

# Prediction of Student Performance using Machine Learning



Anusha M, K Karthik, P Padmini Rani, VSrikanth

**Abstract:** Educational foundations are delivering capable and shrewd understudies and specialists, yet when we think about quality and value of the student's advancement in his profession; it is as yet a challenge or an inquiry. These organizations center on quality in training. Consistently countless alumni from schools and colleges, as for the information gathered from the criticism of students, order an information mining strategy is connected to it. It is a stage to examine the elements influencing the performance of students so as to assess the present understudy execution and make proficient strides in the expectation of the no doubt happening connections between the different parts of learning and to upgrade the nature of instruction in future and help the educational organizers to design in like manner.

**Key Words:** Machine Learning, Prediction, performance

## I. INTRODUCTION

Nowadays, investigate including instructive information is uncommonly extended. Correspondingly, Information mining using diverse instructive information called Educational Data Mining (EDM) is also used. The essential objective of Educational data mining (EDM) consolidates execution conjecture, understudy showing, space illustrating, examination and portrayal of understudy data, recommendation system, gathering understudies, etc. Additionally, look into on understudy execution expectation have been considered from various characteristics in understudies' condition, for instance, understudies' direct, economics, understudies' information, mental and budgetary. Each quality include a couple of segments used in the estimation. Directly off the bat, the understudies' practices were evaluated reliant on the repeat of understudy's access to each part isolated from E-learning usage. The understudies' economics join place of residence, side interests, family size, work and guidance of gatekeepers and others. Next, The understudies' information, like affirmation or enrolment data, fuses name, age, sex, address, etc.. Mental characteristics stress with limits, airs, lead and motivation; in the interval, money related depict about the monetary establishment of the understudy and family and in expansion their insightful establishment. There are different information mining methods, for example, arrangement, relapse, thickness estimation, grouping, and relationship mining had been executed in instructive information inquire about.

Revised Manuscript Received on October 30, 2019.

\* Correspondence Author

**Anusha M\***, Department of CSE, KLEF, Green Fields, Vaddeswaram, Guntur dt, Ap, India

**K Karthik**, Department of CSE, KLEF, Green Fields, Vaddeswaram, Guntur dt, Ap, India

**P Padmini Rani**, Department of CSE, KLEF, Green Fields, Vaddeswaram, Guntur dt, Ap, India

**VSrikanth**, Department of CSE, KLEF, Green Fields, Vaddeswaram, Guntur dt, Ap, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Moreover, techniques like Xg-boost, Naïve Bayes and K-means have additionally been utilized in foreseeing understudies' execution and mining instructive information. Thus, this exploration likewise foresees the last examination of undergrad understudy studying Management Information System. Likewise, this investigation not just anticipate the execution (last examination), yet additionally contrast the two model, which one is better in term of exactness, execution, and blunder rate.

## II. LITERATURE SURVEY

	Reference	Authors	Method Used
1	Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perception	FebriantiWidyahastuti VianyUtamiTjhin	Linear Regression Multilayer Perception
2	Review on Predicting Student's Performance using Data Mining Techniques	A. M. Shahiri and W. A. Husain	Linear Regression
3	Prediction of student academic performance by an application of data mining techniques	S. Sembiring, M. Zalis, D. Hartana, S. Ramliana, and E. Wari,	Linear Regression
4	Enhancing the capabilities of Student Result Prediction System	P. Chaudhury, S. Mishra, H.K. Tripathy, and B. Kishore,	Logistic Regression
5	Comparative analysis of bayes and lazy classification algorithms	S. Vijayarani and M. Muthulakshmi	Naïvebaye classifier
6	Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance	C. J. Willmott and K. Matsuura	Linear Regression
7	Analyzing the Student Performance using Classification Techniques to find the Better Suited Classifier	S. Dominick and Razak TA	Decision tree classifier

Table 1: Literature Survey

### Supervised machine learning

- Regression and classification are ordered under a similar umbrella of supervised machine learning. Both offer a similar idea of using referred to datasets (alluded to as preparing datasets) to make forecasts.

- In supervised learning, a calculation is utilized to take in the mapping capacity from the info variable (x) to the yield variable (y); that is  $y = f(X)$ .
- The goal of such an issue is to estimated the mapping capacity (f) as precisely as conceivable to such an extent that at whatever point there is another information (x), the yield variable (y) for the dataset can be anticipated.
- Lamentably, there is the place the comparability between regression versus classification machine learning closes.

### III. WHAT IS MACHINE LEARNING

AI is an utilization of man-made reasoning (AI) that gives frameworks the capacity to naturally take in and improve as a matter of fact without being unequivocally programmed. Machine learning centers around the advancement of PC programs that can get to information and use it learn for themselves. The way toward learning starts with perceptions or information, for example, precedents, direct involvement, or guidance, so as to search for examples in information and settle on better choices later on dependent on the models that we provide. The essential point is to permit the PCs learn automatically without human intercession or help and modify activities likewise.

#### a. Machine Learning Methods

AI calculations are frequently ordered as supervised or unsupervised. Supervised AI algorithms can apply what has been realized in the past to new information utilizing named guides to foresee future occasions. Beginning from the investigation of a known preparing dataset, the learning calculation creates an induced capacity to make forecasts about the yield esteems. The framework can give focuses to any new contribution after adequate preparing. The learning calculation can likewise contrast its yield and the right, planned yield and discover mistakes so as to alter the model in like manner.

- In contrast, unsupervised AI algorithms are utilized when the data used to prepare is neither ordered nor marked. Unsupervised learning thinks about how frameworks can derive a capacity to portray a concealed structure from unlabeled information. The framework doesn't make sense of the correct yield, however it investigates the information and can draw deductions from datasets to depict concealed structures from unlabeled information.
- Semi-supervised AI calculations fall some place in the middle of administered and unsupervised learning, since they utilize both marked and unlabeled information for preparing – regularly a little measure of named information and a lot of unlabeled information. The frameworks that utilization this strategy can extensively improve learning precision. More often than not, semi-directed learning is picked when the obtained named information requires talented and applicable assets so as to prepare it/gain from it. Something else, acquiring unlabeled information by and large doesn't require extra assets.
- Reinforcement AI calculations is a learning technique that communicates with its condition by creating activities and finds mistakes or rewards. Experimentation seek and deferred compensate are the most applicable qualities of support learning. This technique enables machines and programming operators to consequently decide the perfect

conduct inside a particular setting so as to boost its execution. Straightforward reward criticism is required for the operator to realize which activity is ideal; this is known as the fortification flag.

Machine learning empowers investigation of enormous amounts of information. While it for the most part conveys quicker, progressively exact outcomes so as to recognize productive chances or perilous dangers, it might likewise require extra time and assets to prepare it appropriately. Consolidating AI with AI and subjective advancements can make it significantly progressively viable in handling substantial volumes of data.

#### 3.1 What is big data

While the expression "big data" is moderately new, the gathering and putting away a lot of data for possible examination is ages old. This idea picked up momentum in the mid 2000s when industry examiner Doug Laney enunciated the now-standard meaning of enormous information as the three Vs:

**Volume:** Organizations gather information from an assortment of sources, including business exchanges, online life and data from sensor or machine-to-machine information. Before, putting away it would've been an issue – however new advancements, (for example, Hadoop) have facilitated the weight.

**Velocity:** Information streams in at an extraordinary speed and should be managed in a convenient way. RFID labels, sensors and keen metering are driving the need to manage downpours of information in close continuous.

**Variety:** Information comes in a wide range of organizations – from organized, numeric information in conventional databases to unstructured content reports, email, video, sound, stock ticker information and budgetary exchanges.

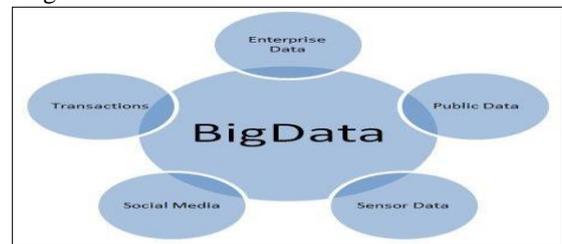


Fig 1: Some Sources of Big Data

#### b. Importance of Big Data

The significance of big data doesn't rotate around how much information you have, yet what you do with it. You can take information from any source and dissect it to discover answers that empower 1) cost decreases, 2) time decreases, 3) new item improvement and enhanced contributions, and 4) brilliant basic leadership. When you consolidate enormous information with powerful examination, you can achieve business-related undertakings, for example,

- Deciding main drivers of disappointments, issues and imperfections in real time.
- Generating coupons at the purpose of sale dependent on the client's purchasing habits
- Recalculating whole risk portfolios in minutes.
- Detecting fraudulent conduct before it influences your association.

**c. Considerations in Big Data**

- Cheap, large storage.
- Faster processors.
- Affordable open source, circulated enormous information stages, Example-hadoop
- Cloud figuring and other adaptable asset assignment game plans.

The algorithms used are k-means, xgboost, linear regression and naive bayes and they are contrasted to find out which of the algorithm gives maximum accuracy.

**d. K-means**

K-Means is one of the least difficult unsupervised learning algorithms that unravel the well known clustering problem. The strategy pursues a basic and simple approach to characterize a given informational index through a specific number of bunches. Clustering as a strategy for discovering subgroups inside perceptions is utilized generally in applications like market division wherein we attempt and discover some structure in the information. Despite the fact that an unsupervised AI system, the groups can be utilized as highlights in a directed AI model.

**ALGORITHM**

**STEP 1:**

- We arbitrarily pick 'K' bunch focuses (centroids).
- Let's expect these are  $c_1, c_2, \dots, c_k$  and we can say that;
- 'C' is set of centroids.

**STEP 2:**

- In this progression, we assign out every data point to nearest focus, this is finished by computing Euclidean separation.
- Where  $dist()$  is the Euclidean separation.

**STEP 3:**

- In this progression, we locate the new centroid by taking the normal of the considerable number of focuses appointed to that bunch.
- 'Si' is the arrangement of all focuses assigned out to the *i*th cluster

**STEP 4:**

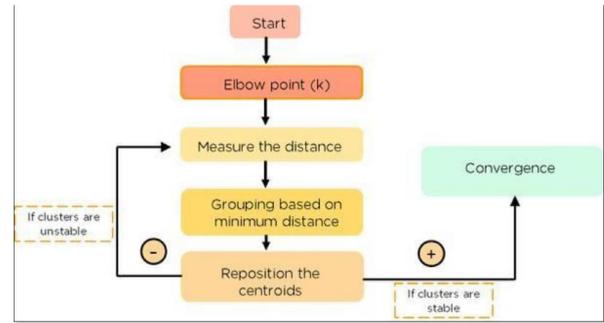
- In this step, we repeat steps 2 and 3 until the cluster assignments do not change .
- That means we repeat the algorithm till our clusters remain stable,

**Advantages**

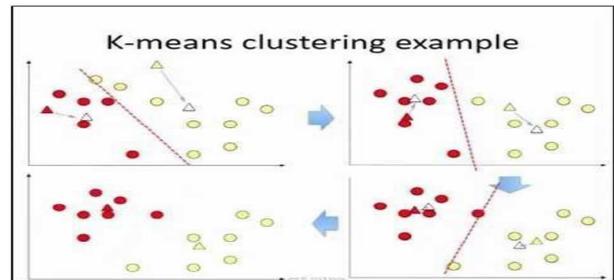
- Easy implementation
- k-means is computationally faster with large number of variables
- k-means produces tighter clusters compared to hierarchical clustering
- If the centroids are re-calculated, this can change the entire cluster.

**Disadvantages**

- It is hard to predict the clusters formed
- Final results get impacted more with initial seeds
- Final results are impacted more by order of data



**Fig 2: K-MEANS Clustering Methodology**



**Fig 3: K-MEANS Clustering**

**e. XG-boost**

XGBoost, short for (extreme) gradient boosting, is a fast, portable, and distributed implementation of the gradient boosting (trees) calculation. XGBoost is a calculation that has as of late been ruling connected AI and Kaggle rivalries for organized or forbidden information. XGBoost is an execution of gradient boosted decision trees intended for speed and execution. XGBoost is a product library that you can download and introduce on your machine, at that point access from an assortment of interfaces.

**• XGBoost Features**

The library is laser centered around computational speed and model execution, in that capacity there are not many laces. All things considered, it offers various propelled highlights.

**• Model features**

The execution of the model backings the highlights of the scikit-learn and R usage, with new increases like regularization. Three principle types of inclination boosting are upheld:

- Slope Boosting calculation likewise called angle boosting machine including the learning rate.
- Stochastic Gradient Boosting with sub-inspecting at the line, section and segment per split dimensions.
- Regularized Gradient Boosting with both L1 and L2 regularization.

**• System features**

The library gives a framework to use in a scope of figuring situations, not least:

- Parallelization of tree development utilizing the majority of your CPU centers amid preparing.

## Prediction of Student Performance using Machine Learning

- Dispersed Computing for preparing exceptionally substantial models utilizing a group of machines.
- Out-of-Core Computing for extremely huge datasets that don't fit into memory.
- Store Optimization of information structures and calculation to utilize equipment.

### Algorithm

The usage of the calculation was designed for effectiveness of figure time and memory assets. A structure objective was to utilize accessible assets to prepare the model. Some key calculation usage highlights include:

- Scanty Aware usage with programmed treatment of missing information esteems.
- Square Structure to help the parallelization of tree development.
- Kept Training with the goal that you can additionally help an officially fitted model on new information.



Fig 4 :Features of XG-boost

### Algorithm used by XG-boost

- The XGBoost library actualizes the gradient boosting decision tree calculation.
- This calculation passes by bunches of various names, for example, gradient boosting, numerous added substance relapse trees, stochastic inclination boosting or angle boosting machines.
- Boosting is an outfit procedure where new models are added to address the blunders made by existing models. Models are included successively until no further enhancements can be made. A prevalent model is the AdaBoost calculation that loads information indicates that are hard foresee.
- Angle boosting is a methodology where new models are made that foresee the residuals or mistakes of earlier models and after that additional together to make the last expectation. It is called inclination boosting in light of the fact that it utilizes an angle plunge calculation to limit the misfortune while including new models. This methodology performs both regression and classification of predictive models.

### f. Naive Bayes

- It is a classification method with a supposition of autonomy among indicators. In straightforward terms, a Naive Bayes classifier accepts that the nearness of a specific element in a class is inconsequential to the nearness of some other element.
- Bayes model is easy to fabricate and especially valuable for huge data. Naive Bayes outperforms even highly sophisticated arrangement strategies.
- Naive Bayes calculations are for the most part utilized in sentiment analysis, spam filtering, suggestion frameworks and so forth.
- Naive bayes is quick and simple to execute however their

greatest detriment is that the prerequisite of indicators to be free. In a large portion of the genuine cases, the indicators are reliant, this ruins the execution of the classifier.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood
Class Prior Probability

Posterior Probability
Predictor Prior Probability

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Fig 5: Probability Function in naive bayes

### Steps to implement naive bayes algorithm

- Load the CSV information into Python and change all strings into floats. Separate all information into preparing information and test information. All things considered, the Naive Bayes approach makes forecasts from the preparation information, while it utilizes the test information to assess the precision of the model. For this situation, we have made 67% preparing information and 33% test information.
- Presently the time has come to examine the information. Every datum will be isolated dependent on the class it has a place with. From that point forward, we have to ascertain the mean and standard deviation for every one of the information inside each trait inside each class.
- Presently it an opportunity to utilize the preparation information to make expectations.
- We have to utilize the Gaussian Probability Density Function here.
- In light of the rundown of information for each class, comprising of the mean and change, in view of the info esteem, we process the probability that the esteem will have a place with a specific class (utilizing the Gaussian Probability Density Function).
- Our forecast will relate to the class with the biggest likelihood.
- At long last, we have to know how exact our gauge was. It is the rate right out of all expectations made.

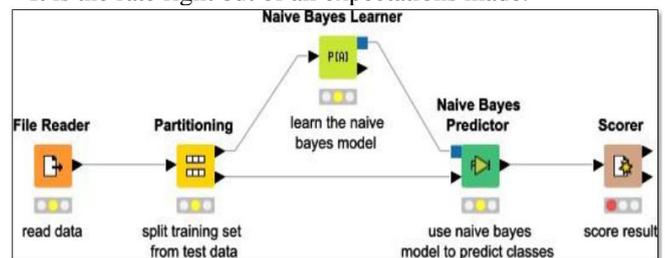


Fig 6: Steps illustrating naive bayes learner

### Advantages

- It is simple and quick to foresee class of test data. It likewise perform well in multi class forecast.
- At the point when suspicion of independence holds, a Naive Bayes classifier performs better contrast with other models like calculated relapse and you need less preparing information.

- It performs well with the case of categorical input variables with numerical variables. For numerical variable, typical appropriation is expected.

**Disadvantages**

- With the case of categorical variables (in test informational collection), which was not seen in preparing training set, then model will assign out a 0 (zero) likelihood and will be unfit to make an expectation. This is regularly known as "Zero Frequency".
- On the other side naive Bayes is otherwise called an awful estimator, so the likelihood yields from predict probabilities are not to be paid attention to as well.
- Another constraint of Naive Bayes is the presumption of free indicators. All things considered, it is practically inconceivable that we get a lot of indicators which are totally free.

**g. Logistic Regression**

- In statistics, the logistic model (or logit show) is a broadly utilized factual model.
- In its fundamental structure it utilizes a strategic capacity to demonstrate a twofold needy variable, albeit lot progressively complex augmentations exist.
- In regression examination, calculated regression (or logit relapse) is assessing the parameters of a strategic model (a type of binomial relapse).
- Numerically, a double strategic model has a needy variable with two conceivable qualities, for example, pass/fizzle, win/lose, alive/dead or solid/liquid; these are spoken to by a marker variable, where the two qualities are named "0" and "1".
- In the strategic model, the log-chances (the logarithm of the chances) for "1" is a linear combination of at least one independent variables ("indicators"); the independent variables can each be a binary factor (two classes, coded by a pointer variable) or a ceaseless variable.

**Advantages**

- We can effectively model x variables as probability function.
- Ols assumptions violations are not to be worried about
- Predictions always range between 0 to 1 only.

**Disadvantages**

- Simple interpretation of linear co-efficients is lost
- The effect is not suitable for non-linear model
- We cannot compute the same statistics( ex- r-square)
- This model is not very easy to interpret and implement

Fig 7: LOGISTIC REGRESSION

**IV. EXPERIMENTAL INVESTIGATIONS**

**Exploratory data analysis**

- Exploratory Data Analysis alludes to the basic procedure of performing introductory examinations on data
- EDA is used to find patterns to spot anomalies, to test theory and to check presumptions with the assistance of outline insights and graphical portrayals.
- It is a decent practice to comprehend the information first and gather the same number of experiences from it.
- EDA is tied in with understanding information in hand,

before getting them messy with it.

- EDA is a way to deal with investigating data sets to abridge their fundamental attributes, frequently with visual strategies.
- A factual model can be utilized or not, yet fundamentally EDA is for seeing what the information can let us know past the formal demonstrating or speculation testing task.
- Exploratory information investigation was elevated by John Tukey to urge analysts to investigate the information, and potentially define theories that could prompt new information gathering and examinations.
- EDA is unique in relation to initial data analysis(IDA) which centers all the more barely around checking assumptions required for model fitting and speculation testing, and taking care of missing qualities and making changes of factors as required. EDA encompasses IDA

**The targets of EDA are**

- Propose theories about the cause for observed phenomena
- Survey assumptions on which statistical inference will be based
- Support the choice of proper statistical instruments and strategies
- Give a basis to promote data collection through overviews or experiments
- Reveal hidden structure;
- Separate vital factors;
- Identify exceptions and anomalies;
- Test hidden suspicions;
- Create tightfisted models; and
- Decide ideal factor settings For information accumulation through studies or experiments

**Techniques**

- Most EDA methods are graphical in nature with a couple of quantitative systems. The explanation behind the substantial dependence on designs is that by its very nature the primary job of EDA is to liberally investigate, and illustrations gives the investigators unparalleled capacity to do as such, luring the information to uncover its auxiliary privileged insights, and being constantly prepared to increase some new, regularly unsuspected, knowledge into the information. In mix with the normal example acknowledgment capacities that we as a whole have, designs gives, obviously, unparalleled capacity to do this.
- The specific graphical methods utilized in EDA are frequently very basic, comprising of different systems of:
- Plotting the raw information, (for example, information follows, histograms, bi histograms, likelihood plots, slack plots, square plots, and Youden plots.
- Plotting straightforward measurements, for example, mean plots, standard deviation plots, box plots, and fundamental impacts plots of the crude information.
- positioning such plots to expand our regular example acknowledgment capacities, for example, utilizing various plots per page.



## 4.1 Philosophy

EDA isn't indistinguishable to statistical graphics in spite of the fact that the two terms are utilized conversely. Measurable illustrations is a gathering of procedures - all graphically put together and all centering with respect to one information portrayal perspective. EDA envelops a bigger setting; EDA is a way to deal with information investigation that defers the standard presumptions about what sort of model the information pursue with the more straightforward methodology of enabling the information itself to uncover its basic structure and model. EDA is certifiably not a minor gathering of methods; EDA is a theory with respect to how we dismember an informational collection; what we search for; what we look like; and how we translate. The facts demonstrate that EDA vigorously utilizes the accumulation of strategies that we call "factual designs", yet it isn't indistinguishable to measurable illustrations essentially.

## 4.2 Focus

The EDA approach is definitely that - a methodology - not a lot of strategies, however a frame of mind/theory about how an information examination is done.

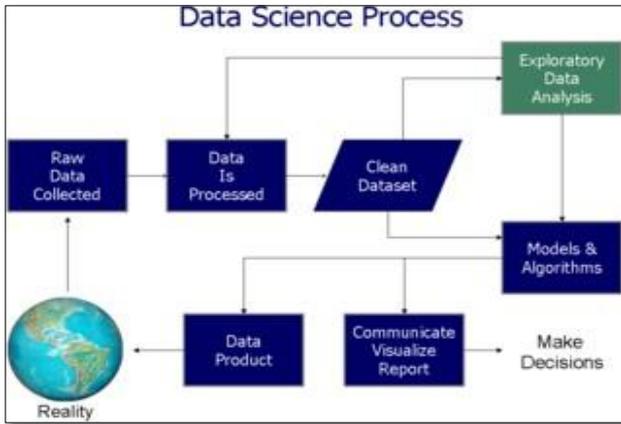


Fig 4: Data Science Process

## V. EXPERIMENTAL RESULTS

### Libraries used

- NumPy is a library for the Python programming language, including support for expansive, multi-dimensional exhibits and grids, alongside a vast gathering of abnormal state scientific capacities to work on these clusters.
- Seaborn is a Python information perception library dependent on matplotlib. It gives an abnormal state interface to drawing appealing and useful measurable illustrations.
- In PC programming, pandas is a product library composed for the Python programming language for information control and examination. It offers information structures and activities for controlling numerical tables and time arrangement
- Matplotlib is a plotting library for the Python programming language and its numerical science augmentation NumPy. It gives an item arranged API to implanting plots into applications utilizing broadly useful GUI toolboxes like Tkinter, wxPython, Qt, or GTK+
- Scikit-learn is a free programming AI library for the Python programming language. It highlights different characterization, regression and clustering calculations including support vector machines

- XGBoost is an open-source programming library which gives a slope boosting system to C++, Java, Python, R, and Julia. It takes a shot at Linux, Windows, and macOS. From the venture portrayal, it plans to give an "Adaptable, Portable and Distributed Gradient Boosting Library"
- In statistics, the logistic model is a generally utilized factual model. In its essential structure it utilizes a strategic capacity to display a binary reliant variable, although a lot increasingly complex extensions exist. In relapse examination, strategic relapse is assessing the parameters of a calculated model
- Naive Bayes classifiers are a gathering of arrangement calculations dependent on Bayes' Theorem. It's anything but a solitary calculation however a group of calculations where every one of them share a typical standard, for example each pair of highlights being grouped is autonomous of one another.

## VI. RESULTS

### 5.1 Import the data and clean the data

```

In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: import plotly.plotly as py
from plotly.offline import init_notebook_mode, init
init_notebook_mode(connected=True)
import plotly.graph_objs as go

In [3]: import os
data_dir = os.path.join('C:/Users/Kalyan/Desktop/Minor project/stu_ds_csv')
data = pd.read_csv(data_dir)

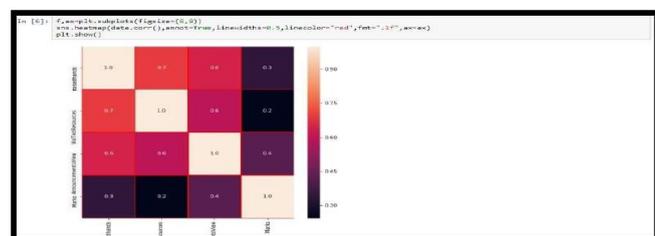
Out[3]:
  gender  nationality  handedness  stageID  genderID  sectionID  topic  knowledge  relation  raisedhands  VisitedResources  AnnouncementsView  Marks  %
0  M  KWI  Right  Intermediate  Q04  A  IT  F  Father  16  18  3  20
1  M  KWI  Right  Intermediate  Q04  A  IT  F  Father  20  20  3  25
2  M  KWI  Right  Intermediate  Q04  A  IT  F  Father  10  7  0  30
3  M  KWI  Right  Intermediate  Q04  A  IT  F  Father  30  25  5  35
4  M  KWI  Right  Intermediate  Q04  A  IT  F  Father  40  50  12  50
    
```

```

In [4]: data.info()
Out[4]:
Out[4]:
  count  480.000000  480.000000  480.000000  480.000000
  mean    46.775000  54.797917  37.918750  43.283333
  std     30.779223  33.080007  26.611244  27.637735
  min      0.000000  0.000000  0.000000  1.000000
  25%     15.750000  20.000000  14.000000  20.000000
  50%     50.000000  65.000000  33.000000  39.000000
  75%     75.000000  84.000000  58.000000  70.000000
  max    100.000000  99.000000  98.000000  99.000000
    
```

	raisedhands	VisitedResources	AnnouncementsView	Marks
count	480.000000	480.000000	480.000000	480.000000
mean	46.775000	54.797917	37.918750	43.283333
std	30.779223	33.080007	26.611244	27.637735
min	0.000000	0.000000	0.000000	1.000000
25%	15.750000	20.000000	14.000000	20.000000
50%	50.000000	65.000000	33.000000	39.000000
75%	75.000000	84.000000	58.000000	70.000000
max	100.000000	99.000000	98.000000	99.000000

### 5.2 Generate Correlation Matrix







## ACKNOWLEDGEMENT

One of the authors Anusha M is very much thankful to Department of Science and Technology (DST), Government of India, New Delhi, for awarding her with a Women Scientist's scheme under DST – WOS (A) File no:- SR/WOS-A/ET- 1071/2015 Programme.

## REFERENCES

1. O. M. N. Alfian E, "Undergraduate Students' Performance: a case of University Malaya," Emerald Quality Assurance in Education vol 13 no 4, vol. 13 2005.
2. R. D. Ibrahim Z, "Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression, 2007," in 21st Annual SAS Malaysia Forum, 5th September 2007, Shangri-La Hotel, Kuala Lumpur, 2007.
3. W. M. C. Heather J, Nicola Price, "Predictors of academic success in a Bachelor of Nursing course," Elsevier Sciences, Nurse Education Today, vol. 23, pp. 246-254, 2003.
4. M. A. Wang T, "Using Neural Networks to Predict Student's Performance," in Proceedings of the International Conference on Computers in Education (ICCE'02), 2002.
5. I. Z. Rusli N M, Janor RM, "Predicting Students' Academic Achievement: Comparison between Logistic Regression, Artificial Neural Network, and Neuro-Fuzzy," in International Conference on Computers, 2008.
6. Arsad P M, Buniyamin N J A Manan, "Neural Network Model to Predict Electrical Students' Academic Performance," in International Congress on Engineering Education, ICEED 2012, Park Royal Penang Malaysia, 2012.
7. SPSS, "SPSS and Linear Regression <https://statistics.laerd.com/spss-tutorials/linear-regression-using-spss-statistics.php>. Accessed 16 August 2013," Accessed 16 August 2013.
8. Arsad P M, Buniyamin N, Rasheid N A, "Evaluation of Academic Performance of Electrical Engineering Bachelor Degree Students from Various Entry Levels: a Case Study," in 9th WSEAS International Conference on Education and Educational Technology (UDU'10), IWATE Japan Dec 2010 Iwate Prefectural University Japan, 2010.
9. Arsad P M, Buniyamin N, Manan J A, Hamzah N, "Proposed Academic Students' Performance Prediction Model: A Malaysian Case Study," in International Congress on Engineering Education (ICEED), 7-8 December 2011, Malaysia 2011 3rd, 2011.