



“Predicting Student Academic Performance of learning in Educational Data Mining Paradigm using KNN-GA”

Anand Tiwari¹, Dr.Dayashankar Pandey²

tiwarianand9131@gmail.com¹, dayashankar.rkdfist@gmail.com²

¹M Tech Scholar Dept.of Information Technology , RKDF IST SRK UNIVERSITY, BHOPAL

² HOD, Dept.of Information Technolog., RKDF IST SRK UNIVERSITY, BHOPAL

How to Cite this Article:

Tiwari, A. (2026). Predicting Student Academic Performance of learning in Educational Data Mining Paradigm using KNN-GA. International Journal of Creative and Open Research in Engineering and Management, <i>02</i>(7), 1-9. <https://doi.org/10.55041/ijcope.v2i7.214>

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.



<https://doi.org/10.55041/ijcope.v2i7.214>

Abstract : The performance of students in higher education is a most challenge task day by day in academic as well as in other curricular activities. Today's everybody can find the huge amount of data stored in the web technology, but they don't relies that what data is suitable? As they all know that internet technology is growing as much as faster, but the learning approach of students are not up to the mark. The emerging research community which helps to find the solution to the said problem is Educational Data Mining. In present scenario, the huge students' data is stored in educational database. That type of database contains widely open or secret information to improve student performance. In our proposed work, we have tested it on reputed dataset, which has been downloaded from a well known organization UCI repository and dataset name is student-mat.csv. This work has investigated the process of classification of plethora of students' data. The work has been divided into two parts. The first part covers the entropy based feature selection, after that classification process has been performed. For the classification, we have used the KNN-GA method and we observed that the accuracy of proposed method is too good when compared with some previous methods. The accuracy of the proposed method was found to be more than 91% whereas the existing method has approx 81% accuracy.

Keywords : Educational data mining , Student Learning Outcome (SLO), Decision tree, SVM, K-Nearest Neighbor, Genetic Algorithm. Moodle; learning management system (LMS); educational data mining (EDM); synchronous and asynchronous learning; student engagement; personalized learning; higher education; machine learning algorithms; online learning platforms

1. Introduction

Educational Data mining is used to recognize, extract and estimate the variables related to the student's learning process. Educational data Mining can be explained as “An emerging discipline, concerned with developing methods to better understand students, and the settings which they learn in”[28]. In EDM various techniques can be used to identify the most effective factor to determine a scores of test attempted by student, and then adjusting these factors to improve the performance of the students. EDM is an interdisciplinary research field inheriting the concepts from various other research areas like Psychometrics, Learning Analytics, Artificial Intelligence, Information Technology, Machine Learning, Statistics, DBMS, Computing and Data Mining as shown in figure 1.1 which provides intrinsic knowledge of teaching and learning process for effective education. Each field contributes a lot in the enhancement of learning process[2].

EDM is a data driven decision making practice which analyze the data of students' in current educational system and make the effective changes for the improvisation of the students' performance.



1.1 Goals of EDM

The goals of educational data mining for the researchers are as follows[3]:-

- Developing a Student Model by studying detailed information of a student such as knowledge, skills, motivation, meta-cognition, attitudes, experiences, satisfaction, learning process and styles and preferences. Anticipate students' future learning performance and outcomes based on data from course activities.
- Analyzing the behavior of learner who guide the student properly.
- Communicate with the stake holders frequently.
- Producing or improving the domain models that characterize the optimal instructional sequences.
- Studying the effects of different kinds pedagogical support that can be provided by learning software.
- Implementation of scientific knowledge which is developed through the

computational models and incorporates the model of the student, the domain and the software pedagogy.

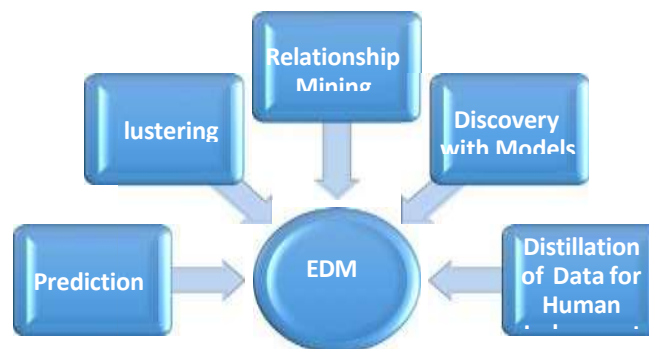


Figure 1.1: Methods to achieve EDM Goals

Prediction implies developing a model that can deduce a single aspect of the predicted variable from some combination of the predictor variable. It assures in developing domain models, such as connecting procedures or facts with the practice session and sequence that gives the optimum teaching procedure and estimates and understand the student learning outcomes which can be achieved by analyzing the success rate on post tests after tutoring [7]. This method has two key uses in educational data mining [8]. One of widely used approach in predicting student educational outcomes without considering intermediate factors first the essential features of the model will be considered which gives information about the underlying construct. Another type of usage is to predict the output value and to evaluate its usage in particular context and to obtain a label.

Generally there are three types of prediction[8]: classification, regression, and density estimation. In *classification*, the predicted variable is a binary or categorical variable. Some commonly used classification methods are logistic regression, decision trees and support vector machines, Bayesian classification. In *regression*, the predicted variable is a continuous variable. Some common regression methods linear regression, neural networks, and support vector machine regression. In *density estimation*, the variable which is to be predicted is a probability density function. Density estimators can be evaluated on different kernel functions, including Gaussian. Bayesian classification provides practical learning algorithms and prior knowledge and includes observed data. It evaluates explicit probabilities for hypothesis considered and is robust to noise in input data.

Clustering refers to the process of forming a group of abstract objects into classes of similar objects. In EDM the goal is to find data points that naturally group together, splitting the complete data set into separate set of clusters. Clustering can be used in discovery of new student behavior patterns and Investigating similarities and dissimilarities between the different institutes. This process postulate that each data point must belong to exactly one cluster or more than one cluster or to no clusters[8].

Relationship Mining refers to developing the relationships between the variables in a dataset and express them as rules. There are usually four types of relationship mining exists such as correlation mining, association rule mining, sequential pattern mining and casual data mining. Association rule mining focuses to discover if-then rules of the form when few set of variables are known and another variable tends to a specific value. Correlation mining targets to find linear correlation between the variables. Sequential



pattern mining is intended to find the temporal associations between the events.

The key use of relationship mining is to discover the association between performance of student and the sequences of course and to seek the most effective pedagogical strategies for robust learning.[8]

1.2 EDM Components



Figure 1.2: Components of EDM

The figure 1.2 shows the key components of EDM are: Stakeholder, Environment, Data, Methods-Tools and Techniques, Task [9]. Each component can be explained as follows:-

1.2.1 EDM stakeholder can be divided into three groups:-

Primary group Students(learners) and Faculties(teachers, learners, educators etc) lies under this category as they are directly involved with the teaching and learning.

Secondary group Alumni and Parents comes under this category as they directly deals with the development of foundation.

Hybrid group Employers, Administrators and Educational planners and experts forms the hybrid group and they deals with the administrative and decision making.

1.2.2 EDM environment can be of three types :-

Formal Environment The situation, when there is direct interaction with primary group stakeholder like in a face to face classroom.

Informal Environment When there is indirect interaction with primary group stakeholder like in web based learning, online supported task.

Computer Supported Environment When there is direct or indirect interaction with all types of stakeholder of education.

1.2.3 EDM Data

In education sector there is massive amounts of data of various universities, schools, and learning development centres in both structured and unstructured format. This type of data can be generated in digital and physical formats and can be uncertain. Thus EDM data can categorized in three ways:

Digital data The data generated during the distance education training of the stakeholder, web based instruction, computer supported collaborative learning used in social networking sites and online group meeting space.

Physical data The data produced during the interaction with the stakeholder in traditional and modern classrooms, interactive learning environments, students' attendance, course information, additional data collected from the academic department of an institution of the learners.

Uncertain data When data is generated either with measurement or scientifically like sensor generated data or summarization of data etc.



Discovery Oriented These methods can be constructed on models discovered with prediction, Classification, Relationship mining, Neural Network, Web mining etc.

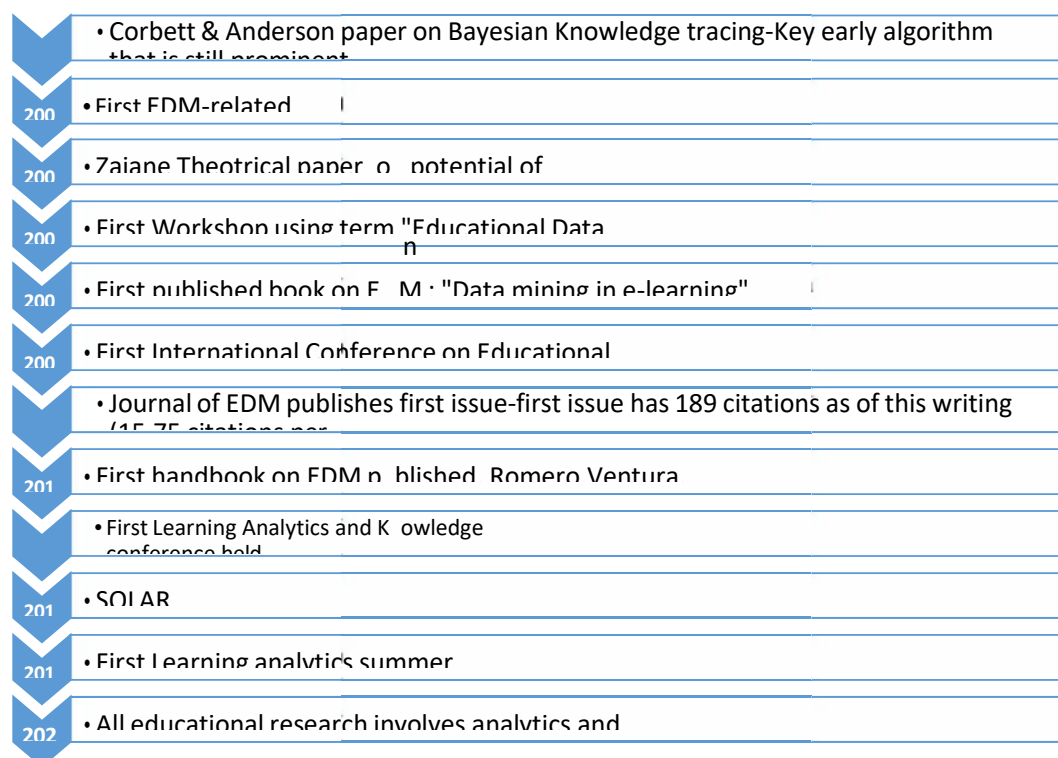
1.2.4 EDM Tools

There are various visualised automated analytical tools which are currently used in Data mining technique as:

- Gephi: It is an interactive visualisation and possesses the exploration platform. It is specifically used for network and dynamic and hierarchical graphs.
- Graphviz: As the name suggested, it uses the graph formation for representing the diagrams of network.
- KNIME: It has a data exploration platform, modular in nature, which explores data flows and examines results.
- Pajek: It can easily visualize, which possesses visualization tool mainly used for social network analysis as it is capable in handling the complex network data.
- R: It is an environment and a programming language which provides the facility of data manipulation, calculation, visualization etc.
- RapidMiner: It is a tool meant for Machine Learning (ML) and used to create data flows.
- VisuaLinks: It is a graphical analysis tool, when heterogeneous data is generated, this tool discovers patterns and hidden networks.
- VizTree: It is a Visual Data Mining system for the time series analysis using tree visualization and interaction.
- Weka: It is a collection of Machine learning and Data mining algorithm meant for analytical process which allows users to create pipelines to visualize.
- XmdvTool: This tool is interactive which is meant for visual exploration of multivariate data sets.
- Matlab: It is an environment and programming language which enables users to perform computationally intensive tasks.

2. Literature Review

EDM, when seen as research community is found to be related to learning analytics. When EDM is viewed as an area of scientific inquiry, is concerned with the analysis of large scale educational data and emphasis on automated methods. EDM was first emerged in workshop series starting in 2005, which turned into an annual conference in 2008 and created a journal in 2009, the International Educational Data Mining society in 2011. The timeline of key events in the formation of the EDM community



has been stated in figure 2.1. **Figure 2.1: History of EDM**



● Ryan S.J.d. Baker (2010) [8] has discussed Data Mining for education in USA where he mentioned the advantages over the Traditional Educational Research Paradigms and introduced the primary categories of educational data mining like prediction, clustering, relationship mining, discovery with models, and distillation of data for human judgement, the former three belongs to the types of data mining whereas the latter two are specifically meant for educational data mining.

In this literature, the main applications of EDM were elaborated such as improvement of current student models and prior information about the learning and learners. He also suggested the method to determine the relative effectiveness of various types of learning meant for students. The various categories of method with its goal and key application of each method has been stated. He also discussed the two data repositories such as PSLC DataShop (one of the largest repository in the world for learning interaction data) and the National Center for Education Statistics (NCES) which provides a foundation for educational data mining highly feasible.

● Pooja Gulati and Dr. Archana Sharma (2012) [26] had expressed their views on Educational Data Mining for Improving Educational Quality in which they focuses on improvement of functional activities of education through learners, educators and the arrangement of the classes. According to the authors, Educational Data mining can provide assistance in the design of the educational content like schedule of the classes and the organization of the way of the teaching classes, to improve the learning process and help the learners in taking decision for their academics.

In this literature, the author emphasize on increasing the number of courses as the area of interest for various students also varies. Different prediction models

● Nidhi Chopra and Manohar Lal (2012) [15] had presented their views on Research Methodology for Educational Data Mining in India. According to the author there can be two possible dimensions of EDM research: One is from educators point of view and other from Administrators/Managerial point of view.

There are various educational data repositories which can be used by the educators such as PSLC (Pittsburgh Science of Learning Centre, CMU(Carnegie Mellon Univeristy) etc. These repositories uses various tools such as:

DataShop: It provides some public or access on demand datasets. Datashop have predesigned methods such as Performance profiler, Error report, Learning curve, WEKA, RapidMiner, SPSS, R and Matlab.

CTAT (Cognitive Tutor Authoring Tools): It is a method to prepare intelligent tutors. The tutors can be of two types: example-tracing tutors and cognitive tutors.

TagHelper: The tool used to apply machine learning technologies to text processing problems.

English LearnLab Data: It is a online data search system for the student of various courses of PSLC.

Tutalk: This tool is used in tutoring and learning research, for authoring and experimenting with natural language dialogue.

EDM for administrators can be viewed by collecting the raw data and preprocess those data to obtain the interesting patterns. Then research methodology for the analysis of educational dataset will be applied.

Siti Khadijah Mohamad, Zaidatun Tasir (2013) [27] presented a review literature on EDM. In this literature, the study is made on the role of data mining in an online learning environment. The authors preferred clustering algorithm to predict the student learning behavior. They also focused on technical aspects of e-learning where they analyzed the module of features which influenced the students' learning process with the data mining analysis.

The tools which make use of e-learning such as WebCT, Moodle and Blackboard made a great development in the educational data mining. The current scenario is more inclined towards the use of social networking tools such as blog, facebook etc. which will gradually replace the e-learning application in order to gather log data. The resultant log data will be forwarded to further process of Educational data mining.

● Rajini Jindal and Malaya Dutta Borah (2013) [9] presented the survey on educational data mining and research trends. In this literature, components, related tools, techniques, educational outcomes and research trends were discussed.

EDM objectives can be viewed in two ways: Administrative objectives and Academic objectives. Academic objectives can be at three levels-Person oriented, Department oriented and Domain oriented whereas Administrative objectives deals with higher authorities perspective.

EDM components, EDM environment, EDM data, EDM tools, and various EDM methods were discussed. The EDM methods proposed were Classification, Statistics, Clustering, Prediction, Neural network, Association rule mining and web mining. The cross faculty can also be appointed. This problem can be solved by Association rule mining.



● Mr. Suhas G. Kulkarni et al. (2013) [13] has understood EDM from different perspective. According to him, the data is hierarchical in educational sector which can be placed in the following manner

Figure 2.2: Data at Different Levels



The author explained different types of classrooms which have different ways of learning and possess distinct data sources for respective objectives like traditional classroom emphasize on face to face teaching, distance education which provides education to the students who are separated by time and space, Learning Content Management System offers wide variety of course content through channels and workspaces to achieve information sharing and communication, Adaptive and intelligent web-based educational system provides an access to the content uploaded on the web approach.

The author explained different types of classrooms which have different ways of learning and possess distinct data sources for respective objectives like traditional classroom emphasize on face to face teaching, distance education which provides education to the students who are separated by time and space, Learning Content Management System offers wide variety of course content through channels and workspaces to achieve information sharing and communication, Adaptive and intelligent web-based educational system provides an access to the content uploaded on the web approach. The aspirations from the EDM were discussed, in which the researcher wants EDM to consider all the social and organizational aspects of education, EDM must consider all the inputs of digital artefacts that is fed into the learning systems. It must abide by the fundamentals of quality data analysis and EDM hopes to provide the information of the various effective pedagogies and outlier detection of the learning system.

3. Edm and Student Learning Outcomes Techniques

3.1 Introduction

Educational data mining is concerned with extraction of hidden information from the crude data gathered from educational system. Educational data mining consist of four phases:

Phase 1: The aim of this phase is to develop the relationship between the data and these relationships can be utilized in the various data mining techniques like prediction, classification, clustering etc. for the further processing.

Phase 2: The objective of this phase is to validate the relationships obtained in the previous phase so that the problem of over fitting can be overcome.

Phase 3: In this phase, the prediction is made for the further processing, over the validated model obtained in the learning environment.

Phase 4: The last phase supports the decision making process using the predictions.

There are distinct and different data mining techniques used in order to achieve the following objectives:

- To discover the students' reaction to a particular pedagogical strategy and group those students according to the similar characteristics persists.
- To detect students' game-playing.
- To find out the misconceptions possessed by students and group those students, who are failure-driven.
- To recognize the students having low motivation and discover some actions to lower the drop-out rates.
- To classify the student, according to the usage of intelligent tutoring system.



.Genetic Algorithm

Genetic algorithm was proposed by John Holland. These algorithms can be considered as the general model of the adaptive processes but the basic objective of this algorithm is optimization. The population of candidate solutions are involved on which series of genetic operators acts to discover new, hopefully better solutions to the particular problem abiding by the rule of natural evolution. At each step, the algorithm randomly selects individuals from the current population, who becomes parent, and produces the offspring for the next generation. The population can be generated in the form of tree and can be termed as "generation tree". The different solutions are applied to the different categories of the problems and then the all the solutions for the specific problems is assigned a reward score so that each solution can be compared with one another for attaining the optimal solution This optimal solution is used in the generation of the next branch of the tree. This solution is inspired by the Darwin's theory of evolution, which states that "the survival of the fittest". All the solutions evaluated has to undergo a process called the fitness function to measure the appropriateness of solution generated by Genetic Algorithm .In general, the solutions generated by the genetic algorithm is termed as chromosomes. These chromosomes are composed of genes whose values can be either numerical, binary, characters or symbols which undergoes the various process like crossover and mutation to produce new offspring.

The difference between Genetic Algorithm and Classical Algorithm can be summarized as follows:

Table 3.1: Difference between GA and Classical Algorithm

S.No.	Genetic Algorithm	Classical Algorithm
1.	At each iteration, the set of points is generated which can be termed as population. The optimal solution can be approximated by selecting the best point in the population.	At each iteration, a single point is generated. The sequence of points approximates the optimal solution.
2.	Random number generators select the next population.	Deterministic computation is used to select the next point in the sequence.

4. Proposed Methodology

4.1 Introduction

This section illustrates the proposed methodology used for feature selection and simultaneously classification approach is also discussed. At the initial stage, GAIN based technique is applied in the raw dataset of student-mat.csv (Math course) and student-por.csv(Portuguese language course), dataset is downloaded from <https://archive.ics.uci.edu/ml/datasets/Student+Performance>, that is particularly used for feature selection. After completion of the process, the other proposed classification approach, named as K-NN with genetic algorithm (KNN-GA) is used to classify the data according to the assigned classes and concurrently evaluates the error rate. Details of the methods are describes in below selections:

4.2 K- Nearest Neighbor (KNN)

This algorithm is one of the most effective algorithm for the classification, and also it is used for prediction and estimation process. KNN follows a non-parametric technique which is also termed as instance based learning. It means that, this process does not make any prior assumptions for the data distribution and not even obeys any rules of generalization for the training dataset. It adapts a very simple and feasible process to undergo a classification process, where it is compared to the most similar records in the training dataset.



This process is also known as lazy learning, as it prolongs the decisions to generalize beyond the training examples till a new query is encountered. K-NN algorithm is one of the easiest to comprehend and performs the following operation :

- Determines the group of k-objects in the training dataset which are closest to the test object.
- It provides the basis for the assignment of a label on the dominance of the particular class in the neighborhood.

4.3 Genetic Algorithm

Genetic Algorithm is adaptive heuristic search algorithm for the implementation of evolutionary ideas of natural selection and genetics. Genetic algorithm is a problem solver which imitates the process of biological progression. Genetic algorithm are applicable to the situations as follows:

- The search space is complex, large or inexpressive.
- Domain knowledge is inadequate or expert knowledge is difficult to encode to narrow the search.
- Unavailability of mathematical analysis.
- Failure of traditional search method.

A simple genetic algorithm follows the same general rule which are:

- Initial population can be generated by selecting the random population of solutions.
- The estimation of fitness of the solution within population is determined.
- Creation of offspring is done by the selection of pairs of solution, based on their fitness and are added to the next generation of the population.
- Mutation, the genetic operator can be applied to the offspring.

Thus, this process can be elaborated as Genetic algorithm, which comprises of the components such as population generator, selector and a fitness estimator. Addition to this, three more genetic operators such as selection, mutation and crossover can be used in this algorithm. The crossover operator generates two off springs, by combining parts of the species of two individuals which in turn replaces the low fitness individuals in the population. Probably, the mutation operator inverts randomly chosen bit(s). After producing the number of generations, the search process will be terminated. For a given specific problem, the input can be deciphered, which is a set of encoded potential solutions to that problem and a metric called a fitness function is used that allows individual candidate for a quantitative evaluation. These randomly generated candidates are termed as solutions, with the motive of enhancement. The algorithm evaluates each candidate according to the fitness function. Generally, in the collection of randomly generated candidates, the most promising candidates, which contributes to the solution of the problem are chosen. The multiple copies, with random changes, are generated from these candidate solutions

5. Results and Discussion

5.1 Introduction

This chapter describes experimental result of implementation process. The comparison between proposed methodology with the existing one has been made using SVM based method and partial ID3 algorithm. This chapter also includes used dataset description and the tools used for simulation purpose.

5.2 Tools used

The tool used for the analysis purpose of the proposed methodology is MATLAB 2012A 64 bit. The system configuration is 8GB RAM, dual core 2.5Ghz I5 processor and windows 8 64 bit operating system.

5.2.1 Outline of MATLAB

MATLAB is an acronym for MAtrix LABoratory. It was originated by Cleve Moler with J.H. Wilkinson, George Forsythe and John Todd. MATLAB is a software package which provides interactive environment and has been developed to deal with various mathematical computations. It uses matrices and vectors for the calculation purpose. It is widely used in engineering, mathematics and research areas. The basic data element for MATLAB is array or a matrix which makes the computation reliable and easier. The availability of wide variety of toolboxes, commands and functions makes it more sophisticated.

MATLAB supports object-oriented programming and platform independence. It can be considered as a powerful "scalar" and "matrix" calculator and at the same time very simple high-level programming language. It facilitates graphic commands which makes the visualization of results possible. This package is helpful for control theory, symbolic computation, signal processing,



optimization and simulation and various other fields of engineering and applied sciences.[24]

There are two kinds of file supported by MATLAB are:[23]

Scripts files: These files are program files having extension ".m". In these files the series of MATLAB commands are composed in one file which can be executed all at once, thus this code is generally initialization code. These files redirect data as input, present on workspace as they cannot get inputs and cannot exhibit outputs therefore it contains certain templates for some callbacks, to control GUI behavior.

5.3 Dataset

For the experimental purpose, the conventional dataset "Student Performance Dataset" has been used. The source of this dataset is <https://archive.ics.uci.edu/ml/datasets/Student+Performance> which contains the performance of student enrolled in secondary education. The description of the used dataset is as follows

Table 5.1: Dataset Information

Dataset Characteristics	Multivariate	Number of Attributes	33
Attribute Characteristics	Integer	Area	Social
Associated Tasks	Classification, Regression	Date Donated	27-11-2014
Number of Instances	649	Number of Web Hits	67207

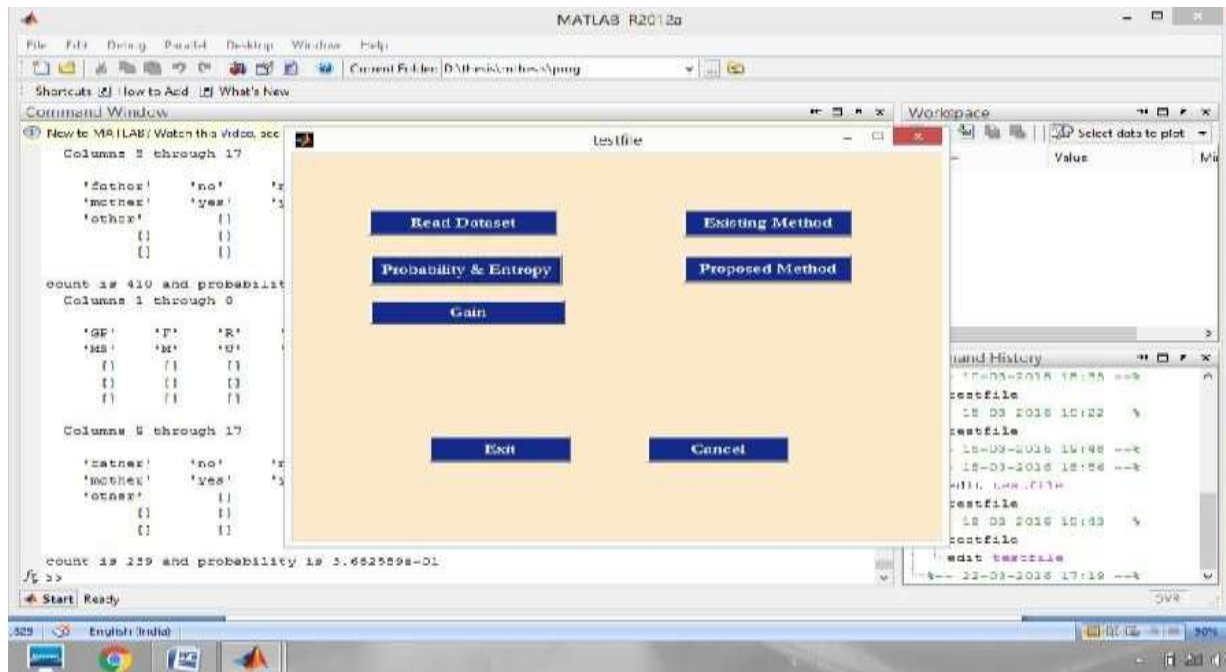


Figure 5.2: Main GUI Environment

6. Conclusion&Future Work

This work is intended to develop a novel approach for the student learning system, and the method is successfully tested on the very reputed data set, that has been downloaded from UCI repository namely student-mat.csv. While observing the downloaded dataset, it has total 32 attributes including numeric and string sets, and contains various information of the different type of student. Here one more attribute has been added before pre-processing the data and which resulted into dataset of total 33



attributes. So in the first step we have solved data selection challenges, by using the entropy based proposed methodology. Here process takes little more time for pre-processing but at the end when we observed the result outcomes, then it was found that the classification accuracy is much better than existing method. After feature normalization and selection, we performed different classification process in normalized selected dataset, and we observed that the SVM based method produced 92.25% accuracy with 4.96 error rate, but in the our proposed scheme we found more than 96% accuracy in every test cases with minimum error rate approx 2.5. This was made possible by applying KNN based approach and for the quality improvement of knn classification we used genetic approach, which has been improved the classification accuracy of the method.

In future the work can be made more efficient, like pre-processing and data selection process can be enhanced with the help of other methodology. Here we observed that the proposed methodology is producing excellent compared result, but takes more pre-processing time if data instances are increases. So in future the pre-processing time can be reduce and the method can be tested for the high instances datasets.

7. REFERENCES

- [1]. Hina Gulati; " Predictive Analytics Using Data Mining Technique";IEEE, 2nd International Conference on Computing for Sustainable Global Development (INDIACom) 978-9-3805-4416-8, pp 1-4, 2025.
- [2]Romero, C., and Ventura S.; " Data Mining in Education"; WIREs Data Mining and Knowledge Discovery., Vol. 3, pp.12-27, 2024.
- [3]Sedigah Abbasnasab Saradeh, Mohd. Rashid Mohd. Saad, Abdul Jalil Othman , RosalamChe Me; "Enhancing Education Quality Using Educational Data"; Scholars Journal of Arts, Humanities and Social Sciences, 2 -3B, pp 440-444, 2024.
- [4]Baker, Romero And Ventura;"Developing Effective Educational Experiences through Learning"; Proceedings of 12th International Conference on Artificial Intelligence in Education, pp330-335. 2010, 2023
- [6]George Siemens, Ryan S.J.d Baker;"Learning Analytics and Educational Data Mining: Towards communication and collaboration"; Proceedings of the 2nd International Conference on Learning Analytics and Knowledge; 978-1-4503-1111-3, pp 252-254; 2022.
- [7]Baker, Gowda, and Corbett; "Artificial Intelligence in Education"; Proceedings of 4th International Conference on Educational Data Mining; pp 179-188; 2021
- [8]Ryan S.J.d. Baker, "DATA MINING FOR EDUCATION", Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, International Encyclopaedia of Education; 3rd Edition, pp 1-10, 2019.
- [9]Rajni Jindal, Malaya Dutta Borah;"A Survey on Educational Data Mining and Research Trends";International Journal of Database Management Systems (IJDMS) Vol.5, No.3, June 2018.
- [10] Dr. S. U. Maheshwari, K.S. Divyaa;"A Study on Educational Data Mining"; International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Vol. 4 Issue 10, October 2018.
- [11] Jan Geryk;"Applied Visual Analytics for Higher Education Decision Making";Masaryk University faculty of informatics; pp 8-9; January 2018.
- [12] Rebecca Cartwright, Ken Weiner, and Samantha Streamer-Veneruso;"Student Learning Outcomes Assessment Handbook"; pp 100-170; Revised may 2018.
- [13] Mr.Suhas G. Kulkarni, Mr.Ganesh C. Rampure, Mr.Bhagwat Yadav ;"Understanding Educational Data Mining";International Journal of Electronics and Computer Science Engineering;2277-1956,Vol.2,pp 773-777, 2018.