**Objective-**

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

**Abstract-**

The field of Learning Analytics is one of the hot topics of the 21st century. The paper explores a part of learning analytics using a Bayesian Networks model from a questionnaire data filled by the students of BITS Pilani to predict if the given course will run successfully on an e-learning platform. Through the simulation results it was found out that the approach can be used to suggest improved methods helping the instructors and students to reform their practices to maintain a synergy for a successful running of the course. The study was conducted on Engineering courses assuming independent input variables and cannot be necessarily generalized to other streams. Further research is necessary for quantifying these results into other fields.

**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 incubators and accelerators 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 many 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].

The application of this type of a procedure needs a directed mathematical model that is probabilistic in nature. The Bayesian Network is easy to train and predict the test data and with features like the choice of inter-relations of input dependencies it performs way better than other algorithms like regression and K-means clustering. The significance of Bayesian networks also lies in the fact that we are dealing with categorical and numerical data. Furthermore, we use a Gaussian Naive Bayes network where the features values are assumed to be in a Gaussian distribution symmetric about the feature value mean in a bell-shaped manner in which each variable is independent and is trained on the student learning characteristics and E-Learning Platform attributes.

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 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].

Bayesian network can be integrated in a variety of applications like Decision Making, Image Processing, System Relaibility, Analysis and PPDM along with this it also finds a wide application in the medical fields. The use of Bayesian Network and it looks promising to be soon used in real-time application for making instant decisions. Although there are a lot of challenges and fine tuning that need to be performed before getting sure that the system is error free along with the algorithm. There were several studies that proposed the use of Bayesian Networks. 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] 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. 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.

The use of Bayesian networks for predictive analysis in online education were more specifically described in 2009 by Kao and Liu where they 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 schoolteachers the system proved to be the most efficient among all the user groups.[13] Their study proposes the Bayesian network classification model so that the relative efficiency of future systems can be foreseen.[13]

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.*” The learning algorithm was further developed and proposed as a system by Zhang and Zhuang in 2007. An ITS (Intelligent Tutoring system) that considers the pedagogy of Adaptive Learning the following system was proposed with the help of Bayesian networks to give appropriate feedback by assessing the students learning states through his studying and academic performance patterns which can be used by the teachers to provide guidance to the students and help students to analyze the next step they need to take.[11] The figure given below shows the different layers of an adaptive E-Learning system.

Diagram

Description automatically generated

*Fig1: An Adaptive E-Learning System*

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 parameters 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].* Further,[37] Francesco and Moreno believe that the opportunities provided through ICT combined with efficient mechanisms like Bayesian Nets provide the students with a great learning opportunity through a way in which students can see their progress real-time and get real-time feedback and capture their learning information, A feature that is not present in most remote E-Learning platforms. In this paper a similar approach was taken by taking into account the interactions between the 3 stakeholders E-Learning systems, Students and the teachers using independent input variable dependencies. The study material recommendation model is also used in present in the correlation matrix that is further discussed in the discussion section of the paper.

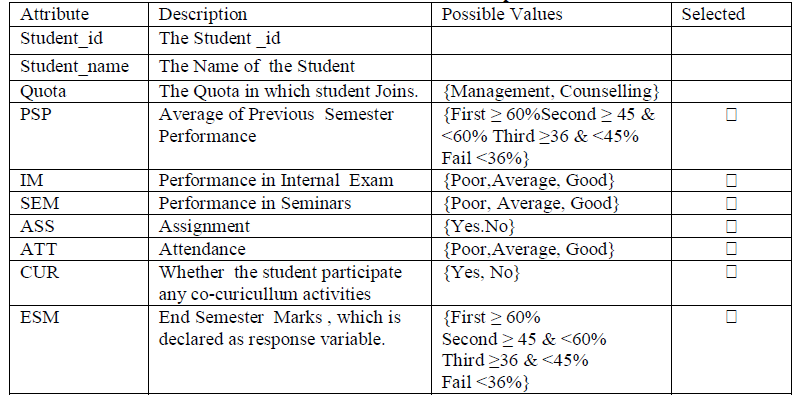
[30] In 2013 Sundar conducted a study for predicting the Academic performance of the students using Bayesian Network Classifiers.He discussed the provision of high-quality education to students is the main purpose and goal of an educational institution. The prediction of unmotivated or students who have shown less interest in the subjects is also discussed. The author used Bayesian Network classifiers to train the model to get the output of predicting the states. Using this the universities can track the students who may drop-out or need personal assistance or guidance. It would also help in the identification of meritorious students. The research was further substantiated by [31] Karim, Sharabiani, Atanasov and Darabi in 2014 in their paper on the Bayesian network for predicting the academic performance of students in engineering programs discussed that the model is based on the Bayesian networks framework. The main objective of their study was to predict second term grade of the students in 3 major courses. The way in which the grades are allotted to the students in particular subjects have major impact on their moral affecting their interest in their subject and resulting in higher dropout rates. They therefore proposed a model to predict the future grades of these students in the subjects and identify the students who would be needing help or counselling. Karim, Sharabiani, Atanasov and Darabi in their paper proposed a model for predicting the grade of engineering students and they concluded that the student’s grade and performance do play an important role in the student’s academic success in engineering. The students’ performance in those courses plays a crucially important role in determining the success that the student can get in his academics. With the accurate prediction of the student grades, a lot can be done to help the students on the borderline cases in the paper authors used “*The demographic and academic data of freshman engineering students in the University of Illinois at Chicago to configure and calibrate the network. The model was developed in the framework of Bayesian Networks and its performance got compared with the conventional statistical and data mining models in the literature and it appeared to be more efficient than them.*” The major purpose and contribution that their paper proposed was to develop a Bayesian Network with by including student details, their number per semester and the level of difficulty of each class in modeling the network to exercise their influence on students' marks in each subject. The outcome of this activity can be improved by including the time difference between the semester the student has taken for each class and the semester he or she takes in class. And incorporating personal, social, and psychological factors into account that affect students' strengths each semester, can grow model accuracy.

[32]Lamia and Hafidi in 2016 suggested a dynamic Bayesian network that could detect the learning states of students. And they found a huge range of literature that supported the fact the learning process can become more effective and improve the student’s performance with the help of the teaching strategies that align with the learning styles of the students. However, the old approaches towards the learning styles are conventional and have become obsolete. Lamia and Hafidi stated that “*Dynamic Bayesian network that represent the matches between LS and teaching strategies in order to determine how much a given strategy is interesting to a student. The LS theory that supports this approach is the LS model proposed by Felder-Silverman's learning styles model (FSLSM). Their approach gradually and constantly adjusts the student model, taking into account students' performances, student's effort, student's intensity, student's resistance and student's attention. Promising results were obtained from experiments.*” [33] Rajper, Shaikh and Mallah in their approach to the problem of detecting the students learning styles to get an idea of the student style of learning and change their style of content delivery stated that with the help of learning theories *“Felder Silverman learning style theory is largely used by researchers on LMS for learning styles’ identification on LMS. E-learning is the use of Information and Communication Technology (ICT) in learning prospects and found as rapidly growing mode of education these days.”* [35] Carmano, Castillo and Millan proposed modelling student learning styles through DBN. The authors discuss a model for Felder and Sylverman based learning styles trained using the dynamic Bayesian networks. The initialization of the model is in construct to the Learning Style Questionnaire Indexes.The classification of the object as appropriate and inappropriate then takes place through interaction with the Bayesian model. The paper proposed objects of learning according to the learner preference and styles with the help of a decision that is probabilistic in nature and determines the preferences matching with the learning styles and object determining the interest of a student to a particular object. There is similarity in the model behavior to the recommender system which is content based. The correlation between the learning styles and objects as an input to the classifiers give the status as to how interesting is the object to the user.

[34] Ueno and Okamoto in their paper on Bayesian agent in E-Learning discussed an agent acquiring knowledge about a particular domain through the database of the logs of the learning history and give messages that can motivate the students to perform better. The proposed model builds a Bayesian network to predict the final status of the learner and then with the help of database logs in its record compares its progress and processes with the outstanding learner’s and accordingly generates appropriate motivation messages to the learner tailored according to his needs.

*The following variables were employed by the authors in their model:*

* *The average learning time for each topic.*
* *The average learning time for each course which consists of fifteen lectures*
* *The number of times the learner accessed the e-learning system.*
* *The average of the degree of understanding of each topic (This is measured by the response to the question which is corresponding to each topic)*
* *The number of topics which the learner has learned.*
* *The average number of times the learner has completed each topic. (This implies the time the learner repeated each topic.)*
* *The average learning time for each lecture, which consists of several types of contents and runs 90 minutes*
* *The number of times which the learner has posted opinions or comments to the discussion board.*
* *The average learning time for each course which consists of fifteen lectures*
* *The final status: (1) Failed (Final examination score below 60); (2) Abandon (The learner withdraws before the final examination), (3) Successful ((Final examination score is more than 60 but less than 80); and (4) Excellent (Final examination mark is more than 80.)*

The study found that the Bayesian network model performs better than the decision tree model. The results found out clearly implied that the learning gets enhanced by this model and the motivational adaptive messages play an important role in enhancing a positive effect on the status of the learner. The attributes selected by Ueno and Okamoto were used in our paper as a reference to build upon the information that we could extract through the stakeholders in the E-Learning value chain. Further Rajper, Shaikh and Mallah conducted a survey to identify their E-Learning activities and predict their learning styles based on the attributed like [33] “*login time on LMS, immediate contact person in case of difficulty, frequently used tool to contact their preferred person in case of difficulty, participation activities on Discussion Board (DB), reading behavior, participation in chat, assignments submission*”. 

*Fig2: Student attributes Data Table*

The table presented above shows the dataset description with the attributes taken for prediction of the student’s performance by Sundar.et.al [30]. He concluded that “*Informations like Previous semester marks,Internal Marks,Performance on Seminars,Assignment,Attendance, Co-Curricular Activities were collected from the student‟s database, to predict the performance of the end semester marks. This study will help the students improve their performance and it helps teacher to identify those students which needs a special attention to reduce failing ration and taking appropriate action at right time.”*

The attributes use by the authors are the Platform attributes we have tried to take this study further by taking into the instructor feedback attributes. The only disadvantage of E-Learning is the lack of assistance and monitoring from a mentor or a teacher which has been tried to fill with the help of Bayesian networks in our study.

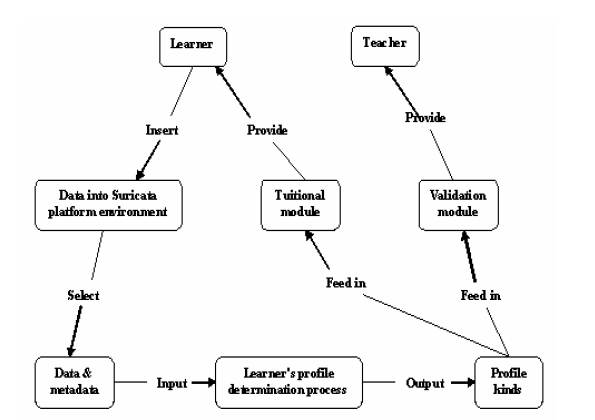
Carmano, Castillo and Millan described the Learning Styles “as the way a person collects, processes and organizes information.”

There are 4 dimensions according to which FSLSM classifies the students:

* *Active / Reflective (Processing). Active people consider having understood a piece of information only if they have discussed it, applied it or tried to explain it to other people. Reflexive people, on the other hand, prefer reflecting about the issue before assuming a practical posture.*
* *Sensing / Intuitive (Perception). Sensing people are meant to learn from tasks related to problems and facts that could be solved by well-behaved methods, with no surprises or unexpected effects. Besides, this style usually refers to students that are fond of details and very good memorizers of facts and practical applications. Conversely, intuitive students are meant to discover alternate possibilities and relationships by themselves, working with abstractions and formula, which allows them to understand new concepts and to quickly and innovatively perform new tasks.*
* *Visual / Verbal (Input). Visual-driven people find no difficulties in interpreting, for an example, pictures, diagrams, timelines or movies. Distinctly, verbal students’ personal learning processes are driven by written or spoken explanation.*
* *Sequential / Global (Understanding). Sequential people structure their learning process by logically, successively chained steps, each one of them related to the search for solutions. On the other hand, global students learning processes are distinguished by random jumps: they often are able to solve a complex problem, although they do not know how they arrived at the solution.*

Whenever a student selected a learning object the learning style changed, and the initial learning style also got refined but on a change in a preference object selection that did not match with the current learning state the model self-learned and accordingly updated and modified itself. A design of modelling student learning styles using a dynamic Bayesian network was presented by Carmano, Castillo and Millan [35]. The changes in students’ preferences have been accounted for and after validation can include more objects for the students. The dynamic adaptation was further discussed by [37] Francesco and Moreno in their paper on Using Bayesian Networks in the Global Adaptive E-learning Process where they showed how Bayesian networks can be used to provide a personalized process of learning known as adaptive E-Learning which adapts itself according to the needs and choices of the students and present contents to him according to his preferences through the information the model has of him. A directed acyclic graph is used for calculating the probabilities made by the learner with each activity.

They stated that is possible to optimize the model and the process through learning metrics which is all the assessments and information relating to all types of information and ways of learning and development.



*Fig3: Student’s profile determination*

[37] Francesco and Moreno concluded that a learner’s profile is created through the activities that he selects and realizes using this it is possible to visualize the progress of a learner in each of the subjects. The figure above shows the way in which a student’s profile is determined through his past and current actions. Teachers have access to the learning profiles of their students which can then be used by them to analyze draw conclusions and make reformations onto their teaching and they can even club people with same learning patterns and assign them special tasks and assignments according to the way in which they can understand the maximum with minimal input. They suggested that “*Bayesian Networks could help to improve the current systems and, current self-learning models.*”

The literature stated mostly had a very trivial and common approach to problem solving. The Bayesian nets provided good accuracy but most of the data that they were trained upon was small and in case of large data. A great model was suggested by [36] Carmona, Castillo and Milan to get better results to discover the student preferences. They proposed a whole sort and filter types process to determine a learner’s process. The entire process was broken down by them into the 3 processes of filtering, prediction, and adaptation. The authors concluded this paper with a model that discover the student preferences of educational materials over a period. The proposed model is capable of filtering large chunks of information available to enable a better use of the resources available. A DBN allowed the detection of student’s preferences available, so that their users can make a better use of it. In their conclusion they mentioned initialization of a decision model for every student through the set of rules generated data matching the multimedia resource with the learning styles. *“The model was also able to adapt itself to changes in the student’s preferences.”*

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 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.

A close up of text on a white background

Description automatically generated

*Fig4: 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 on the official Electrical Electronics Department communication groups 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.

Graphical user interface, application, pie chart

Description automatically generated *Fig5: The Collected Gform Data*

The collected data about the students and the platform attributes were then refined using methods of cleaning, encoding and normalizations,

[11] 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. We use a data centric model in our study.

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, duplicate, typos, NA values or Empty Spaces to ensure uniformity, completeness, consistency, minimum error margin and efficiency. 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 using scaling to a range technique and made sure that all the ranges were equally scaled to make it compatible with the algorithm to be trained upon. The Data was then manually as well as algorithmically scanned in In order to remove ambiguous or extreme values this type of

technique is known as feature clipping and does not disturb the model predictions and give accurate results. In the end all data was converted into similar d-types. 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.

Table

Description automatically generated

*Fig4: The pre-processed Data Table* Table

Description automatically generated 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 have 3 different types of Bayesian classifiers. The first one being the Naïve Bayes classifier which considers normal distribution of data through a normal distribution generated through a Gaussian distribution. The second one being the Multinomial classifier which is used when there is multinomial modelling distribution of events. This typed of classifier is mostly used in case of features with frequencies. The third type is Bernoulli classifier used when the features are independent and generated through a Bernoullian process. In our study we have used a Gaussian classifier to create a Gaussian distribution model without any co-variance with the help of standard deviation and mean.

At every data point, the z-score distance between that point and each class-mean is calculated, namely the distance from the class mean divided by the standard deviation of that class. Thus, we see that the Gaussian Naive Bayes has a slightly different approach and can be used efficiently. We then Created a Gaussian Classifier with-model = GaussianNB(), then trained the model using the training sets and created the model 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%**.

Text

Description automatically generated

*Fig5: Final Accuracy (Jupyter Notebook)*

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

Chart, treemap chart

Description automatically generated

*Fig6: 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.

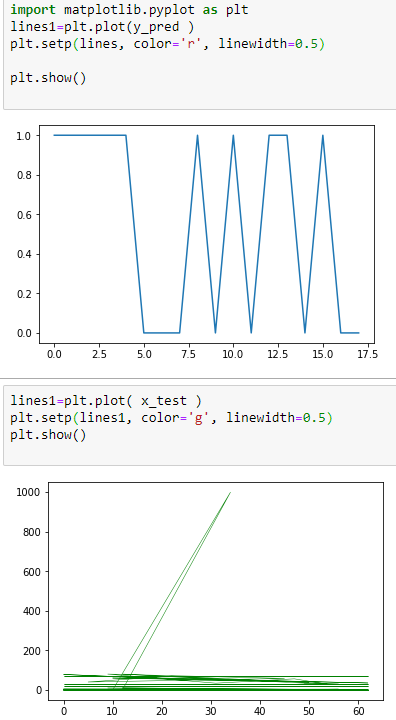
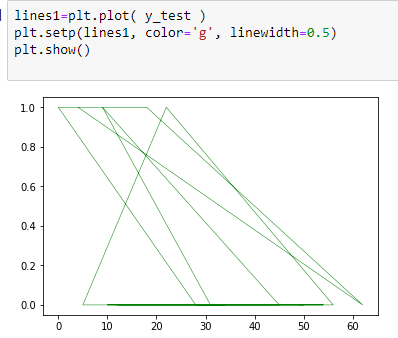
The provision of Take-Home Assignment 'Q' decreased the browsing time of the student on the platform. Providing Assignments decreased the interest of students in the course in turn affecting their browsing activity on the platform.

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 were 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.

The figure given below in the left describes the output of the predicted values and the figure to the right shows us the output over the 60 students. The values are discrete in the binary form of 0’s and 1’s and are plotted as a continuous line and hence we can tally theses figures to find out that the model is quite robust and the classifier graphs and maps the output precisely well.

*Fig7: The final output against the student number*

**Discussions**-

The study was 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.

[32]Lamia and Hafidi in their paper had concluded that the learning process can become more effective and improve the student’s performance with the help of the teaching strategies that align with the learning styles of the students. Similarly, through our results it was found that the people who got content delivered according to their preference performed better and showed more participation and interest in the course as compared to the people whose preferred methods of accessing the contents were completely different from those delivered by the instructor. These results were strengthened when [35] Castillo and Millan in their study proposed that the objects of learning according to the learner preference and styles with the help of a decision that is probabilistic in nature and determines the preferences matching with the learning styles and object determining the interest of a student to a particular object. There is similarity in the model behavior to the recommender system which is content based. The correlation between the learning styles and objects as an input to the classifiers give the status as to how interesting is the object to the user.

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. These results can be substantiated through the study by [2] Kondo and Hatanaka and Sundar.et.al [30] to provide feedback to the instructors of the students and identify students that were likely to get a lower grade, drop out or fail in the course. In our case the students who were unhappy with the course would end up not studying the course diligently and in turn getting a low grade. Our results were further proved by [31] Karim, Sharabiani, Atanasov and Darabi in 2014 who stated that the way in which the grades are allotted to the students in particular subjects have major impact on their moral affecting their interest in their subject and resulting in higher dropout rates. They further stated that by including student details, their number per semester and the level of difficulty of and its influence on students' marks in the subject improved their model. The trained model in our study performed better on including the attributes mentioned above.

Then it was found that the current instructors heavily rely on live+recorded lectures along with PowerPoint Presentations. Our results were in synergy with those made by [4] Baisakhi Chakraborty and Meghamala Sinha with the study material recommendation by evaluating the learning style of the students from the materials browsed and his test performance. 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. Most 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.

Chart, pie chart

Description automatically generated

*Fig8: 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.

**Limitations of the Study-**

This study has potential limitations. We had assumed the input variables to be independent in the model, but we can make interconnected and accordingly take the weights of one variable on another and find their interdependency relations. We had conducted this study in an Engineering college course so this study cannot be substantiated to field like arts. Further research is necessary for quantifying these results into other fields. The study was conducted on 80% of the people who had taken the course. The students were trusted to had have filled the questionnaire with integrity. The empirical results reported herein should be considered in the light of some limitations the algorithm hyperparameters were used in accordance with the datatypes attributes and sample sizes and would change with a change in the sample size. These gaps can be filled in with the help of a using a dynamic Bayesian network which takes Dynamic inputs in an inter-related way and allocates weight with each preference [37][35]. The main advantage of the model is that is requires minimum human intervention as it self-learns and accordingly modifies and updates itself.

**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 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 does not 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 many 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. In an E-learning course people hardly track all their progress individually and it is hard to get feedback all by yourself. The learning metrics can generate a paradigm shift in the teaching and e-learning systems to significantly improve the research system which can be possible with the help of technology and algorithms that can provide real-time feedback like Bayesian Nets.

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*](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-****Chanaa, A., & Faddouli, N. E. (2018). Deep learning for a smart e-Iearning system. 2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech), 1–8.* [*https://doi.org/10.1109/CloudTech.2018.8713335*](https://doi.org/10.1109/CloudTech.2018.8713335)

***7-****Shao-Zhong Zhang, Hong Yu, Hua Ding, Nan-Hai Yang, & Xiu-Kun Wang. (2003). 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), 1, 153-156 Vol.1.* [*https://doi.org/10.1109/ICMLC.2003.1264461*](https://doi.org/10.1109/ICMLC.2003.1264461)

***8-****Chamba-Eras, L., Arruarte, A., & Elorriaga, J. A. (2016). Bayesian Networks to predict reputation in Virtual Learning Communities. 2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI), 1–6.* [*https://doi.org/10.1109/LA-CCI.2016.7885721*](https://doi.org/10.1109/LA-CCI.2016.7885721)

***9-****Guan, M., Jia, J., Yang, Y., Yuhua, & Qingzhang Chen. (2013). Research on adaptive e-Learning system using technology of learning navigation. 2013 8th International Conference on Computer Science Education, 24–29.* [*https://doi.org/10.1109/ICCSE.2013.6553877*](https://doi.org/10.1109/ICCSE.2013.6553877)

*10-Zhang, N. L., & Poole, D. (1996). Exploiting Causal Independence in Bayesian Network Inference. ArXiv:Cs/9612101.* [*http://arxiv.org/abs/cs/9612101*](http://arxiv.org/abs/cs/9612101)

*11-Zhang, L., Zhuang, Y., Yuan, Z., & Zhan, G. (2007). Auto diagnosing: An intelligent assessment system based on Bayesian Networks. T1G-7.* [*https://doi.org/10.1109/FIE.2007.4417872*](https://doi.org/10.1109/FIE.2007.4417872)

*12-Colace, F., De Santo, M., Pietrosanto, A., & Troiano, A. (2006). Work in Progress: Bayesian Networks for Edutainment. 13–14.* [*https://doi.org/10.1109/FIE.2006.322573*](https://doi.org/10.1109/FIE.2006.322573)

***13-****Kao, H.-, Liu, M.-, Huang, C.-, & Chang, Y.-. (2009). E-learning Systems Evaluation with Data Envelopment Analysis and Bayesian Networks. 2009 Fifth International Joint Conference on INC, IMS and IDC, 1207–1210.* [*https://doi.org/10.1109/NCM.2009.285*](https://doi.org/10.1109/NCM.2009.285)

***14-****Deshmukh, R., & Wangikar, V. (2011, January 30). Data Cleaning: Current Approaches and Issues.*

***15-****Tae, K. H., Roh, Y., Oh, Y. H., Kim, H., & Whang, S. E. (2019). Data Cleaning for Accurate, Fair, and Robust Models: A Big Data - AI Integration Approach. ArXiv:1904.10761 [Cs].* [*http://arxiv.org/abs/1904.10761*](http://arxiv.org/abs/1904.10761)

***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*](https://www.ijana.in/papers/33.pdf)

***17-*** *Gaur, Poonam. (2015). Research Trends in E-Learning. Media Communique NIU Journal of Media Studies, ISSN 2395-3780. 1. 29-41.*

*18- Stathacopoulou, R., Magoulas, G. D., Grigoriadou, M., & Samarakou, M. (2005). Neuro-fuzzy knowledge processing in intelligent learning environments for improved student diagnosis. Information Sciences, 170(2), 273–307.* [*https://doi.org/10.1016/j.ins.2004.02.026*](https://doi.org/10.1016/j.ins.2004.02.026)

*19-**H, J. M. Z., Hossen, J., Sayeed, S., Ho, C. K., K, T., Rahman, A., & Arif, E. M. H. (2018). A Survey on Cleaning Dirty Data Using Machine Learning Paradigm for Big Data Analytics. Indonesian Journal of Electrical Engineering and Computer Science, 10(3), 1234–1243.* [*https://doi.org/10.11591/ijeecs.v10.i3.pp1234-1243*](https://doi.org/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. (2019). Causal Analysis of Learning Performance Based on Bayesian Network and Mutual Information. Entropy, 21(11), 1102.* [*https://doi.org/10.3390/e21111102*](https://doi.org/10.3390/e21111102)

***22-*** *Nguyen, H. T., Zhao, W., & Yang, J. (2010). A Trust and Reputation Model Based on Bayesian Network for Web Services. 2010 IEEE International Conference on Web Services, 251–258.* [*https://doi.org/10.1109/ICWS.2010.36*](https://doi.org/10.1109/ICWS.2010.36)

***23-*** *Chamba-Eras, Luis, & López-Faican, L. (2014). Bayesian networks to predict the learning style of student in virtual environments. Novas Práticas Em Informação e Conhecimento, 3, 107–115.*

***24-*** *B Daniel, B. K., Zapata-Rivera, J.-D., & McCalla, G. (2003). A Bayesian Computational Model of Social Capital in Virtual Communities. In M. Huysman, E. Wenger, & V. Wulf (Eds.), Communities and Technologies (pp. 287–305). Springer Netherlands.* [*https://doi.org/10.1007/978-94-017-0115-0\_15*](https://doi.org/10.1007/978-94-017-0115-0_15)

***25-*** *Qi, J.-J., Li, Z.-Z., & Wei, L. (2005). A Trust Model Based on Bayesian Approach. In P. S. Szczepaniak, J. Kacprzyk, & A. Niewiadomski (Eds.), Advances in Web Intelligence (pp. 374–379). Springer.* [*https://doi.org/10.1007/11495772\_58*](https://doi.org/10.1007/11495772_58)

***26-*** *Li, D., Li, L., & Yang, H. (2010). Trust Model Based on Bayesian Network in E-commerce. 2010 International Conference on E-Business and E-Government, 4993–4996.* [*https://doi.org/10.1109/ICEE.2010.1253*](https://doi.org/10.1109/ICEE.2010.1253)

***27-*** *Jøsang, A., Ismail, R., & Boyd, C. (2006). A Survey of Trust and Reputation Systems for Online Service Provision. Decision Support Systems, 43, 618–644.* [*https://doi.org/10.1016/j.dss.2005.05.019*](https://doi.org/10.1016/j.dss.2005.05.019)

***28-*** *Patel, J., Teacy, W. T. L., Jennings, N. R., & Luck, M. (2005). A Probabilistic Trust Model for Handling Inaccurate Reputation Sources. In P. Herrmann, V. Issarny, & S. Shiu (Eds.), Trust Management (pp. 193–209). Springer.* [*https://doi.org/10.1007/11429760\_14*](https://doi.org/10.1007/11429760_14)

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

***30-*** *Sundar, P. (2013). A COMPARATIVE STUDY FOR PREDICTING STUDENT’S ACADEMIC PERFORMANCE USING BAYESIAN NETWORK CLASSIFIERS. IOSR Journal of Engineering 2250-3021, 3, 2250–3021.* [*https://doi.org/10.9790/3021-03213742*](https://doi.org/10.9790/3021-03213742)

***31-*** *Sharabiani, A., Karim, F., Sharabiani, A., Atanasov, M., & Darabi, H. (2014). An enhanced bayesian network model for prediction of students’ academic performance in engineering programs. 2014 IEEE Global Engineering Education Conference (EDUCON), 832–837.* [*https://doi.org/10.1109/EDUCON.2014.6826192*](https://doi.org/10.1109/EDUCON.2014.6826192)

***32-*** *Mahnane, L., & Hafidi, M. (2016). Automatic detection of learning styles based on dynamic Bayesian network in adaptive e-learning system. International Journal of Innovation and Learning, 20(3), 289–308.* [*https://doi.org/10.1504/IJIL.2016.079067*](https://doi.org/10.1504/IJIL.2016.079067)

***33-*** *Rajper, S., Shaikh, N. A., Shaikh, Z. A., & Ali Mallah, G. (2016). Automatic Detection of Learning Styles on Learning Management Systems using Data Mining Technique. Indian Journal of Science and Technology, 9(15).* [*https://doi.org/10.17485/ijst/2016/v9i15/85959*](https://doi.org/10.17485/ijst/2016/v9i15/85959)

***34-*** *Ueno, M., & Okamoto, T. (2007). Bayesian Agent in e-Learning. Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007), 282–284.* [*https://doi.org/10.1109/ICALT.2007.82*](https://doi.org/10.1109/ICALT.2007.82)

***35-*** *Carmona, C., Castillo, G., & Millán, E. (2008). Designing a Dynamic Bayesian Network for Modeling Students’ Learning Styles. 2008 Eighth IEEE International Conference on Advanced Learning Technologies, 346–350.* [*https://doi.org/10.1109/ICALT.2008.116*](https://doi.org/10.1109/ICALT.2008.116)

***36-*** *Carmona, C., Castillo, G., & Millán, E. (n.d.). Discovering Student Preferences in E-Learning. 10.*

***37-*** *Moreno, F. J. T., & Moreno, M. J. G. (n.d.). Using Bayesian Networks in the Global Adaptive E-learning Process. 4.*

Appendix:

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