

# EMPLOYABILITY OF KNN, NAÏVE BAYES, AND RANDOM FOREST ALGORITHMS BY LEVERAGING MACHINE LEARNING TECHNIQUES TO ENHANCE THE APPROPRIATE SELECTION OF SUBJECTS FOR STUDENTS OF ICSE BOARD SYSTEM

ROMHARSH MITTAL

## ABSTRACT

Objective: Academic advising requires a lot of expertise, time, and responsibility. To assist the human advisors efficiently, the upcoming computerized advising system is a necessity.

Methods/Statistical Analysis: Course Advisory System has been implemented using the WEKA tool to recommend subjects for 8th class students of the ICSE board. Machine learning algorithms – Naïve Bayes, J48, PART, Random Forest, and KNN, have been modeled and tested on the data set. The performance of each classifier has been compared and analyzed.

Findings: It is inferred that no advising system has been developed to assist school students in subject selection. Research work based on Indian students' requirements is minimal. Research work based on students' data caters more to binary class problems, whereas the addressing of multiclass issues is minimal. This work proposes an advising system for the school students of the 8th standard of ICSE board to choose their electives.

Application/Improvements: This work focuses on the Indian educational system of school students. The approach takes care of the school students, which will add its advantage to the existing systems. As school students are more vulnerable by making the wrong decisions, the course Advisory system will assist them in analyzing their academic history and help them choose their electives wisely. The classification algorithms might give better accuracy with increasing instances. The Course advisory system can be enhanced using an ensemble approach.

## 1. INTRODUCTION

In the education system, students have to select the subjects that they to learn from a pool of various options. It is a crucial decision-making process as the future opportunities largely depend on the topics that they had studied at the school level. This research work focuses on assisting human advisors to advice 8th class students of ICSE board to choose their electives. Indian Certificate of Secondary Education (ICSE) is a board of education in India. The subjects offered in the 8th class are divided into three groups. The first group consists of the compulsory subjects- English, History & Civics, Geography, and Indian language. The second group consists of nine options- Mathematics, Science, Environmental Science, Mathematics, Commercial Studies, Technical Drawing, A modern foreign language, A classical style, Computer science, Environmental science, and Agricultural Science, out of which two subjects have to be chosen. The third group consists of three options – Economic

Application, Computer Application, and Commercial Application, out of which one question has to be selected. If the choice of items is inappropriate, then the student might get barred throughout life from reading that particular subject. For example, if a student doesn't choose math in the 9th and 10th, then the student might find it challenging to study in the later years, so advising the students must be done carefully.

The advisor generally takes decisions based on their previous experiences and the current rules of the school. This suggests a model of the computerized advisory system which uses the concept of training the system with the past data and testing the system with the new data. To make advising more profound, the current research study proposes the use of machine learning algorithms to train the network. WEKA efficiently supports machine learning algorithms as academic problems are liable to repeat themselves, so the solutions given to past issues can be used to provide solutions to current similar matters.

The need for computer-based academic advising systems is elaborated. The decisions taken by human advisors can sometimes be biased. To overcome this issue, the computerized system can assist human advisors in making academic decisions. Generally, social advising is based on past experiences and the current rules of the institution. Course advisory expert system<sup>1</sup> is designed as a case-based and rule-based reasoning system to recommend courses to a student. The recommendation is made based on the academic information, academic details, and family background details of students. Generally, students choose courses depending on the risk level. According to the risk theory, there are three dimensions of course selection- the lecture's style of teaching, the difficulty of the course, and the learning value. There are various other reasons which can decide the choice like- graded performance, classroom environment, and motivation. Students' data is collected using a questionnaire, and risk perception are studied. In employed the Intelligent Academic Advising using Genetic algorithms and decision trees to track the progress of the students and help the human advisors and students to achieve their academic goals. Suggests an online course recommender that suggests courses to the students depending on their learning styles. The information about the learning styles and learning patterns of students is collected through the 44 item questionnaire. In, software tools can help faculty and staff who advise undergraduate students and are used in course planning. In online-based educational advising, the system is designed based on the educational curriculum of each institution, grading system, and university policies using a Decision tree. In, a distance education advisor is designed based on the interest and academic performance of students using user-based collaborative filtering. In machine learning techniques like IB1, Naïve Bayes, C4.5, and RBF are used to improve the learning experience for online students. The indecision tree algorithm is used to analyze the data and enhance the student's learning process. In student's performance is analyzed, and the grade is predicted using classification algorithms- Naïve Bayes, K-means, Expectation-Maximization, K- nearest neighbors, and Decision tree.

In Artificial Neural Network and Web mining is used to identify the learning pattern of the individual student. In classification model is designed to predict student academic performance. In classification algorithms are used to determine the college students who are at the risk of failing in their first year. In early warning system is designed to estimate the student's success in exams using a naïve Bayes

algorithm. A web-based advising system is designed for undergraduate students. From the literature, it is inferred that the availability of the advising system to assist school students in subject selection is very minimal. Also, the research work based on Indian student's requirements for academic advising is minimal. It is also evident that the research work based on student's data caters more to binary class problems, whereas the addressing of multiclass issues is minimal. This current paper proposes an advising system that advises the school students of the 8th standard of ICSE board to choose their electives.

## 2. METHODOLOGY

The steps involved in implementing the Course Advisory System are given in Figure 1.

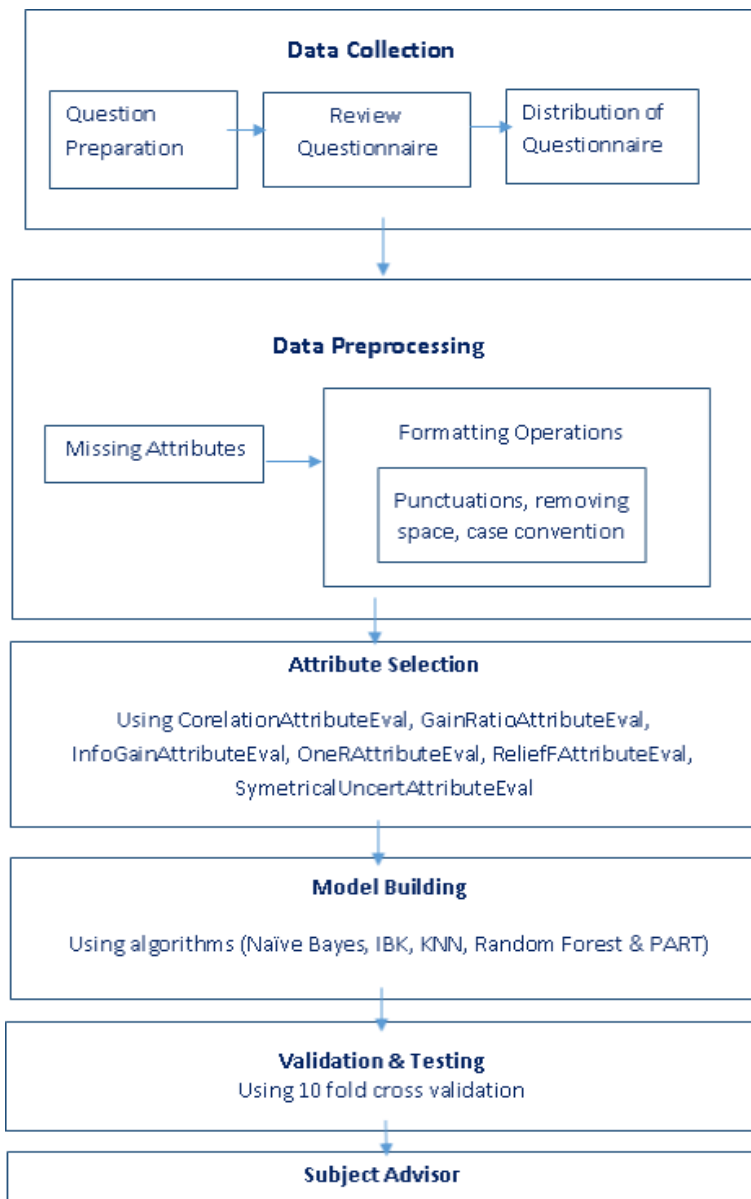


Figure 1. Advisory system.

**Figure 1: Steps involved in the process**

- Data collection- a questionnaire was prepared to collect data of students within the age group of 14-18 belonging to ICSE board who are studying/completed 8th standard. The questionnaire consists of 34 questions. A pilot study was conducted to review the questionnaire. The questionnaire targets collecting detailed information about 6th and 7th class academic records, focusing particularly on the subjects- Mathematics, Science, and English. The questionnaire also aims at capturing detailed information about the family and the financial condition of the student. The questionnaire was modified to make sure that people of every age group can understand and fill the questionnaire. After drafting the final questionnaire, it was put into social media sites for the collection of real-time data. Three hundred twenty-four student instances are considered for this research work.
- Data pre-processing- some data were not filled properly with more than five missing attributes, so those data were not included in the final data set. Punctuations and space were removed. The case was converted to lowercase.
- Attribute selection- using attribute selection algorithms like- CfsSubsetEval, 25 attributes were found to be important.
- Other attribute selection algorithms like-  
CorelationAttributeEval,  
GainRatioAttributeEval,  
InfoGainAttributeEval,  
OneRAttributeEval,  
ReliefFAttributeEval,  
SymmetricalUncertAttributeEval, the ranking of all the attributes, were found. Using the majority voting technique, the final ranking of all the attributes was listed.
- Model building- with the selected attributes, classification algorithms were applied to get the outcome. The algorithms used were- Naïve Bayes, J48, IBK, Random Forest, and PART. Naïve Bayes algorithm assumes that the value of each attribute contributes equally to the outcome. KNN assigns weight to the neighbours. The nearest neighbours contribute more to the outcome than the more distant ones. PART generates a decision list; the outcome is based on the best attribute. IBK helps to select the value of K based on cross-validation and also do distance weighting. Random Forest is used to creating a forest of random trees.
- Validation and testing- 10- fold cross-validation was used to measure the accuracy of the dataset.

Description of each attribute:

Awareness- this attribute asks whether the student is aware of a web-based advisory System. From the collected data, we can infer that 116 students are aware of the advisory systems, and 208 students are

not aware of the advisory systems. Feel to opt for any other course- this attribute checks the number of students happy with their choice and whether they feel to opt for any other course after joining. From the collected data, we can infer that 140 students wanted to change the subjects, and 184 students do not want to change the questions.

Gone through this attribute checks whether the students have consulted any advisory system before choosing the items. From the collected data, we can infer that 84 students took help in selecting the details, and 240 students did not take any assistance in choosing the topics. Ready- this attribute enquires whether the students are prepared to accept a new web-based course advisory system that is proposed in the present work. After joining- this attribute enquires whether the students have consulted any web-based advisory system after joining the course. Course Advisor to be- this attribute identifies the type of course Advisory system like a machine interface, human interface, face to face counseling, or telephonic. If the machine- if the Course Advisory System is a machine, then the students may use it as a web-based system or automated voice-based system. Meeting expectation- This attribute asks whether the course taken meets their expectations. Source of information- this attribute enquires from the source from which they got inspired to choose a particular subject. Out of all the options, the internet seemed to have been the primary source of information.

Parent's qualification- this attribute collects the highest academic degree of the parent of a student. This attribute helps us to analyze parental skills. Family status- these attributes contribute to understanding the family status of the student. Physical disability- this attribute takes into account whether the student has some physical disability. Level of attendance- this attribute accounts whether the student is punctual in class. Class participation- this attribute determines the attentiveness of the student on a scale of 5. The medium of education before 6th- this attribute helps to know the medium of training before the 6th. This attribute helps the advisor in deciding whether the student can perform well in a particular recommendation of a subject. If the medium of education before 6th was some other language other than English, then the student might find it hard to perform well with difficult issues.

Previous board - ICSE board being the strictest board of education, anybody with any other board before 6th standard might find it difficult to adjust with the problematic subjects if recommended. Favourite items- this attribute helps in knowing the interest of the student in a particular course and advise accordingly. Least favourite thing- this attribute helps the advisor to understand the subjects that the student is not interested in, so that the advisor may never recommend those subjects to the student. Subject best performed in – this attribute helps the advisor in knowing the question that the student is best in. In turn, it helps to recommend a combination having that subject. Subject struggled in- this attribute helps the advisor in knowing a particular topic that the student is weak in. Overall 6th percentage- this attribute records the mark of the student in portion. This attribute helps the advisor in knowing whether the student is below or above average. Overall 7th percentage- this attribute collects the mark of the student in proportion. This attribute, along with the 6th percentage, helps in deciding the overall performance of the student.

Area of interest - this attribute focuses on the field in which the student wants to build a career so that while recommending a particular subject, the field of the student can be kept in mind to get better outcomes.

Factors - this attribute lists out the reasons for choosing a specific theme. This attribute helps in understanding the mindset of the student while selecting a particular subject and leaves the responsibility on the hands of the advisor to direct the students on the path of their choice.

Role- this attribute helps the advisor in understanding the role played by parents and teachers in helping the student choose the subject.

Long term interest – this attribute helps the advisor in understanding where the student wants to see himself in the upcoming years. The guided recommendation can help them reach their goals. Extracurricular activities this attribute accounts for the extracurricular activities that the student is interested in apart from academics. As that sometimes might help the advisor in knowing the hidden talent of the student and recommend accordingly.

Connection- this attribute tries to connect the extracurricular activities with academics. Help- this attribute helps the advisor to know the extent to which the student needs help from the advisor. *Group 1* - lists out the electives liable to be chosen by the student. There are a total of nine subjects available, out of which two issues have to be selected. *Group 2* - There are a total of three questions available under this attribute, out of which only one question has to be chosen. Class- this attribute lists out all the combinations possible, including two questions from *Group 1* and two items from *Group 2*.

### 3. IMPLEMENTATION

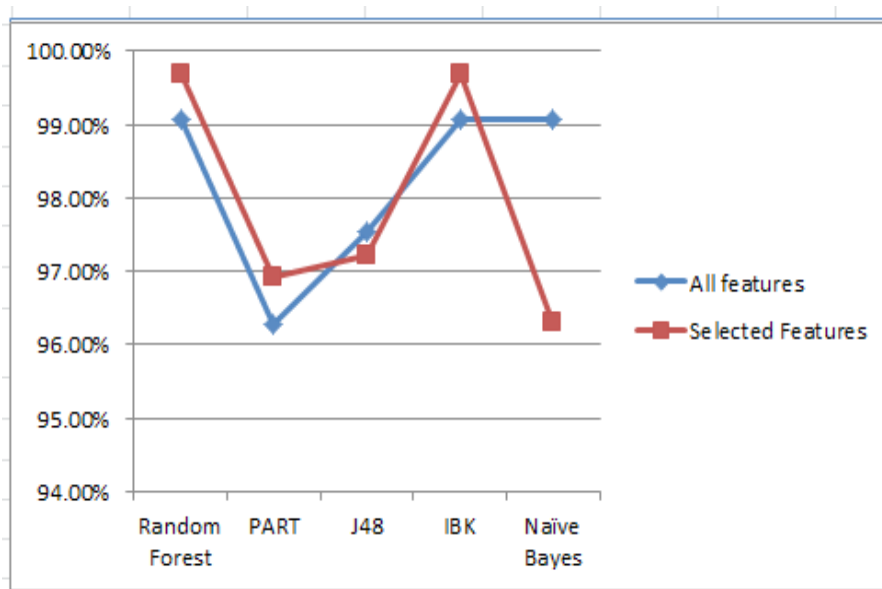
The procedure for implementing the course advisory system includes the following activities:

- The student should be in the 8th class.
- The students should be of the ICSE board.
- The students should fill the questionnaire.
- The questionnaire should take detailed information about the previous class's academic record (Math, Science, and English).
- The survey should capture detailed information about the subjects best performed and the subjects worst performed.
- The survey should mention the combination of items to choose from.
- The poll should check the family background of the student.
- The students are choosing Computer Science from the group I should not select Computer Applications from group II.
- After submitting the questionnaire, the system should be able to recommend subjects to the student based on the trained records.

Initially, 400 instances were recorded, out of which only 324 cases were considered for mining. Seventy-six examples had incomplete information. Initially, 34 attributes about the students were recorded, out of which only 25 attributes were finalized based on attribute selection algorithms.

## 4. RESULTS

Accuracy is a performance measure considered for various multiclass classifiers. Figure 2 is the graphical comparison of the Classification algorithms with and without Feature Selection. It can be inferred that with feature selection, the efficiency of each classification algorithm has increased. The accuracy of the classification algorithm Random Forest has increased from 99.07% to 99.69%. In the case of PART, the skill has increased from 96.29% to 96.91%. IBK also shows an increase in efficiency from 99.07% to 99.69%. Overall, it can be inferred that the classification algorithms-Random Forest and IBK are equally suitable for the current data set.



**Figure 2.** Comparison of performance of all classifiers using feature extraction methods.

## 5. CONCLUSIONS

8th class students generally choose the elective due to parental pressure, peer pressure, class environment, grade, and motivation, which might not end up giving them the outcome that they expect in the future. Course Advisory The system helps the students to know their strengths and weakness academically, thus helping the students to choose an elective which will help realize their future expectations. The approach is taken in the current paper takes care of the school students, which will add its advantage to the existing systems. As school students are more vulnerable to making wrong decisions, the course The advisory system will assist them in analyzing their academic history and help them choose their electives wisely. The classification algorithms might give better accuracy with increasing instances. The Course advisory the system can be enhanced using an ensemble approach.