International Journal on Future Revolution in Computer Science & Communication Engineering
Volume: 4 Issue: 4
ISSN: 2454-4248
311 – 320

Intelligent Recommendation System for Higher Education

Nikita Sawarkar
M.Tech (CSE) Student,
Department of Comp Tech, YCCE,
Nagpur

Dr M M Raghuwanshi
Professor, Department of Comp Tech, YCCE,
Nagpur

Dr. K. R. Singh
HOD, Department of Comp Tech, YCCE,
Nagpur

***Abstract*:-**Education domain is very vast and the data is increasing every day. Extracting information from this data requires various data mining techniques. Educational data mining combines various methods of data mining, machine learning and statistics; which are appropriate for the unique data that comes from educational sector. Most of the education recommendation systems available help students to choose particular stream for graduate education after successful schooling or to choose particular career options after graduation. Counseling students during their course of graduate education will help him to comprehend subjects in better ways that will results in enhancing his understanding about subjects. This is possible by knowing the ability of student in learning subjects in past semesters and also mining the similar learning patterns from the past databases. Most educational systems allow students to plan out their subjects (particularly electives) during the beginning of the semester or course. The student is not fully aware about what subjects are good for his career, in which field he is interested in, or how would he perform. Recommending students to choose electives by considering his learning ability, his area of interest, extra-curricular activities and his performance in prerequisites would facilitate students to give a better performance and avoid their risk of failure. This would allow student to specialize in his domain of interest. This early prediction benefits the students to take necessary steps in advance to avoid poor performance and to improve their academic scores. To develop this system, various algorithms and recommendation techniques have to be applied. This paper reviews various data mining and machine learning approaches which are used in educational field and how it can be implemented.

Keywords: *Choice based credit system, Machine Learning, Recommendation system.*

*****

I. Introduction

India's higher education system is ranked as third largest education system in the world; with Unites States and China being the first two [21]. The University Grants Commission (UGC) is the central governing body which is responsible for advising the Government, coordinate between the centre and the state and defining and enforcing its standards. Although India has many prestigious colleges like IITs, IIMs, IIITs, central universities and many more; the country still lags behind internationally [1].

The traditional education system in India, also known as teacher centric education system, the focus is all on the teacher and the student's participation is discouraged. The teacher is in full charge of the activities carried out in classroom, which ensures that classroom remains orderly and students can concentrate more. This approach makes the students independent of taking their decision as collaboration with each other is discouraged. However, the communication skills of the students may get hamper in this system as they are not allowed to talk or work together for any activity. Moreover, the biggest con this approach has is that students are not allowed to express their opinions, ask questions and direct their own learning.

The outcome based education system, also known as student centric education system, shares the focus on both teacher and student. The students and teacher participate actively in the discussion or in the learning equally. Unlike teacher centric education system, this approach encourages student's collaboration and group activities which generates student's interest in learning. Since students are also involved, they are free to express themselves, ask questions and learn independently; which does not hamper their communication skills. However, it is the difficult task for teacher to manage all students' activities at once, which can be difficult when students are working on different stages of the same activity. This educational system is based round goals or outcomes. By the end of the learning process, each student should have achieved the goal. The role of faculty, adapts into instructor, trainer, facilitator, and/or mentor based on outcomes targeted.

Recently, the University Grants Commission (UGC) has proposed Choice Based Credit System (CBCS) to have a uniform grading system across India and abroad. As the name indicates, CBCS allows students to have a choice to choose from prescribed courses; referred as core, elective, minor or soft skill courses. This system permits students to study at their own pace and the assessment is entirely based on credits the students earn. The CBCS is the one step taken ahead to redefine higher education by developing the curriculum keeping pace with liberalization and globalization in education sector. This system aims at taking measures to enhance the quality of higher education in India by improving and bringing innovation in curriculum, teaching learning process and examination and evaluation schemes. The CBCS is already being implemented in top institutions in India and abroad. In most of the existing systems, students are supposed to plan out their subjects at the beginning of the program (with the help of career

 ISSN: 2454-4248
311 – 320

counselor) or at the beginning of the semester without accessing the students interest and his performance evaluation [9].

II. Choice Based Credit System in Engineering

The All India Council for Technical Education (AICTE) is the statutory body and a national-level council for technical education, under Department of Higher Education, Ministry of Human Resource Development. AICTE is responsible for proper planning and coordinated development of the technical education and management education system in India. Time to time AICTE provide guidance and recommendations for under graduate (UG) studies, post graduate (PG) studies and research in engineering and technology. As per AICTE approval process handbook 2018-19 [22] there are 4397 approved institutions offering UG program with an annual intake of 1581372.

The CBCS can be seen as a "cafeteria" approach wherein students can select the subjects of their choice, understand them and learn at their own speed [1]. The subjects are three types of main courses namely the Core, elective and Foundation courses. The non-credit courses are also included in the program which are assessed either "satisfactory" or "unsatisfactory" and do not include in the performance evaluation. The core subjects are compulsory for each student to get the degree while elective subjects are those which a student can choose from a pool of subjects and are included for the final result of the student. The three categories of elective courses include discipline specific elective (DSE) course, dissertation/project, and generic elective course. The discipline specific elective courses are offered by the main discipline or subject of study, which can also include discipline related elective courses. The dissertation or project course is designed for the student to acquire specialized knowledge to a project. This course is generally studied by the student himself with guidance from faculty allotted. Generic elective course is chosen from undisciplined course to give the student an exposure to other disciplines. Foundation courses are offered to students to enhance their knowledge in subjects such as Environmental Science, Communication Skills and others which are mandatory for all disciplines [1].
The basic features of CBCS includes following [1][9]:

1. A semester pattern is followed in which a year is divided into two semesters: Odd and Even. The progress of student is calculated on the basis of courses taken on each semester. Each semester consists of academic training of fifteen to eighteen weeks; though there can be flexibility in this depending on the hours of teaching. Generally, 90 hours per semester are allotted for teaching.
2. Each course/subject is assigned a credit. The student earns this credit by passing in that particular discipline/subject. However, the student has liberty of earning these credits by his own pace and taking his own time.
3. The student is given freedom of choosing the number of subjects per semester and earns credits for those subjects. The remaining subjects can be compensated in the next semester. The credits can be earned in other college too. This gives the student a provision for credit transfer.
4. A continuous evaluation of student is done by teachers and student himself through assignments, tests and end semester examination.
5. A 10-point grading system has been introduced by UGC instead of marks.

The CBCS is in compliance with the global grading systems.

III. Issues Related to CBCS

CBCS is more advantageous over conventional methods of teaching [9]. However, some limitations are faced in practical applications. In most of the existing systems, the students are supposed to plan out their subjects in the beginning of the program; which can be done with the help of program counselor. The students or the counselor plans out the subject of the semester at the beginning of the program or at the beginning of each semester. However, student's performance or interest is not accessed before planning out the subjects. This affects the student's performance and the risk of failure. The student is not aware about what he has to choose to set his career path. That is, there is a lack of a mentor who can guide the student as to what should be his next steps towards his goals. The student himself is not aware of his capabilities and his career goals.

Though currently existing education recommendation system helps students to choose a graduating stream after junior college or after schooling, they do not provide for counseling or mentoring students during the course of education.

If a continuous mentor is available to a student throughout the program, then it is easier for him to plan his career more efficiently and improve his learning ability.

 ISSN: 2454-4248
311 – 320

IV. Related Work:

Data mining techniques and machine learning algorithms are of use when massive amount of data is to be classified and analyzed. In the past few years, some efficient works have focused on the use of machine learning algorithms and data mining techniques in the context of educational environment.

Grewal DS and Kaur K, proposed a multidimensional approach to the recommendation system in addition to the typical information of user and items [4]. It makes recommendations based on multiple dimensions, profiles, and aggregation hierarchies.

Cesar Vialardi, Javier Bravo, Leila Shafti, Alvaro Ortigosa, proposed that building a recommendation system in educational sector requires proper organization of data [6]. One way of preprocessing this educational data is use of clustering algorithm. K-means clustering algorithm has shown good outcomes to predict student's results.

A lot of work has been done in the field of predicting student's performance [14] [15] [16] [17]. A novel approach which uses recommender system techniques for educational data mining, especially in predicting student performance, also proposes how to map the educational data to items in recommender systems, as proposed by Nikitaben Shelke, Shriniwas Gadage [14]. To validate this approach, recommender system techniques are compared with traditional regression methods such as logistic regression by using educational data and the results claimed by Shruthi P, Chaitra B P [16] showed that the proposed approach have better efficient prediction outcomes.

Mining educational data can aid the students in selecting their course major. However, more research is required as to which data mining techniques and machine learning algorithms can give precise results to support student's selection.

To reduce the risk of failure, a course enrollment recommender system have been developed by Cesar Vialardi, Javier Bravo, Leila Shafti, Alvaro Ortigosa [6], which reminds students of their duties, warns them against difficult courses, and recommends them potentially beneficial courses [6]. It helps students to plan their subjects at the beginning of each semester. In addition to these existing systems, a collaborative recommendation system was engaged using association rules algorithm to recommend university elective courses to a target student based on what other similar students have taken.

An intelligent recommendation system for course selection of graduate courses, proposed by Grewal DS and Kaur K, provided a solution to the student's biggest bewilderment of choosing their correct field after the higher secondary examination [4]. This system provides for an effective counseling and guidance to the 10+2 students.
Another intelligent system, proposed by Reddy JM, Wang T, acts as a course coordinator which is not achieved by the faculties of particular college or universities. This system helps students to select their course majors [7].

V. Resolving Issues:

The course recommendation system plays a significant role in managing the curriculum and counseling students on academic matters, with a view to fostering their academic progress. The students can plan the subjects of higher semester based on their performance in lower semester. This would help in his better performance as the student would choose only those elective subjects in whose core subjects he has performed well. Better selection of electives would lead to a better and clear choice of career and thus can help in comprehending the subjects. The selection of electives can be based on marks supported by various other contributing factors such as student's personal interest in the domain of subject, his extra circular activities in the campus, understanding of the core subject and so on.

The prediction of selection of electives for upcoming higher semester will be based on past learning of the data. Mining scoring patterns can help to discover learning abilities of students from past databases; which would help in better prediction. Students can consider this prediction and choose electives accordingly to specialize in his particular domain of interest. The advantage of this early prediction can benefit student in taking necessary steps towards having and reaching his goal and thus avoiding poor performance. Moreover, the academic scores of the student would also improve as he would study willingly for his subject of interest.

The selection of electives and how the student performs in those throughout the course of programs shows a great impact on his career path. Choosing path of interest shows an influence in one's performance.

 ISSN: 2454-4248
311 – 320

Based on the choice and prediction of electives, a career path can be recommended to the student. Having known career options in advance, the student can take necessary steps in developing his personality towards that particular field. The future career options may include pursing higher educational courses like post graduation in engineering or management; having a job in specialized area; entrepreneur etc.

The main problem faced by students in universities is to take the right decisions for their career [1] [9]. The election of subjects plays a key factor in shaping the future of students.

The current system forces students to enroll for the subjects prior to the beginning of semester. These courses include professional courses, elective courses, compulsory courses and skill courses. The compulsory courses do not leave a choice, however the students have liberty of choosing their elective subjects according to their interest. Being ignorant as a student, they face difficulties in having a conclusion as to which subjects are going to be helpful in their path of career as well as suits their interest. This takes the chances of misshaping the career of students.

VI. Machine Learning

Machine learning can be defined as an application of Artificial Intelligence (AI) that provides systems to learn and improve automatically from experience without being explicitly programmed. This works in the same way as human brain: the more number of times humans read, the more we learn and remember.

The principle objective of machine learning is to allow computers to learn automatically from experience without any human intervention or aid and act accordingly. The process of learning involves observation of data, searching and evaluation of patterns, learning from examples, and direct experiences. This helps in making proper decisions or predictions for future.

Machine learning can be categorized in four ways:

1. Supervised Learning:

Labeled data is the data consisting of a set of training examples, where each example is a pair consisting of an input and a desired output value. Such data is used in supervised learning. Supervised learning occurs when the algorithm learns from example data which already has labels. This learning can be thought of as human learning under the supervision of teacher. The students are provided with good examples by the teacher to memorize and by which students draw general rules.

The main type of problems for this algorithm includes regression and classification problems. Supervised learning may include Nearest Neighbor, Naïve Bayes, Decision Trees, Linear Regression, Support Vector Machines (SVM) and Neural Networks.

2. Unsupervised Learning:

In contrast to supervised learning, the unsupervised learning consists of data which has plain examples and no labels. The information which is used to train data is neither classified nor labeled. This allows the algorithm to determine patterns on its own without any associated response. This type of algorithm tends to restructure the data into something else, such as new features that may represent a class or a new series of uncorrelated values. This kind of learning can be thought of the methods that humans use to find that certain objects or events belong to same class by observing some similarity between objects.

3. Semi-supervised learning:

These algorithms fall between supervised and unsupervised learning as both labeled and unlabeled data is used for training. Generally a small amount of labeled data and a large amount of unlabelled data is used. Semi-supervised algorithms serve as the best candidates for the model building. These methods exploit the idea that even though the group memberships of the unlabeled data are unknown, this data carries important information about the group parameters.

4. Reinforcement Learning:

It is a learning method that interacts with its environment by producing actions and discovers errors or rewards. Reinforcement learning occurs when you present the algorithm with examples that lack labels, as in unsupervised learning. However, an example with positive or negative feedback according to the solution the algorithm proposes. This method allows machines and software agents to automatically determine the ideal behavior within a specific context in order to maximize its performance.

Q-Learning, Temporal Difference (TD), Deep Adversarial Networks are some of the common reinforcement algorithms that are used widely.

 ISSN: 2454-4248
311 – 320

The following algorithms are most widely used in machine learning:

1. Linear Regression:

Linear regression is a way to explain the relationship between a dependent variable and one or more explanatory variables using a straight line. Linear regression models try to make the vertical distance between the line and the data points (the residuals) as small as possible. This is called "fitting the line to the data." Often, linear regression models try to minimize the sum of the squares of the residuals (least squares), but other ways of fitting exist. They include minimizing the "lack of fit" in some other norm (as with least absolute deviations regression), or minimizing a penalized version of the least squares loss function as in ridge regression [21].

Linear regression is straightforward to understand and explain, and can be regularized to avoid over fitting. In addition, linear models can be updated easily with new data. However, linear regression performs poorly when there are non-linear relationships. They are not naturally flexible enough to capture more complex patterns, and adding the right interaction terms or polynomials can be tricky and time-consuming.

The results of linear regression are human interpretable. It has fast training and prediction speed and performs well with lesser amount of data. Nevertheless, it suffers from lower average predictive accuracy.

2. Logistic Regression:

Logistic regression is another technique borrowed by machine learning from the field of statistics. It is the method for binary classification problems. Predictions are mapped to be between 0 and 1 through the logistic function, which means that predictions can be interpreted as class probabilities. It measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function, which is the cumulative logistic distribution.

Surprisingly, logistic regression falls under classification problem ad not under regression problems. It has fast training and prediction speed and performs well with lesser amount of data. However, it also suffers from lower average predictive accuracy and tends to underperform when there are multiple or non-linear decision boundaries. The outputs have a nice probabilistic interpretation, and the algorithm can be regularized to avoid over fitting. Logistic models can be updated easily with new data.

3. K-Nearest Neighbor (KNN):

KNN is a supervised learning algorithm in which the result of new instance query is classified. This classification is based on majority of K- nearest neighbor category. The goal of this algorithm is to classify new data object based on its attributes and training samples. This classifier does not fit to any model, unlike regression problems. If given a query point, we can find k number of objects or training points which are closer to the query point. K Nearest Neighbor classification algorithm uses neighborhood classification as the prediction value of the new query instance. These algorithms are memory-intensive, perform poorly for high-dimensional data, and require a meaningful distance function to calculate similarity. In practice, training regularized regression or tree ensembles are almost always better uses of your time.

4. Naïve Bayes:

Naïve Bayes is the most powerful algorithm used for predictive modeling. It is based on conditional probability; which means the probability of occurrence of event A, given that some event B has already occurred. The model is essentially a table that gets updated through the training data. For prediction of new observation or query, the table is simply “looked up” for the class probabilities in the probability table based on its feature values. The name "naive" comes because of its core assumption of conditional independence; which means that all input features are independent from one another. Though this assumption rarely holds true in the real world, Naïve Bayes model performs surprisingly well practically. It is very easy to scale with database. However, being very simple to implement, this model is often beaten by models properly trained and tuned using the previous algorithms listed.

5. Support Vector Machine:

Most commonly used for classification problems, SVM is a supervised learning algorithm which is based on the idea of hyperplane that best divides a dataset into two classes. Hyperplane can be thought of as a line that linearly separated as divides a set of data. The goal is to choose a hyperplane with the greatest possible margin between the hyperplane and any point within the

 ISSN: 2454-4248
311 – 320

training set, giving a greater chance of new data being classified correctly. Support vectors are the data points nearest to the hyperplane, the points of a data set that, if removed, would alter the position of the dividing hyperplane. Because of this, they can be considered the critical elements of a data set. SVM's are fairly robust against overfitting, especially in high-dimensional space. However, currently in the industry, random forests are more preferred, because they don't scale well to larger data set.

6. K-Means Algorithm:

K-means is a clustering problem which falls under unsupervised algorithm. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity. The centroids of the K clusters are used to label new data.

6.1 Recommendation systems:

Recommendation systems are intelligent agents that provide it's users with item recommendation or suggestions [4] [5]. There exists a dependency between user and item-centric activity. This can be explained with a very simple example of a customer buying a specific book. The customer can be interested in similar books on the basis of authors or the kind of book (inspirational, spiritual, romantic, fantasy, sci-fi, comedy). Studying the user history, we can recommend the customer similar kinds of books.

Collaborative filtering recommendation system makes the aggregation of customer's preferences and recommends them to other customers with similar behavioral pattern.

Content based recommendation systems make use of unsupervised machine learning algorithms to induce a classifier to distinguish between interesting and uninteresting items for the customer.

Knowledge based recommendation system collects data about customers and their products chosen to reason what meets the customer's requirements by making use of discrimination tree, decision support tools and case based reasoning.

In an education environment, recommendation systems act as an intelligent agent or a continuous mentor to the students, having as starting point previous actions from other students with approximately the same characteristics which includes academic performance of the student, his extra-curricular activities and other personal information.

Building a recommender system requires utilization of many machine learning algorithms. These systems requires learning from past user data or history. The system "learns" from this data to give predictions by filtering only useful data from pool of information. These predictions are generally co-related to suit user's interest and their needs which are learnt from past data.

The user history provides for "training" the system. The system learns from it to give predictions on test data. For this learning, various data mining and machine learning algorithms are used; for example SVM, Clustering algorithms, decision trees, Apriori algorithm, K-means, K-Nearest Neighbors etc.

VII. Proposed Work:

The proposed system aims at building a recommendation system for the prediction of future career paths for the students of undergraduate. The career path of the students is based on the selection of the electives throughout the course of program.

Selection of electives is the most crucial part of the career. First, the prerequisites of the elective subjects are analyzed. The performance in the core subjects decides the understanding of the core elements of related subject to the student. Thus, performance analysis of the student is crucial.

This can be explained with a very simple example. Suppose we have a dataset of 3rd semester passed out students. For testing, we take a new data of the student. The nearest neighbors of the new test data are found (in ascending order) using the KNN algorithm. Suppose we consider first five neighbors. If we find the performance of these five neighbors (students) in the fourth semester, we can estimate the performance of new student in the fourth semester; or at least we can provide with the performance range of student by simply considering the minimum and maximum marks amongst the five students. This would help in warning the student and thus can avoid the risk of failure.

The recommendation of electives can simply be done on the basis of performance in core subjects. The selection of electives ultimately decides the career path of student. By this, the student achieves specialization

 ISSN: 2454-4248
311 – 320

in a particular area. Thus, the proposed system helps in predicting the future career path of the student.

VIII. Work Carried Out:

8.1 Warning System:

The dataset of passed out students of session 2016-17 of Computer Technology Department, YCCE College, Nagpur is taken. The data is cleaned and divided into different sets of files according to semesters.

The first phase aims at creating a warning system for the student. The warning system would provide the student an estimate of range of marks he can score in each subject for upcoming semester. If, according to estimate, the student scores relatively poor in maximum number of subjects, then he has the scope to improve those scores. Thus, the risk of failure of the student can be prevented here.

To implement this warning system, KNN algorithm is used. The following is the KNN algorithm [25]:

1. Determine the value of K i:e number of neighbors
2. Calculate the distance between query instance and all the training samples.
3. Sort the distances and determine the nearest neighbors based on K^{th} minimum distance
4. Gather the category of nearest neighbors
5. Use simple majority of category of nearest neighbors as the prediction value for query instance.

According to algorithm, we have assumed the value of K as five.

The dataset that has been used to carry out the objective is taken from Computer Technology Department, YCCE College, Nagpur. The dataset consists of academic details of passed out students of the year 2016-2017.

Initially, the data for 3rd semester is filtered. A test data of new student is passed. The nearest neighbors of this test data is calculated using Manhattan Distance. The distance is calculated by the following formula:

$$D= |x_1 - x_2| + |y_1 - y_2|$$

Here (x_1, y_1) and (x_2, y_2) are coordinates of the planes P1 and P2 at right angles [26].

The nearest neighbors are sorted in ascending order and the first five are chosen as the most nearest neighbor to the new student (test data). The registration numbers of these neighbors are looked up in the data of 4th semester students. This gives their complete data of each subject. Through this we can find the minimum and maximum range of each subject. Thus, we can predict the performance range of the new student for the 4th semester.

Consider the following example:

The test data entered is as follows for each subject in 3rd semester.

Sub Code	CT203	CT205	EE218	GE201	CT204	CT206	CT210	EE217
Marks	65	70	84	54	87	84	80	82

On calculating the Manhattan Distance of the test data i:e, the query instance, with the dataset of passed out students, we get the following output:

```
The registration numbers with their mahattan distances are as follows:
(42.0, 12010545.0)
(57.0, 13010571.0)
(58.0, 13010650.0)
(63.0, 13010619.0)
(64.0, 13010609.0)
The 5 Nearest Neighbor (registration numbers) are:
[12010545.0, 13010571.0, 13010650.0, 13010619.0, 13010609.0]
```

The figures to the left in the bracket indicate the Manhattan Distance which is calculated and the figures to the right indicate the respective registration numbers which are nearest neighbors.

The data of 4th semester for these neighbors is retrieved:

 ISSN: 2454-4248
311 – 320

	Regno	CT207	CT208	CT221	ET213	GE206	CT209	CT222	ET214
4	12010545	54	46	69	51	58	66	64	67
14	13010571	81	76	79	77	91	88	85	80
44	13010609	75	95	78	83	87	90	87	84
52	13010619	87	89	90	86	95	75	86	81
68	13010650	75	86	82	61	79	88	87	84

After retrieving the data for the nearest neighbours in 4th semester, we calculate the minimum and maximum marks for each subject. This gives us the range in which the new student will lie.

```
The minimum marks in each subject are as follows:
Regno     12010545
CT207           54
CT208           46
CT221           69
ET213           51
GE206           58
CT209           66
CT222           64
ET214           67
dtype: int64

The maximum marks in each subject are as follows:
Regno     13010650
CT207           87
CT208           95
CT221           90
ET213           86
GE206           95
CT209           90
CT222           87
ET214           84
dtype: int64
The student would score in the range of min to max in each subject.
```

8.2 Recommendation of Electives:

This phase aims at building a system to help improve student regarding their choice of electives. Students are often confused while planning their subjects for upcoming higher semester and may end up choosing electives which can harm their career. The choice of electives is important because it specifies the area of specialization of the student. Though there are many factors contributing to the selection of electives, the marks scored in core subject play an important role. If the student has not performed well in the core subject, it indicates that his understanding in the subject lacks (if other factors are not considered). Therefore, the performance in the core subject is the basis for selection of electives.

A simple Rule based Learning System has been implemented for recommendation of electives. The core subject is entered and the marks for core subject are displayed for particular registration number. To choose or recommend a particular elective we have assumed criteria of scoring greater than 50 marks in its core subject. If the student scores marks greater than 50 in core subject, then the corresponding elective will be recommended.

In the example, the student has scored more than 50 marks in the core subject CT203 (Data Structure), therefore he is eligible to take elective subjects such as DAA (Design and Analysis of Algorithm) or ADS (Advanced Data Structure).

```
Please enter registration Number:-11010633
0    54
Name: CT203, dtype: int64
      Regno  CT203  CT205  EE218  GE201  CT204  CT206  CT210  EE217
0  11010633     54     45     43     51     59     66     55     68
Please select ADS or DAA
```

 ISSN: 2454-4248
311 – 320

The following example shows the selection of elective for the student who has scored less than 50 marks in the core subject. Such student is **not** applicable to choose electives based on the core subjects.

```
Please enter registration Number:-11010633
0    24
Name: CT301, dtype: int64
     Reg No  Sem  CT301  CT302  CT330  CT340  GE303  CT303  CT331  CT341  \
0  11010633    5     24     48     47     16     51     56     67     63

   CT328
0     65
Not applicable to select Language Processor
```

IX. Validation Results:

The test data of the student for example for warning system was taken from 3rd semester passed out student's data. The test data was removed from original file for accurate results. The test data is manually validated and it is found that out of eight subjects, predictions for six subjects are accurate. Apart from above example, validation for fifteen test data were carried out and out of eight subjects, predictions for almost seven subjects were accurate. The subjects which missed accuracy had a difference of one to two marks in most of the cases.

The following table shows the minimum and maximum range in each subject for 4th semester according to computational results:

Sub Code	CT207	CT208	CT221	ET213	GE206	CT209	CT222	ET214
Min marks	54	46	69	51	58	66	64	67
Max marks	87	95	90	86	95	90	87	84

The following table shows the data of 4th semester (of test data):

Sub Code	CT207	CT208	CT221	ET213	GE206	CT209	CT222	ET214
Marks obtd	80	87	84	77	95	90	91	85

Thus, it can be clearly seen that six predictions are accurate out of eight.

The example for recommendation of electives, the student who has scored less than 50 marks in CT301 (TOC), is recommended not to take LP in upcoming semester. However, the student had taken LP subject in the next semester and had not performed well by scoring 42 marks in CT318 (as validated from the available data of passed out students).

1	Reg No	Sem	CT314	CT316	CT318	CT324	ET331	CT317	CT319	CT325	ET332	CT329
2	11010633	6	0	44	42	0	37	20	53	28	77	50

X. Conclusion:

This paper reviews various research carried out in the field of educational data mining, to improve and predict student's performance; and the implementation work carried out. This system is designed for the students who can have an intelligent agent throughout the course of program and thus know their future in a better way. This would help the students to avoid risk of failure. The subject of focus is to improve student mentoring system which would help student to excel in his area of interest and to predict a definite career path.

XI. Future Scope:

The contributing factors for prediction can be expanded to student's domain of interest, accessing his data of extra-curricular and co curricular activities, his understanding and preparation of subject and other qualitative factors. This system can be expanded to all streams of engineering and other educational fields.

References:

[1] P. S. Aithal, P.M. Suresh Kumar, Analysis of Choice Based Credit System in Higher Education, MPRA Paper No. 71743, posted 6 June 2016 07:23 UTC

[2] Tripti Dwivedi, Diwakar Singh, Analyzing Educational Data through EDM Process: A

[3] Survey, International Journal of Computer Applications (0975 – 8887) Volume 136 – No.5, February 2016

[4] José del Campo-Ávila, Ricardo Conejo, Francisco Triguero, Rafael Morales-Bueno, Mining Web-based Educational Systems to Predict Student Learning Achievements, International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 3, No 2.

[5] Grewal DS and Kaur K, Developing an Intelligent Recommendation System for Course Selection by Students for Graduate Courses, ISSN: 2151-6219 BEJ, an open access journal, Volume 7, Issue 2

[6] Youngseok Lee and Jungwon Cho, An Intelligent Course Recommendation System, mart Computing Review, Volume 1, Issue 1, October 2011

[7] Cesar Vialardi, Javier Bravo, Leila Shafti, Alvaro Ortigosa, Recommendation in Higher Education Using Data Mining Techniques, Educational Data Mining 2009

[8] Reddy JM, Wang T (2014) Online Study and Recommendation System. Final report ACM 1-8.

[9] Yu Lou, Ran Ren, Yiyang Zhao, A Machine Learning Approach for Future Career Planning, Standford

[10] Ms. Maitri Jhaveri, Dr. Jyoti Pareek, Ms. Jasmine Jha, Assemblage of Recommendations with Constraints - A Choice Based Credit System Perspective, 2013 IEEE Fifth International Conference on Technology for Education

[11] P.Veeramuthu, Dr.R.Periyasamy, V.Sugasini, Analysis of Student Result Using Clustering

[12] Techniques, P.Veeramuthu et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5092-5094

[13] Amirah Mohamed Shahiria, WahidahHusaina, Nur'aini Abdul Rashid, A Review on Predicting Student's Performance using Data Mining Techniques, The Third Information Systems International Conference, Procedia Computer Science 72 (2015) 414 – 422, Elsevier

[14] Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan, A Systematic Review on Educational Data Mining, DOI 10.1109/ACCESS.2017.2654247, IEEE

[15] Dr. M. Thangamani, T.Selvakumar, Exploring Educational Dataset using Data Mining Technique, International Journal of Advanced Research in Biology, Engineering, Science and Technology (IJARBEST) Vol.2, Issue.2, February 2016

[16] Nikitaben Shelke, Shriniwas Gadage, A Survey of Data Mining Approaches in Performance Analysis and Evaluation, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015 ISSN: 2277 128X

[17] P. Kavipriya, A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 12, December 2016 ISSN: 2277 128X

[18] Shruthi P, Chaitra B P, Student Performance Prediction in Education Sector Using Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 3, March 2016 ISSN: 2277 128X

[19] Suchita Borkar, K.Rajeswari, Attributes Selection for Predicting Students' Academic Performance using Education Data Mining and Artificial Neural Network, International Journal of Computer Applications (0975 – 8887), Volume 86 – No 10, January 2014

[20] C. Romero and S. Ventura, Educational data mining: a review of the state of the art, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions, vol. 40, pp. 601-618, 2010.

[21] K. Chu, Designing a course recommendation system on web based on the students' course selection records, in Proc. of World Conference on Educational Multimedia, Hypermedia and Telecommunications, pp. 14-21, 2003.

[22] Kongsakun K, Fung CC, Chanakul T (2010) Developing an intelligent recommendation system for a private university in Thailand. Issues in Information Systems, 11: 467-472.

[23] www.wikipedia.com

[24] AICTE approval process handbook 2018-19 (https://www.aicte-india.org/sites/default/files/APH%202018-19%20Modified.pdf)

[25] ALL INDIA SURVEY ON HIGHER EDUCATION (AISHE) report 2016-17 (http:// www.aishe.gov.in/aishe/viewDocument.action?documentId=239)

[26] Aspiring Minds' National Employability Report-Engineers 2016.

[27] www.evoledu.com

[28] A Programmer's Guide to Data Mining: The Ancient Art of Numerati by Ron Zacharski