



**International Journal of Intelligent Enterprise**

ISSN online: 1745-3240 - ISSN print: 1745-3232

<https://www.inderscience.com/ijie>

---

**Analysis of dependence of grade point average over psychological factors incorporating advanced data mining**

Leena Khanna, Shailendra Narayan Singh, Rita Kumar, Mansaf Alam

**DOI:** [10.1504/IJIE.2023.10051667](https://doi.org/10.1504/IJIE.2023.10051667)

**Article History:**

Received:	19 September 2018
Last revised:	12 November 2019
Accepted:	17 February 2020
Published online:	30 November 2022

---

## Analysis of dependence of grade point average over psychological factors incorporating advanced data mining

---

Leena Khanna\*, Shailendra Narayan Singh  
and Rita Kumar

ASET,  
Amity University Uttar Pradesh, India  
Email: leena09.khanna@gmail.com  
Email: Snsingh36@amity.edu  
Email: rkumar16@amity.edu  
\*Corresponding author

Mansaf Alam

Department of CS,  
JMI New Delhi, India  
Email: malam2@jmi.ac.in

**Abstract:** Academic performance of any individual is dependent upon numerous aspects regarding the day to day life of the individual under consideration. Academic performance is measured in terms of the grade point average or GPA as it is called. Grade point average is dependent not only on the faculty but also on various cognitive and non-cognitive factors. The non-cognitive or the psychological parameters include factors like study habits, social anxiety, extraversion and allied concepts. In this study, a detail analysis of numerous psychological factors that can impact the grade point is studied. Bases on the various psychological factors studied, the performance of an individual is classified and in forthcoming examination can be forecasted.

**Keywords:** forecasting; grade point average; achievement motivation scale; study habits; extraversion; conscientiousness; social anxiety; emotional competence; self-efficacy for children.

**Reference** to this paper should be made as follows: Khanna, L., Singh, S.N., Kumar, R. and Alam, M. (2023) 'Analysis of dependence of grade point average over psychological factors incorporating advanced data mining', *Int. J. Intelligent Enterprise*, Vol. 10, No. 1, pp.31–47.

**Biographical notes:** Leena Khanna received her MTech in Computer Science from Amity University. She has both industrial and academic experience. Her areas of interest include software engineering, data mining and data warehouse.

Shailendra Narayan Singh is a Professor of Computer Science and Engineering having experience in teaching and administration. He received his BTech, MTech, PhD in the field of Computer Science and Engineering and his research area includes software reusability for aspect oriented and object-oriented software. He has more than 15 years of teaching experience in the academic field of computer science and engineering in Indian and abroad universities.

Rita Kumar completed her PhD in Psychology in 1988 from St. John's College. She is currently working as a Professor in Amity Institute of Psychology and Allied Sciences. She has a rich experience of more than 20 years and worked in different institution as a lecturer, counsellor, and vice principal. Her areas of interest include motivational psychology, emotional intelligence, career counselling. She has more than 20 research papers in national and international journals of repute.

Mansaf Alam received his Doctoral in Computer Science from Jamia Millia Islamia New Delhi in 2009. He is currently working as an Associate Professor in the Department of Computer Science, Jamia Millia Islamia. He is also the Editor-in-Chief, *Journal of Applied Information Science*. He is in editorial board of some reputed *International Journals in Computer Sciences* and has published about 50 research papers. His areas of research include cloud database management system (CDBMS), object-oriented database system (OODBMS), genetic programming, bioinformatics, image processing, information retrieval and data mining.

---

## 1 Introduction

In the modern era, the outcome of academic initiatives which are initiated by parents or the academic institutes depends upon various factors. These factors vary a great extent ranging depending upon the psychological condition of an individual. A change in any of the psychological indicator can increase or decrease the performance of a student in any examination. This in turn will have impact on the career of the student as well as the annual result of the institution. Thus, making it of a prime concern for any education institute to map the psychological wellbeing of a student to his performance to the grade point average (GPA) of the student.

The individual differences which have an impact on the grade point equivalent have been extensively researched. There had been numerous studies which tried to find out the various psychological parameters impacting the outcome on grade point average. One highly important study was done where it was found that there are rational and realistic ground that the academic performance of an individual can be forecasted using the various personality traits of the individual.

In the present condition, where the dynamics of the global job markets are changing every day, the requirement of quality work force coming from higher educational institutions is extremely high. With the advent of sophisticated AI techniques and especially, the advancement in the field of artificial intelligence has resulted in the various tools being available for analysing the psychological status of a student and thus, depicting the impact of this over his grade point average.

The research conducted by various academic institutions in the field of data mining has contributed a great deal of knowledge in the domain of forecasting technologies. The utilisation of the any data mining technique is based upon the fact that the data holds essential features which can be utilised for predicting the future with a certain amount of certainty and thus, assist in decision making.

In this paper, we put present a method which can analyse the impact various psychological factors impacting the performance of a student in higher education institutions. The design of system incorporates the use of various psychometric scales in

harmony with different techniques from data mining to generate the grade point average outcome of the student. The inference system depicts the approach taken by an expert clinical psychologist. The final forecast generated by the system is designed to help the academicians to help the institutes in focusing upon the individual students who because of various psychological parameters have a decreased performance in the examination and thus, increasing the overall institution results as well as help students fight against various psychological issues and aid in the counselling of students.

The paper is structured in seven sections. In Section 2, we are presenting the literature review of the domain in consideration. In Section 3, we're explaining various psychological scales which are used in this study. In Sections 4 and 5, the various data mining techniques and research methodology is discussed, respectively. Finally in Section 6, we put forward the result of the study. In Section 7, we conclude the study.

## **2 Literature review**

There is widespread use of data mining in the higher education system. Work has been carried by numerous researchers in this field. They have gone through various aspects of data mining in the domain of education. In this section we present a holistic survey of various psychological as well as data mining techniques as existing in the research domain and the methodology followed to complete the literature survey.

### *2.1 Literature review methodology*

The framework for conducting the literature review is separated in three parts: scope of survey, identification of research literature and Reporting the literature of interest and is depicted in Figure 1.

In part 1, we have defined the problem under consideration. This phase has been divided into three phases. In this phase, the study starts by describing the area of the research which is then followed by the formulation a research goal describing the process of the survey for the literature review. This process ends up by producing a unique statement of problem which becomes the basis upon which the future work will be conducted. The domain of research happens in this study to be the psychological parameters, the data mining techniques and their intersection over higher educational domain. The aim of this research work is to design a method in which can be utilised by higher educational institutions to help students perform better in academics.

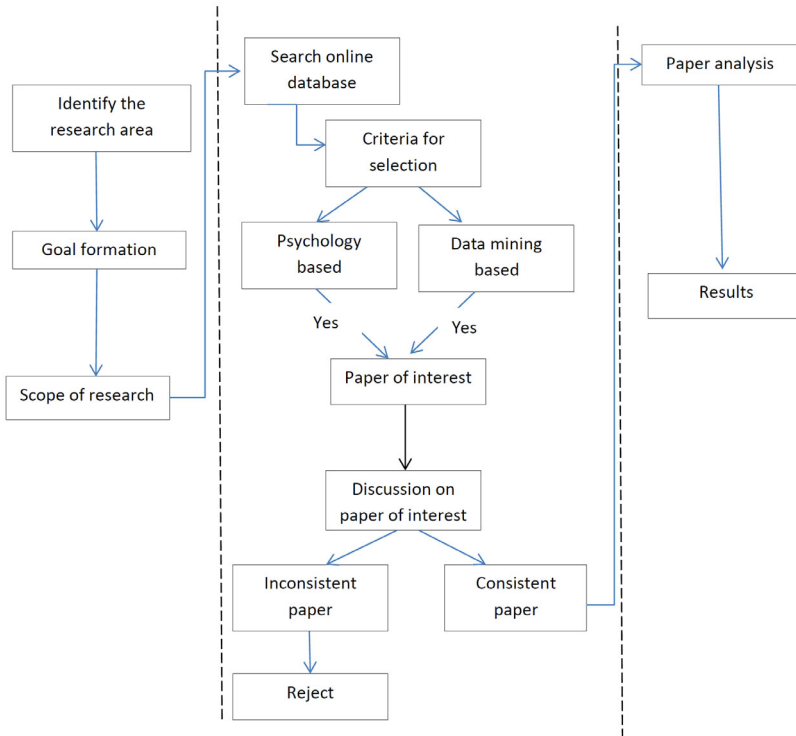
In part 2, we have framed the conditions for choosing of research literature pertaining to the domain of study. Seven online databases were examined to figure out the details for work being done in domain of study. These databases are:

- Elsevier
- IEEE
- Taylor & Francis
- ACM
- Oxford

- Science Direct
- Springer.

The research databases were searched using the keywords ‘forecasting’, ‘grade point average’, ‘achievement motivation scale’, ‘study habits’, ‘extraversion’, ‘conscientiousness’, ‘social anxiety’, ‘emotional competence’, ‘self-efficacy for children’. To search for different permutations of the keywords Boolean operators were used.

**Figure 1** Procedure for literature review (see online version for colours)



## 2.2 Classification of under research literature

This section provides a holistic view over the research literature being considered in this study. A total 44 research papers were consulted during this study. This review of this research literature is classified in this section owing to the findings, year as well as the major data mining technique which has been used in paper. These are as follows:

### 2.2.1 Classification of research articles based upon the major finding

This section presents the major finding/designed developed mapped to the research literature in Table 1.

**Table 1** Classification based upon key findings

<i>S. no.</i>	<i>Reference number</i>	<i>Major findings</i>
1	Sharma et al. (2016)	Method for literature survey
2	Saranya et al. (2014)	Representation of institutional growth perspective based upon the performance of students.
3	Cairns et al. (2014)	Analysed the students data and generated clusters of strong, medium and weak performance
4	Arora and Badal (2014)	Detected the set of weak students
5	Osmanbegović and Suljić (2012)	Designed a predicting model for performance in academics
6	Sukanya et al. (2012)	Detected the outlier of low academic achievers
7	He and Zhang (2011)	Designed a comprehensive style of student evaluation
8	Sakurai et al. (2011)	Forecasted the chances of success of curricula
9	Aher and Lobo (2011)	Student performance analysis
10	Ayesha et al. (2010)	Detected the outlier of weak students
11	Kovacic (2010)	Mapped attributes of enrolment with the academic success
12	Al-shargabi and Nusari (2010)	Examined the academic accomplishment of students to their financial growth
13	Gao (2010)	Drop out forecasting
14	Knauf et al. (2010)	Analysed e-learning success rate
15	Wu et al. (2010)	Designed a system to classify the grades
16	Bian and Gong (2010)	Evaluated the effectiveness of study in students
17	Liu and Zhang (2010)	Designed prediction model to predict the performance on the basis of behaviour.
18	Zhu et al. (2009)	Forecasted achievements of students
19	Shi et al. (2009)	Improvised apriori to mine data tables
20	Ramasubramanian et al. (2009)	Detected students weak in studies
21	Selmoune and Alimazighi (2008)	Found out the factors impacting the success as well as failure among students
22	Dai and Zhang (2008)	Forecasted the final grade depending upon the historical performance
23	Hien and Haddawy (2007)	Socio-economic variables can influence the academic outcome from students
24	Al-Radaideh et al. (2006)	Designed system to predict the final grades
25	Rasmani and Shen (2005)	Academic performance classification
26	Delavari et al. (2005)	Outlier detection of weak students
27	Salazar et al. (2004)	Gave the parameters for the analysis of performance in academic institutes
28	Minaei-Bidgoli and Punch (2003)	forecasted the final grade
29	Kotsiantis et al. (2002)	Predicted the performance of an examination

**Table 1** Classification based upon key findings (continued)

<i>S. no.</i>	<i>Reference number</i>	<i>Major findings</i>
30	Sharma and Mavani (2011)	Comparison of algorithms in terms of forecasting accuracy
31	Fadzilah and Abdoulha (2009)	Comparison of algorithms in terms of forecasting accuracy
32	Wook et al. (2009)	Comparison of algorithms in terms of forecasting accuracy and concluded that classifier is best for classification
33	Romero et al. (2008)	Comparison of algorithms in terms of forecasting accuracy
34	Bakar et al. (2013)	Graduates lack critical thinking, team work and analytical approach
35	Jantawan and Tsai (2013)	Comparison of algorithms in terms of forecasting accuracy
36	Dejaeger et al. (2012)	Designed a satisfaction criterion for a course under consideration
37	Srimani and Patil (2012)	Evaluation of faculty staff
38	Yadav and Pal (2012)	System for assistance in enrolment
39	Pandey and Pal (2011)	Found out that multi-lingual class is more suited than unilingual class
40	Balakrishnan (2010)	Forecasted the learning disabilities
41	Ahmed et al. (2009)	Patterns in matching associations
42	Pumpuang et al. (2008)	Predicted the sequence for course registration
43	Bresfelean (2007)	Analysis of continuation of education
44	Aksenova et al. (2006)	Prediction of enrolment head count

### *2.2.2 Classification of research articles based upon the data mining technique used*

To figure out the key algorithms utilised for forecasting in the domain of higher educational research, we have done an analysis of higher education and the data mining techniques. Eleven data mining techniques have been utilised in the analysis of higher educational domain. It is clear from this analysis presented in Table 2 that decision tree was the most widely used technique in this domain followed by the Naive Bayesian algorithm.

### *2.2.3 Classification of research articles based upon the year of publication*

Figure 2 depicts the spread of articles referred in this study as per their year of publication.

**Table 2** Classification based upon the technique used

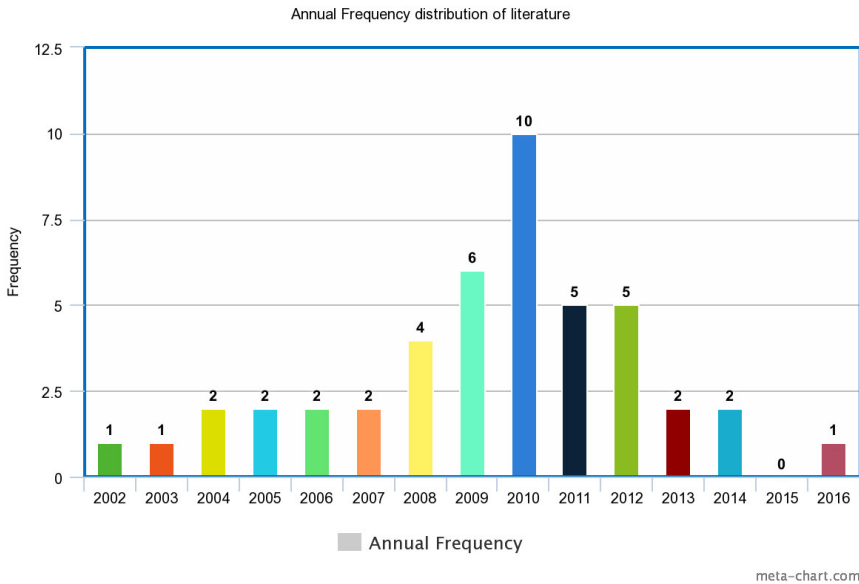
<i>S. no.</i>	<i>Data mining technique</i>	<i>Reference number</i>
1	Naive Bayes	Saranya et al. (2014), Osmanbegović and Suljić (2012), Sukanya et al. (2012), Hien and Haddawy (2007), Al-Radaideh et al. (2006), Kotsiantis et al. (2002), Sharma and Mavani (2011), Jantawan and Tsai (2013), Pumpuang et al. (2008), Talavera and Gaudioso (2004)
2	K-means clustering	Cairns et al. (2014), Aher and Lobo (2011), Ayesha et al. (2010), Al-shargabi and Nusari (2010), Salazar et al. (2004), Fadzilah and Abdoulha (2009), Wook et al. (2009), Bakar et al. (2013)
3	Association rule mining	Arora and Badal (2014), He and Zhang (2011), Al-shargabi and Nusari (2010), Zhu et al. (2009), Selmoune and Alimazighi (2008), Pandey and Pal (2011), Ahmed et al. (2009)
4	Decision trees	Osmanbegović and Suljić (2012), Sakurai et al. (2011), Kovacic (2010), Al-shargabi and Nusari (2010), Knauf et al. (2010), Wu et al. (2010), Bian and Gong (2010), Liu and Zhang (2010), Al-Radaideh et al. (2006), Delavari et al. (2005), Salazar et al. (2004), Kotsiantis et al. (2002), Sharma and Mavani (2011), Fadzilah and Abdoulha (2009), Wook et al. (2009), Romero et al. (2008), Bakar et al. (2013), Jantawan and Tsai (2013), Dejaeger et al. (2012), Yadav and Pal (2012), Balakrishnan (2010), Pumpuang et al. (2008), Bresfelean (2007)
5	Artificial neural networks	Aher and Lobo (2011), Gao (2010), Kotsiantis et al. (2002), Wook et al. (2009), Romero et al. (2008), Dejaeger et al. (2012), Srimani and Patil (2012)
6	Rough set theory	Gao (2010), Ramasubramanian et al. (2009), Al-Radaideh et al. (2006)
7	Apriori algorithm	Shi et al. (2009)
8	Genetic algorithms	Dai and Zhang (2008), Minaei-Bidgoli and Punch (2003)
9	Fuzzy logic	Rasmani and Shen (2005), Romero et al. (2008)
10	Support vector machine	Dejaeger et al. (2012)
11	Random forest	Srimani and Patil (2012), Aksenova et al. (2006)

### 2.3 Methodology and analysis

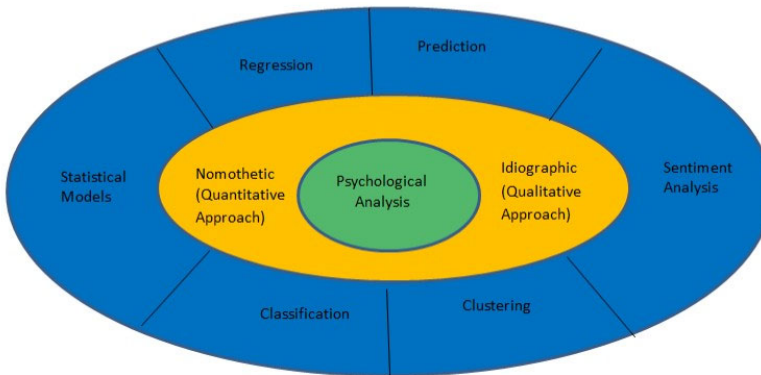
In the history of human civilisation forecasting has been done since ages. In this section, we classify the existing work in the domain of consideration into broad themes of the data mining and its applicable techniques. The classification as depicted in Figure 3 shows the interaction of psychological approach with the data mining technique. Psychological analysis is a highly interdependent part thus sits in the centre of analysis.



**Figure 2** Classification based upon the year of publication (see online version for colours)



**Figure 3** Forecasting framework of data mining for psychological analysis (see online version for colours)



The proposed overview of data mining for psychological analysis is based upon the theoretical literature available in the domain under consideration.

### 2.3.2 Psychological analysis

In psychology the analysis is done in two different ways which are as follows:

#### 2.3.2.1 Nomothetic (quantitative approach)

The word ‘nomothetic’ is derived from the Greek word ‘nomos’ which translates to law. This is utilised by the psychologists while trying to study what the subject under consideration shares with other people around him. This approach is used for the analysis

and interpretation of data gathered by an investigator incorporating the use of inferential as well as descriptive statistics. This is utilised while incorporating the use of a psychological test.

### *2.3.2.2 Idiographic (qualitative approach)*

The word ‘idiographic’ is derived from the Greek word ‘idios’ which means ‘private’. Psychologists who are interested in this aspect of study seek the experience of each individual which makes them unique. This approach is dependent upon the qualitative methods of collecting the data such as diaries, interview and other allied written materials which have either been obtained from the subject under consideration or is expected to have been belonging to the same.

## **3 Psychological scales**

In this study we are using a total of seven psychological scales were used which depict various attributes of personality. These were selected on the basis of various personality traits which were under consideration and had shown influence over academic performance in the research literature. These personality traits are as follows.

### *3.1 Achievement motivation*

It is a type of psychological test designed to evaluate a wide set of work-related achievement motivation. In this study, it is being used to find out the motivation to achieve a higher competence in academic performance.

### *3.2 Study habits*

These are the sum total of all the practices that are utilised by any individual to learn any aspect of life or subject. Study skills are mostly defined as the ability of the student to manage all the resources including time to fulfil an academic task with high amount of certainty. Thus, study habits happen to be the different forms of routines which are followed by a student during any period in a conducive environment.

### *3.3 Extraversion*

It is defined as a state in which one obtains gratification not from the inside but from outside oneself. People with this trait enjoy human communications and tend to be talkative, enthusiastic as well as assertive. These people are always energised and take pleasure in activities that incorporate large number of people. These people also tend to work best in a group. In this study, Extraversion as defined in L.R. IPIP scale measures in utilised.

### *3.4 Conscientiousness*

It is the personality trait which depicts how careful/vigilant a person is. The word ‘conscientiousness’ itself means an aspiration to perform a task well and take into serious

consideration the obligations by others. People with this personality trait tend to perform more in terms of efficiency and are highly organised. They depict an inclination to display self-discipline, dutiful acts as well as target for achievement in their respective field. They display a planned behaviour and are generally reliable. In this study, the conscientiousness as defined in Goldberg IPIP scale measures is used.

### *3.5 Social anxiety*

Social anxiety is the nervousness faced by some people in social situations. Individuals which are in social anxiety depict behaviour of avoiding interaction in form of gazes, lesser number of facial expressions as well as difficulty in starting and maintaining conversations. Most of the people have faced social anxiety at some point of time in their life. In this study, social anxiety scale named as SIAS is used.

### *3.6 Emotional competence*

Emotional Competence is one's capability to show his/her inner feelings to others in a comprehensive form. This shows the easiness of a person when around other people to express. It is one of the essential social skills to identify, understand and react to emotions of yourself as well as others around you. In this study, the emotional competence scale is utilised.

### *3.7 Self-efficiency*

Self-efficiency is the belief of an individual in his/her inborn abilities to achieve the desired goals. Psychologists have been studying self-efficiency from numerous perspectives. But in this study, we stick to the utilisation of self-efficiency to the domain of achievement in higher education. To deduce this self-efficiency questionnaire SEQ-C was utilised.

## **4 Data mining techniques**

Data mining is the analysis of existing data to find out new information out it. Usually, it is done over tremendous amount of data to figure out the required patterns of interest. The large-scale implementation information technology has grown leap and bounds generating different systems for transactions and different ones for analytics. Data mining provides the bridge between the two. In this study, as eminent from the literature review, we will be using the following methods to analyse the data. These methods are:

### *4.1 Naïve Bayes classifier*

Naïve Bayes classifier belongs to the family of simple probabilistic classifier which utilises the Bayes theorem and implements a strong assumption of independence among the features being classified. These are highly saleable classifiers which need a linear number of variables to the predictors in the learning problem.

## 4.2 *Decision trees*

As evident from the literature review, decision tree is the most commonly used technique in this domain. A decision tree happens to be a decision support tool which implements a graph structure and its leaf nodes depict the outcome and the intermediate nodes represents the state in which the machine is presently is. It follows a sequence of decisions to conclude a possible outcome. It happens to be one of the ways to depict an algorithm which has conditional control statements.

## 4.3 *K-means clustering*

It is a method of quantising the input vectors and is quite popular in the domain of cluster analysis. The aim of k-means is to divide the number of observations say  $n$  into a particular number of clusters say  $k$  and the observations belong to one of the  $k$  clusters depending upon their mean. The partitions generated in this process are known as Voronoi cells. As per Table 2, it has been widely used in the research literature.

## 4.4 *Artificial neural networks*

These are a set of computational methods which are inspired from the biological neural networks which make up the brain of animal brains. These are designed to learn the anticipated behaviour from the examples which are provided to them. These are deployed in used in the research literature for their power to forecast unknown conditions based upon previous data. In this study, back propagation algorithm is used as the training method for feed forward network.

# 5 **Methodology**

This research is carried out in three phases. In the first phase, the various psychological questionnaires were arranged in a random order and were distributed in 14 institutions of higher education for data collection from graduate and post-graduate courses. In phase 2, the information was processed to give it a digital shape and noise was removed. In phase 3, these were analysed using the data mining methods as mentioned in Section 4 of this paper and present the achieve results. In details, these phases are as follows:

## 5.1 *Data collection*

This was one of the initial phases of the research. Once the psychometric scales and the questionnaire were decided, the final set of questions was distributed across 14 institutions of higher education and was filled by over 1,000 students from the graduation as well as post- graduation courses. These institutions from where the data was collected were basically from the Delhi/NCR Regions which covered a mixed of private and public colleges. This was done in batches and was arranged and translated into the electronic format. Except for the psychological scales, the grade point average of students was collected along with their type of course.

## 5.2 Data pre-processing

As soon as the data collection was completed, it was noticed that few students has not filled the questionnaire fully. These entries were deleted from the data set. This bought the effective number down to 837.

The data collected over the psychometric scales was arranged and then min-max normalisation was utilised over psychometric data. This helped us to range bound the data and thus, helped range bound the data improving the final outcome of the study.

The grade point equivalent followed in different educational institutions was different. Few instructions followed 10-point scale while some followed 4-point scale. To bring both on level plane field and comparable, the 10-point scale was reduced to 4-point scale. Thus, all the entries in the database were translated in a 4-point scale which ranged from 0–4. Since, these are defined as grade which ranged from A to F, we arranged these grades into clusters of EXCELLENT, GOOD, AVERAGE and BAD. The range of vales these belongs is as in Table 3.

**Table 3** Nomenclature of clusters

<i>S. no.</i>	<i>Range</i>	<i>Name</i>
1	3.67–4	Excellent
2	2.67–3.67	Good
3	1.67–2.67	Average
4	0–1.67	Bad

## 5.3 The achieved results

The main objective of this research was to predict the class in which a student will belong in his coming examination based upon various psychological inputs which are retained in the model. Several different models as explained in Section 4 of this paper are applied to the dataset to build a classification model and a final comparison of all these models is provided. To build these models MATLAB was used. The data was split into training data, testing data and validation data. The ratio of the split was 7:1.5:1.5, i.e., 70% of data was used for training, 15% each for testing and validation of the trained model. To present the results, in this section we'll be using true positive rate and the precision. True positive rate is the percentage of examples which are classified to their respective class. In other words, a result is said to be true positive if its output if mapped to the class it belonged. Precision is defined as all the examples which are mapped to a class and they belonged to that class as well. In other words, the level of accuracy achieved by the classifier.

### 5.3.1 Decision tree classifier

These are very powerful as well as popular tools in the domain of classification. These happen to be a tree like structure which starts with a starting node belonging to the root attributes and finish their journey at the leaf node depicting a class. In our case, the input vector was mapped to four classes. Thus, there happen to be 4 leaf nodes representing four classes. The advantage of the decision tresses is this that generates and represents

rules in their intermediate nodes which are easy to understand and interpret by a human user.

The data was fed into MATLAB work space and decision tress was implemented over the same. The results are presented in Table 4.

**Table 4** Decision tree classifier results

<i>S. no.</i>	<i>Predicted class</i>	<i>True positive rate</i>	<i>Precision</i>
1	Excellent	0.89	0.91
2	Good	0.83	0.52
3	Average	0.63	0.78
4	Bad	0.81	0.67
5	Average	0.79	0.72

### 5.3.2 *Naïve Bayes classifier*

This is a class of statistical classifier based upon the Bayes theorem which predicts the class membership by utilising the probabilities of different sets involved. This algorithm assumes that the effect an attribute produces over a class is independent of the attributes. In other words, this means that in the input vector, the impact of each unit is different and is independent of each other. Thus, these are graphical models which can describe the joint conditional probability of the different elements in the input vector. These are known for high accuracy in certain domains like text analysis. The results of the analysis are provided in Table 5.

**Table 5** Naive Bayes classifier results

<i>S. no.</i>	<i>Predicted class</i>	<i>True positive rate</i>	<i>Precision</i>
1	Excellent	0.523	0.325
2	Good	0.412	0.345
3	Average	0.675	0.648
4	Bad	0.798	0.823
5	Average	0.602	0.53525

### 5.3.3 *K-nearest neighbour classifier*

This is a method of classification in which the training set is classified based upon the difference of vector from nearest matching cluster. It is a type of instance-based learner which is also known as lazy learners. In our case, the numbers of clusters are four. Thus, k in our case becomes four each cluster representing one of the desired states. The results are presented in Table 6.

### 5.3.4 *Artificial neural networks*

These are a biologically inspired set of algorithms which are widely used in the domain of machine learning. These are used clustering, classification and forecasting quite often. In our case, we are using these for classifying the input vector into one of the four desired classes. The architecture which provided local optima was 7-15-4. A feed-forward back

propagation neural network was trained using this architecture and the results of the experiments are given in Table 7.

**Table 6** K-NN classifier results

<i>S. no.</i>	<i>Predicted class</i>	<i>True positive rate</i>	<i>Precision</i>
1	Excellent	0.423	0.652
2	Good	0.78	0.714
3	Average	0.512	0.538
4	Bad	0.813	0.896
5	Average	0.632	0.7

**Table 7** Artificial neural networks results

<i>S. no.</i>	<i>Predicted class</i>	<i>True positive rate</i>	<i>Precision</i>
1	Excellent	0.689	0.648
2	Good	0.712	0.735
3	Average	0.789	0.641
4	Bad	0.856	0.914
5	Average	0.7615	0.7345

### 5.3.5 Performance comparison

The performances of all the techniques used in this study are compared in Table 8. This figure shows that decision tree-based classifier provided the results with the highest certainty.

**Table 8** Performance comparison of classifiers

<i>S. no.</i>		<i>Decision tree</i>	<i>K-NN</i>	<i>Naïve Bayes</i>	<i>ANN</i>
1	True positive	0.79	0.632	0.602	0.7615
2	Precision	0.72	0.7	0.53525	0.7345
3	Overall	0.755	0.666	0.568625	0.748

## 6 Research findings

In this study, we have been able to predict the performance of students based upon their mental well-being. It had been established by numerous psychological studies and cross verified by our study that performance of students can vary based upon their mental well-being. In this study, we have been able to predict the movement in the GPA. We have analysed that in this study using various classification techniques of data mining and decision trees has shown the highest accuracy with True positive rate as 0.79 and precision as 0.72 as depicted in Table 8. Decision tree performed best out of all the supervised learning techniques under consideration and hence this architecture can be used to implement decision support system for GPA forecasts.

## 7 Conclusions and future work

This paper utilises various data mining techniques. The study is consistent with the previous researches where it is identified that the accuracy of decision tree is better as compared to other techniques. In future, a general method to forecast the group a student belongs to in his forthcoming examination based upon numerous psychological factors will be proposed. This study helps to understand that psychological parameters can cause results of a student vary greatly. There are multiple factors which are taken into consideration to analyse their dependence on GPA, but which factor plays the most important role is yet to be studied. If these factors are taken into consideration well in time, the results of students as well as academic institutions will improve greatly, and overall quality of Indian educational system can be enhanced to a greater extent.

## References

- Aher, S.B. and Lobo, L.M.R.J. (2011) 'Data mining in educational system using weka', *International Conference on Emerging Technology Trends (ICETT)*, Vol. 3.
- Ahmed, A.M., Norwawi, N.M. and Ishak, W.H.W. (2009) 'Identifying student and organization matching pattern using Apriori algorithm for practicum placement', *International Conference on Electrical Engineering and Informatics, 2009. ICEEI'09*, Vol. 1, IEEE.
- Aksenova, S.S., Zhang, D. and Lu, M. (2006) 'Enrollment prediction through data mining', *2006 IEEE International Conference on Information Reuse and Integration*, IEEE.
- Al-Radaideh, Q.A., Al-Shawakfa, E.M. and Al-Najjar, M.I. (2006) 'Mining student data using decision trees', *International Arab Conference on Information Technology (ACIT'2006)*, Yarmouk University, Jordan.
- Al-Shargabi, A.A. and Nusari, A.N. (2010) 'Discovering vital patterns from UST students data by applying data mining techniques', *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, Vol. 2, IEEE.
- Arora, R.K. and Badal, D. (2014) 'Mining association rules to improve academic performance', *International Journal of Computer Science and Mobile Computing*.
- Ayesha, S. et al. (2010) 'Data mining model for higher education system', *European Journal of Scientific Research*, Vol. 43, No. 1, pp.24–29.
- Bakar, N.A.A., Mustapha, A. and Nasir, K.M. (2013) 'Clustering analysis for empowering skills in graduate employability model', *Australian Journal of Basic and Applied Sciences*, Vol. 7, No. 14, pp.21–24.
- Balakrishnan, J.M.D. (2010) *Significance of Classification Techniques in Prediction of Learning Disabilities*, arXiv preprint arXiv:1011.0628.
- Bian, X. and Gong, Y. (2010) 'The application of data mining technology in analysis of college student's performance', *2010 2nd International Conference on Information Science and Engineering (ICISE)*, IEEE.
- Bresfelean, V.P. (2007) 'Analysis and predictions on students' behavior using decision trees in Weka environment', *Proceedings of the ITI*.
- Cairns, A.H. et al. (2014) 'Towards custom-designed professional training contents and curriculums through educational process mining', *The Fourth International Conference on Advances in Information Mining and Management*.
- Dai, S. and Zhang, P. (2008) 'A data mining algorithm in distance learning', *12th International Conference on Computer Supported Cooperative Work in Design, 2008. CSCWD 2008*, IEEE.
- Dejaeger, K. et al. (2012) 'Gaining insight into student satisfaction using comprehensible data mining techniques', *European Journal of Operational Research*, Vol. 218, No. 2, pp.548–562.



- Delavari, N., Beikzadeh, M.R. and Phon-Amnuaisuk, S. (2005) 'Application of enhanced analysis model for data mining processes in higher educational system', *6th International Conference on Information Technology Based Higher Education and Training, 2005. ITHET 2005*, IEEE.
- Fadzilah, S. and Abdoulha, M. (2009) 'Uncovering hidden information within university's student enrollment data using data mining', *Proceedings of the Third Asia International Conference on Modelling & Simulation Conference*, IEEE Computer Society.
- Gao, N. (2010) 'Proposing data warehouse and data mining in teaching management research', *2010 International Forum on Information Technology and Applications (IFITA)*, Vol. 1, IEEE.
- He, Y. and Zhang, S. (2011) 'Application of data mining on students' quality evaluation', *2011 3rd International Workshop on Intelligent Systems and Applications (ISA)*, IEEE.
- Hien, N.T.N. and Haddawy, P. (2007) 'A decision support system for evaluating international student applications', *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual*, IEEE.
- Jantawan, B. and Tsai, C-F. (2013) *The Application of Data Mining to Build Classification Model for Predicting Graduate Employment*, arXiv preprint arXiv:1312.7123.
- Khanna, L., Singh, S.N. and Alam, M. (2016) 'Educational data mining and its role in determining factors affecting student's academic performance: a systematic review', *2016 1st India International Conference on Information Processing (IICIP)*, IEEE.
- Khanna, L., Singh, S.N., Kumar, D. and Alam, M. (2017) 'Genetic algorithm approach to optimize factors determining students academic performance', *11th INDIACom IEEE International Conference on Computing for Sustainable Global Development*, New Delhi (India).
- Khanna, L., Singh, S.N., Kumar, D. and Alam, M. (2018) 'Optimal feature selection using filter and wrapper based approach for predicting students' academic performance', *12th INDIACom IEEE International Conference on Computing for Sustainable Global Development*, New Delhi (India).
- Knauf, R. et al. (2010) 'Empirical evaluation of a data mining method for success chance estimation of university curricula', *2010 IEEE International Conference on Systems Man and Cybernetics (SMC)*, IEEE.
- Kotsiantis, S., Pierrakeas, C. and Pintelas, P. (2002) *Efficiency of Machine Learning Techniques in Predicting Students' Performance in Distance Learning Systems*, Educational Software Development Laboratory Department of Mathematics, University of Patras, Greece.
- Kovacic, Z. (2010) 'Early prediction of student success: mining students' enrolment data'.
- Liu, Z. and Zhang, X. (2010) 'Prediction and analysis for students' marks based on decision tree algorithm', *2010 3rd International Conference on Intelligent Networks and Intelligent Systems (ICINIS)*, IEEE.
- Minaei-Bidgoli, B. and Punch, W.F. (2003) 'Using genetic algorithms for data mining optimization in an educational web-based system', *Genetic and Evolutionary Computation Conference*, Springer, Berlin, Heidelberg.
- Osmanbegović, E. and Suljić, M. (2012) 'Data mining approach for predicting student performance', *Economic Review*, Vol. 10, No. 1, pp.3–12.
- Pandey, U.K. and Pal, S. (2011) *A Data Mining View on Class Room Teaching Language*, arXiv preprint arXiv:1104.4164.
- Pumpuang, P., Srivihok, A. and Praneetpolgrang, P. (2008) 'Comparisons of classifier algorithms: bayesian network, C4. 5, decision forest and NBTree for course registration planning model of undergraduate students', *IEEE International Conference on Systems, Man and Cybernetics, 2008. SMC 2008*, IEEE.
- Ramasubramanian, P., Iyakutti, K. and Thangavelu, P. (2009) 'Enhanced data mining analysis in higher educational system using rough set theory', *African Journal of Mathematics and Computer Science Research*, Vol. 2, No. 9, pp.184–188.

- Rasmani, K.A. and Shen, Q. (2005) 'Subsethood-based fuzzy rule models and their application to student performance classification', *The 14th IEEE International Conference on Fuzzy Systems, 2005. FUZZ'05*, IEEE.
- Romero, C. et al. (2008) 'Data mining algorithms to classify students', *Educational Data Mining*.
- Sakurai, Y., Tsuruta, S. and Knauf, R. (2011) 'Success chances estimation of university curricula based on educational history, self-estimated intellectual traits and vocational ambitions', *2011 11th IEEE International Conference on Advanced Learning Technologies (ICALT)*, IEEE.
- Salazar, A. et al. (2004) 'A case study of knowledge discovery on academic achievement, student desertion and student retention', *2nd International Conference on Information Technology: Research and Education, 2004. ITRE 2004*, IEEE.
- Saranya, S., Ayyappan, R. and Kumar, N. (2014) 'Student progress analysis and educational institutional growth prognosis using data mining', *International Journal of Engineering Sciences & Research Technology*.
- Selmoune, N. and Alimazighi, Z. (2008) 'A decisional tool for quality improvement in higher education', *ICTTA 2008. 3rd International Conference on Information and Communication Technologies: From Theory to Applications, 2008*, IEEE.
- Sharma, A., Bhatnagar, V. and Bansal, A. (2016) 'Implementations of data mining techniques in stock market forecasting: a generic framework', *International Journal of Intelligent Enterprise*, Vol. 3, Nos. 3–4, pp.311–326.
- Sharma, M. and Mavani, M. (2011) 'Accuracy comparison of predictive algorithms of data mining: application in education sector', *Advances in Computing, Communication and Control*, pp.189–194, Springer, Berlin, Heidelberg.
- Shi, L., Bai, J.N. and Zhao, Y.L. (2009) 'Mining association rules based on apriori algorithm and application', *Proc. of IFCSTA*, Vol. 3, pp.141–145.
- Srimani, P.K. and Patil, M.M. (2012) 'A classification model for edu-mining', *PSRC-ICICS Conference Proceedings*.
- Sukanya, M. et al. (2012) 'Data mining: performance improvement in education sector using classification and clustering algorithm', *International Conference on Computing and Control Engineering (ICCCE 2012)*, Vol. 12.
- Talavera, L. and Gaudioso, E. (2004) 'Mining student data to characterize similar behavior groups in unstructured collaboration spaces', *Workshop on Artificial Intelligence in CSCL. 16th European Conference on Artificial Intelligence*.
- Wook, M. et al. (2009) 'Predicting NDUM student's academic performance using data mining techniques', *Second International Conference on Computer and Electrical Engineering, 2009. ICCEE'09*, Vol. 2, IEEE.
- Wu, X., Zhang, H. and Zhang, H. (2010) 'Study of comprehensive evaluation method of undergraduates based on data mining', *2010 International Conference on Intelligent Computing and Integrated Systems (ICISS)*, IEEE.
- Yadav, S.K. and Pal, S. (2012) 'Data mining application in enrollment management: a case study', *International Journal of Computer Applications*, Vol. 41, No. 5.
- Zhu, L., Li, Y. and Li, X. (2009) 'Research on early-warning model of students' academic records based on association rules', *2009 WRI World Congress on Computer Science and Information Engineering*, Vol. 4, IEEE.