

# Paper

*by* Jay Maru

---

FILE	UPDATED_TECHNICAL_PAPER_1.DOCX (220.64K)		
TIME SUBMITTED	17-MAY-2017 11:35AM	WORD COUNT	3333
SUBMISSION ID	815296928	CHARACTER COUNT	17945

# Performance Prediction and All Round Profile Strengthening

1<sup>3</sup>hwini Dalvi 2Jay Maru, 3Nupoor Joshi, 4Saureen Parekh, 5Shreyank Masrani

1Assistant Professor Department of Information Technology, K J Somaiya College of Engineering, Mumbai  
2345Student, Department of Information Technology, K J Somaiya College of Engineering, Mumbai

**Abstract:** In recent years, a lot of emphasis is being given to the marks scored by a student and a lot of different mechanisms are being designed and implemented to predict a student's performance so that various measures can be taken as to control and improve his performance. However, the education system has seen a recent shift in the trends from what it was previously. Nowadays, not only academic scores are considered to help a person get into a good university for further studies, but equal importance is being given to his profile. Hence we aim to develop a system which would not only predict a student's marks, but also will help him build a strong profile to gain a competitive edge.

## I. INTRODUCTION:

The field of education has seen a recent shift in trend where more and more importance is being given to a student's graduation scores and profile along with his marks in entrance exams. Be it in the case of higher education or job application, graduation marks and a student's profile are now playing a significant role in deciding where the student will stand. In such a scenario, a lot of emphasis is given to profile building and improvement of academic records of the students. Though hard work is an essential factor, but it is smart work which helps students boost their performance. Understanding this, we aim to design an Android application which will provide the student with invaluable insights which would help him to score better and enhance his performance in his engineering exams.

Our application aims at providing students with functionalities from a simple CGPI calculator, Profile improvement tips, Trends of marks scored by the previous batch in their exams, Prediction of the marks a student is likely to score in his upcoming semester. The application will be making use of Multiple Linear Regression to predict a student's marks based on his academic records.

## Chapter 2: Related Research

### A. Papers predicting performance from academics

Many attempts were made to predict the performance of a given student depending upon his academic records of which some of the good approaches are discussed below.

Chew Li Sa, DayangHananib<sup>6</sup>.Abang Ibrahim, Emmy DahlianaHossain, Mohammad bin Hossin [14] developed a **Student Performance Analysis System (SPAS)** to predict if a student will pass or fail in System Analysis and Design depending upon his previous grades using a set of algorithms like J48, CART, BF Tree, Random Tree and J48graft and discovered that BF Tree turned out to be the best predictor among them all.

A similar kind of approach was made by Ms.Tismy Devasia ,Ms.Vinushree T , Mr.VinayakHegde [11]who used Nave Bayes algorithm and classified students into performance groups of good, average, poor and excellent which were supposed to be their results in final exams.

Similarly PoojaPathak, NehaBansal, Shivani Singh [7] developed Mulnayakan- An artificial Neural Network model to predict a student's marks in Maths depending upon his academic records using LM, BFGS and Resilient Backpropagation algorithms and have concluded that LM algorithm had the maximum accuracy of 89%.

Owing to the fact that the performance of students is directly proportional to the difficulty of the subject, KamaljitKaur and KuljitKaur[3] tried exploring how the difficulty of a subject affects the performance of students in the subject. They discovered that the accuracy of J48 and CART algorithms were about 98% in case of the easy subject but were comparatively lesser in the case of difficult subjects and M5P algorithm turned out to be the best algorithm in terms of accuracy for predicting final scores.

There has been a significant amount of research on the dependence of fundamental subjects' marks and concepts on the final grades by PauziahMohdArsad, NorlidaBuniyamin, Jamalul-lailAbManan[6] who tried using a neural network approach to identify the relation between fundamental subject marks on the final grade of the student using Levenberg Marquardt model and realized that the diploma students reached the higher pointers faster than the matriculation students.

PauziahMohdArsad, NorlidaBuniyamin and Jamalul-Jail AbManan[1] in another paper talk about how the marks scored in fundamental subjects affect the performance of students in their final semester. They concluded that the co-efficient of Co-relation (R) turned out to be much higher in the case of neural network approach as compared to the multiple regression approach.

The entire idea of performance prediction using academics was taken a step ahead by KrinaParmar, Prof. DineshkumarVaghela, Dr PriyankaSharma[12] who tried implementing a distributed data mining approach to design a high accuracy model. They implemented J48 and Random Tree algorithms on multiple independent datasets and generated various rules from each. These rules were then combined to form a global model.

B. Papers considering Academics along with other factors:

However the performance of a student does not only depend upon his academics but also a lot of various factors and there was no stone left unturned in exploring those factors and understanding how did those factors affect the performance.

Md. FahimSikder, Md. Jamal Uddin, SajalHalder [10] considered factors consisting a student's academic records as well as his social, personal and family details and implemented Levenberg Marquardt algorithm on MATLAB and realized that the accuracy of prediction turned out to be higher when the student's graders were lower. Also they concluded that a student's performance depends not only upon his academics but a lot of other factors like his family education, extra curricular activities, social media interaction, etc.

PriyankaAnandraoPatil, R. V. Mane [4] also considered various factors like personal and family social and economic information and academic progress using GSP and FP Tree and realized that GSP algorithm is more effective when applied to large datasets but poor when applied to smaller datasets. GSP algorithm was used to predict the frequent patterns and FP Tree algorithm predicted how the student will perform.

PrabhjotKaur and WilliamjeetSingh[13] presented a SGPA Prediction system which used j48, LMT, Random Tree, REP Tree algorithms on 20 attributes consisting of his early life and family information along with his marks. The accuracy of RepTree turned out to be the highest(61.70%) in classifying students into classes of performance (Below Average, Average, Good, Excellent).

The number of factors coming into consideration were increased multifold when S Chaitanya Kumar, E Deepak Chowdary, Venkatramaph Kumar S, K V Krishna Kishore [5] considered Online-Learning Skills, Problem-Solving Efficiency, Time Management, Intention of Doing Higher Studies, Adaptable Nature, Sports Participation, Day-Scholar, Self-Learning, Time spent on Social Networks, Versatility Nature, Library Usage, Practical Knowledge and Aggregate and implemented M5P on them to predict the performance of system. The accuracy of M5P turned out to be 97.17%.

Emotional stability also plays a role in student's performance which was taken into consideration by Tripti Mishra, Dr. Dharminder Kumar, Dr. Sangeeta

Gupta [2] who tried to predict the 3<sup>rd</sup> semester results of MCA students depending upon their academics and emotional characteristics by identifying the emotional traits of a student using Emotional Skills Assessment Process using J8 and Random Tree Algorithms. The accuracy of J48 algorithm turned out to be 88.71% and Random tree algorithm was 94.418%.

Comment mining is a concept which is used to predict and identify the characteristics of a student depending upon his comments on various topics.

Shaymaa E. Sorour, Tsunenori Mine [9] proposed a system which mines a student's characteristics depending upon his comments on a set of questions asked to him after a lecture to get an idea of which factors affect learning and thereby predicting their final score using C4.5 and Random Forest algorithms.

A lot of people tried predicting a student's performance considering a variety of factors like his academic records, his emotional traits, social media interaction, family background, etc. However, not much light has been put on the fact that every student is unique and his abilities and capabilities are different than his colleagues and hence any standalone model would not be able to predict a student's performance accurately unless it considers a student's personality traits as well.

Hence, we are developing a model which maps a student's personality traits using the BIG 5 model to identify a student's behavioral traits and thereby predicting his performance based on his personality and his academic records.

We too have used a similar kind of approach to design a system which will predict the marks a student is likely to score in his final semester.

#### Data Collection and Transformation:

- **Data Collection:** This includes the result mark-sheets of exams given by students of K. J. Somaiya college of Engineering. The result include the marks scored by a student in his End Semester Exams (ESE), Continuous Assessment (CA) exams, Term work(TW) exams and Practical/ Oral examinations along with the credits earned and the pointers scored.  
A student's mobile phone usage logs was collected with the help of RescueTime, a third party app which collects and displays

the time one has spent over applications over the past month.

A peer review test was taken of 15 students which were to be filled in by their classmates. This test would help us identify and validate a student's personality.

- **Data Transformation:** This step included firstly conversion of data from PDF to Excel sheets and then cleaning of data by removing of redundant fields, completion of incomplete fields, replacing erroneous values and modifying the results to fit a standardized pattern.
- **Data Loading:** The data once cleaned was then loaded and hosted onto Xampp server for further processing and retrieval.

#### Multiple Linear Regression:

The Multiple Linear Regression model is used to predict the pointer for a particular semester based on marks of previous semesters.

**Step 1:** The relationship being captured is  
 $\text{targetmrk} = f(\text{sem1mrk}, \text{sem2mrk}, \dots, \text{targetmrk}-1)$   
 targetmrk = marks obtained by student in target semester (value to be predicted).

$f(\text{sem1mrk}, \text{sem2mrk}, \dots) =$  marks obtained by a student in previous semesters.

**Step 2:** Makes use of training dataset.

4 = total number of students in training dataset.

Writing mathematically using linear function for prediction

$X=BY$

Where X= matrix of order [ i+1,j]

Where i = N; j = ts-1;

Where ts = target semester;

X<sub>ij</sub> represents marks obtained by 'ith' student in 'jth' semester in training dataset.

$$X = \begin{matrix} & \begin{matrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} & \dots & \dots & \text{upto } ts-1 \\ 1 & x_{31} & x_{32} \\ \vdots & & \end{matrix} \\ \text{Upto } N; \end{matrix}$$

Y = matrix of order [ 1\*N] which contains marks obtained in the target semester by ' N' students in the training data set.

$$Y = \begin{matrix} y_0 \\ y_1 \\ y_2 \end{matrix}$$



$\vdots$   
 $\vdots$   
 Upto  $y(N-1)$ ;  
 4

When used for training data to obtain  $b$ ,  $y$  denotes training target semester marks.

$B$  = Vector of unknown parameters given by  
 $b_0$   
 $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \end{bmatrix}$   
 Upto  $b(ts-1)$ ;  
 1

### Step 3:

This is overdetermined system as number of equations greater than number of unknowns and hence has no solution. So the goal is to find the coefficients  $b$  which fit the equations best in the sense of minimising mean square error ( $= E[(\text{actual value} - \text{predicted value})^2]$ ).

$\hat{b} = \arg \min S(b)$ .

The object 5 is to minimize  $S(b)$

$$S(b) = y - Xb = (y - Xb)^T (y - Xb) = y^T y - 2b^T X^T y + b^T X^T X b.$$

Differentiating with respect to  $b$  and equating to zero gives

$$-X^T y + (X^T X) \hat{b} = 0$$

$$\hat{b} = (X^T X)^{-1} X^T y$$

Using the above equation, the vector  $b$  (which contains unknown parameters) is found. In this way unknown parameters are estimated with the help of training data.

Once the unknown parameters are found, we can predict target semester marks for test data as

$$\text{Target sem marks} = b_0 + b_1 \text{sem1} + b_2 \text{sem2} + b_3 \text{sem3} + \dots$$

### Profile Building:

Understanding that each student is having different abilities and capabilities, it is never a good idea to have a standard set of advices for every student in order to strengthen his profile.

A generic set of suggestions might turn out to work amazingly for some students but might be disastrous to some other students.

Example: If a particular student is planning to become a UI/UX developer, suggesting him to do courses and internships on Java would be completely useless, but for another student who wished to

become an android developer, learning and understanding the basics of Java would be the best advice one can give him.

So the best and the most reliable profile building tips would be the one which have been specially tailored for a given student. We must understand what the student is willing to build his profile for and what factors are going to affect the same.

We plan to help students first strengthen their fundamentals in the core subjects by pointing out the subjects from the syllabus which might be directly or indirectly associated to his plans.

For example: An IT student who wishes to be a DBA needs to know his subjects DBMS, ADMS and DMBI at the tip of his fingers.

The next step would be seeking out internships and doing various courses which would be related to his interests. For example: If our student aspiring to be a DBA does courses on Cross Platform Database Connectivity and some internship which includes ETL process and generating insights about the clusters and correlations would put him at a very strong preference than his competitors.

Hence our System aims to understand the future plans/ goals of the user and give him the subjects he needs to focus on, the courses he can consider and the internships he must choose such that they are in alignment with his goals.

Finally, in order to help him ace the game, it is necessary to ensure that he's completely equipped and updated about the latest advancements and innovations in his field. So the app aims to integrate a RSS feed module which will help him get the latest trends and affairs in his hand.

### Implementation:

- CGPI Calculator:

This can be referred to as a simple CGPI Calculator which identifies a student's current semester depending upon the roll number entered by him. This includes a simple user friendly interface wherein a student is supposed to enter the marks scored by him in Continuous Assessment Tests, End Semester Exams and the Term Works in the subjects mentioned. This module simply calculates the CGPI on the basis of the entered marks.

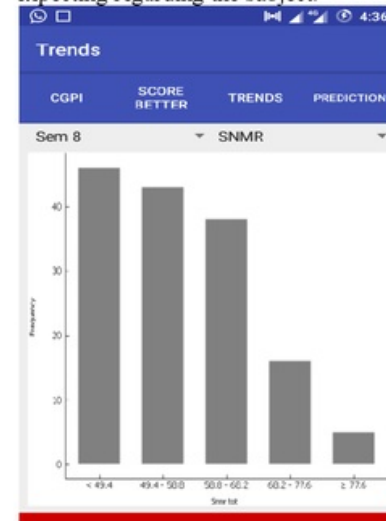
MainActivity			
CGPI	SCORE BETTER	TRENDS	PREDICTION
Course	CA	ESE	TW
BDA	26	74	40
SNMR	23	70	29
CSM	30	65	35
Electives	29	83	32
Project 2	45	0	0
--	40	100	25/50
SUBMIT		5.92	RESET

- Score Better:**  
 This module is basically based on the rules of the University for Calculation of CGPI and assignment of credits depending upon the marks. The system is such that one has to score between ranges of marks in order to receive the associated credits. For example, in a particular subject depending upon the marks scored in CA tests, a person can score anywhere between 51-67 to secure a 7 multiplier, irrespective of whether he scores 51 or 67, he will be getting the same 7 multiplier. So this module one has to input his marks scored in the CA tests and in turn get the minimum marks required to score the associated credit multiplier as the output.

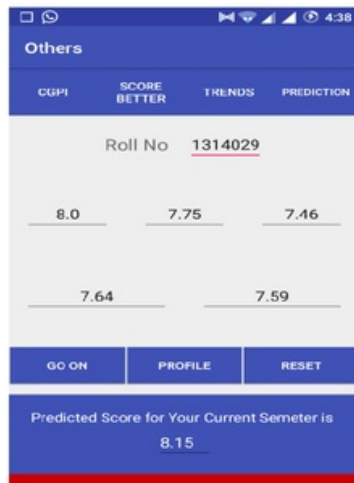
ScoreBetter			
CGPI	SCORE BETTER	TRENDS	PREDICTION
Example : 50-59/100 gains 6 multiplier but 60 gains 7 multiplier, thus we need to know imp. levels marks			
CA Marks	29	Passing Marks	30.0
Marks for 4 Multiplier			30.0
Marks for 5 Multiplier			30.0
Marks for 6 Multiplier			35.0
Marks for 7 Multiplier			52.0
Marks for 8 Multiplier			68.0
Marks for 9 Multiplier			77.0
Marks for 10 Multiplier			93.0
SUBMIT		RESET	

- Trends:**

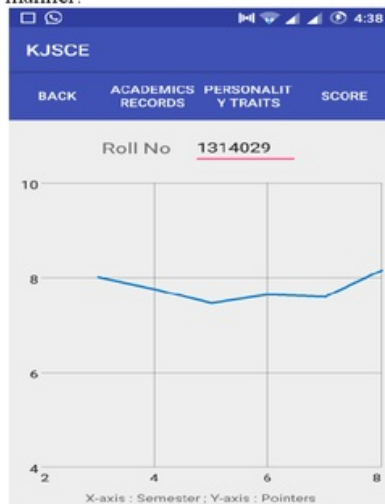
This section aims to provide a student with vital information depending upon the performance of the previous batch. This section gives a brief idea of the number of students falling in different range of marks in the subject selected by the user. One gets an idea of the marks scored by majority of the students and also the number of students passing/failing in the subject so that one can roughly have an idea of what he should be expecting regarding the subject.



- Prediction:**  
 This section forms the heart of our application which predicts the score a student is going to score in the upcoming semester giving him enough time to plan ahead and prioritize and strategize his efforts in order to try and score better. The application automatically retrieves the marks scored by the student from the roll number entered by him from the server and predicts the marks he will score in the upcoming semester. This module uses multiple linear regressions which takes the marks scored in his previous semesters as independent variables to predict the marks he will score in the current semester.



- **Academic Records:**  
This module displays a simple line graph of the marks scored by student in his previous semesters enabling him to keep a track of his progress in a really simple yet effective manner.



#### Conclusion:

Hence in order to conclude we can say that, if we want to help the student achieve his goals, we have to make sure that he does not only concentrate on academics but also on profile building. Hence any standalone model would be considered enough to help student achieve his goals only if it is concentrating not only on scoring better but also

providing him a way to build his profile to suit the best to his goals.

#### References:

1. Arsad, Pauziah Mohd, and Norlida Buniyamin. "Prediction of engineering students' academic performance using Artificial Neural Network and Linear Regression: A comparison." *Engineering Education (ICEED), 2013 IEEE 5th Conference on*. IEEE, 2013.
2. Mishra Tripti, Dharminder Kumar, and Sangeeta Gupta. "Mining students' data for prediction performance." *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on*. IEEE, 2014.
3. Kaur, Kamaljit, and Kuljit Kaur. "Analyzing the effect of difficulty level of a course on students performance prediction using data mining." *Next Generation Computing Technologies (NGCT), 2015 1st International Conference on*. IEEE, 2015.
4. Pati, Priyanka Anandrao, and R. V. Mane. "Prediction of Students Performance Using Frequent Pattern Tree." *Computational Intelligence and Communication Networks (CICN), 2014 International Conference on*. IEEE, 2014.
5. Kumar, S. Chaitanya, et al. "M5P model tree in predicting student performance: A case study." *Recent Trends in Electronics, Information & Communication Technology (RTEICT), IEEE International Conference on*. IEEE, 2016.
6. Arsad, Pauziah Mohd, and Norlida Buniyamin. "A neural network students' performance prediction model (NNSPPM)." *Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on*. IEEE, 2013.
7. Pathak, Pooja, Neha Bansal, and Shivani Singh. "Mulyankan: A prediction for student's performance using Neural Network." *Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on*. IEEE, 2015.
8. Bunkar, Kamal, et al. "Data mining: Prediction for performance improvement of graduate students using classification." *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on*. IEEE, 2012.
9. Sorour, Shaymaa E., and Tsunenori Mine. "Building an Interpretable Model of Predicting Student Performance Using Comment Data Mining." *Advanced Applied Informatics (IIAI-AAI), 2016 5th IIAI International Congress on*. IEEE, 2016.
10. Sikder, Md Fahim, Md Jamal Uddin, and Sajal Halder. "Predicting students yearly performance using neural network: A case study of BSMRSTU." *Informatics, Electronics and Vision*

(ICIEV), 2016 5th International Conference on. IEEE, 2016.

11. Devasia, Tismy, T. P. Vinushree, and Vinayak Hegde. "Prediction of students performance using Educational Data Mining." *Data Mining and Advanced Computing (SAPIENCE), International Conference on.* IEEE, 2016.

12. Parmar, Krina, Dineshkumar Vaghela, and Priyanka Sharma. "Performance prediction of students using distributed Data mining." *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on.* IEEE, 2015.

13. Kaur, Prabhjot, and Williamjeet Singh. "Implementation of student SGPA Prediction System (SSPS) using optimal selection of classification algorithm." *Inventive Computation Technologies (ICICT), International Conference on.* Vol. 2. IEEE, 2016.

14. Sa, Chew Li, Emmy Dahliana Hossain, and Mohammad bin Hossin. "Student performance analysis system (SPAS)." *Information and Communication Technology for The Muslim World*

(ICT4M), 2014 The 5th International Conference on. IEEE, 2014.

15. Wald, Randall, Taghi Khoshgoftaar, and Chris Sumner. "Machine prediction of personality from Facebook profiles." *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on.* IEEE, 2012.

16. Kalghatgi, Mayuri Pundlik, Manjula Ramannavar, and Nandini S. Sidnal. "A neural network approach to personality prediction based on the big-five model." *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* 2.8 (2015): 56-63.

17. Hamid, Norsiah Abdul, Mohd Sobhi Ishak, and Siti Syamsul Nurin Mohmad Yazam. "Facebook, YouTube and Instagram: Exploring Their Effects on Undergraduate Students' Personality Traits." *The Journal of Social Media in Society* 4.2 (2015).

18. Xu, Runhua, et al. "Towards Understanding the Impact of Personality Traits on Mobile App Adoption-A Scalable Approach." *ECIS*. 2015.



# Paper

---

## ORIGINALITY REPORT

---

%**8**

SIMILARITY INDEX

%**6**

INTERNET SOURCES

%**4**

PUBLICATIONS

%**5**

STUDENT PAPERS

---

## PRIMARY SOURCES

---

**1**

Submitted to Indian Institute of Technology,  
Bombay

Student Paper

%**4**

---

**2**

S Chaitanya Kumar, E Deepak Chowdary, S  
Venkatramaphanikumar, K V Krishna Kishore.  
"M5P model tree in predicting student  
performance: A case study", 2016 IEEE  
International Conference on Recent Trends in  
Electronics, Information & Communication  
Technology (RTEICT), 2016

Publication

%**2**

---

**3**

[ijircce.com](http://ijircce.com)

Internet Source

%**1**

---

**4**

[www.it.iitb.ac.in](http://www.it.iitb.ac.in)

Internet Source

%**1**

---

**5**

Submitted to Cardiff University

Student Paper

%**1**

---

**6**

[www.rroj.com](http://www.rroj.com)

Internet Source

<%**1**

---

"Implementation of student SGPA Prediction System(SSPS) using optimal selection of classification algorithm", 2016 International Conference on Inventive Computation Technologies (ICICT), 2016

Publication

---

---

EXCLUDE QUOTES      OFF

EXCLUDE MATCHES      OFF

EXCLUDE  
BIBLIOGRAPHY      ON