

APPLICATION OF DATA MINING TECHNIQUES AND ALGORITHMS IN EDUCATIONAL SECTOR: A SYSTEMATIC LITERATURE REVIEW

Shital C.Kadam, Sadhu Vaswani Institute of Management Studies For Girls, Pune, India
Prof. Rahul Patil, Abhinav Education Society's Institute of Management & Research, Pune.,
India

Abstract

Numerous fields of application have made significant use of data mining for prediction or classification. Data mining has seen widespread adoption in the fields of science, bioinformatics, and business because it greatly increases the productivity of its analysts by turning their vast, unmanageable, and prone to ignorable information into useful pieces of knowledge. The study of applying data mining, machine learning, and statistics to information produced in educational contexts is known as educational data mining (EDM). It makes use of a variety of methods to enhance educational outcomes and clarify educational processes for future decision-making. Clustering, Classification, Regression, Neural Networks, Artificial Intelligence, Association Rules, Genetic Algorithms, Decision Trees, and other techniques and algorithms are used to discover knowledge from databases. The study offered a systematic review of the literature on EDM that spans articles published between 2011 and 2023. Analysis is done by exploring objectives, algorithms used and Findings of research study on EDM. Examining these documents reveals that there is more room for research into EDM in course prediction.

Keywords:

EDM prediction, classification, machine learning, Neural Networks, Artificial Intelligence

I.INTRODUCTION

The burgeoning inter-disciplinary study field known as Educational Data Mining involves both computer science and education. (EDM). EDM makes use of data mining tools and techniques to draw out relevant and useful information from sizable educational datasets [1]. It can be viewed as an interdisciplinary study area that offers fundamental understanding of the teaching and learning process for quality education [2]. The goal of Educational Data Mining (EDM) is to analyze different educational data sets using methods and techniques from machine learning, statistics, Data Mining (DM), psycho-pedagogy, information retrieval, cognitive psychology, and recommender systems in order to address educational issues. There are numerous tools like Weka and Rapid Miner, among others, that use a combination of DM algorithms to assist

Researchers and stakeholders in finding solutions to particular issues. [3]. In [6] providing students with the information and skills they need to transition into successful careers within a set time frame is one of the main objectives of any educational system. Prediction of students' performance has long been a fascinating field of study in educational data mining, and it aids in identifying the weaker or more vulnerable students.[8]In knowledge discovery, data mining has been used to find patterns based on user requirements.[6]. If properly applied, educational data mining, a comparatively new field, has a lot of potential to benefit society [10].In [6] lists the following categories for the work in EDM: prediction, clustering, relationship mining, data distillation for human evaluation, and discovery with models. In [13] shows that the field of educational data mining is one that has a lot of potential. It focuses on creating methods for examining particular kinds of educational data and using those methods to learn more about students' strengths and flaws. The study looks at the goals of the research, the algorithms used to carry it out, and the research's findings in order to examine and analyze research on EDM. Standard systematic reviews of the EDM literature will be used to achieve this goal.

Following are the research objectives:

1) In this survey work focuses on objectives, algorithms used and Findings of research on EDM.

2) We have done a thorough, systematic review of the literature that spans between (2011–2023) areas of education where we can data mining techniques used for prediction and decision making. The remaining section of the paper is structured as follows: Section II discusses the literature, Section III explains the methodology and Section IV highlights the results and discussion followed by Section V conclusion and future enhancements

II. LITERATURE SURVEY

Given that this is a review article, it is crucial to describe the search criteria for the literature and the underlying methodology.

The research question for this study is to aggregate the use of data mining algorithms to educational data.

We have outlined some of the key EDM works.

Section A:

Student performance analysis

In [4] predicted student performance in distributed environment using classification algorithm in data mining. The aim of the study was to create global rules by merging many local rules created by many classification algorithms. java using .net beans 7.0 has used in the study. Weka tool is used for the study. Data set collected from parul institutes students contains 1000 records. Random decision tree is applied on the dataset. Rules generated from each notes were combined to create global model with the help of specific classification method. Result shows that global rules can be beneficial for their individual results and also for academic institution profile. developed a decision tree model to predict student performance in engineering dynamics – a high-enrollment, high-impact, and core engineering course. Result of the study revealed that key significant factors are student's score in Statics and cumulative GPA which has impact for student performance in engineering dynamics. It turns out that decision tree predictions are more precise than those made using the more conventional multivariate linear regression method. The study include five predictor/independent variables including the student's cumulative GPA and scores in four prerequisite courses: Engineering Statics, Calculus I, Calculus II, and Physics And one outcome/dependent variable (i.e., a student's score in dynamics) the data set consists of 750 data records from 125 students in Semesters A (45 students) and B (80 Students) having majors were: mechanical engineering (61), civil engineering (40), and bioengineering and other majors (24). Rules have been established and the relationship between the courses that the students failed has been revealed using an apriori algorithm on the data of students from Istanbul Eyup I.M.K.B. Vocational Commerce High School. For studying algorithms on data, a software programme for schools has been developed. This general-purpose programme was designed to extract association rules from databases. All of the student and school data has been uploaded to a database. Using these data association rules has obtained [5]. In [6] the data and methodology employed in those investigations is included in this paper's assessment of pertinent studies in the EDM field. The development of techniques for the extraction of intriguing, comprehensible, practical, and original knowledge is aided by the data mining of educational data (EDM), which could result in a better understanding of students and the environments in which they learn. Study observed that EDM can be used in a wide range of situations, such as detecting at-risk students, prioritizing the learning requirements of various student groups, raising graduation rates, accurately evaluating institutional performance, allocating campus resources, and improving topic curriculum renewal. [7] Compared various performance of feature selection techniques used for predicting students performance to identify best feature selection and classification algorithm. For the study algorithms used are J48, Naïve Bayes, Bayes Net, IBk, OneR, and JRip. The algorithms CfsSubset evaluations, Chi-Squared Attribute Evaluation, Information Gain Attribute Evaluation and Relief attribute evaluation used in this study. Two testing options such as cross validation and percentage split were applied on the classifiers. ROC values and F-Measure were generated. In this study Classifiers was implemented Without Feature Selection algorithm (WFS) and with Feature Selection algorithm (WFS). And finally

comparison has made so from study found that CFS Subset Evaluator performed better than other three feature selection algorithms. WEKA software used in the study. Data collected from Tamil Nadu state of India. Student's performance's report of final examination of Bachelor degree program in 2014 was considered for the study. Total 610 students with 21 attributes data collected by using questionnaire. Attributes like demographic characteristics, student's admission data and pre-collegiate features of the students and examination performance were considered. From the study it is concluded that IBK shows high accuracy of 99.680%. than other classifiers and Subset Evaluator performed better than other three feature selection algorithms. So this study will help to identify the slow learners so that faculty will focus on them for their improvement. [8] Used Improved Naive Bayesian mining technique for proposed system. Proposed system will predict the academic performance of the students based on their academic history and predict upcoming semester performance. For the study 300 engineering students' of Gulzar Group of Institutes has collected. Total 21 attributes were considered for the study. FP Rate, Precision, Recall, TP Rate, F-measure, Kappa Statistics and Accuracy parameters were considered to make comparison between Naive Bayes and Improved Naive Bayes. Study concluded that Improved Naive Bayes is shows accuracy of 86.66 % which is better than only Naive Bayes. Build the classification model by using Support Vector Machines, Decision Tree and Gradient Boosting algorithms to predict the student's performance at last semester external exams. For the study data collage in JNTUA region in Andhra Pradesh were collected. In the study accuracy of selected classifiers were calculated and found that every classifiers showed the accuracy above 75%. Accuracy varies upon the attributes. Results show that prediction of distinction class was not good as compared to first class. Decision Tree algorithm, Support Vector Machine and Boosting are the algorithms applied on the dataset. Data was collected from the college are of students data collected during admission process. Students Sex, Students Branch, Students category, Students grade in High School, Intermediate %, Day scholar-Living Location of Student, Hostler- Student stay in hostel or not, Last Semester Mark, Internal Exams Grade, Seminar Performance, Assignment Exam, Present Attendance, External Exams are attributes included in the study. Researcher conclude that Boosting algorithm perform well than other classifiers [9]. In [10] objective of the study was to comparative study of marks predictions using data mining techniques. For the study researchers has used data set of four hundred and eighty students with 16 parameters such as gender, nationality, place of birth, marks etc. Students performance was classified in three category such as low-level ,mid-level and high-level.for the study researcher has selected attributes like gender, nationality, place of birth, educational stages ,grade levels, section ID, Topic covers, semester, parent raise hand, visited resources, Viewing announcements, Discussion groups , Parent Answering Survey , Parent School Satisfaction, Student Absence Days. All theses parameters are used to predict the performances category of the students. The algorithms used were J-48, Random forest, naive bayes, naive bayes multinomial, K-star, IBK. Researcher found that J48 (Decision Tree) giving 75.8333 % accuracy , Random Forest - 76.6667 % , Naive Bayes - 67.7083 % , Naive Bayes Multinomial - 43.9583 % , K-star - 73.75 % , IBK - 72.91%.From the result is concluded that random forest algorithms performs best that the others. Designed the model which predicts the student's performance in end semester examination by using decision tree method and probabilistic classification method. This can predict the early dropout. Also compared two classification algorithms ID3 and Naive Based algorithm. Data such as Attendance, Class test, Seminar and Assignment marks were collected from the student's management system [11].

Student placement

In [12] studied the employability factors that can be enhancing employability skills which deal with the adequate teaching methodologies. Study mainly focuses the employability of engineering graduates. Questionnaire was used to collect data from 300 records of Graduates studied from various engineering colleges. Analysis was done by SPSS tool using Chi square Test. Study shows that Engineering Branch and the working specialization shows significant Association, Engineering Branch And Students placed in campus interview, Graduates Working on the Specialization and job

satisfaction, Placement opportunities in campus interview and CGPA of a student, Salary received per annum and satisfaction with job profile affects are important employability factors which can improve the chances of employability. Predicted student placement in software companies using educational data mining. For the study classification algorithms has used. Data collected from students undergoing higher education of Delhi state university's affiliates. For the prediction students' personal, pre-university and operational characteristics were considered. Total data from 1000 students were collected. Algorithms like Naïve Bayes and Bayes Net, Kstar, IBk, Decision table, OneR, JRip, J48, LMT, Random Forest, Random Tree were implemented on the data set using with 10 fold cross validation and Percentage split. Study showed that J48 Decision Tree Algorithm is the best compared to the other algorithms with accuracy 76.41%. From the study it is observed that students using both Internet based (Facebook, Twitter) and Mobile based (Whatsapp, wechat, line) networking methods get placed as compared to students who are using only one method. Study shows that whether student will employed or unemployed and can be guided to improve upon the features that contribute towards his placement [13].

Career selection

In [14] the aim of the study was to predict the graduation of students on time using artificial neural networks. For the study data collected were from 2009 to 2013 of 193 students academic and student affairs administration (BAAK) in STMIK Widya Cipta Dharma Samarinda. matlab tools with artificial neural network were applied on the dataset. Grade point, Cumulative semester of credits, economic status, job status, goal attributes considered for the study. From the study it is examined that the neural network with one input layer (5 neurons), one hidden layer (5neurons), one output layer (one neuron), learning rate = 0.001, activation function trained, learning function trained can be used to predict with precision.

Proposed prediction and recommendation systems using use of mining and statistical Algorithms. Study focused on accuracy approach for prediction. In the study Pattern mode, K-Cross Validation [10], Training and validation methods were used for the study. Proposed prediction and recommendation systems was build using C4.5 algorithm. Java and SQLServer 2008, Statistical tool and Weka tool were used for the study. Data set contain 200 students having 12 attributes. The proposed model shows 86% accuracy. Strengths, Reliability, decision Making, Attitude, Discipline, learning Ability, social Ability, Adaptability, Leadership, Team Work, Verbal Communication, and Persuading attributes considered for the study. Data mining algorithms such as C4.5, Naïve Bayes, K Star, simple cart were considered and their accuracy were compared. From the study it is revealed that C4.5 performs better than others. The proposed system predicts the career choice using student's psychological condition [15]. In [16] proposed a model from students' enrollment data to discover highly positive association rules. We assess the suggested approach using computer science students who entered the programme in 2008/2009. The information was received in text file and Microsoft Excel format from the Division of Academic at University Malaysia Terengganu. There were 160 participants. Successful extraction of 4,177 association rules in total. Only three of these association regulations are reasonable and acceptable. The study concluded that proposed model can discover the association rules with highly correlated. In [17] developed knowledge based decision technique by applying new attribute selection measurement function (heuristic) to the existing C4.5 DT algorithm for selecting proper branch of engineering. Study also analyzed the accuracy between decision tree algorithm (C5.0) and back propagation algorithm (ANN). for the study AIEEE2007 Database were considered on which a modified DM technique was applied. Result showed that accuracy rate using C5.0 and ANN algorithms for different partition sizes is 12 % higher. Designed the model to select an institution and program or course. For course prediction attributes such as field of interest, personality, trait, and mental ability of students were considered for the study. 6, 50,000+ datasets were collected study. For the study the proposed analytics focused on analysis of counseling guidance by using big data by using hadoop-hive software. The aim of the study to analyze historical datasets of various

colleges affiliated to university and accordingly performed comparison of their performance. For the comparison factors considered are placements, student- staff ratio, affiliations, students' performance and student rating. Proposed system has various modules like Student Population , Seat Filling by Category ,Staff student ratio ,Performance ,Branches are Choosing Mostly , Colleges are Choosing Mostly ,Placements ,Staff Info ,NBA accredited colleges , NAAC accredited colleges , Autonomous .this study designed web based application using big data analytics which reduce efforts' required for college selection[18].In [19] presented decision support system/expert system for Guiding Fresh Students in Selecting a Faculty in Gomal University which is based on student capabilities like intelligence, understanding, comprehension, mathematical concepts plus his/her past academic record .in the study authors converted the knowledge into facts and rules in CLIPS syntax, and store them in the knowledge base of the CLIPS language. The developed expert system is rule based which identify y the most suitable faculty or major for the student based on his abilities and capabilities extracted from the test module results. compared various data mining algorithms to predict the the career decision for students who have passed 12thstd.For the study Random Forest learning technique has used.200 samples data collected of class 10th and 12th standard students of career counseling. Various data mining algorithms such as ID3, K-Means, Naïve Bayes and Random Forest were applied on the data set. Student Id, Gender, Student category, Medium of teaching,stream,10th Grade,12th Grade, Type of coaching, Scholarship, Admission Type, Type of Coaching Material ,efficiency, Father's occupation, Mother's occupation, Parent's income status are the attributes considered for the study. The study concluded that as data size increases Random forest gives better result or accuracy [20].In [21] focused on application of data mining techniques for student enrollment in higher education and to find out which factor create the loyal students. Also identify areas in which data mining be applied. Researches explained how associations, classification and clustering will be used in educational data mining in the area like to predict faculty retention, resources optimization, also in recruitment. Focused on to find out the effect of parameters such as "parent's profession," "mass media" and "personal choice" on students" career selection. The first semester students from three of Lahore's top business schools participated in an online poll to provide the data, which was then collected using a process called systematic random sampling. This research supports the idea that "parenting profession" has impact on a student's career decision. Similar to this, "personal choice" has a significant and immediate influence on students' job decisions. This study also asserts that media is a crucial factor in career choice. Additionally, the three factors—parents' occupations, the media, and students' own preferences—all interact to impact students' career decisions. According to the study, "personal choice" and "mass media" have a greater influence on students' career decisions than parents' occupations do. Data analysis was done using AMOS 18 and SPSS 17 software. This study supports the idea that parents' occupations have a significant and direct impact on their children's career choices. Similar to this; "personal choice" is strongly related to the career choices made by student. Additionally, "mass media" has a direct and indirect impact on the career choices of student. Additionally, each of the three variables—parents' occupations, the media, and pupils' own preferences—influences the other two [22].In [23] proposed system for career option predications based on their personality trait, interest and their capacity to take up the course. Study performed the comparison of performance between the data set using C5.0 with adaptive boosting and C5.0 on dataset with personality, interest and capacity. Adaptive boosting algorithm is implemented which showed 94% accuracy which is better than5.0 without adaptive boosting algorithm .In [24] developed the Student Career Prediction System using The K-Nearest Neighbor and Certainty Factor algorithms. Exam results, student interests, and aptitude were utilized to establish each student's career-appropriate decision. Dataset consists of 102 students in which training data of 78 students and testing data of 24 students. Questionnaire was used to collect the data. The findings indicate that the K Nearest Neighbor approach was successful in predicting careers because of a value derived from the Certainty Factor. When these two techniques were combined, the accuracy value for K-Nearest Neighbour Value $k = 3$ was 70%, and the combined Certainty factor was 0.99%.The K-Nearest Neighbor and Certainty Factor

technique combined yielded a value of 93.83%. Therefore, it can be said that the K nearest Neighbor approach and the certainty factor method have good accuracy. It was anticipated that the two-way analysis would give students greater information to help them choose a career.

Developed the Decision Support System Education Career Choice Using Fuzzy Model. The created expert system provides insight into the body of knowledge that the upcoming specialty can be attributed to. An expert system was implemented in the initial step, allowing you to select one of five preferred models for the study's direction from a list of options based on the Klimov approach. Then, at the second level, a fuzzy direction of training was implemented with the aid of a fuzzy apparatus, specifically the Mamdani algorithm, in which the terms of the output variable were the names of specializations. Five fuzzy choice models that are based on the input linguistic variables are used to examine the impact of the input parameters on the output variable of the training direction choice. For each model, a knowledge base has been created that consists of production rules that are expressed as fuzzy linguistic assertions. The efficiency of the developed models has been confirmed by experimental research. Making an informed decision will be improved by the capability of testing using the system and receiving results in the in form of the specialized name. The data were compared with the output of the generated system after the fuzzy system was modeled using FUZZY TECH. The result was a system that was divided up into many modules. The user is required to complete a test to determine his preferences in the first section. The system gathers and evaluates the results when the test is passed. The mechanism of imprecise choosing of the direction of inquiry is triggered the second section based on the findings. This strategy will greatly boost the accuracy of specialty selection and enhance the indicators for graduates to enter the field they have chosen [25].

Educational Data Mining Application

In [26] reviewed on educational data mining, its advantages and barriers. Examined work done in education data mining area, its applications and algorithms [27]. In [28] identified role of data mining in educational field. Reviewed different data mining algorithms such as classification, clustering, association and genetic algorithms used in educational sectors [29]. Analyzed the different data mining techniques and algorithms for patten finding used in education institutes. Also examined the goal of educational data mining in higher education [30]. In [31] improved decision tree model by using ID3 to predict whether student will continue or drop the Course.

Section B: Potential Research Gaps Identified

Following is the summarized list of potential research gaps identified through extensive literature from section A

- 1) It is observed [4],[5],[6],[7],[8],[9],[10],[11],[12],[13] that most of the research in educational data mining is focused on student performance prediction, students placement and few on career selections.
- 2) The research work conducted previously [16] , [18] , [23] , [47] , [51] , [53] does not considered all factors at a time from demographic, personal, academic and social area while course selection.
- 3) The systems [18],[25],[33],[38],[52],[45] fall short of providing a general solution aimed at a larger population because they are tailored systems limited to a particular geographic region, where students from particular cultural backgrounds and educational backgrounds have been taken into account..
- 4) Some Models [24], [32], [34], [38], [40],[44],[45],[46],[47],[49], [52], suggest only specific course or Jobs using time consuming aptitude test.
- 5) The authors found that the Identifying and implementing influencing features from demographic, personal, academic and social area can plays a significant role in the prediction of course stream using various Machine learning algorithms with best accuracy.

III. METHODOLOGY

The following are the main stages in conducting a literature search:

A. Constructing Search Terms

The information below will aid in clarifying the search terms we used to answer our research query. Educational data mining, Educational data mining algorithms, use of data mining algorithms in EDM, students performance prediction, students retentions prediction , carrier guidance, career recommendation system, course recommendation system, predictive model in Emplacement prediction using data mining , ML used in EDM. learners' behavior, Use of Machine learning algorithms in EDM predict drop out ration using ML, Course selection Model using ML. career guidance model.

B. Search Strategy

By determining the educational attribute and algorithm for the career or course section, we created the search terms. We also looked for different keywords and alternatives. In our search phrases, we used boolean operators like AND, or, and not. The pertinent papers were found by searching and sorting through various databases. Table I lists the datasets.

C. Publication Selection

a. Inclusion Criteria

The following factors are used to determine what constitutes appropriate literature, including journal and magazine articles, conference papers, technical reports, books and e-books, early access articles, standards:

- Studies that examined educational characteristics in relation to a application of data mining techniques and tool.

b. Exclusion Criteria

The following parameters were used to weed out any literature that wasn't pertinent for this study.

- Studies those are not pertinent to the research question.
- Studies without descriptions or analyses of data mining algorithms in educational area.

c. Selecting Primary Sources

or this review process, we have chosen 62 papers.

d. Range of Research Papers

The published research from the years 2011 to 2023 is covered by the review of the literature conducted for this study

IV. RESULTS AND DISCUSSION

This paper reports on systematic literature review on application data mining in educational sector and implication of various machine learning algorithms to develop predictive models for taking various decisions. Study analyzed the articles published between 2011 to 2023. Figure 1 shows the articles count from 2011 to 2015 and 2016 to 2023.

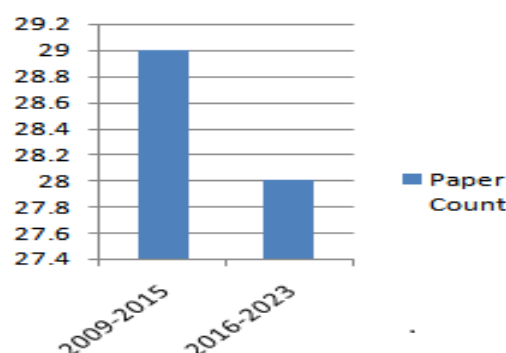


Figure 1: distribution of articles Year Wise

The study aim to explore the objectives of articles, algorithms used and key finding in the articles. The finding of the study shows that most of the researcher showed in the interest in the area such as finding students performance in academic, placement, learning behavior and career recommendation. We studied various articles on career recommendation but not proper study focused on considering various such as demographic, personal, social, academic and economical factors of students for single course

selection from different fields such as Medical, Engineering, Education, Commerce and management, Science, Arts, Fashion Technology and architecture etc.

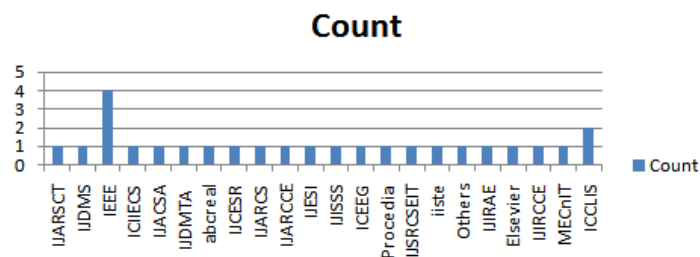


Figure 2: Distribution of articles journal wise

Figure 2 shows the distribution of articles journal wise. various articles are available which shows that most of the researchers now interesting EDM.

Figure 3 shows frequency of Machine Learning algorithms used in various articles which focus on course or career recommendation. Algorithms such as Fuzzy logic, Random forest tree and K-nearest neighbor are mostly considered for the study.

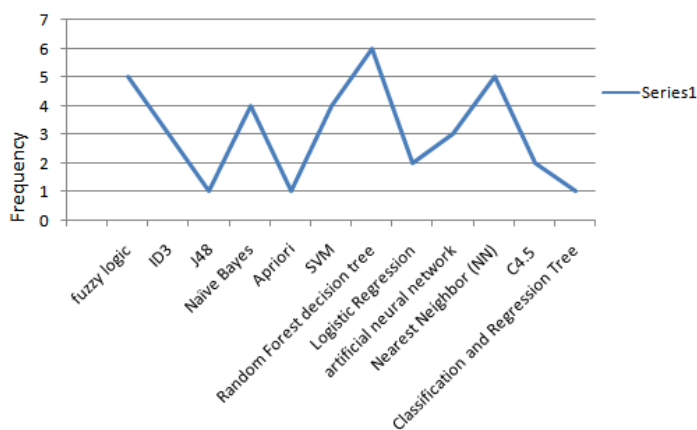


Figure 3: frequency of Machine Learning algorithm

Reference	Problem/ Objective	Algorithm/Method	Key Findings
[14]	Prediction Student Graduation on Time Using Artificial Neural Network on Data Mining Students STMIK Widya Cipta Dharma Samarinda	Artificial Neural Network	To predict the graduation of students on time using artificial neural networks.
[16]	Novel Professional Career prediction and recommendation method for individual through analytics on personal Traits using C4.5 Algorithm	C4.5 Algorithm	proposed prediction and recommendation systems using students psychological condition
[18]	Counselling Guidance Using Big Data Analytics	Big data by using hadoop-hive software	designed the model to select an institution and program or course

[19]	A Proposed Decision Support System/Expert System for Guiding Fresh Students in Selecting a Faculty in Gomal University	Rule based	presented decision support system/expert system for Guiding Fresh Students in Selecting a Faculty in Gomal University
[20]	A Higher Education Predictive Model Using Data Mining Techniques	ID3, K-Means, Naïve Bayes and Random Forest	To predict the career decision for the 12 th passing out students
[23]	Career Counselling Using Data Mining	C5.0 with adaptive boosting and C5.0	proposed system for career option predications
[24]	Analysis Optimization K-Nearest Neighbor Algorithm with Certainty Factor in Determining Student Career	K-Nearest Neighbor and Certainty Factor algorithms	Designed Student Career Prediction System
[25]	A Fuzzy Inference based Decision Support System for Solving the University-Course Admission Choice Problem	Fuzzy Logic	create a decision support system that would help university applicants who were faced with the admissions choice problem
[32]	Predicting Career Using Data Mining	ID3, Classification and Regression Tree, Random Forest, Support Vector Machines, Neural Networks CART and Multi-Layer Perceptron (MLP)	prediction of student's estimated career
[33]	DESIGN OF AN ONLINE EXPERT SYSTEM FOR CAREER GUIDANCE	Rule based knowledge-base	web based Expert System Career Guidance
[34]	Development of a Comprehensive and Integrated Expert System for Career Assessment based on Cognitive models	Cognitive models	designed An expert system which investigate and find an appropriate job related to software career(s) in the software industry for the students
[35]	THE SIGNIFICANCE OF AWARENESS ABOUT SELECTION AND RECRUITMENT PROCESSES IN STUDENTS' CAREER DECISION MAKING	statistical analysis	study aims to determine the value of selection process awareness in students' career decisions

[36]	Scrutinizing Artificial Intelligence based Career Guidance and Counselling Systems: an Appraisal	Artificial Intelligence (AI)-	Scrutinized AI based career guidance systems.
[37]	Career Selection: Role of Parent's Profession, Mass Media and Personal Choice	statistical analysis	find out the effect of parameters such as "parent's profession," "mass media" and "personal choice" on students' career selection personal choice" and "mass media" have a greater influence on students' career decisions than parents' occupations do
[38]	Developing an Intelligent Recommendation System for Course Selection by Students for Graduate Courses	Clustering Technology, Feed-forward back propagation probabilistic neural network and Classification using Fuzzy Logic and Rough Set.	Developed recommendation system for students to select course engineering Stream
[39]	Career Prediction Model Using Data Mining And Linear Classification	Binary classification , Logistic Regression, K-Nearest Neighbors and Stochastic Gradient Descent, random forest	developed a Intuitive Career System to predict the career
[40]	Machine Learning Approach for Student Career assessment in the modern world	Artificial Neural Network	proposed an expert system using a machine learning approach for assessing student's career

Reference	Problem/ Objective	Algorithm/Method	Key Findings
[41]	Expert system for Career Selection: A Classifier Model	ID3 , J48 , IB1 fuzzy logic	designed a framework of expert system for career selection
[42]	Prediction of Student Enrolment Using Data Mining Techniques	Apriori and Naïve Bayes algorithm	predicts Student's Branch Prediction and second is Stream Analysis of engineering course
[43]	Design and implementation of framework for higher	statistical analysis and data visualization	framework for higher education enrollment

	education enrolment through ICT enabled application using JEE		
[44]	A Framework For Analytical Services Using Data Mining Techniques to Predict the Enrollment of Student At A University – A Case Study	Naive Bayes" algorithm	Model to predict whether the student will enroll to university or not.
[45]	Smart Career Guidance and Recommendation System	SVM, Random Forest decision tree, One Hot encoding, XG boost Logistic Regression and Linear Discriminate Analysis	Recommendation System in Choosing a right field in CSE/IT stream
[46]	An Intelligent Career Counselling Bot A System for counseling	artificial intelligence algorithms	Intelligent Career Counseling Bot
[47]	Career Guidance through Multilevel Expert System Using Data Mining Technique	framework for Multilevel Expert System to advice Career path Nearest Neighbor (NN)	framework for Multilevel Expert System to advice Career path Nearest Neighbor (NN)
[48]	Attitude Towards Career Guidance and Counselling Among Higher Secondary School Teachers Under Present Scenario in West Bengal	statistical analysis	Studying the higher secondary school teachers' attitude towards Career Guidance and Counseling. Study concluded that there is no significant difference between attitude with respect to their gender, location of school and subject streams.
[49]	EM&AA: An Algorithm for Predicting the Course Selection by Student in e-Learning Using Data Mining Techniques	Clustering and Apriori Association Rule Algorithm	a framework for Course Recommendation System
[50]	Prediction of Student Enrolment Using Data Mining Techniques	REP Tree algorithm ,neural network	designed a model for branch selection while taking the admission in the engineering

[51]	An incremental ensemble of classifiers as a technique for prediction of student's career choice	Proposed algorithm used to predict the career choice of a student accurately	Naïve Bayes, K-Star and SVM using the majority voting methodology worked together in incremental ensemble
[52]	recommendation System for UiTM Perlis Students Using Fuzzy Logic	a fuzzy logic	a fuzzy logic-based career recommendation system
[53]	Online Career Counsellor System based on Artificial Intelligence: An approach	SVM ,Decision tree	Online Career Counselor System based on Artificial Intelligence

Table 1

V. CONCLUSION AND FUTURE ENHANCEMENTS

In this study, systematic literature review conducted application of education data mining and mainly focused on implementation of machine learning algorithms to course recommendation system.

Total 53 articles between 2011 to 2023 were considered for the review. The finding of the study shows that most of the researcher showed in the interest in the area such as finding students performance in academic, placement, learning behavior and career recommendation. We studied various articles on career recommendation but not proper study focused on considering various such as demographic, personal, social, academic and economical factors of students for single course selection from different fields such as Medical, Engineering, Education, Commerce and management, Science, Arts, Fashion Technology and architecture etc. The development of numerous policies in career guidance and counselling depends on this understanding, which is fundamental and important. All of the proposed methods were discovered to have flaws and were therefore lacking in one or more contexts.

In the future, the author wants to identify influencing factors for career selections and developed predictive model for career selection using same.

REFERENCES

- [1]. Mr. Pradeep Nayak¹ , Mohammed Sufiyan² , Mohan Raju. V³ , Monisha. N. S. ⁴ , Moollya Gautami Bhaskar ,” Review Paper on Educational Data Mining”, nternational Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 3, Issue 2, January 2023
- [2] Rajni Jindal and Malaya Dutta Borah ,” A SURVEY ON EDUCATIONAL DATA MINING AND RESEARCH TRENDS “,International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2013
- [3] Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan,” A Systematic Review on Educational Data Mining” 2169-3536 (c) 2016 IEEE,DOI 10.1109/ACCESS.2017.2654247, IEEE Access
- [4] Krina Parmar, Prof. Dineshkumar Vaghela, Dr Priyanka Sharma(2015),”PERFORMANCE PREDICTION OF STUDENTS USINGDISTRIBUTED DATA MINING”, IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ,ICIIECS'15
- [5].
- [6]. Abdulmohsen Algarni (2016) "Data Mining in Education", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016456 | P a g e www.ijacsa.thesai.org
- [7]. T.Velmurugan, C. Anuradha (2016) “Performance Evaluation of Feature Selection Algorithms in Educational Data Mining”, International Journal of Data Mining Techniques and Applications Volume: 05 Issue: 02 December 2016 Page No.131-139 ISSN: 2278-2419

- [8]. Khoo, K., Ban, T.K., Neng, C.Y., Hooi, B.K. & Joan, C.Y. (2015). Students Choices of choosing Colleges and Course of Study in Penang. ABC Research Alert, 3(1), 1-7. Retrieved from <http://www.abcreal.weebly.com>
- [9]. Kolluru Venkata Nagendra, K.Sreenivas, P.Radhika()STUDENT PERFORMANCE PREDICTION USING DIFFERENT CLASSIFICATION ALGORITHMS”, INTERNATIONAL JOURNAL OF CURRENT ENGINEERING AND SCIENTIFIC RESEARCH (IJCESR), ISSN (PRINT): 2393-8374, (ONLINE): 2394-0697, VOLUME-5, ISSUE-4, 2018
- [10] Bhrigu Kapur, Nakin Ahluwalia,” Comparative Study on Marks Prediction using Data Mining and Classification Algorithms” International Journal of Advanced Research in Computer Science, Volume 8, No. 3, March – April 2017, ISSN No. 0976-56
- [11]. Ditika Bhanushali, Seher Khan, Mohommad Madhia, Shoumik Majumdar(2018),”Student Performance Prediction And Analysis”, International Journal of Advanced Research in Computer and Communication Engineering,ISO 3297:2007 Certified Vol. 7, Issue 3, March 2018
- [12] Selvi.R, Anitha.P, Padmini.P (2016)” A Study on Employability of Engineering Graduates Using Data Mining Techniques”, International Journal of Engineering Science Invention (IJESI) ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726 www.ijesi.org || PP: 12-16
- [13]. Kavita Pabreja (2017),”Comparison of Different Classification Techniques for Educational Data”, International Journal of Information Systems in the Service Sector,Volume 9 • Issue 1 • January-March 2017
- [14] Asep Nurhuda, Dewi Rosita(2017),”Prediction Student Graduation on Time Using Artificial Neural Network on Data Mining Students STMIK Widya Cipta Dharma Samarinda” , *ICEEG '17*, June 14-16, 2017, Turku, Finland. Copyright 2017 ACM 978-1-4503-5248-2/17/06...\$15.00.,DOI: <http://dx.doi.org/10.1145/3108421.3108431>
- [15] Lokesh S. Katore, Bhakti S. Ratnaparkhi, Dr. Jayant S. Umale(2015),”Novel Professional Career prediction and recommendation method for individual through analytics on personal Traits using C4.5 Algorithm”,proceedings of 2015 global conference on communication technologies(GCCT 2015),978-1-4799-8553-1/15/\$ 31.00@ 2015 IEEE
- [16] Zailani Abdullaha*, Tutut Herawanb, Noraziah Ahmadb, Mustafa Mat Derisc “Mining significant association rules from educational data using critical relative support approach”, *Procedia - Social and Behavioral Sciences* 28 (2011) 97 – 101
- [17] Malaya Dutta Borah, Rajni Jindal , Daya Gupta Ganesh Chandra Deka (2011) “Application of knowledge based decision technique to Predict student enrollment decision”, 2011 International Conference on Recent Trends in Information Systems, 978-1-4577-0792-6/11/\$26.00 ©2011 IEEE
- [18] Jinka Thirunarayana (2018)“ Counselling Guidance Using Big Data Analytics”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology © 2018 IJSRCSEIT | Volume 3 | Issue 3 | ISSN : 2456-330.
- [19]. Muhammad Zaheer Aslam*, Nasimullah, Abdur Rashid Khan
A Proposed Decision Support System/Expert System for Guiding Fresh Students in Selecting a Faculty in Gomal University, Pakistan ,Industrial Engineering Letters www.iiste.org ISSN 2224-6096 (Print) ISSN 2225-0581(Online) Vol 1, No.4, 2011
- [20]. Subhalaxmi Panda , P. A Pattanaik , Tripti Swarnkar (2017)” A Higher Education Predictive Model Using Data Mining Techniques”, individual papers by the papers’ authors
- [21]. P.Veeramuthu, R.Periasamy ,” Application of Higher Education System for Predicting Student Using Data mining Techniques”, International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163,Volume 1 Issue 5 (June 2014)
- [22] Esra Mahsereci Karabulut* , Selma Ayşe Özelb , Turgay İbrikçib,c,” A comparative study on the effect of feature selection on classification accuracy” 2212-0173 © 2012 Published by Elsevier Ltd. doi: 10.1016/j.protcy.2012.02.068
- [23] Nikita Gorad, Ishani Zalte, Aishwarya Nandi, Deepali Nayak(2017)”Career Counselling Using Data Mining”, International Journal of Innovative Research in Computer and Communication

Engineering (*An ISO 3297: 2007 Certified Organization*) Vol. 5, Issue 4, April 2017, ISSN(Online): 2320-9801 ISSN (Print): 2320-9798

[24] Nunsina, Tulus, Zakarias Situmorang (2020) “Analysis Optimization K-Nearest Neighbor Algorithm with Certainty Factor in Determining Student Career” 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT)

[25]. Olha Pronina and Olena Piatykov(2021),” The Decision Support System Education Career Choice Using Fuzzy Model”, COLINS-2021: 5th International Conference on Computational Linguistics and Intelligent Systems, April 22–23, 2021, Kharkiv, Ukraine EMAIL: pronina.lelka@gmail.com (O. Pronina); piatykov_o_ye@pstu.edu (O. Piatykov) ORCID: 0000-0001-7085-8027 (O. Pronina); 0000-0002-7731-3051 (O. Piatykov) © 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org)

[26] B.Namratha, Niteesha Sharma(2016) “Educational Data Mining –Applications and Techniques” International Journal of Latest Trends in Engineering and Technology (IJLTET) ISSN: 2278-621X, Vol 7 issue 2 July 2016, <http://dx.doi.org/10.21172/1.72.575>

[27] Dipesh Walte 1 , Hari Reddy 2 , Vivek Ugale 3 , Amol Unwane 4 (2014)Overview of algorithms in Educational DataMining for Higher Education: An Application Perspective, International Journal of Engineering Research & Technology (IJERT),Vol. 3 Issue 2, February – 2014 IJERTIJERT ,ISSN: 2278-01

[28] P. Meena Kumari, SK.Abdul Nabi and Puppal Priyanka(2014),”Educational Data Mining and its role in Educational Field” ,(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 2458-2461,ISSN-NO0975-9646

[29] Saddam Khan, Sunny Gupta, Yogesh Kumar Sharma, Dr. Radhakrishna Rambola (2015) “A Study on Data Mining Techniques and Genetic Algorithm in Education Sector”, IJCSMC, Vol. 4, Issue. 3, March 2015, pg.681 – 683, ISSN 2320–088X

[30]. Nupur Bhagoriya, Priyanka Pande (2017)“EDUCATIONAL DATA MINING IN THE FIELD OF HIGHER EDUCATION-A SURVEY”,IC™ Value: 3.00, ISSN: 2277-9655, April, 2017] Impact Factor: 4.116,IC™ Value: 3.00 CODEN: IJESS7

[31]. Subitha Sivakumar, Sivakumar Venkataraman and Rajalakshmi Selvaraj (2016)“Predictive Modeling of Student Dropout Indicators in Educational Data Mining using Improved Decision Tree”, Indian Journal of Science and Technology, Vol 9(4), DOI: 10.17485/ijst/2016/v9i4/87032, January 2016 ,ISSN (Print) : 0974-6846 , ISSN (Online) : 0974-5645

[32]. Md. Yeasin Arafath, Mohd. Saifuzzaman(2018),”Predicting Career Using Data Mining”, 2018 International Conference on Computing, Power and Communication Technologies (GUCON) Galgotias University, Greater Noida, UP, India. Sep 28-29, 2018

[33]. S. Saraswathi1 , M. Hemanth Kumar Reddy2 , S. Udaya Kumar3 , M. Suraj4 , Sk. Khaja Shafi5,” DESIGN OF AN ONLINE EXPERT SYSTEM FOR CAREER GUIDANCE”, Volume: 03 Special Issue: 07 | May-2014, Available @ <http://www.ijret.org>, IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-730

[34]. V. Chandra Prakash1 , J.K.RSastry2*G.Reeshmika3 , M.Pavani4 , P.Chikitha Sree5 , J.S Ravi Teja6, Development of a Comprehensive and Integrated Expert System for Career Assessment based on Cognitive models International Journal of Emerging Trends in Engineering Research, 7(11), November 2019, 617- 627

[35]. THE SIGNIFICANCE OF AWARENESS ABOUT SELECTION AND RECRUITMENT PROCESSES IN STUDENTS’ CAREER DECISION MAKING, Rashid Ali Shar Baloch Dr. Naimatullah Shah Department of Public Administration, University of Sindh, Jamshoro, Sindh, Pakistan, European Scientific Journal May 2014 edition vol.10, No.14 ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431

- [36]. Scrutinising Artificial Intelligence based Career Guidance and Counselling Systems: an Appraisal, Tehseen Mehraj, 2Asifa Mehraj Baba, International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online) Vol. 7, Issue 1, pp: (402-411), Month: January - March 2019, Available at: www.researchpublish.com
- [37]. Career Selection: Role of Parent's Profession, Mass Media and Personal Choice, Noshina Saleem*, Mian Ahmad Hanan Irfan Saleem and Rao Majid Shamshad Bulletin of Education and Research December 2014, Vol. 36, No. 2 pp. 25-37
- [38]. Grewal DS and Kaur K(2016) "Developing an Intelligent Recommendation System for Course Selection by Students for Graduate Courses", Business and Economics Journal, Bus Eco J 2016, 7:2DOI: 10.4172/2151-6219.1000209
- [39]. Rucha Hemant Rangnekar, Khyati Pradeep Suratwala, Sanjana Krishna, Dr. Sudhir Dhage(2018), "Career Prediction Model Using Data Mining And Linear Classification", 978-1-5386-5257-2/18/\$31.00 2018 IEEE
- [40]. V. Sai kuswanth, Gogineni Krishna Chaitanya, B.Sekhar Babu, Uppuluri Lakshmi Soundharya (2019), "Machine Learning Approach for Student Career assessment in the modern world", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6S, March 2019 699
- [41]. M.L.Waghmode, Jamsandekar(2016), "Expert system for Career Selection: A Classifier Model", Volume 4, Issue 1, January 2016 International Journal of Advance Research in Computer Science and Management Studies, ISSN: 2321-7782, Research Article / Survey Paper / Case Study Available online at: www.ijarcsms.com
- [42]. Heena Sabnani, Mayur More, Prashant Kudale, Prof. Surekha Janrao (2018) "Prediction of Student Enrolment Using Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072.
- [43]. SVinayak Hegde et al 2019 "Design and implementation of framework for higher education enrolment through ICT enabled application using JEE" IOP Conf. Ser.: Mater. Sci. Eng. 561 012120
- [44]. A Framework For Analytical Services Using Data Mining Techniques to Predict the Enrollment of Student At A University – A Case Study
- [45]. J.lakshmi prasanna, DR.D.Haritha (2019) "Smart Career Guidance and Recommendation System", © IJEDR 2019 | Volume 7, Issue 3 | ISSN: 2321-9939
- [46]. Archana Parab, Siddhesh Palkar; Satish Maurya; Sonal Balpande(2017), "An Intelligent Career Counselling Bot A System for counseling", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072
- [47]. Dr. Gufran Ahmad Ansari (2017) "Career Guidance through Multilevel Expert System Using Data Mining Technique" I.J. Information Technology and Computer Science, 2017, 8, 22-29 Published Online August 2017 in MECS (<http://www.mecs-press.org/>) DOI: 10.5815/ijitcs.2017.08.03
- [48]. Suman Kalyan Panja, Kamal Krishna De(2015) "Attitude Towards Career Guidance and Counselling Among Higher Secondary School Teachers Under Present Scenario in West Bengal", Volume : 5 | Issue : 12 | December 2015 | ISSN - 2249-555X
- [49]. Sunita B. Aher (2014), "EM&AA: An Algorithm for Predicting the Course Selection by Student in e-Learning Using Data Mining Techniques", J. Inst. Eng. India Ser. B (January–March 2014) 95(1):43–54, DOI 10.1007/s40031-014-0074-3
- [50]. Heena Sabnani, Mayur More, Prashant Kudale, Prof. Surekha Janrao (2018) "Prediction of Student Enrolment Using Data Mining Techniques", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072.

- [51]. Ms. Roshani Ade, Dr. P. R. Deshmukh (2014),” An incremental ensemble of classifiers as a technique for prediction of student’s career choice”, 978-1-4799-3486-7/14/\$31.00_c 2014 IEEE
- [52]. Career Path R Tajul Rosli Razak¹ , Muhamad Arif Hashim² , Noorfaizalfarid Mohd Noor³ , Iman Hazwam Abd Halim⁴ , Nur Fatin Farihin Shamsul recommendation System for UiTM Perlis Students Using Fuzzy Logic, 978-1-4799-4653-2/14/\$31.00 © 2014 IEEE
- [53] Kartikey Joshi, Amit Kumar Goel, Tapas Kumar 2020, Online Career Counsellor System based on Artificial Intelligence: An approach IEEE 7th International Conference on Smart Structures and Systems ICSSS 2020

मध्य भारती
मानविकी एवं समाजविज्ञान की द्विभाषी शोध-पत्रिका



MADHYA BHARTI
(UGC CARE Group-1, Multi disciplinary)

CERTIFICATE OF PUBLICATION

This is to certify that the article entitled

**APPLICATION OF DATA MINING TECHNIQUES AND ALGORITHMS IN EDUCATIONAL SECTOR:
A SYSTEMATIC LITERATURE REVIEW**

Authored By

Prof. RAHUL PATIL

Abhinav Education Society's Institute of Management & Research, Pune.

University Grants Commission

Published in

Madhya Bharti -Humanities and Social Sciences

: ISSN **0974-0066** with IF=6.28

Vol. 84, No. 29, January - June : 2023

UGC Care Approved, Group I, Peer Reviewed, Bilingual, Biannual, Multi-
disciplinary Referred Journal



Chief Editor
प्रो. अश्विकादित्य शर्मा