Bayesian Networks in E-Learning

Virendra Singh Nirban a\*, Tanu Shukla b and Daksh Dave c

a Humanities Department, BITS Pilani, Pilani, India; b Humanities Department, BITS Pilani, Pilani, India; c Electrical Electronics Department, BITS Pilani, Pilani, India

full correspondence details here including e-mail for the \*corresponding author

short biographical notes on all contributors here if the journal requires them.

Bayesian Networks in E-Learning

In the field of educational research 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 at an educational institute to predict if the given course will run successfully on an e-learning platform. Through the simulation results it was found that the BN approach can be used to suggest improved online instruction delivery 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 for generalized inferences.

Keywords: bayesian networks; e-learning; learning analytics; probability

Subject classification codes: include these here if the journal requires them

# Introduction

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.

(Chanaa and Faddouli 2018) discussed the 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 authors deeply discuss the factors like the compatibility of the students with instructor’s course delivery method and the student’s prior preparation and eagerness to learn. (Chanaa and Faddouli 2018) used Deep Learning to provide a personalized E-Learning Resource Platform according to the user preference.

(West et al. 2012) stated that the application of modelling a learning progression 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 BN performs way better than other algorithms like Decision Tree (Ueno and Okamoto 2007), Regression and K-means clustering. The significance of Bayesian Networks also lies in the fact that it deals with categorical and numerical data.

The learning sector possesses vast amounts of records of the student data. (Kondo and Hatanaka 2019) believed that 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 which has been discussed by (Carmona, Castillo, and Millán 2008).

The modernization in our current education system has given rise to E-Learning systems. 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. The loopholes based in this industry gave us the motivation to carry out this research. There was a huge shift from offline to an online mode of education in the past 5 years. However, after the COVID times the offline mode is no more a choice and making the maximum use of the E-Learning platform lies of utmost importance to us.

Our study aims to use machine learning algorithms for analyzing the smooth operation of a course by extracting the characteristics of a student and the E-Learning Platform on the 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 to ensure that there is a synergy that is maintained between the instructor and the student and there is maximum utilization of each other’s effort.

The paper is further classified as follows section 2 contains review of literature related to current trends in E-Learning Analytics. Section 3 discusses the method of study used. Section 4 presents the results of the output. Section 5 presents the discussions containing the interpretation and description of the significance of the findings. The section 6 provides the conclusions.

1. **Literature Review**

(Chanaa and Faddouli 2018) stated that research is valuable on the use of Machine learning models which is currently gaining popularity in E-Learning. Educational stakeholders 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.

Bayesian Networks can be integrated in a variety of applications. There were several studies that proposed the use of Bayesian Networks. (Nguyen, Zhao, and Yang 2010), (Chamba-Eras and López-Faican 2014), (Stathacopoulou et al. 2005), (Daniel, Zapata-Rivera, and McCalla 2003)and (Qi, Li, and Wei 2005).

The use of Bayesian Networks for predictive analysis in online education were more specifically described by (Kao et al. 2009) where they proposed 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. Their study proposes the Bayesian Network classification model so that the relative efficiency of future systems can be foreseen.

An Adaptive Learning algorithm for course learning system was built by (Guan et al. 2013) where “*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 et al. 2007). In his paper 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. The figure 1 given below shows the different layers of an adaptive E-Learning system.

Diagram

Description automatically generated

*Figure1: An Adaptive E-Learning System*

In the empirical analysis done by (Kondo and Hatanaka 2019) used a Bayesian Network to find out the learning states of the students which provided instructor with the feedback of the students who 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 the process transparent. 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 et al. 2012) in their paper to get the optimal Bayesian Network score. To prove this point further (Chakraborty and Sinha 2016) in their research paper worked further in learning analytics to come up with an effective 4 component model. Their model improved study material recommendations 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.*”. The parameters used to collect the data and directing the nodes and usage were also very impactful while running the Network, (García et al. 2007) 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”.* Further,(F.Moreno and M.Moreno 2005) believe that the opportunities provided through ICT( Information Communication Technology ) 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.

(Sundar 2013) 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 which is the main purpose and goal of an educational institution and help the institution to identify the students that can potentially drop out. The research was further substantiated by (Sharabiani et al. 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. Further (Sharabiani et al. 2014) 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. The major purpose and contribution that their paper proposed was to develop a Bayesian Network by incorporating 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 the BN algorithm 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.

(Mahnane and Hafidi 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. (Mahnane and Hafidi 2016) 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.*” (Rajper et al. 2016) 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 discussed the Felder Silverman learning theory and its importance in E-Learning research. He stated that *“Felder Silverman learning style theory is largely used by researchers on LMS for learning styles’ identification on LMS.**”* (Carmona, Castillo, and Millán 2008) 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.

(Ueno and Okamoto 2007) 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 study found that the Bayesian Network model performs better than the decision tree model. The results 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. Further, (Rajper et al. 2016) conducted a survey to identify their E-Learning activities and predict their learning styles.

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. (Carmona, Castillo, and Millán 2008) described the Learning Styles “as the way a person collects, processes and organizes information.” 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 (Carmona, Castillo, and Millán 2008). 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 (F.Moreno and M.Moreno 2005) 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.

(F.Moreno and M.Moreno 2005) 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. The authors 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. Teachers have access to the learning profiles of their students which can then be used by them to analyze and draw conclusions and make reformations into 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 review highlights that most studies used a common approach to problem solving. The Bayesian Networks provided good accuracy but most of the data that they were trained upon was small. A great model was suggested by (Carmona, Castillo, and Millán, 2005) for training the Bayesian Networks on huge datasets 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.  They concluded that *“The model was also able to adapt itself to changes in the student’s preferences.”*

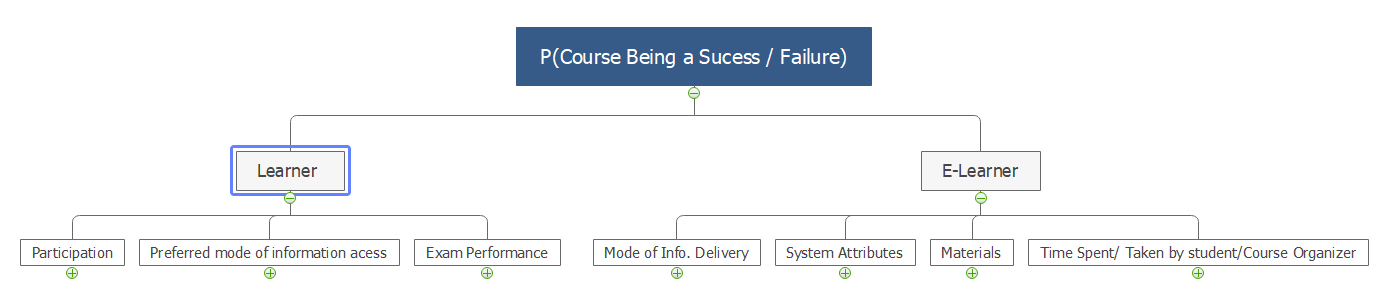
Nonetheless the usage of Learning Analytics remains one of the rigorously pursued areas in educational research providing huge opportunities for further exploration.

1. **Methodology**

For this study, we propose to use a Bayesian Network which is a graphical model comprising of nodes and edges with probabilities. We first design a Bayesian Network using observed relationships in between different variables and then calculate the probabilities given the occurrence of those events. We propose a fully independent model with random independent variables. In the model Nodes have been treated as random variables and edges as relationships between the random variables while making the graphical model.

(Chamba-Eras, Arruarte, and Elorriaga 2016) used a similar type of correlation to measure the reputation of the community members 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.*”

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. The attributes selected by (Ueno and Okamoto 2007) 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. We would be using similar type of sub-attributes of the E-Learning System collected from the Students making all the edges of the variables in the graph pointing directly towards the outcome variable i.e. (whether the course will be successful?). The figure 2 given below shows the attributes and the interconnections proposed in this paper for feeding into BN model.



*Figure2: Bayesian Network Model*

After preparing the model of the graph, we use conditional probabilities where given a certain activity performed or selected by the student predict the probability for the further outcome because of the selected activity.

For this study one undergraduate course Analog and Digital VLSI Design was selected for the learning analytics. For the students, the data relating to the student’s preferences, attitude and perceptions was collected and categorized based on Learner and E-Learner attributes through a questionnaire on the official Electrical Electronics Department communication groups which had 7 main attributes Participation, Preferred mode of information access, Exam performance, Mode of information delivery, System attributes, Materials, Time spent/taken by student/Course organizer as shown in the figure 2. These main attributes were spit into several sub-attributes for breaking down the data for simplifying the analysis. The sub-attributes were declared as shown in the Table-1 below:



*Table1: Sub-Attributes*

In the Table-1 given above the first row shows the main attributes and the rows below the attribute’s corresponding columns show the list of the sub-attributes based on which the test data was formed to train the network and draw inferences. The model is trained using the naïve Bayes algorithm which uses the Bayes Theorem on each class to predict the probability that the data points belongs to a particular class and finally selects the class having the highest probability. The figure 3 given below shows the summary of the collected questionnaire data of 2 participation sub-attributes.

Graphical user interface, application, pie chart

Description automatically generated *Figure3: The Collected Gform Data*

According to (Zhang et al. 2007) there are 3 types of models which can be used to train 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 conducted to assess the learning styles of different students. The questionnaire consisted of a set of questions that measure the learning styles of the students and their interest in that course. Further the sum and averages of all the questionnaire items were computed to draw conclusions.

The data was collected and analyzed to learn the student’s learning style by knowing how the student interacts with the system. 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. The data is pre-processed in excel after collection. The figure4 given below shows the pre-processed data that was collected after executing the model in Jupyter Notebooks. Initially for analysis purposes describe function is used to remove any unwanted, duplicate, typos, NA values or Empty Spaces to ensure uniformity, completeness, consistency, minimum error margin and efficiency. Further for consistency purposes One Hot Encoding, Binary Encoding and Label Encoding on the Categorical Variables Dataset is used to make the dataset consistent for processing. The dataset is then normalized using scaling to a range technique so that all the ranges are equally scaled to make it compatible with the algorithm to be trained upon. Using a technique known as feature clipping all the ambiguous and extreme values are removed through algorithmic as well as manual scanning. In the end all data is converted into similar d-types. The dataset is then split into two sets 25 percent test dataset and 75 percent training dataset. Further fundamental Naive Bayes algorithm was applied to our dataset with the assumption that each feature has an independent contribution. Each variable is taken to be equally contributing to the output while processing.

Graphical user interface, application

Description automatically generatedGraphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generated*Figure4: The pre-processed Data Table*

We have 3 different types of Bayesian classifiers. The first one being the Naïve Bayes classifier 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 which is 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. In the Gaussian model 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 for the purpose of normalizing the data and getting a bell-shaped Gaussian curve. Thus, it can be observed that the Gaussian Naive Bayes has a slightly different approach and can be used efficiently. The Gaussian Classifier is then created and using the existing training sets the model is trained.

Following that there is a sanity check to clean the dataset of the unwanted values. The check is performed for values like NaN(Not a Number), Inf(Infinity), and missing cells (for skewed datasets). 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 (A parameter with a value that can control the learning process) and get the optimal accuracy for our model.

1. **Results**

The Final Accuracy achieved after hyperparameter tuning was found out be **72.22%** as shown in the figure-5 given below.

Text

Description automatically generated

*Figure5: Final Accuracy (Jupyter Notebook)*

The Figure 7 given below shows how the independent variables are inter-related and their effect on the dependent variables. In the figure given below the Column AC is the dependent variable and the columns A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, AB, AC and AD are the independent variables.

Given Below are the Encoded names of the columns in the Correlation Matrix-  *Figure6: Encoded Column names*

In the correlation matrix presented in figure-7, 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 of the cell 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.

Chart, treemap chart

Description automatically generated

*Figure7: Correlation Matrix*

The variable, 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.

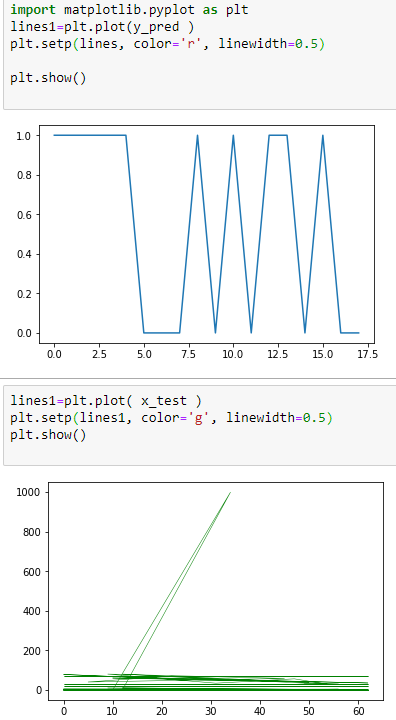
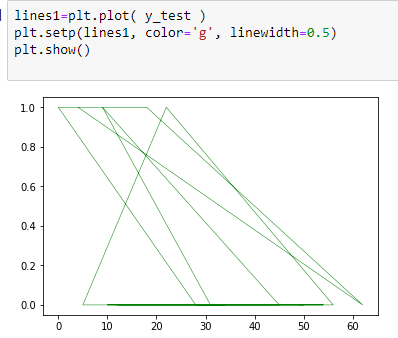
We further found out that the ease of accessing the platform ‘N’ was increasing as the Platform got a more intuitive User Interface ’O’. A better User Interface makes the platform more appealing to use and results in an increased retention rate.

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-8a describes the output of the predicted values and the figure-8b 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 from the figures given below it can be observed that the model is quite robust and the classifier graphs and maps the output precisely well.

*Figure8a: Predicted output Figure8b: Test dataset output*

1. **Discussion**

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 observed that 57% of the students 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 geographical areas there is a poor internet connection.

(Mahnane and Hafidi 2016) 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 learners 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 (Carmona, Castillo, and Millán 2008) in their study concluded that according to the learner’s preferences and styles it is possible to determine the matching preferences and interest of a student relating to a particular object with the help of Bayesian Networks. There is similarity in the model behaviour 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 % respondents found the course to be hard 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 align with the study by (Kondo and Hatanaka 2019) and (Sundar 2013) to provide instructors with feedback of the students and identify students who 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 corroborated by (Sharabiani et al. 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 (Chakraborty and Sinha 2016) 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 assignments were recorded 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 Chart, pie chart

Description automatically generatedindication of the interest of the student in that course.

*Figure9: Pie Chart-Course Progression*

To conclude the final survey, from the figure-9 given above it was discovered that a staggering 77.3% of people were dissatisfied with the way the current course is progressing. A better approach can be taken to increase these numbers by reducing the Evals and THA’s and making a student friendly paper. The main determiners for the success of E-Learning courses are the difficulty level of the paper, the number of take home-assignments and the synergy of the student with the instructor’s method of teaching.

1. **Acknowledgements**

We acknowledge all the students and instructors whose data has been used in these analyses.

1. **Disclosure Statement**

There is no conflict of interest among the author(s).

1. **Notes on Contributors**

**Dr. Virendra Singh Nirban**

***Dr. Virendra Singh Nirban,*** Department of Humanities and Social Sciences, BITS Pilani. He is an Associate Professor; his research interest includes Educational Informatics**,** Technology Adoption Behavior and Discourse Analysis of Computer-Mediated Communication.

**Dr. Tanu Shukla**

***Dr. Tanu Shukla***, Department of Humanities and Social Sciences, BITS Pilani. She is an Associate Professor; her research interests are linked to Educational Policy and Planning, Educational Psychology, Technology-Behaviour.

**Daksh Dave**

***Daksh Dave***, Department of Electrical Electronics, BITS Pilani. He is a third-year undergrad student pursuing his majors in Electronics.

1. **References**

Carmona, C., G. Castillo, and E. Millán. 2008. “Designing a Dynamic Bayesian Network for Modeling Students’ Learning Styles.” In *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, 346–50.

Carmona, Cristina, Gladys Castillo, and Eva Millán. n.d. “Discovering Student Preferences in E-Learning,” 10.

Chakraborty, B., and Meghamala Sinha. 2016. “Student Evaluation Model Using Bayesian Network in an Intelligent E-Learning System.” *IIOAB Journal* 7 (January): 51–60.

Chamba-Eras, L., A. Arruarte, and J. A. Elorriaga. 2016. “Bayesian Networks to Predict Reputation in Virtual Learning Communities.” In *2016 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 1–6.

Chamba-Eras, Luis, and Lissette López-Faican. 2014. “Bayesian Networks to Predict the Learning Style of Student in Virtual Environments.” *Novas Práticas Em Informação e Conhecimento* 3 (November): 107–15.

Chanaa, A., and N. E. Faddouli. 2018. “Deep Learning for a Smart E-Iearning System.” In *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*, 1–8.

Chen, Jing, Jun Feng, Jingzhao Hu, and Xia Sun. 2019. “Causal Analysis of Learning Performance Based on Bayesian Network and Mutual Information.” *Entropy* 21 (11). Multidisciplinary Digital Publishing Institute: 1102.

Colace, F., Massimo De Santo, Antonio Pietrosanto, and Alfredo Troiano. 2006. “Work in Progress: Bayesian Networks for Edutainment.” In , 13–14.

Daniel, Ben Kei, Juan-Diego Zapata-Rivera, and Gordon McCalla. 2003. “A Bayesian Computational Model of Social Capital in Virtual Communities.” In *Communities and Technologies*, edited by Marleen Huysman, Etienne Wenger, and Volker Wulf, 287–305. Dordrecht: Springer Netherlands.

Deshmukh, Ratnadeep, and Vaishali Wangikar. 2011. “Data Cleaning: Current Approaches and Issues.” In .“Estimation of Students’ Learning States Using Bayesian Networks and Log Data of Learning Management System | International Journal of Institutional Research and Management.” 2021.

Guan, M., J. Jia, Y. Yang, Yuhua, and Qingzhang Chen. 2013. “Research on Adaptive E-Learning System Using Technology of Learning Navigation.” In *2013 8th International Conference on Computer Science Education*, 24–29.

García, Patricio, Analía, Amandi, Silvia, Schiaffino, and Marcelo, Campo. "Using Bayesian Networks to Detect Students' Learning Styles in a Web-based education system".*Proceedings of ASAI 2005, Argentine Symposium on Artificial Intelligence* (2005).

Hand, David J. 2018. “Aspects of Data Ethics in a Changing World: Where Are We Now?” *Big Data* 6 (3). Mary Ann Liebert, Inc., publishers: 176–90.

H, Jesmeen M. Z., J. Hossen, S. Sayeed, C. K. Ho, Tawsif K, Armanur Rahman, and E. M. H. Arif. 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): 1234–43.

Jøsang, Audun, Roslan Ismail, and Colin Boyd. 2006. “A Survey of Trust and Reputation Systems for Online Service Provision.” *Decision Support Systems* 43 (November): 618–44.

Kao, H.-, M.- Liu, C.- Huang, and Y.- Chang. 2009. “E-Learning Systems Evaluation with Data Envelopment Analysis and Bayesian Networks.” In *2009 Fifth International Joint Conference on INC, IMS and IDC*, 1207–10.

Kondo, N. and T. Hatanaka. “Estimation of Students' Learning States using Bayesian Networks and Log Data of Learning Management System.” (2019).

Li, D., L. Li, and H. Yang. 2010. “Trust Model Based on Bayesian Network in E-Commerce.” In *2010 International Conference on E-Business and E-Government*, 4993–96.

Lim, Sungsoo, and Sung-Bae Cho. 2006. “Online Learning of Bayesian Network Parameters with Incomplete Data.” In *Computational Intelligence*, edited by De-Shuang Huang, Kang Li, and George William Irwin, 309–14. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer.

Nguyen, H. T., W. Zhao, and J. Yang. 2010. “A Trust and Reputation Model Based on Bayesian Network for Web Services.” In *2010 IEEE International Conference on Web Services*, 251–58.

Mahnane, Lamia, and Mohamed Hafidi. 2016. “Automatic Detection of Learning Styles Based on Dynamic Bayesian Network in Adaptive E-Learning System.” *International Journal of Innovation and Learning* 20 (3). Inderscience Publishers: 289–308.

F. Moreno, A. Carreras, M. Moreno and E. R. Royo, Using Bayesian Networks in the Global Adaptive E-learning Process, EUNIS 2005, Manchester, pp. 1-4, 2005.

Patel, Jigar, W. T. Luke Teacy, Nicholas R. Jennings, and Michael Luck. 2005. “A Probabilistic Trust Model for Handling Inaccurate Reputation Sources.” In *Trust Management*, edited by Peter Herrmann, Valérie Issarny, and Simon Shiu, 193–209. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer.

Qi, Jian-Jun, Zeng-Zhi Li, and Ling Wei. 2005. “A Trust Model Based on Bayesian Approach.” In *Advances in Web Intelligence*, edited by Piotr S. Szczepaniak, Janusz Kacprzyk, and Adam Niewiadomski, 374–79. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer.

Rajper, Samina, Noor A. Shaikh, Zubair A. Shaikh, and Ghulam Ali Mallah. 2016. “Automatic Detection of Learning Styles on Learning Management Systems Using Data Mining Technique.” *Indian Journal of Science and Technology* 9 (15).

Roh, Yuji, Geon Heo, and Steven Euijong Whang. 2019. “A Survey on Data Collection for Machine Learning: A Big Data -- AI Integration Perspective.” *ArXiv:1811.03402 [Cs, Stat]*, August.

Shao-Zhong Zhang, Hong Yu, Hua Ding, Nan-Hai Yang, and Xiu-Kun Wang. 2003. “An Application of Online Learning Algorithm for Bayesian Network Parameter.” In *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, 1:153-156 Vol.1.

Sharabiani, A., F. Karim, A. Sharabiani, M. Atanasov, and H. Darabi. 2014. “An Enhanced Bayesian Network Model for Prediction of Students’ Academic Performance in Engineering Programs.” In *2014 IEEE Global Engineering Education Conference (EDUCON)*, 832–37.

Stathacopoulou, Regina, George D. Magoulas, Maria Grigoriadou, and Maria Samarakou. 2005. “Neuro-Fuzzy Knowledge Processing in Intelligent Learning Environments for Improved Student Diagnosis.” *Information Sciences* 170 (2): 273–307.

Subhaswini, C J, A Sharmila, G Shobana, and Raina Miriam Jose. n.d. “Data Cleaning and Visualization Using Machine Learning,” 4.

Sundar, Praveen. 2013. “A COMPARATIVE STUDY FOR PREDICTING STUDENT’S ACADEMIC PERFORMANCE USING BAYESIAN NETWORK CLASSIFIERS.” *IOSR Journal of Engineering 2250-3021* 3 (March): 2250–3021.

Tae, Ki Hyun, Yuji Roh, Young Hun Oh, Hyunsu Kim, and Steven Euijong Whang. 2019. “Data Cleaning for Accurate, Fair, and Robust Models: A Big Data - AI Integration Approach.” *ArXiv:1904.10761 [Cs]*, April.

Ueno, M., and T. Okamoto. 2007. “Bayesian Agent in E-Learning.” In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, 282–84.

West, Patti, Daisy Wise Rutstein, Robert J. Mislevy, Junhui Liu, Roy Levy, Kristen E. Dicerbo, Aaron Crawford, Younyoung Choi, Kristina Chapple, and John T. Behrens. 2012. “A Bayesian Network Approach To Modeling Learning Progressions.” In *Learning Progressions in Science: Current Challenges and Future Directions*, edited by Alicia C. Alonzo and Amelia Wenk Gotwals, 257–92.

Zhang, Liang, Yue-ting Zhuang, Zhen-ming Yuan, and Guo-hua Zhan. 2007. “Auto Diagnosing: An Intelligent Assessment System Based on Bayesian Networks.” In , T1G-7.

Zhang, N. L., and D. Poole. 1996. “Exploiting Causal Independence in Bayesian Network Inference.” *ArXiv:Cs/9612101*, November.

1. **Appendices**
2. **Table with Captions**

Table 1. Type your title here. Obtain permission and include the acknowledgement required by the copyright holder if a table is being reproduced from another source.

1. **Figures**

Figure 1. Type your caption here. Obtain permission and include the acknowledgement required by the copyright holder if a figure is being reproduced from another source.

1. **Figures Captions**