Decision Support System Model for Student Performance Detection using Machine Learning

Prashant Dixit  
Research Scholar,  
Sangam University

Dr. Harish Nagar  
Associate Professor,  
Sangam University

Prof. Sarvottam Dixit  
Dean(E),  
Mewar University

Abstract: Any educational institution's primary aim is to give students the highest educational opportunity and skills. To accomplish the purpose is necessary to recognize students who require additional assistance and take effective steps to increase their results. Educational programs require creative ways of improving school efficiency in order to produce the best outcomes and lower the risk of failure. Educational Decision Support System (EDSS) has recently boomed in education systems, enabling pupil outcome monitoring and assessment to be carried out in progress. Inadequate information systems face difficulty and obstacles to profit adequately from EDSS due to the lack of precision, improper study of the characteristic characteristics and inadequate database. A detailed literature review and collection of the highest predictive methodology are very critical for improving the prediction process. Machine learning methods were used for this study in order to construct a classifier that can forecast students' success in the economic field. This paper propose a Knowledge base DSS model to evaluate to evaluate the student performance via mid-term exam and final-term exam using machine learning techniques.

Keywords: Educational Knowledge Base System (EKBS), Educational Expert System, Artificial Neural Network, Logistic Regression, Support Vector Machine

1.1 INTRODUCTION
Enhancing student success and standard of education at all educational institutes is of vital importance. A profound study of the learners' prior history will play a crucial role in delivering quality education to students. EDSS provides study and optimization of student outcome prediction approaches. Based on the prediction findings, the average outcome and success will improve for both mid and final stages if the student needs are met in time. Important characteristics and prior data of students are obtained for the objectives of performance review and estimation. In order to get a better understanding and prediction, various machine learning methods and classification algorithms are applied later. The aim of the EDSS is to decrease the failure rate, strengthen the education environment and assess the core features and take the achievement of students into account. It also helps us to build predictive models that are useful to forecast results. It not only helps to improve students at risk directly, but also offers information and insights on preparing the academic process for the next year. Different machine-learning types of algorithms such as Naïve Bayes, decision tree, neural networks, detections by outliers and sophisticated statistics have in recent years been used. These strategies are used to obtain information on the student information, to support decision support systems, and to extract trends, etc. The academic achievement of the student is usually calculated via previous intermediate exams, but there are other important attributes which influence the student's overall performance. Several methodological and statistical research experiments on student databases have recently been published. The Bays and judgment being used by Kabakchieva and his colleagues [1]. In Bayes and decision-making classifications, Kabakchieva and his colleagues [1] predicted student success on pre-university results. Any of the other suggested literary approaches are based on neural networks, computational models and ID3 [2-7] protocols.

1.2 LITERATURE SURVEY
The student success estimation using and machine learning methods for the advancement of the new world are a wide range of research available. Osmanbegović and his collaborators in the [2, 3] project are contrasting various data mining methods with the Bayesian Classification Network, and with the aid of decision-making bodies. Classification has been solved by the neural network. It offers different methods for model precision and is recognized by the Bayesian classifiers and compares estimated prediction functions. The work has been focused on data sets provided by Tuzla students and economic personnel, and their data sets have been used for this purpose between 2010 and 2011. The weka program is used; multiple algorithms offer varying results; average is the best solution to take the results into account. MLP, NB and J48 yield good results for this reason. This document does not take account of student backgrounds and characteristics. For reliable and consistent data, further tests are essential. In the proposal for a system, Marquez-Vera and his colleagues, and Sergi Rovira [4, 5] have employed real figures from six hundred seventy Zacatecas students, Mexico school and Barcelona University results. They used computer study, whitebox classification system, algorithms for decision tree and inference rules. Three studies in hypotheses were performed to detect the loss ratio and the school dropout rate. They used the approach of feature collection, only 15 of the 77 qualities that are deemed

IJERTV10IS050017 www.ijert.org

(This work is licensed under a Creative Commons Attribution 4.0 International License.)
the best of education systems were considered. This work is also carried out using dimensional simulation and computational techniques. Weka Tool is used and findings for improved comprehension are not seen in a visual format. Successful resolution of the data-unbalanced crisis. To forecast the success of the student on the basis of previous college and individual qualities. Two rules for learning were used to evaluate data in Kabakchieva and his colleagues[1] by data mining algorithms. Of which two major categories, a decision-taker and the next neighboring category is employed. Weka is used for deployment, data are taken into account from 10330 students and 20 criteria. Weka filters are used by an algorithm in data sets of J48 and JRip reliably and excellent performance. The kNN classification and Bayes classification systems do not work. The techniques ID3 have been used as a separate approach suggested by Ahmed and his colleagues[6] for prediction of student results, ranking methods, clustering, artificial intelligence, neural networks, regression, related rules, generics and decision tree. Any characteristics have been obtained to forecast student outcomes. Data compilation of 1547 documents used for success estimation for these purposes. Weka is used in the implementation of the decision tree. The characteristics such as attendance, temperamental and environmental conditions are not monitored. Another approach presented is based on longitudinal information obtained from the evidence from the Gwinnett County Public Schools, which was made by Tamhane and his collaborators[7] in class 8 evaluation in Math and science. The strategies for implementing missed principles include imputation of means, the logistics regression, decision-making trees and the naive Bayes. They use weka and SPSS models and are taking into account many demographic variables. Logistic regression offers the best outcomes from all these approaches. There was a problem with the process of filling missed values. It makes predictions noisy, and for students' whole data one classification is not enough. The classification should be a viable choice for providing the student's optimum risk predictions. The behavioural details and causes of registration should also be taken into consideration. A neural network approach has been used to forecast student marks by Arora and her collaborators[17]. For mapping Gaussian method the inputs are based on a radial equation. In order to prediction the current marks, the knowledge relating to grade details for students between 2010 and 2011 and 2011 is useful. The record comprises around two hundred and fifty subjects, consisting of more than thousand student details.

The data sets provide scholarly, educational, enrollment and personal information during studies. Weka data mining toolkit is hired. This paper serves to develop a classification model to check student achievement; the variable of gender, GPA, credit and assessment score are also used as a significant feature. Using data sets from higher education institutions is more effective to obtain a blueprint for students' results. Artificial intelligence methods are better used to produce better outcomes. In a system to track the student learning process, Grivokostopoulou and his colleagues[20] employed semantic the rules; It helps to devise students' results. Semantic laws and ontologies are also used to enhance schooling, to have benefits and to read. For adaptive learning a number of artificial intelligence techniques are used. A powerful decision tree technique is used to react to the artificial intelligence lost or passed percentages beforehand; c4.5 and k means algorithm, the removal of classification rules and clustering is applied. In this way, students should choose the field of concern for the best subject, as this platform is not too effective. Lime survey was used, with very smaller participants in this survey around 277. Due to these shortcomings, the student performance cannot be found successfully. There was another better alternative. The NB tree classification model is used by Christian and his colleagues to estimate the success of the student[20].

The thesis focused on the creation of a computer-based prediction method was proposed in Mgala and his colleagues [24]. A model was based on a broad dataset from graduates eight, who took part in the Kenya primary school exam diploma. The total number of students was 2426. The average fee is used to process lost values. In preprocessing steps filters have been used for prediction model construction in feature selection technology, machine learning techniques and data mining algorithms. The logistic regression, multilayer perceptron, minimum optimization sequentially algorithm, Bayesian network classification schemes, naïve Bayes classificatory, random forest classification and J48 algorithms are used. The datasets were taken from rural areas, as the characteristics are based on rural variables and cannot yield efficient effects on urban education systems. In certain cases, as in urban student analyzes, college or university students, this approach does not provide the right results.
Student personal descriptions and familial spending data are deemed to be very useful in figuring out how they do in advance [8]. The classification should be a viable choice for providing the student's optimum risk predictions. The behavioral details and causes of registration should also be taken into consideration. A neural network approach has been used to forecast student marks by Arora and her collaborators [17]. For mapping Gaussian method the inputs are based on a radial equation. In order to prediction the current marks, the knowledge relating to grade details for students between 2010 and 2011 and 2011 is useful. The record comprises around two hundred and fifty subjects, consisting of more than thousand student details.

1.3 EDUCATIONAL KNOWLEDGE BASE

In this knowledge base enrichment of all the activities of student related information.

1.3.1 Machine Learning

This section introduces a concise overview of the methods of machine learning used in this study. Computer systems were absolutely given new skills in machine learning. Data from multiple sources sometimes referred to as training data was obtained. How good this model can be determined by the consistency and quantity of data obtained. Data for machine learning training should be prepared to clean up the data, randomize it and simulate the relationship between the concentrations[27]. The strength of the machine learning is to forecast responses, using an automatic learning algorithm, rather than a human assessment, in order to create a model that is stronger than a basic principle.

1.3.1.1 Artificial Neural Networks:- The Artificial Neural Network describes the sequence of balanced ties between input and output units. In order to predict the correct goal label in certain input cases, ANN learns by adjusting the weight of connections[28]. Back propagation algorithm is one of the well-known learning algorithms used for training the ANN. ANN has many benefits, such as its high noisy data tolerance and its good efficiency in classification designs which has not been trained on, so it is used where the relationship between the class mark and the data set has no information.

1.3.1.2 Logistic Regression:- Logistic regression provides a mechanism of mathematical modeling that explains the connection between many explanatory variables, X1...Xk, and a variable dependent, D. The logistic model uses the logistics function as a statistical structure with an input scale from 0 to 1. The logistic model will explain the likelihood of an occurrence that is always approximately 0 and 1[29].

1.3.1.3 Naïve Bayes:- The classification paradigm of Naïve Bayes is known to be Bayesians simplest variety. This model implies that each attributes of the function is separate from other properties of the destination. Per x instance in the dataset includes a1, a2,... ai. F(x) is equivalent to any value of the predetermined V= (v1, v2, ... vn) set [30].

1.3.1.4 Decision tree: It was used in several applications in real term, such as financial research, surgery, molecular biology, processing and astronomer. The technique requires a selection measure for the attribute or function to pick the attributes or to split best the training presented instances into different target groups during the construction of the decision tree. The information benefit, the profit ratio and the Gini index are such metrics. ID3, CART, and C4.5 are common decision trees[31].

1.3.2 Educational Decision support System

When we understand that education is one of the world's most pressing problems. The vision is an all-encompassing state. Decision making may be characterized as an unbelievably necessary duty for managers to obtain and scrutinize the relevant knowledge regarding management problems in order to ensure that alternate options of behavior are selected as needed. All higher education administrators need to be aware of the decision-making process because the education industry, like all prescribed bodies, has a decision-making system in general. Decision aspects in all areas face complex situations, excessive weight of data, data transmitted within the organization, risk and unstable conditions. Powered training institutions or universities feel the need to strengthen instruments and resources in persuasive decision-making, to quickly and definitely support them in any management framework. Any informational systems have long had judgment backup modules, but in the broadest context have extra money and administrative data surveys. One of the first initiatives to establish a DSS in advanced instruction is to create valid conceptual instruments/academic instruments for gathering, organizing and analyzing imperative data and knowledge for the decision-making process [24].

1.3.3 Educational Knowledge management

The dynamic system of components, processes and relationships between human capital, facilities, technologies, information and knowledge staff is educational knowledge management. This partnerships are in a 'environment' where information management behavior can be encouraged or hindered. An systematic and evidence-based methodology is the prevalent approach to the study of organizational km, focusing on information & technology. Systemic methodology simultaneously explores and studies the cultural influences within the enterprise and knowledge management system – humanist, techno-structural and awareness and environment. A formal description of KM is therefore required. The management of knowledge is, generally, "an integrated, regular, targeted and ongoing socio-technical function to promote, capture, gain, generate, organize, store, retrieve, share, distribute, transfer, reuse and evaluate experiences, knowledge and assets (tacit and explicit) to achieve competitive advantage by improving the quality of knowledge and knowledge."
1.3.4 Educational Dialog management
This work in the field of graphics, interaction styles, software tools, dialog systems, language translation software and "widgets" gui was important to pave the way for realistic programmers to construct interaction designs in computers.

Dialog is...
• synonym for communicating, talking or talking
• a characteristic human accomplishments
• an end in itself,
• a medium for educational education
• including nonverbal, human contact and multimodal interactions
• a platform for the continuing co-construction of new definitions
• a logistically and politically confined educational dialogue, curriculum and appraisal framework

1.3.5 Educational decision management
Decision-making is considered to be the most critical management process. Policy-making challenges, like curriculum administration, are very popular in many fields. The bulk of decisions taken in the educational issue are made from some very simple knowledge or from an intuitive viewpoint. Decisions are a mechanism from which a variety of options to reach the desired outcome are chosen.

1.3.6 Educational expert system
An expert system is a software computer that tries to function like a human being in a certain field. It uses a human knowledge base to resolve issues or to explain uncertainty, where one or more human experts should usually be consulted. Specialist systems based on expertise, or only specialist systems use human knowledge to address topics which are usually human ingenuity.

1.3.6.1 User Interface:- All communicate with the system in an integrated interface in which user requests are generated about a specific topic and the system is then advised to compute and settle on a question response or guidance. The uniqueness of the features enables clients to query how, why and what layout.

1.3.6.2 Knowledge base:- In specific experts are required to do analysis and gather evidence in the field of expert concepts in areas where they fall out of supply to complete a knowledge base within an expert's scheme. If the system matures, current regulations can be introduced and old ones can be changed by using the acquisition module or suppressed from the knowledge base. All these are usually carried out to account for improved technologies with inventions, developments and discoveries.

1.3.6.3 An inference engine:- This works as a search engine, which searches at the base of data that corresponds to the demand of the customer. The deduction engine uses the question to look up the knowledge base and gives the user a response or advice.

1.3.6.4 Knowledge engineering:- Knowledge engineering is an informatics scientist with the knowledge to develop and execute artificial intelligence techniques systems. The Domain Expertise Province is the concentration of expertise in a mission area. Domain knowledge includes both systematic knowledge in the textbooks and experiential information – expertise.

1.4 TRAINING DATA SETS
New dataset is used for building proposed model. Table 1: The Dataset description. For training we have 268 both (MT and FT) instances of economics students’ data with their academic performance from mid-term and final-term yearly exams (type A), courses related exam, number of question, evaluate correct and wrong question and scoring the result presented information in table 1.
Table 1 Description of Data Sets

<table>
<thead>
<tr>
<th>Features</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>UID</td>
<td>Student Unique Identification (Roll No)</td>
</tr>
<tr>
<td>Name</td>
<td>Name of student</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, Female</td>
</tr>
<tr>
<td>Course</td>
<td>Course ID</td>
</tr>
<tr>
<td>Subject</td>
<td>Subject name and subject code</td>
</tr>
<tr>
<td>Exam duration</td>
<td>Mid-term (MT), Final-term (FT)</td>
</tr>
<tr>
<td>Question</td>
<td>Number of Question</td>
</tr>
<tr>
<td>Correct</td>
<td>Correct answer</td>
</tr>
<tr>
<td>Wrong</td>
<td>Wrong answer</td>
</tr>
<tr>
<td>Total score</td>
<td>Gain percentage</td>
</tr>
<tr>
<td>Psychometric Factor</td>
<td>To evaluate the exam stresses</td>
</tr>
</tbody>
</table>

Algorithms based parameter: The following are the algorithms based parameter:

a. Accuracy: describes the correctness of value
b. Probability Threshold: Presents the True Positive and True Negative rates.
c. Execution Time: Time of running the algorithm on dataset.
d. Precision: (number of true positive)/(number of true positive + False positives)
e. Recall: [(number of true positive)/(number of true positive/ number of false negatives)]
f. Number/Size of Rules

1.5 EXPERIMENT

1.5.1 Student Performance flow chart

Predicting the academic success of students predicts how well a student or community does a research work. It will offer informed instruction, recommendations and early input to enhance the student's acquisition of information, systematic appraisal and educational satisfaction. It may also be used to find students at risk who are likely to influence examination results. In the process of defining optimal student curriculum paths and possible replacement of structured exams, a successful and reliably predictive stage may be used to reduce the test burden and workload that impact teacher and student both adversely. Furthermore, learning success prediction will assist universities in selecting the best individuals from the newcomer pool. Quality estimation may also be used in order, for classes, courses and even individual courses, to have more knowledgeable instructions and guaranteed decisions, the most likely to optimize the ability of students and minimize considerably the likelihood of errors that can affect continuing education encouragement in all levels substantially or continuously.

1.5.2 Sample Data

Machine learning techniques have been used to build a classifier that can predict the performance of the students in an economics subject that is offered by Orta Dogu Teknik Universitesi (https://data.mendeley.com). Table 2 consists of training data set of ten students. Input data for the classifiers are provided from CSV file includes student’s marks from yearly exam, data regarding obtained marks in mid-term and final-term, exam type, student ID, gender, course and scoring percentage marks presented details. Table II: The sample input data. Every input data is in alphanumeric form.
1.5.3 Analysis
The highest objective of study is to predict the stresses of mid-term and final-term exams. We applied over a test data set consisting of (268 Final term+268 mid-term) student records. The results show that, among (268 Final term+268 mid-term) students, 68 students have more than 26% chances of being same performance in both term exam. Similarly, 100 students are having more than 37% chances to good performance in MT compare than FT and remaining student are having more than 37% chance to good performance in vice-versa.

In this graph we analysis 25 question attempted by student some are correct and some are wrong but the calculate percentage based on correct questions. And, student 1 got good marks in MT in compare FT, Student 2 got as same as student 1, Student 3 got good marks in FT in compare MT as same as student 4 but student 5 got equal marks in both the terms (FT and MT). It means according to graph analysis student very confident in MT exam but they do not do expected performance in FT vice-versa. And some student as maintain exam respect to perform equally and try to work hard to improve their parameters.

CONCLUSION
It is also critical that the student success is predicted beforehand. After comprehensive research, we concluded that various student datasets give different outcomes with various qualities. The scope of this research is to foresee the prospect of an undergraduate exhibition and structure that is an exceptional alert in the context of advanced learning. The scope of this analysis is to examine the consistency of assumptions about the likelihood of an understudy role and assess the presentation of this instance of knowledge mining methods. In the KDSS to analysis the actual student performance to take better decision and to evaluate any type of examination process improve the efficiency of student to reduce the examination phobia. Expert system play an important role in to solve student related problem superbly any type of gap between two types of examination to remove the exam pressure they ready for new struggles. In future work with the help of machine learning techniques according to data sets to set MT exam tuff this is for practicing exam and some toughness reduce ET exam.

REFERENCES


