format

3. Classify the test set using the training set

The classified training dataset is reevaluated using the testset.

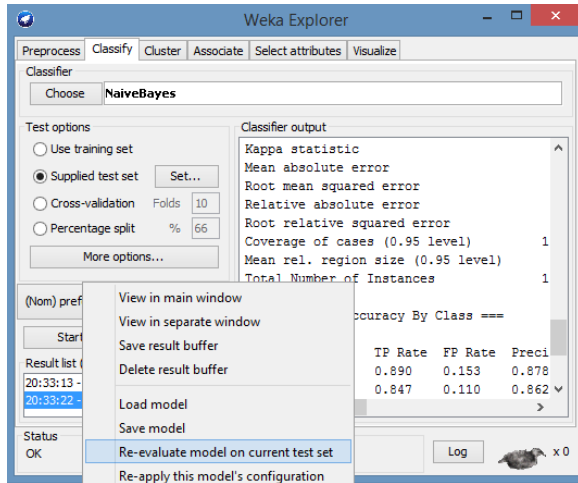


Figure 6: Choosing the Classified Training model and preprocessed Test set

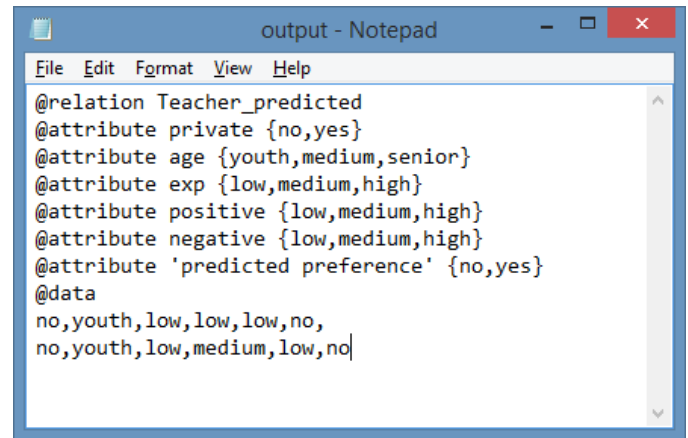


Figure 8: Classified Testset in arff format

The classified test set is further stored in a .arff file. The last attribute 'predicted preference' shows the classified records of tutor.

4 CONCLUSION

In this paper, we introduce the development of Android Application for finding tutor using data mining technique. This system is used by all i.e., technical/non-technical people, parents/children's or tutor. With the help of this application parents can find tutor for their on their smartphone and unemployed tutor can register him/herself as an employed tutor. For future scope, project may not be limited to only small primary and secondary students may expand with more features like client can find tutor for dance class, French class, singing class, etc.,

ACKNOWLEDGMENT

We would like to extend our sincere thanks to all of them who helped us in making our project successful. We are highly indebted to our project guide Prof. Ms. Elizabeth George for her guidance, support and constant supervision as well as providing the necessary information regarding the project. We would like to express our special gratitude and thanks to our project coordinators .Mrs. Shree Jaswal and Mrs. Prachi Raut for arranging the necessary facilities to carry out the project work. We are also highly grateful to Mr. Pramod Shanbag Head of Department (INFT), and Principal Dr. A.K. Sen and Director

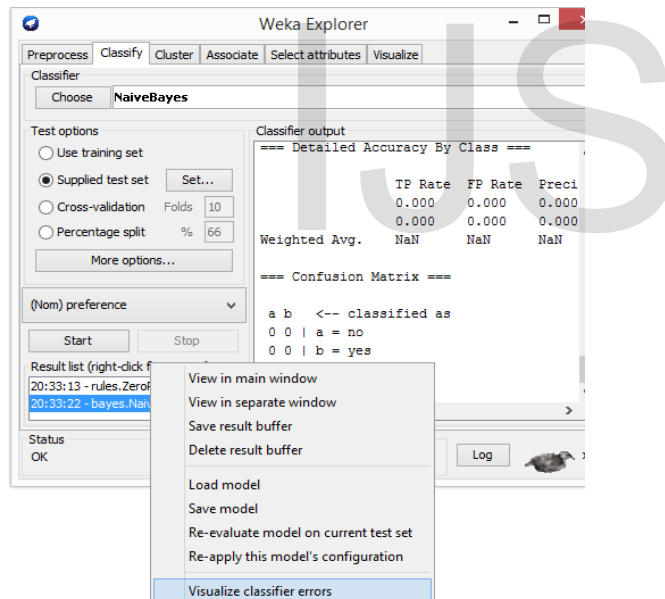


Figure 7: Exporting the result to separate file.

Bro. Melchior Tom for providing the facilities, conducive environment and encouragement

REFERENCES

- [1] S.Anita, "Implementation of Android Voice Recognition for Smart Home Application Using Bluetooth, *International Journal of Innovative Science and Modern Engineering (IJISME)*," volume 3, issue 5
- [2] Mark Hall, " *The WEKA Data Mining Software: An Update*"
- [3] Ming Xu, "Researchment and realization based on database application Technology Proceedments of the 2nd International *Symposium on Computer, Communication, Control and Automation*
- [4] Sumit Garg, "Comparative Analysis of Data Mining Techniques on Educational Dataset", *International Journal of Computer Applications*, Volume 74
- [5] ASSOCHAM India, "Private Coaching poaches mainstream education"
- [6] Reza, Arash and Behrouz, "Comparision of Classification Methods Based on the type of attribute and Sample Size"
- [7] V.Vaithyanathan, K.Rajeshwari, "Comparison of Different Classification Techniques using Different Datasets" May 2013
- [8] T.F.Michael Raj, "Domain Specific Web Service Composition by parameter Classification Using Naïve Bayes Algorithm", *Worlds Applied Sciences Journal*, 2014.
- [9] <http://www.laits.utexas.edu/~anornman/BUS.FOR/course.mat/>
Alex
- [10] <http://www.d.umn.edu/~padhy005/Chapter5.html>
- [11] <http://thearling.com/text/gmwhite/dmwhite.htm>
- [12] http://courses.cs.washington.edu/courses/csep521/07wi/prj/leonardo_fabricio.pdf