Objective-: Given a character of learner and e-learning platform determine the probability that the course will run successfully.

Introduction- E-learning refers to "All the computer-management-based teaching and learning activities carried out in the information technology environment constructed by different transmission networks with communication function" [29].
The Development in technology, internet, infrastructure, and service sectors have made E-Learning very popular. The introduction of E-Learning has instilled a curiosity in the minds of young and adult-learners and the possibilities it can help us achieve breaking the traditional teaching methods that are followed saving money, space and time of all the stakeholders in the value chain.

There are different ways in which students prefer to learn. In the last decade we have seen a boom in the space of Ed-Tech startups wanting to democratize learning. In the future we would find that a degree may no longer stand the same importance as it does today, Education with the same content would be accessible to everyone and available online. Among this the method to detect the optimal way in which the information can be delivered by making analogies through dependencies on each other is getting prominence and helping people to make better responsive decisions.
The modernization in our current education system has given rise to a new era of E-Learning systems built through Software Technology. The past years have seen a rising trend in E-Learning Sector as well as the Learning industry. The grants given by governments and acknowledgements by incuabtors and accelarators have further boosted the cause of people wanting to pursue a career in Ed-Tech industry. This has led to a rise in the herd mentality products, startups and systems in this space that only give importance to meny without realizing whether they are able to tap into the student's mind and able to deliver and provide educational content to the student in a way that suits his interests making his learning smooth and efficient. By giving importance to his learning and behavioral patterns which sadly a remote teacher may not be able to understand. The loopholes based in this industry gave rise to a new generation of systems known as adaptable learning systems.

There are a lot of different features provided by an E-Learning platform and different ways in which the teachers can make its maximum utilization to deliver content to students. The method of delivery of the teacher and compatibility of the student with that along with the student's prior preparation and eagerness to learn along with other factors have been deeply discussed in this paper. The recent rends show an increase in the research areas of Students and Learning Analytics. Recent Studies such as using Deep Learning to provide a personalized E-Learning Resource Platform according to the user preference have been automating the old traditional processes [6].
According to Luis, Anna and Jon in 2016 *"The research carried out in the field of Virtual Learning Centre at higher education showed the potential of this learning community and, in addition, identified the new roles that the members involved in the community play "[8].*
They used a Bayesian Network for estimating the reputation in the Virtual Learning Centre.
By using the data available to them and predicting the trust and reputation relationships values.
The learning sector possesses vast amounts of records of the student data. The usage of learning analytics by the educational institutions has gained prominence due to its effectiveness in helping the institute make well-informed decisions. The advent of Artificial Intelligence has provided

humongous scope in improvement of E-Learning Platforms by providing intelligent and interactive environments to the students. It has made it possible to capture the data in real-time. This study aims to use machine learning algorithms for analyzing the smooth operation of a course given a character of a student and E-Learning Platform on standardized data collected from numerous students to predict the outcome of a course running successfully and helping the instructor to design the course in a better way. The application of this type of a procedure needs a directed mathematical model that is probabilistic in nature. Furthermore, we use a Naive Bayes network which each variable to be independent and is trained on the student learning characteristics and E-Learning Platform attributes.

The paper is further classified as follows In Section 2; We review Literature related to Current trends in E-Learning Analytics. In Section 3, we discuss the method of study used. In section 4, we present the results of the output. In section 5, Discussions we interpret and describe the significance of our findings. The section 6 provides the Conclusions.

Literature Review- The Use of Machine learning models are currently gaining popularity in E-Learning. People are looking forward to making efficient time saving models that save the instructors as well as the students effort by automating the feedback process and make the entire process smoother [6].

In 1996 Zhang and Poole stated "*Bayesian networks aid in knowledge acquisition by specifying which probabilities are needed. Where the network structure is sparse, the number of probabilities required can be much less than the number required if there were no independencies. The structure can be exploited computationally to make inference faster*"[10]

According to Luis, Anna and Jon "*A BN in general is a relationships network that uses statistical methods to represent probability relationships between different nodes. It is a compact representation of the joint probability distribution to reason under uncertainty*" [9].

Given a set of probabilities and finding the set of probabilities which would dependent on other variables also known as conditional probabilities was explained by Shao-Zhong Zhang, Hong Yu, Hua Ding, Nan-Hai Yang and Xiu-Kun Wang in 2003 " *A Bayesian network with a sets of variables {x1x2,…,xn} is consist of two parts. 1) A network denoting conditional independent supposition X. 2) a local probabilistic distribution set P, which contacts with each variable. S is a directed acyclic graph. The node in S corresponds each variable in X. the area between two nodes represents conditional independence.*"

The joint probabilistic distribution given certain conditional independences is given in the figure below:

$$p(x_1, x_2, \ldots, x_n|\zeta) = \prod_{i=1}^{n} p(x_i|x_1, x_2, \ldots, x_{i-1}, \zeta)$$

Fig1:A local probabilistic Distribution

Zhang and Poole described the conditional dependence such that "A *BN can be viewed as representing a factorization of a joint probability. For example, the Bayesian network in Figure*

1 factorizes the joint probability P (a; b; c; e1; e2; e3) into the following list of factors: P (a); P (b); P (c); P (e1ja; b; c); P (e2ja; b; c); P (e3je1; e2)."[10]
The use of Bayesian network as a forecasting mechanism was briefly given in the paper written by Luis, Anna and Jon "Bayesian Networks (BN), known as probabilistic models or belief networks, have been investigated due to a growing interest in predicting future events. BN is circumscribed, as forecasting technique, whose main characteristic is the valuation or qualification observed facts or data. Its role as a forecasting mechanism is very important as it allows inferences about the probability of occurrence of a given event on the basis of observed evidences."

Zhang and Poole also defined a process of computing posterior probability which was known as inference The theory of inference was "In theory, P (XjY =Y0) can be obtained from the marginal probability P (X; Y), which in turn can be computed from the joint probability P (x1; x2; : : : ; xn) by summing out variables outside X[Y one by one. In practice, this is not viable because summing out a variable from a joint probability requires an exponential number of addition".[10]

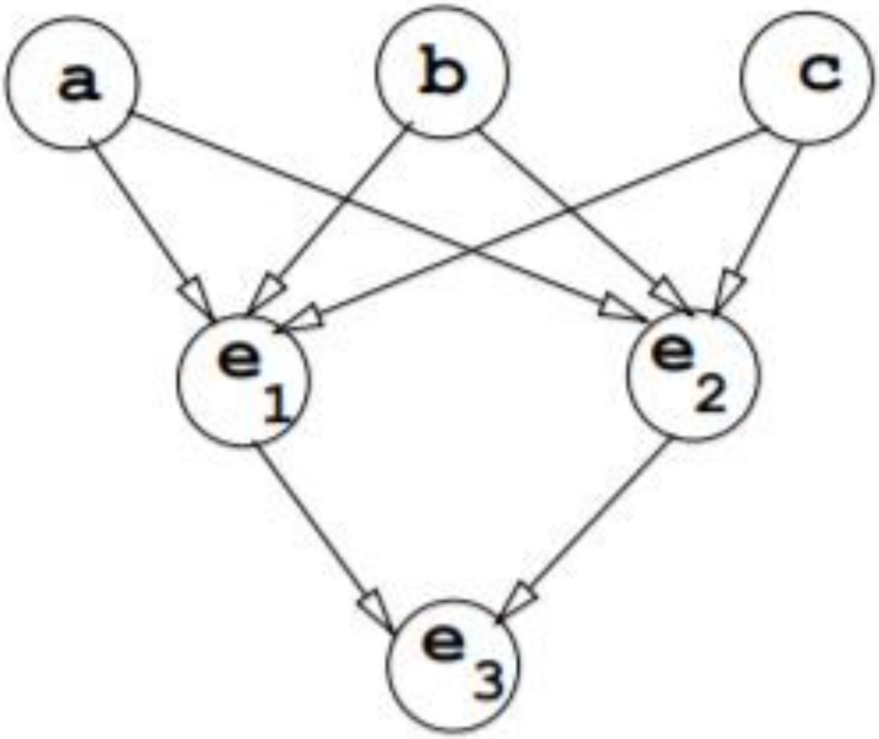


Fig2: A Bayesian Network

A Microsoft paper published in March 1995 by Heckerman explains as to why use a Bayesian network and why it would be a perfect fit for the data collected by us. According to Heckerman "*A Bayesian network is a graphical model that encodes probabilistic relationships among variables of interest. When used in conjunction with statistical techniques, the graphical model has several advantages for data analysis. One, because the model encodes dependencies among all variables, it readily handles situations where some data entries are missing. Two, a Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding* *about a problem domain and to predict the consequences of intervention. Three, because the model has both a causal and probabilistic semantics, it is an ideal representation for combining prior knowledge (which often comes in causal form) and data. Four, Bayesian statistical methods in conjunction with Bayesian networks over an efficient and principled approach for avoiding the overtting of data.*” [1].

In a Bayesian Network there are 2 cases that can happen condition dependence and conditional independence. In the condition dependence case if there is a change in the node information it would also lead to a change in the distribution of the probabilities it relates to. Whereas in the case of Conditional Independence it is mutually exclusive to the events happening in the other nodes. Figure given below gives us the mathematical relationships for independent probabilities. “*For each variable xi, let πi⊆{x1,x2,…,xi−1} is the parent's node of xi, and {x1,x2,…,xi−1} is conditional independence, then:*”[7]

$$p(x_i|x_1, x_2, \ldots, x_{i-1}, \zeta) = p(x_i|\pi_i, \zeta)$$

Fig3: Condition Probability formula

In the paper published in 1996 by Zhang and Poole a more efficient way of analyzing and finding an inference was stated “*in the concept of factorization. A factorization of a joint probability is a list of factors (functions) from which one can construct the joint probability. A factor is a function from a set of variables into a number. We say that the factor contains a variable if the factor is a function of that variable; or say it is a factor of the variables on which it depends.”*

Colace and Desanto in 2006 described a Bayesian Network specialization that was dynamic. According to them “**A** specialization of Bayesian networks are those named dynamic. They work with two copies of standard Bayesian networks. In particular, one represents the network in the instant in consideration (T), while the other one represents the network at the following slot time (T+1). When a dynamic Bayesian network records new evidence, the latter is added to the slot time T and through the inference process, node's values at the slide at time T+1 are calculated and the “roll-up” happens. During the “roll-up”, the slide at time T is erased, the slide at time T+1 becomes the new relative slide at time T and a new copy of the network is created, which identifies itself with the slide at time T+1. By this way, a DBN is able to model some changes during the passing of time.” [12]

In the empirical analysis done in 2019, Kondo and Hatanaka used a Bayesian Network to find out the learning states of the students which provided feedback to the instructors of the students that were likely to get a lower grade. Using Learning analytics for this type of information was beneficial for both the student and the instructor and made things transparent [2]. However, for correctly analyzing any model, Data is the most important aspect. The revolutionary breakthrough with the help of a 4 level learning progression model was used by West, P., Rutstein, D. W., Mislevy, R. J., Liu, J., Choi, Y., Levy, R., … Behrens, J. T. (2010) in their paper to get the optimal Bayesian network score[5]. To prove this point further Baisakhi Chakraborty and Meghamala Sinha (2016) worked further in this domain to come up with an effective 4 component model. Which gave out a huge scope for study material recommendation by evaluating the learning style of the students from the materials browsed and his test performance. A major innovation proposed by them is "*Extending the proposed Bayesian network to a Dynamic Bayesian Network (DBN) which can update student's knowledge over long time spans.*" [4]. The paramteres used to collect the data and directing the nodes and usage were also very impactful while running the Network, Patricio García, Analía Amandi, Silvia Schiaffino, Marcelo Campo (2005) listed down the attributes that could be considered to predict the students learning state with a very interesting approach. In their problem," *random variables represent the different dimensions of Felder's learning styles and the factors that determine each of these aspects. These factors are extracted from the interactions between the student and the web-based education system." [3].*

An Adaptive Learning algorithm for course learning system was built by Guan, Jia "*Constructed by Bayesian Network; and then the prior probability table of influence degree between nodes is obtained deductively through the learners' user profile and Bayesian Network; lastly, adaptive learning path suitable for different learners is generated according to learners' ability diagnosing algorithm, so as to achieve adaptability learning.*"

Zhang and Zhuang in 2007 proposed an ITS (Intelligent Tutoring system) that takes into account the pedagogy of Adaptive Learning the following system was proposed "*for achieving the adaptive learning, assessment results should offer accurate and detail feedback in accordance with student's aptitudes and learning results [1] [2][3] [4]. However, it is difficult and time consuming to assess student's knowledge level or learning status for the teachers manually. Thus, how to automatically diagnose the cognitive state of students from observable data (test results) becomes an interesting issue. Moreover, how to provide students with learning guidance and help after knowing their learning status is also worth further research.*

Most conventional assessment systems measure how much a student knows. Our assessment system determines what a student knows by BNs. This information is useful for an assessor to make decisions for next step education or learning." [11]

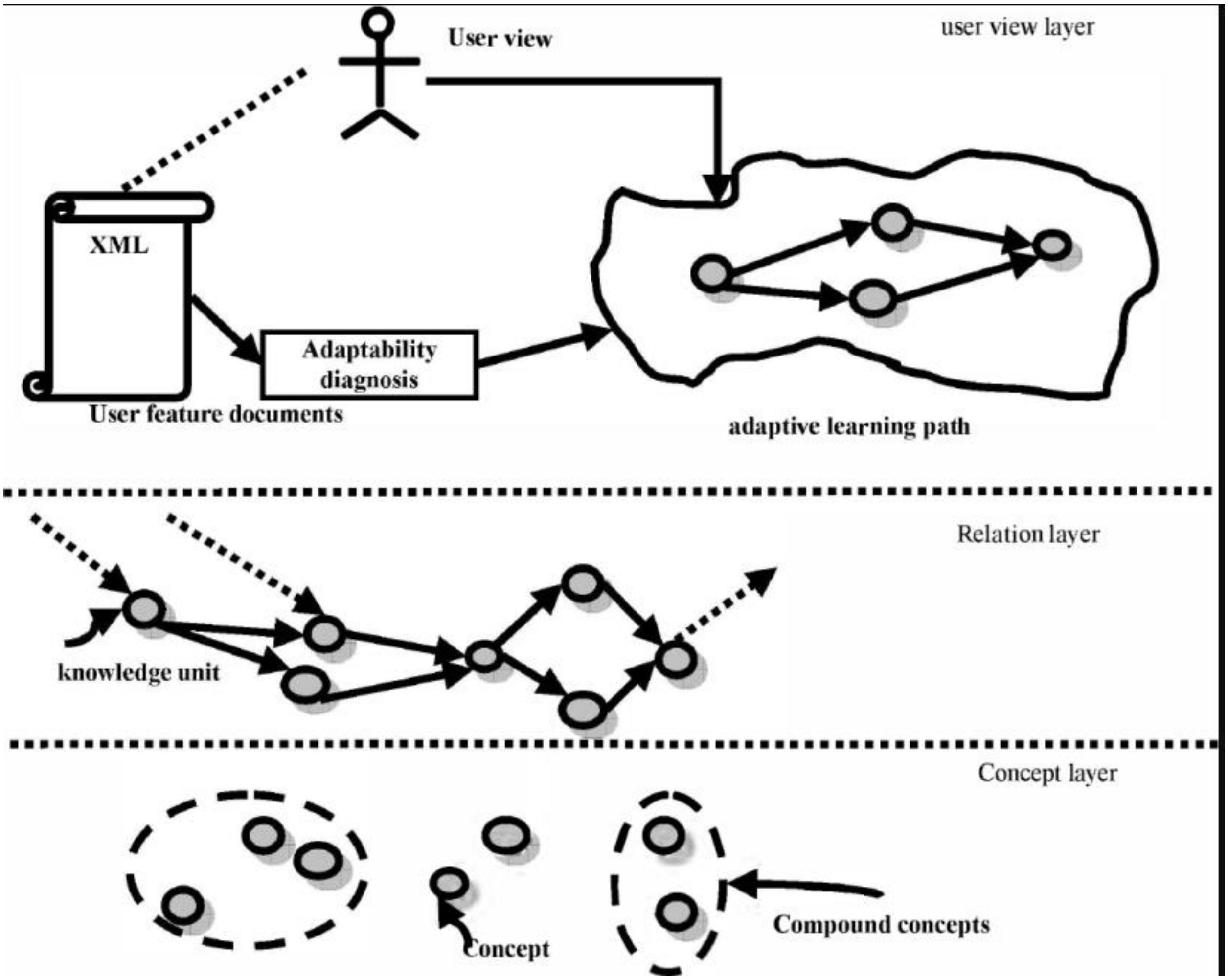


Fig4: An Adaptive E-Learning System

Trang Nguye *et al*. [22] propose a Bayesian trust model to evaluate the performance or availability offered by the reputation Web Services. They use a subjective point of view based on a user scoring system and the service quality monitoring as an objective point of view.

López-Faican *et al*. [23] describe the use of BN to implement a model of uncertainty to predict the student learning style through interaction in a Virtual Learning Environment based on the Felder-Silverman model. The uncertainty model is designed and developed for Moodle Learning Management System.

Regina Stathacopoulou [18] proposes a neural network implementation for a fuzzy logic–based model of the diagnostic process. The neuro–fuzzy synergy allows the diagnostic model to some extent imitate teachers in diagnosing student's characteristics and equips the intelligent learning environment with reasoning capabilities.[18]

In 2009 Kao and Liu proposed an analysis of the Bayesian Networks in E-Learning System where the relative efficiency of the system was reviewed rather than the output. All the grouped users also helped to obtain the efficiency of the entire system. The results obtained from the study was that on the middle class school teachers the system proved to be the most efficient among all the user groups.[13]

This study proposes data envelopment analysis and Bayesian networks in e-learning systems evaluation and classification, respectively. There are several distinguishing features of this study. First, we evaluate the e-learning system from the perspective of relative efficiency, instead of that of output, which is more compatible with investors' standpoint. Second, in the DEA model, the DMUs are the reviewers, whose outcomes (efficiencies) are used to estimate the relative efficiency of the e-learning systems. In this approach, the systems' efficiency can be obtained by all users or grouped users. We find that for middle school teachers, the systems are comparatively efficient. Third, this study proposes the Bayesian network classification model so that the relative efficiency of future systems can be foreseen.[13]

Daniel *et al*. [24] defined a Bayesian computational model in the field of social capital theory that generates conditional probability tables to be evaluated and improved by experts in the application of social capital in Virtual Communities.

Qi *et al*. [25] propose a novel trust model based on Bayesian approach for web-based systems. The relationships between entities are classified into 4 kinds according to what if there are recommendations and/or direct interactions.

Li *et al*. [26] propose a trust model using a BN for e-commerce using conditional probability tables to assess the impact of the factors considered in e-commerce transactions.

Jøsang *et al*. [27] present an overview of existing systems and proposals that can be used to derive measures of trust and reputation of Internet transactions. They propose a use the aggregated ratings about a given party to derive a trust or reputation score, which can assist other parties in deciding whether or not to transact with that party in the future.

Patel *et al*. [28] developed TRAVOS (Trust and Reputation model for Agent-based Virtual Organizations) which models an agent's trust with an interaction partner. Specifically, trust is calculated using probability theory based on past interactions between agents. When there is a lack of personal experience between agents, the model draws upon reputation information gathered from third parties.

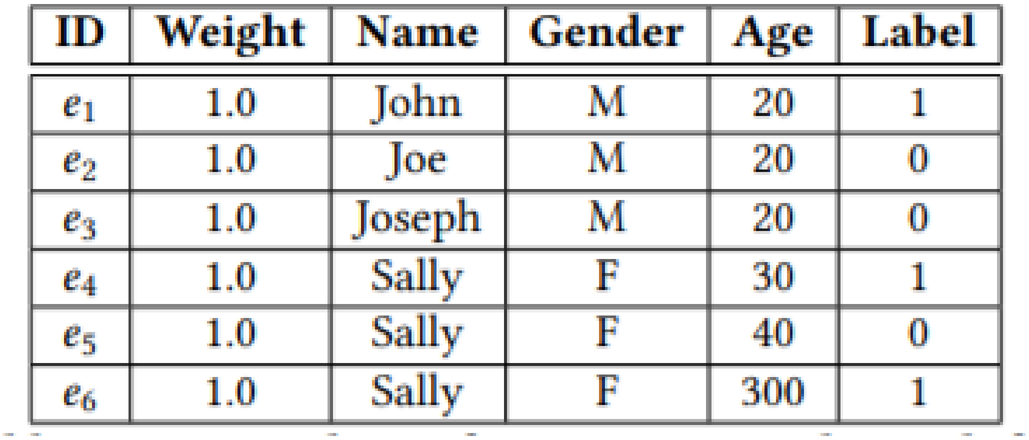
Aciar *et al*. [29] described a recommender system where the user recommendations are made considering the degree of knowledge and user availability to answer questions from other users. The reputation is calculated based on past interactions, more precisely the satisfaction of the user who made the question.

Shubaswini and Sharmila focused on proposing a system that minimizes the time spent on data pre-processing by making the entire process simpler, intuitive and easy in turn giving better results on the models They stated that success of a M.L. algorithm depends on the "*amount of good quality data that is given to it. But this process of cleaning may not be considered as a main area in processing and most often they are not mentioned but it is critical when comes to providing predictions based on the data. The system that uses powerful algorithms to process the noisy data can yield bad results if irrelevant or wrong training of data is given. Machine learning comes into picture when the whole process of splitting the corrupted data from the good data is done in a large amount of time.*"[16]

Deshmukh and Wangikar in 2011 tried analyzing the best algorithmic method for data cleaning but were unsuccessful in coming up with a concrete universal method that would be applicable for everyone. According to them it would be an intuition-based trial and error method to find the best fit for your algorithm.

According to them the Anomalies and common data problems included " *Common data quality problems(anomalies) include inconsistent data conventions amongst sources such as different abbreviations or synonyms; data entry errors such as spelling mistakes inconsistent data formats, missing, incomplete, outdated or otherwise incorrect attribute values, data duplication, irrelevant objects or data. Data that is incomplete or inaccurate is known as dirty data. The various types of anomalies occurring in data that must be eliminated. The type of anomalies can be classified under several types of it. Based on this classification we evaluate and compare existing approaches for data cleansing with respect to the types of anomalies handled and eliminated by them.*" [14]

Tae and Roh in 2019 mentioned the importance of the data preprocessing and the less understanding due to less research in this area. The discussion of a framework for cleaning the data in a unified way to get robust algorithms was discussed by them.[15]



ID	Weight	Name	Gender	Age	Label
e_1	1.0	John	M	20	1
e_2	1.0	Joe	M	20	0
e_3	1.0	Joseph	M	20	0
e_4	1.0	Sally	F	30	1
e_5	1.0	Sally	F	40	0
e_6	1.0	Sally	F	300	1

Fig5: An Uncleaned Dataset Table

The table in Fig-8 shows an uncleaned dataset where e2 and e3 are duplicates which affects fairness and introduces a bias that affects training of the model. Then further e6 has a unrealistic age and has to be discarded since a person can never be 300 years old.

An initial set of training examples with features for predicting whether a person will have high income. The data is not clean (e2 and e3 are duplicates), which may introduce bias that affects model fairness. In addition, e6 has an anomalous age.

In the method of traditional data cleaning, "duplicates must be removed, and values need to be corrected to be within certain ranges or to exist in external data sources. More recently, there are efforts to improve machine learning accuracy and data validation techniques for machine learning pipelines. However, these techniques do not resolve the pressing issues of model fairness or model robustness against adversarial data."[15]

The method of dealing with unprocessed data was discussed briefly in the paper authored by Shubaswini and Sharmila in a set of 5 modules which were to followed in a numeric order while preprocessing and could be applied to almost to any dataset "*In the first module, some columns may contain less information or no information at all, which makes it hard to rely on such columns for analysis and so such columns can be removed provided that they don't cause significant damage to the process. In the second module, some rows may contain empty fields which will again tamper with the proper preprocessing of the dataset. Hence such values are identified and removed. In the third module, the dataset will contain categorical features ranging from numerical to non-numerical values. This application requires only numerical data which is used for analysis and prediction. So, the fields containing numeric values are identified.*

In the fourth module, we deal with missing values which occur for a multitude of reasons — ranging from human errors during data entry, incorrect sensor readings to software bugs in the data processing pipeline. It is probably the most widespread source of errors and the reason for most of the exception-handling. If you try to remove them, you might reduce the amount of data you have available dramatically. So, these fields need to be filled in with appropriate values. In the fifth module, we deal with outliers which are those data points that are really far from the rest of your data points. Mathematically, an outlier is usually defined as an observation more than three standard deviations from the mean. They can show up due to errors in data entry or measurement, or just because there's a variation in the population. Identifying and handling outliers is an important part of data cleaning."

The proposed system could save a lot of time in cleaning the unclean raw data and in case of big data is very effective in optimizing the entire preprocessing step.

The paper published by Zhang and Zhuang briefly discussed the 4 types of architecture models present in a given system comprising of a monitoring, grading monitoring and grading monitor that worked as whole by complementing each other made the students give their tests, automatically assigned scores to them and then analyzed the points and concepts they were lacking in.

Nonetheless the usage of Learning analytics remains one of the hot topics of the 21st century gaining popularity among researchers yet remains to be explored fully.

Methodology- We would be using a Bayesian Network “A type of probabilistic graphical model comprised of nodes and directed edges.” [1]. We first design a Bayesian Network given relationships in between different variables and calculate the probabilities given the occurrence of those events. The model is developed assuming random independent variables. We assume a fully independent model. Now we used Nodes as random variables and edges as relationships between the random variables while making the graphical model.

Luis, Anna and Jon used a similar type of correlation to measure the reputation by considering the activities and resources used by the students “*An aggregation algorithm [31] adapted to the VLC area, calculates the direct experience considering the interaction of members of the VLC with resources and learning activities managed in an LMS. Concretely, the algorithm considers the “I like” actions (positive reinforcement) and “I don't like” actions (negative reinforcement) that each member performs on the resources/activities used and managed by the LMS.*”[8]

A lot of student data can be extracted from the Learning Management System of an institute in the form of demographics, Performances up to date, login, and registration data. We would be using similar type of sub-attributes of the E-Learning System collected from the Students.

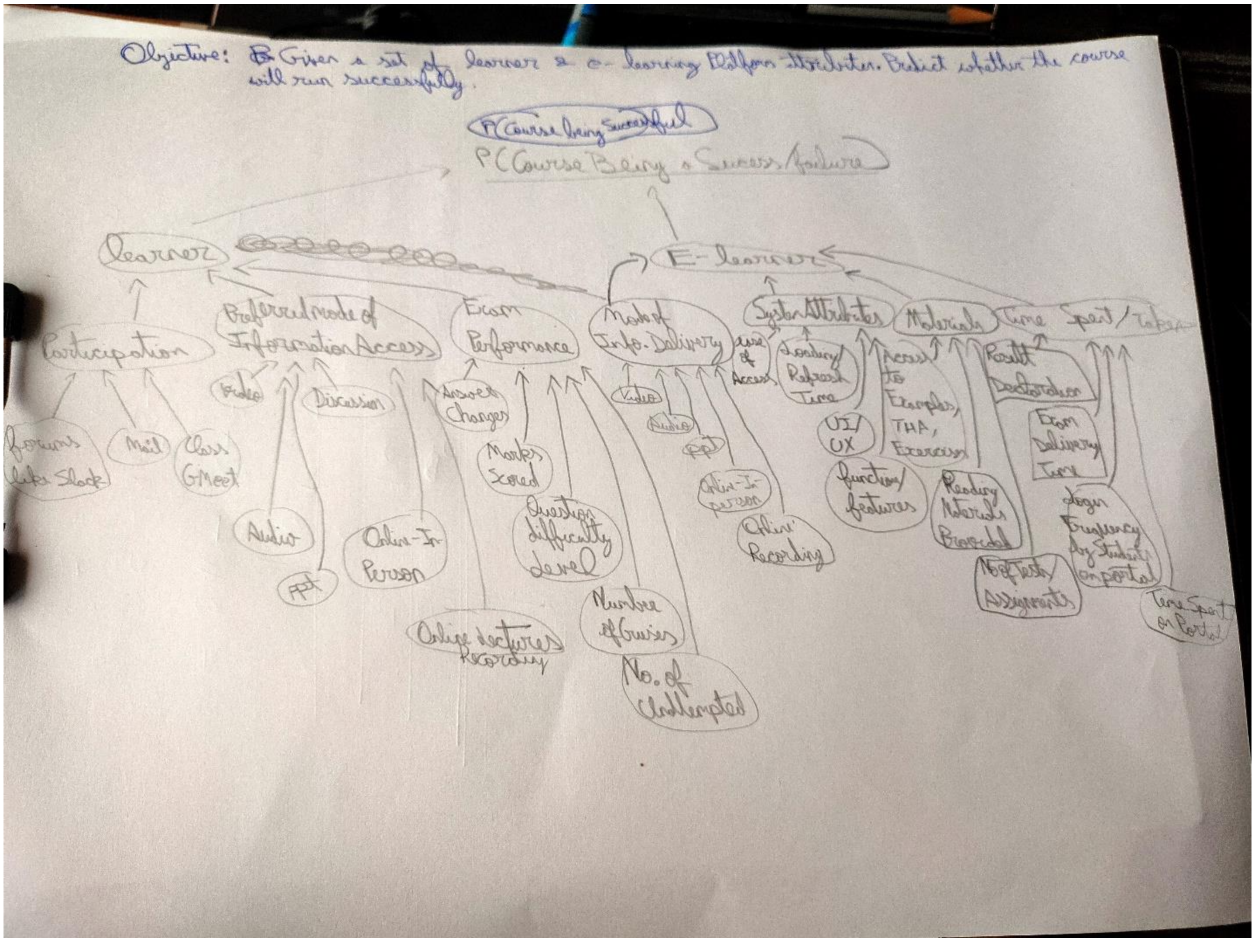
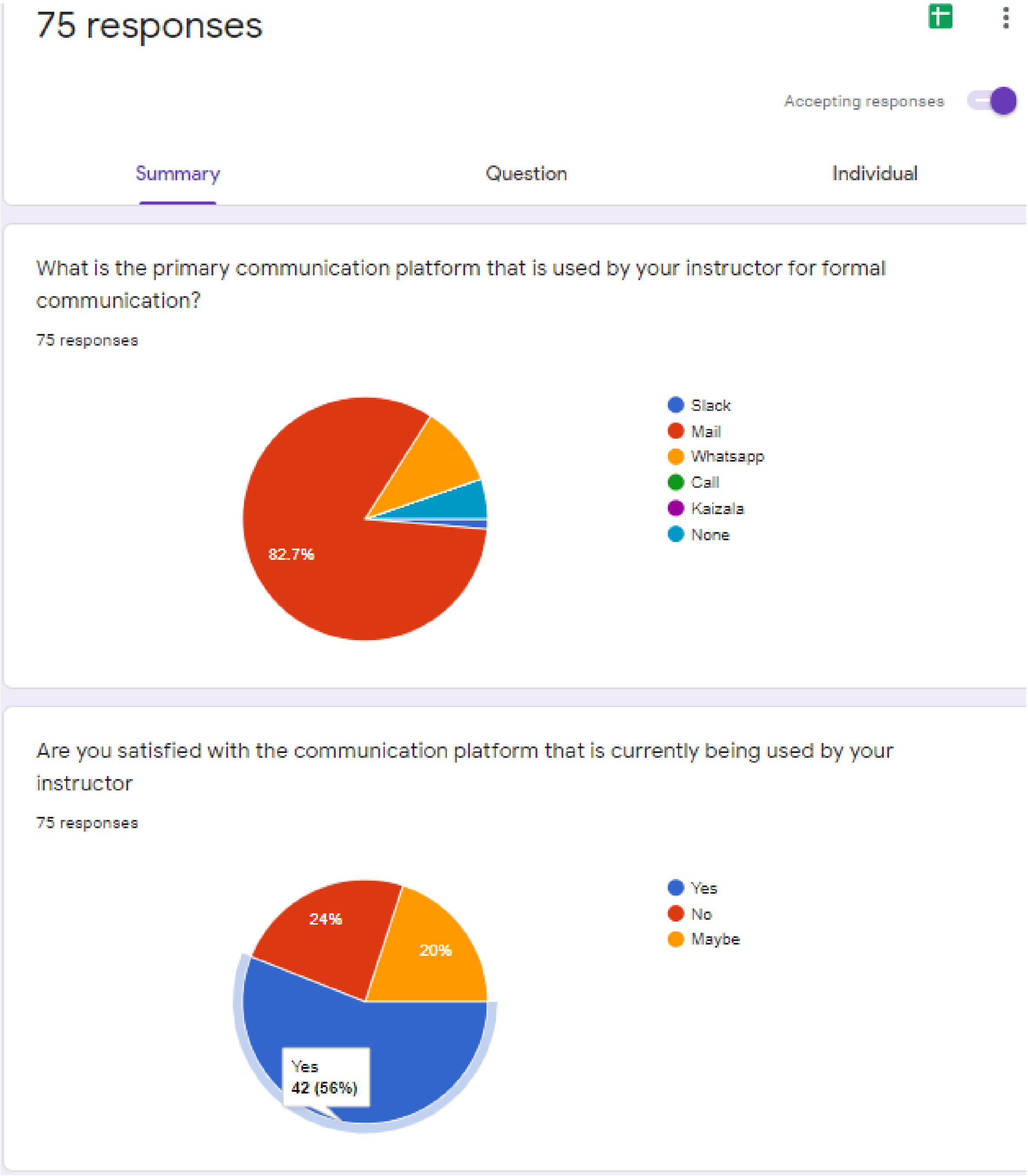


Figure6-Graphical model for a Bayesian Network

After preparing the model of the graph, it is used for the reasoning purposes given the occurrence of certain instances within the model predict the probability for the further outcome.

For this study we picked 1 Electrical Electronics Department course ADVD for the study to be conducted upon. For the students the data was categorized and collected through floating a google form which had 4 main attributes as shown in the figure-1 followed by several sub-attributes on the basis of which the network was trained and inferences drawn.



75 responses

What is the primary communication platform that is used by your instructor for formal communication?

75 responses

Are you satisfied with the communication platform that is currently being used by your instructor

75 responses

Fig7: The Collected Gform Data

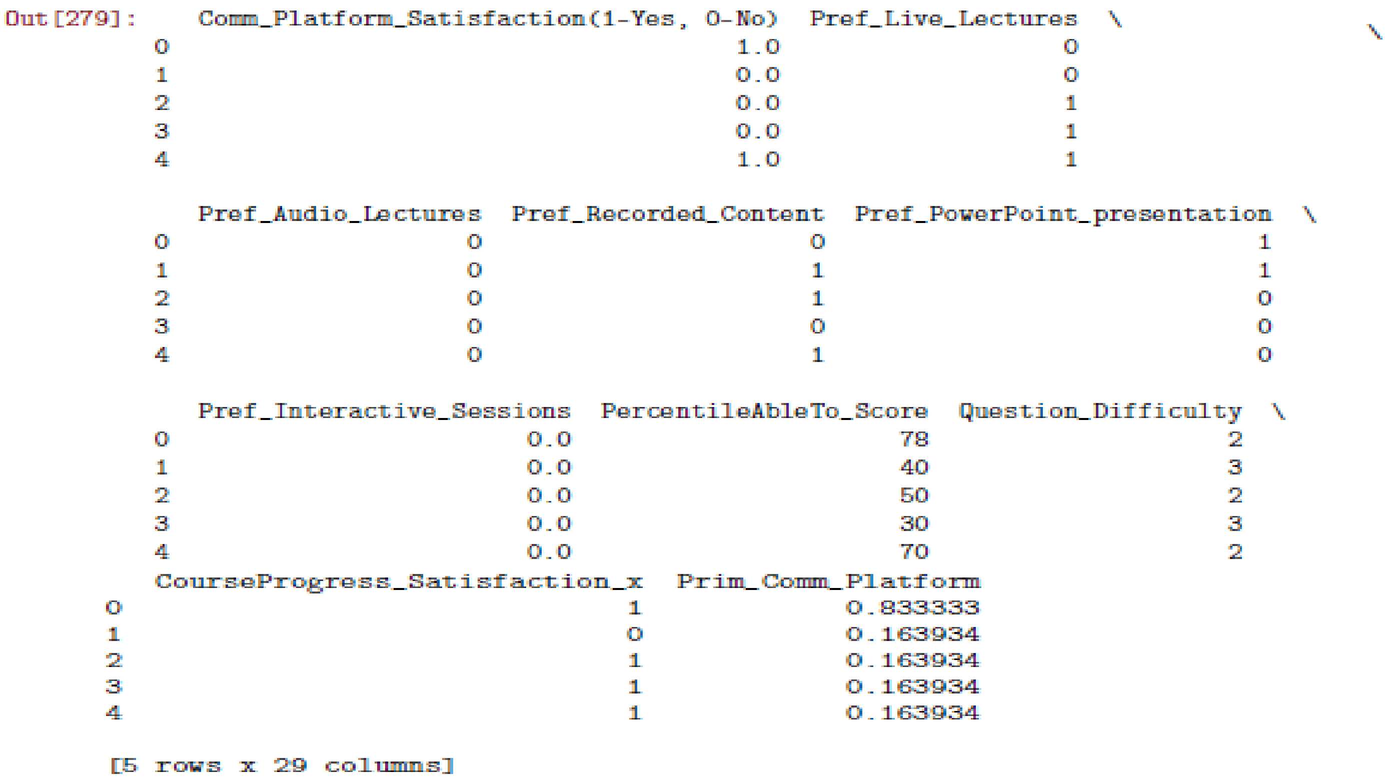
The collected data about the students and the platform attributes would be refined using several methods of cleaning, encoding and normalizations, According to Zhuang and Zhang 3 types of models were proposed that could be trained by the Bayesian Networks the first one being data centric model, second one being an efficiency centric model and the third one being an expert centric model.

A Survey was circulated to assess the learning styles of different students consisting of certain questions that tested the learning styles of the students and his interest in that course. Further the sum and averages of all the questionnaire answers were computed to draw conclusions.

We analyze and collect data as to how the student interacts with the system to learn the student's learning style. The algorithm then gives the probabilities providing useful assistance to the student and instructor through suggestions of more Take Home Assignments, Reading exercises/problems according to his/her preferred learning styles. After collecting the data, we preprocess it in excel. We then used Jupyter Notebook for running this model. We analyzed our dataset using Describe function to remove any unwanted, NA values or Empty Spaces. We then used One Hot Encoding, Binary Encoding and Label Encoding on the Categorical Variables Dataset to make the Dataset consistent for processing. We normalized our data and made it compatible with the algorithm to be trained upon. The Data was then manually as well

algorithmically scanned to remove ambiguous or extreme values that could disturb the model predictions.

We then renamed the columns to make the code cleaner. The Dataset was then split into 25 percent test and 75 percent train. Further The fundamental Naive Bayes algorithm was applied to our dataset with the assumption that each feature has an independent and equal outcome contribution. Each variable is taken to be equally contributing to the output while processing



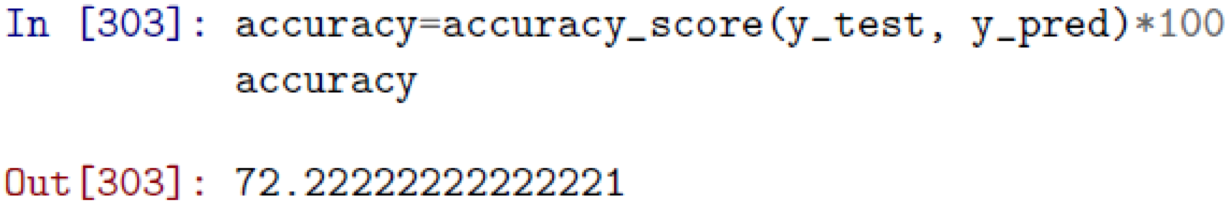
```
Out[279]:    Comm_Platform_Satisfaction(1-Yes, 0-No)  Pref_Live_Lectures  \                 \
          0                                      1.0                   0
          1                                      0.0                   0
          2                                      0.0                   1
          3                                      0.0                   1
          4                                      1.0                   1

             Pref_Audio_Lectures  Pref_Recorded_Content  Pref_PowerPoint_presentation  \
          0                    0                      0                             1
          1                    0                      1                             1
          2                    0                      1                             0
          3                    0                      0                             0
          4                    0                      1                             0

             Pref_Interactive_Sessions  PercentileAbleTo_Score  Question_Difficulty  \
          0                        0.0                      78                    2
          1                        0.0                      40                    3
          2                        0.0                      50                    2
          3                        0.0                      30                    3
          4                        0.0                      70                    2
          CourseProgress_Satisfaction_x  Prim_Comm_Platform
       0                              1            0.833333
       1                              0            0.163934
       2                              1            0.163934
       3                              1            0.163934
       4                              1            0.163934

       [5 rows x 29 columns]
```

Fig8:The precprocessed Data Table

Making all the edges of the variables in the graph pointing directly towards the outcome variable i.e. (whether the course will be successful?).

We then Created a Gaussian Classifier with-*model = GaussianNB(), then* trained the model using the training sets and the model created above.
Following that we checked whether any of the element is NaN, and not whether the return value of the any function is a number to clean the dataset of nan, Inf, and missing cells (for skewed datasets). The Dimensions of the input array also were skewed, as the input csv had empty spaces. Finally, a conversion of data frames X and Y into matrices was required. To compare our final output the predict function for target values of X was used, which returned a matrix of predicted values to be compared against with the ground truth labels that is the y_test and hence, the final accuracy score measured. We then wanted to increase the predicted score, so we then generated a correlation matrix for fine tuning our hyperparameters and get the optimal accuracy for our model.

Results- The Final Accuracy achieved after hyperparameter tuning was found out be 72.22%.



```
In [303]: accuracy=accuracy_score(y_test, y_pred)*100
          accuracy

Out[303]: 72.22222222222221
```

Fig9-Final Accuracy (Jupyter Notebook)

The Figure 1 given below shows how the input variables are inter-related and their effect on the output.

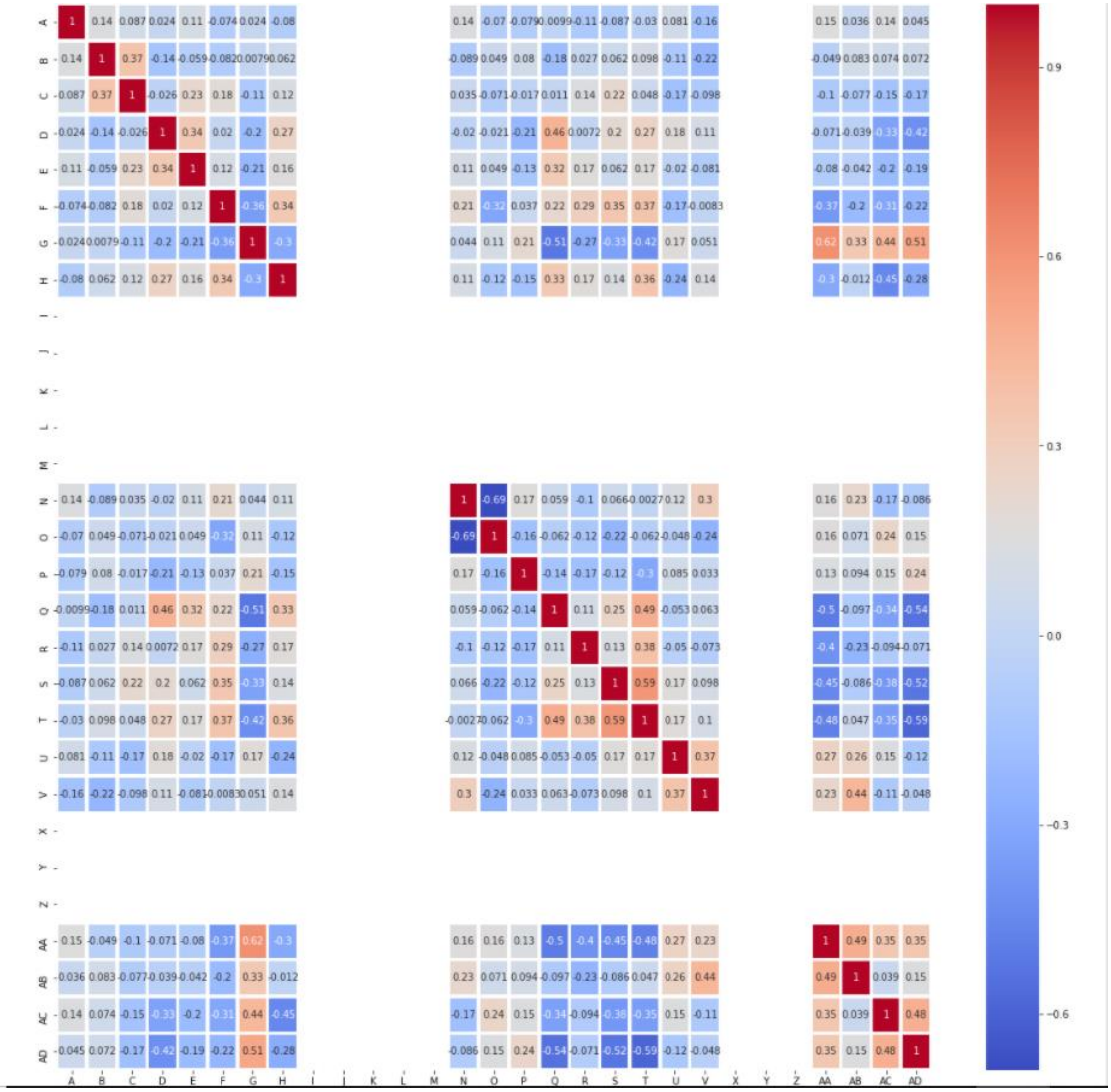


Fig10-Correlation Matrix

Given Below are the Encoded names of the columns in the Correlation Matrix-

Columns = {' Comm_Platform_Satisfaction(1-Yes, 0-No)':'A', 'Pref_Live_Lectures':'B', 'Pref_Audio_Lectures':'C', 'Pref_Recorded_Content':'D', 'Pref_PowerPoint_presentation':'E', 'Pref_Interactive_Sessions':'F', 'PercentileAbleTo_Score':'G', 'Question_Difficulty':'H', 'Curr_Live_Lectures':'I', 'Curr_Audio_Lectures':'J', 'Curr_Recorded_Content':'K', 'Curr_PowerPoint_presentation':'L', 'Curr_Interactive_Sessions':'M', 'PlatformAccessEase(1-Easy, 2-Medium, 3-Hard)':'N', 'platform_UI_intuitive _easy-to-use':'O', 'E-Learning_feature_incorporate':'P', 'THA':'Q', 'PracticeExercises':'R', 'ReadingMaterials':'S', 'QuizSolutions':'T', 'TextBooks':'U', 'Research Papers':'V', 'Materials_provision_Platform':'W', 'Tests_Assignments_in_course':'X', 'Result_Show_Time':'Y', 'AvgDur_Tests':'Z', 'WeeklyHours_Browse_Platform':'AA', 'Portal_Login_Freq':'AB', 'CourseProgress_Satisfaction_x':'AC', 'Prim_Comm_Platform':'AD' }

In the above correlation matrix, a warm-cool color scheme has been used where the warmth of the color increases the positive correlation between the 2 variables. The number inside that column is the impact of the increase in 1 input variable on another input variable.

As the color scheme turns towards dark blue it gives us the negative correlation, and the number inside the box denotes that amount of decrement of one variable due to increment of another variable.

The current ways used by the instructor to deliver the lectures i.e.('I', 'J', 'K', 'L', 'M', 'X', 'Y', 'Z') had zero correlation and impact on the other input variables and output, hence could easily be dropped. The input delivery mode does not matter much to a student.

The variable that is the percentage scored by the student was found to decrease with an increase in the Take-Home Assignments provided, and found to increase with the amount of time spent by the student in browsing the E-Learning Platform 'AA'. This can be attributed to the fact that increase in assignments force the student to study the subject and meet the deadlines a more flexible approach would suit the student better as can be observed with the time spent voluntarily in browsing through the course materials on the platform.

We further found out that the ease of accessing the platform ‘N’ was increasing as the Platform got a more intuitive User Interface ’O’. A better User Interface makes the platform more appealing to use and results in an increased retention rate.

The provision of Take-Home Assignment 'Q' decreased the browsing time of the student on the platform. Providing Assignments decreased the interest of students in the course in turn affecting their browsing activity on the platform.

It was found out that when the Reading Materials ’S’ are provided the instructors mostly upload the Quiz Solutions ‘T’ along with that as well.

The choice of the primary communication platform used by the instructor to communicate greatly impacted the Percentile Score of the Student 'G', and also affected the choice of instructor in the provision of Reading Material 'S' , Quiz Solutions 'T' and Take Home Assignments ‘Q’.

Our final results was that the decision variable was most positively affected by amount of time spent by the Student in browsing the E-Learning Platform ‘AA’, the Primary Communication Platform used by the instructor ‘AD’, the Percentile Score of the Student 'G' and most negatively affected by the Question Difficulty settings 'H', the number of Take-Home Assignments present in the course 'Q', The Number of Reading Materials 'S', and the provision of Quiz Solutions 'T'. The fact that a good score defines a lot about the conceptual clarity as well as the interest, along with the effort put by the student in that course, a good score does define a higher interest of the student in that specific course and higher probability of that course being successful which is also visible through the question difficulty pattern set up for students, the students tend to score a less percentage when they get a hard paper in turn affecting their confidence and making them less interested in the course. The increase in provision of the materials puts a lot of deadlines on the students and kills the freewill of the student to study at his time of discretion, Reducing his time of revision, Grades and in turn his interest in the course.

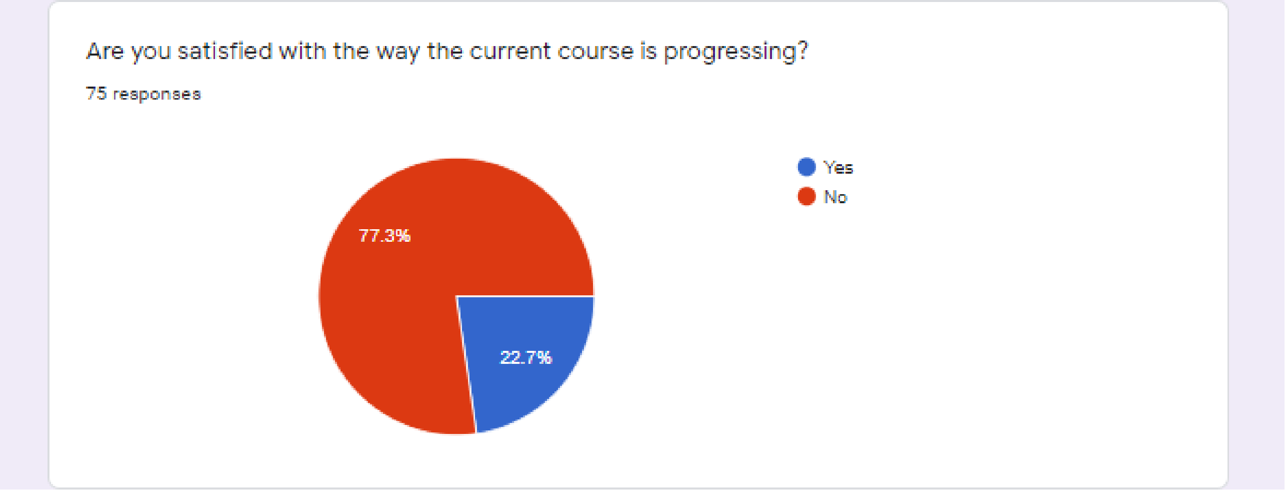
Discussions-The study conducted across a group of 74 students and it was found that 83.8 % instructors used mail as their primary communication platform for interaction with student as it is the most standard platform that is being used since decades and 55.4 % of the students were satisfied with the platform used by the instructor. It was further found out most of the students (57%) prefer recorded content for online delivery of their lectures as they can access the recording at their time and even helps them to download the videos to be viewed later as in a lot of parts there is a poor internet connection. It was found that the average percentile scored by the student in the course is 37.6%. 66.2 % people found out the course to be easy and only 3% of them felt that the course is easy. The low percentage can be attributed to the difficult paper pattern setup by the instructor to make the students study more and prevent them from getting over-confident.

Then it was found that the current instructors heavily rely on live+recorded lectures along with PowerPoint Presentations.

Around 50% found the E-Learning platform used by the instructor easy to use and 78% found it to be intuitive. Approximately 65% of students were satisfied with the current features provided by the E-Learning Platform and only 31% want addition of more features. A majority of the people were satisfied with the E-Learning Platform and some new features can be added to make the workflow smoother.

Around 20 odd Assignments are present there in the course. The average result show time for test results is 72 hours on the platform. The average test duration in the course is 30 mins. The instructor is following a high number of evals with less marks and less time given to the students to make sure that the students follow all the lectures diligently and are up to date with what is being taught in the class.

Approximately 3 hours is the weekly average time spent by the students in browsing the E-Learning Platform. Average Portal login frequency of the students is approx. 2.6 times. Which is not good and can be improved, the time spent in browsing and viewing material is a direct indication of the interest of the student in that course. A better approach can be taken to increase these numbers by reducing the Evals and THA’s and making a student friendly paper.


Are you satisfied with the way the current course is progressing?

75 responses

Fig11: Pie Chart - Course Progression

To conclude the final survey, it was discovered that a staggering 78.4% of people were dissatisfied with the way the current course is progressing.

Conclusions- For running the course successfully the instructor must not put a lot of pressure on students by providing them with a lot of materials with deadlines. A more reformed approach would be to decrease the number of evaluative and materials provided. The main focus should not be on covering width but rather depth in that subject resulting in a better conceptual clarity to the students. Setting up a difficult paper just so that the student doesn't get overconfident and is up to date with the course can backfire and be counter-productive an increase in the score of the student does increase his overall confidence which is showcased through his increased activity browsing the online portal. Getting sufficient time to cover backlog and revise the previous concepts is also important and a large number of evals makes it difficult for an average student to cope up with the course progression on missing one evaluative or a single class and makes the course more demanding.

References-

1-A Tutorial on Learning With Bayesian Networks (March 1995). Microsoft Research MSR-TR-95-06(David Heckerman) (https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-95-06.pdf)

2-Estimation of Students' Learning States using Bayesian Networks and Log Data of Learning Management System (Nobuhiko Kondo, Toshiharu Hatanaka)

3-Using Bayesian Networks to Detect Students' Learning Styles in a Web-based education system (Patricio García, Analía Amandi1, Silvia Schiaffino1, Marcelo Campo1)

4-Chakraborty, B. & Sinha, Meghamala. (2016). Student evaluation model using bayesian network in an intelligent E-learning system. IIOAB Journal. 7. 51-60.

5-West, P., Rutstein, D. W., Mislevy, R. J., Liu, J., Choi, Y., Levy, R., … Behrens, J. T. (2010). A Bayesian network approach to modeling learning progressions and task performance. (CRESST Report 776). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

6-A. Chanaa and N. E. Faddouli, "Deep learning for a smart e-Iearning system," 2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech), Brussels, Belgium, 2018, pp. 1-8, doi: 10.1109/CloudTech.2018.8713335.

7-Shao-Zhong Zhang, Hong Yu, Hua Ding, Nan-Hai Yang and Xiu-Kun Wang, "An application of online learning algorithm for Bayesian network parameter," Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693), Xi'an, 2003, pp. 153-156 Vol.1, doi: 10.1109/ICMLC.2003.1264461.

8-L. Chamba-Eras, A. Arruarte and J. A. Elorriaga, "Bayesian Networks to predict reputation in Virtual Learning Communities," 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), Cartagena, 2016, pp. 1-6, doi: 10.1109/LA-CCI.2016.7885721.

9-M. Guan, J. Jia, Y. Yang, Yuhua and Qingzhang Chen, "Research on adaptive e-Learning system using technology of learning navigation," 2013 8th International Conference on Computer Science & Education, Colombo, 2013, pp. 24-29, doi: 10.1109/ICCSE.2013.6553877.

10-Zhang, N. L., & Poole, D. (1996). Exploiting Causal Independence in Bayesian Network Inference. Journal of Artificial Intelligence Research, 1–28. https://arxiv.org/pdf/cs/9612101.pdf

11-Liang Zhang, Yue-ting Zhuang, Zhen-ming Yuan and Guo-hua Zhan, "Auto diagnosing: An intelligent assessment system based on Bayesian Networks," 2007 37th Annual Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, Milwaukee, WI, 2007, pp. T1G-7-T1G-10, doi: 10.1109/FIE.2007.4417872.

12- F. Colace, M. De Santo, A. Pietrosanto and A. Troiano, "Work in Progress: Bayesian Networks for Edutainment," Proceedings. Frontiers in Education. 36th Annual Conference, San Diego, CA, 2006, pp. 13-14, doi: 10.1109/FIE.2006.322573.

13- H. -. Kao, M. -. Liu, C. -. Huang and Y. -. Chang, "E-learning Systems Evaluation with Data Envelopment Analysis and Bayesian Networks," 2009 Fifth International Joint Conference on INC, IMS and IDC, Seoul, 2009, pp. 1207-1210, doi: 10.1109/NCM.2009.285.

14- Deshmukh, Ratnadeep & Wangikar, Vaishali. (2011). Data Cleaning: Current Approaches and Issues.

15- Ki Hyun Tae, Yuji Roh, Young Hun Oh, Hyunsu Kim, & Steven Euijong Whang. (2019). Data Cleaning for Accurate, Fair, and Robust Models: A Big Data - AI Integration Approach.

16- Loyola-ICAM College of Engineering and Technology, Chennai, India, Subhaswini, C.J.Sharmila, A., Shobana, G., & Jose, R. M. Data Cleaning and Visualization using Machine Learning. Int. Jnl. Of Advanced Networking & Applications (IJANA). https://www.ijana.in/papers/33.pdf

18- Stathacopoulou Regina, D. Magoulas George, Maria Grigoriadou and Maria Samarakou, "Neuro–fuzzy knowledge processing in intelligent learning environments for improved student diagnosis", Information Sciences, vol. 170, no. 2–4, pp. 273-307, February 2005.

19- Jesmeen, M.Z.H. & Hossen, Jakir & Sayeed, Shohel & Ho, C.K. & Tawsif, K. & Rahman, Md. Armanur & Hossain, Md. (2018). A Survey on Cleaning Dirty Data Using Machine Learning Paradigm for Big Data Analytics. Indonesian Journal of Electrical Engineering and Computer Science. 10. 1234-1243. 10.11591/ijeecs.v10.i3.pp1234-1243.

20- Cohen, I., Bronstein, A., & Cozman, F.G. (2001). Online Learning of Bayesian Network Parameters.

21- Chen, J.; Feng, J.; Hu, J.; Sun, X. Causal Analysis of Learning Performance Based on Bayesian Network and Mutual Information. Entropy **2019**, 21, 1102.

22.*H. T. Nguyen, W. Zhao and J. Yang, "A Trust and Reputation Model Based on Bayesian Network for Web Services", 2010 IEEE Int. Conf. Web Serv., pp. 251-258, Jul. 2010.*

23.*L. G. López-Faican and L. Chamba-Eras, "Bayesian networks to predict the learning style of student in virtual environments", AtoZ Novas práticas em informação e conhecimento, vol. 3, no. 2, pp. 107-115, 2014.*

24.*B. Daniel, J. Zapata-Rivera and G. McCalla, "A Bayesian Computational Model of Social Capital in Virtual Communities", Communities and Technologies, pp. 287-305, 2003.*

25.*J.-J. Qi, Z.-Z. Li and L. Wei, "A Trust Model Based on Bayesian Approach", Adv. Web Intell., pp. 374-379, 2005.*

26.*D.-Q. Li, L.-L. Li and H. Yang, "Trust Model Based on Bayesian Network in E-commerce", 2010 International Conference on E-Business and E-Government, pp. 4993-4996, 2010.*

27.*A. Jøsang, R. Ismail and C. Boyd, "A Survey of Trust and Reputation Systems for Online Service Provision", Decis. Support Syst., vol. 43, no. 2, pp. 618-644, Mar. 2007.*

28.*J. Patel, W. T. L. Teacy, N. R. Jennings and M. Luck, "A Probabilistic Trust Model for Handling Inaccurate Reputation Sources", Third Int. Conf. Trust Manag., pp. 193-209, 2005.*

29.*Yingfen Ma: Teaching and learning under network environment-Teaching Models of Network [M]. Beijing Science Press, 200S. IO (In Chinese)*

Appendix:

- Google Form Survey
- Final Bayesian Report (Jupyter Notebook)
- Bayesian Dataset