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

Introduction- E-learning refers to “All the computer-management-based teaching and learning activities carried out in the information technology environment constructed by different transmission networks with communication function” [29].

The Development in technology, internet, infrastructure, and service sectors have made E-Learning very popular. The introduction of E-Learning has instilled a curiosity in the minds of young and adult-learners and the possibilities it can help us achieve breaking the traditional teaching methods that are followed saving money, space and time of all the stakeholders in the value chain.

There are different ways in which students prefer to learn. In the last decade we have seen a boom in the space of Ed-Tech startups wanting to democratize learning. In the future we would find that a degree may no longer stand the same importance as it does today, Education with the same content would be accessible to everyone and available online. Among this the method to detect the optimal way in which the information can be delivered by making analogies through dependencies on each other is getting prominence and helping people to make better responsive decisions.

The modernization in our current education system has given rise to a new era of E-Learning systems built through Software Technology. The past years have seen a rising trend in E-Learning Sector as well as the Learning industry. The grants given by governments and acknowledgements by incuabtors and accelarators have further boosted the cause of people wanting to pursue a career in Ed-Tech industry. This has led to a rise in the herd mentality products, startups and systems in this space that only give importance to meny without realizing whether they are able to tap into the student’s mind and able to deliver and provide educational content to the student in a way that suits his interests making his learning smooth and efficient. By giving importance to his learning and behavioral patterns which sadly a remote teacher may not be able to understand. The loopholes based in this industry gave rise to a new generation of systems known as adaptable learning systems.

There are a lot of different features provided by an E-Learning platform and different ways in which the teachers can make its maximum utilization to deliver content to students. The method of delivery of the teacher and compatibility of the student with that along with the student’s prior preparation and eagerness to learn along with other factors have been deeply discussed in this paper. The recent rends show an increase in the research areas of Students and Learning Analytics. Recent Studies such as using Deep Learning to provide a personalized E-Learning Resource Platform according to the user preference have been automating the old traditional processes [6].

According to Luis, Anna and Jon in 2016 *“The research carried out in the field of Virtual Learning Centre at higher education showed the potential of this learning community and, in addition, identified the new roles that the members involved in the community play ”[8].*

They used a Bayesian Network for estimating the reputation in the Virtual Learning Centre. By using the data available to them and predicting the trust and reputation relationships values.

The learning sector possesses vast amounts of records of the student data. The usage of learning analytics by the educational institutions has gained prominence due to its effectiveness in helping the institute make well-informed decisions. The advent of Artificial Intelligence has provided humongous scope in improvement of E-Learning Platforms by providing intelligent and interactive environments to the students. It has made it possible to capture the data in real-time. This study aims to use machine learning algorithms for analyzing the smooth operation of a course given a character of a student and E-Learning Platform on standardized data collected from numerous students to predict the outcome of a course running successfully and helping the instructor to design the course in a better way. The application of this type of a procedure needs a directed mathematical model that is probabilistic in nature. Furthermore, we use a Naive Bayes network which each variable to be independent and is trained on the student learning characteristics and E-Learning Platform attributes.

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

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

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

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

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

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

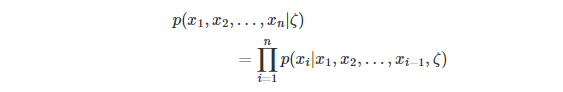
**

Fig1:A local probabilistic Distribution

Zhang and Poole described the conditional dependence such that “A *BN can be viewed as representing a factorization of a joint probability. For example, the Bayesian network in Figure 1 factorizes the joint probability P (a; b; c; e1; e2; e3) into the following list of factors: P (a); P (b); P (c); P (e1ja; b; c); P (e2ja; b; c); P (e3je1; e2).*”[10]

The use of Bayesian network as a forecasting mechanism was briefly given in the paper written by Luis, Anna and Jon “Bayesian Networks (BN), known as probabilistic models or belief networks, have been investigated due to a growing interest in predicting future events. BN is circumscribed, as forecasting technique, whose main characteristic is the valuation or qualification observed facts or data. Its role as a forecasting mechanism is very important as it allows inferences about the probability of occurrence of a given event on the basis of observed evidences.”

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

A picture containing watch

Description automatically generated

Fig2: A Bayesian Network

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

In a Bayesian Network there are 2 cases that can happen condition dependence and conditional independence. In the condition dependence case if there is a change in the node information it would also lead to a change in the distribution of the probabilities it relates to. Whereas in the case of Conditional Independence it is mutually exclusive to the events happening in the other nodes. Figure given below gives us the mathematical relationships for independent probabilities.

“*For each variable xi, let πi⊆{x1,x2,…,xi−1} is the parent's node of xi, and {x1,x2,…,xi−1} is conditional independence, then:*”[7]



Fig3: Condition Probability formula

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

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

Bayesian Network usage has also gained a lot of popularity in other fields other than E-Learning. It’s properties, open-source algorithms along with the ability of inter-relating variables have made it a lot more popular to be used among the Genz innovators and researchers.

[30]Iqbal, Yin , Hong, Ilyas and Ali discussed the use of Bayesian networks in domains that are uncertain and it’s ability to give accurate results and perform well in these types of domains due to its unique natures of taking discrete and continuous random variables, assess risks and formulate important decisions. “*Uncertainty is a major barrier in knowledge discovery from complex problem domains. Knowledge discovery in such domains requires qualitative rather than quantitative analysis. Therefore, the quantitative measures can be used to represent uncertainty with the integration of various models. The Bayesian Network (BN) is a widely applied technique for characterization and analysis of uncertainty in real world domains. Thus, the real application of BN can be observed in a broad range of domains such as image processing, decision making, system reliability estimation and PPDM*

*(Privacy Preserving in Data Mining) in association rule mining and medical domain analysis. BN techniques can be used in these domains for prediction and decision support.”*

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.

Further Iqbal, Yin, Hong, Ilyas and Ali also discussed the use of BN in decision making in a variety of applications discuss by considering causal maps that can be used in making decision in variety of applications. We have considered causal maps for representing BN for making decision. The construction of Bayesian Network is divided into 2 stages that are mainly probabilistic and qualitative.

“*Qualitative stage eliminates the limitations of modeling and makes causal maps compatible with the BN. This procedure has four major modeling issues. First, conditional independencies mean that a network can either be Dependence Map (D-Map) or Independence Map (I-Map). I-map is a graph in which the entire set unconnected of nodes corresponds to independent variables. On the other hand, D-map is a graph that represents the entire set of connected nodes corresponding to dependent variables. A model that follows both I-maps and D-maps is known as a perfect map. Second, reasoning underlying cause-effect relations can either be abductive or deductive.*

*Therefore, a reason is a rational ground or motive for an action or decision. Thus, abductive reasoning is a process that justifies a rational ground from effects to causes. The deductive reasoning is a process that justifies a rational ground from causes to effects. The difference between both reasoning processes is shown in figure below. Third, distinguishing between direct and indirect relationships means to identify the conditional independencies as depicted in the below given figure. The purpose is to draw attention to the conditional independencies’ assumption affects in causal*

*maps. Also, causal maps lack distinguishing in direct and indirect relationship.”*

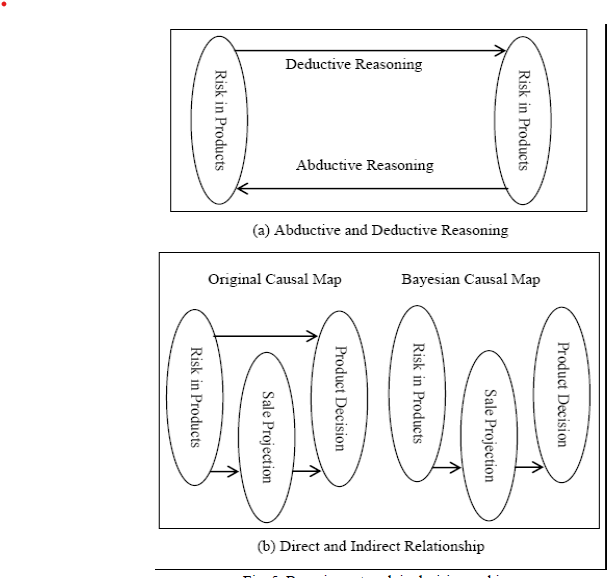


Fig4: Bayesian network in decision making.

The other part in decision making is inferencing in which we have to consider some uncertain nodes with background noises represented using a lot of causal variables. Hence, we can easily conclude that the “Bayesian causal maps are the combination of causal maps and BN. The limitations of causal maps are eliminated with the use of BN. Furthermore, qualitative representation of causal maps is transformed into quantitative representation with more precision. The main advantage of combining qualitative and quantitative representation of causal maps with support of BN is to have features (uncertain variables) with their numerical knowledge. Thus, Bayesian causal maps can extract knowledge, helping in comparative analysis while validating qualitative information in decision making.”

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

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

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

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

Diagram

Description automatically generated

Fig5: An Adaptive E-Learning System

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

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

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

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

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

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

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

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

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

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

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

[41] 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 table presented below shows the dataset description. In his paper Sundar mentions that the “*Current Education System in India, a student's Performance is determined by their performance based on Internal marks and semester marks. The internal marks is carried out by the teacher based upon students’ performance in educational activities such as seminars taken, assignments, co-curricular activities and performance in Internal exams. The end semester examination is one that is scored by the student in semester examination. each student has to get minimum marks to pass a semester in internal as well as end semester examination.*”

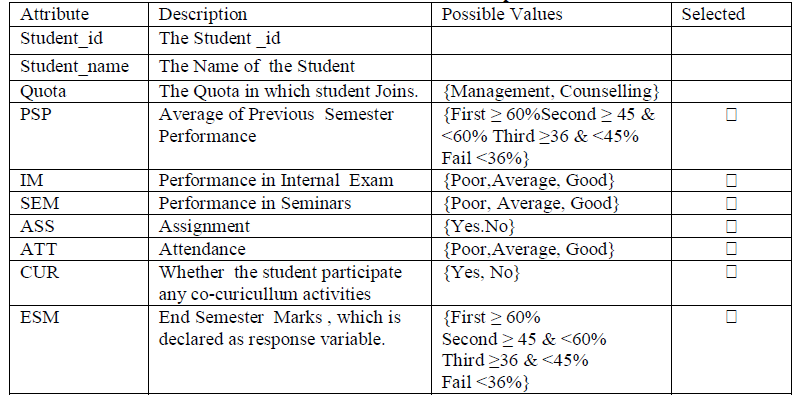


Fig6:Student attributes Data Table

After training the model and running it over the dataset to check it’s values the precision and recall scores are checked upon. The precision should be high and recall minimum for great performance. The naïve Bayes model was able to successfully classify 55 correct prediction and made 45 wrong predictions with an accuracy of 56.3%. Sundar 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 also it helps teacher to identify those students which needs a special attention to reduce failing ration and taking appropriate action at right time.”

[42] 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 papers 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.

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 through “*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.*”

[44] 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 according to the students learning styles was discussed by them. They stated that “*When it is required to understand the students’ learning requirements in terms of learning preferences, the learning styles theories are used. Incorporating learning styles in personalized E-learning systems are found prolific for enhancing the learning of students. Many learning style theories are given by various educationists and research-ers. But the 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. However, E-learning is abundant of advantages but on the other hand learners suffer from lack of supervision and assistance of the teacher/e-teacher Therefore, personalized E-learning systems provide the learning objects and support as per the students’ requirements and needs. In this regard learning styles are incorporated in personalized E-learning systems. For incorporating the learning styles on LMS it is important to map the classroom learning style theory on LMS using any related attributes, i.e., synchronous and asynchronous activities of e-learners*”

[44]A KLSM model that is Kolb’s learning styles model was used to categorize learners into 4 categories according to their learning styles. The model is shown to be used by a lot of research and is proved to be successful.. The KLSM model consists of 4 types of Learning Styles that are Diverger, Assimilator, Converger and Accomodator.

A survey was conducted by the authors to map the activities performed by the student on the institute’s LMS platform. A questionnaire consisting of 9 questions was designed to identify their E-Learning activities and predict their learning styles based on the attributed like [44] “*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 behaviour, participation in chat, assignments submission*”

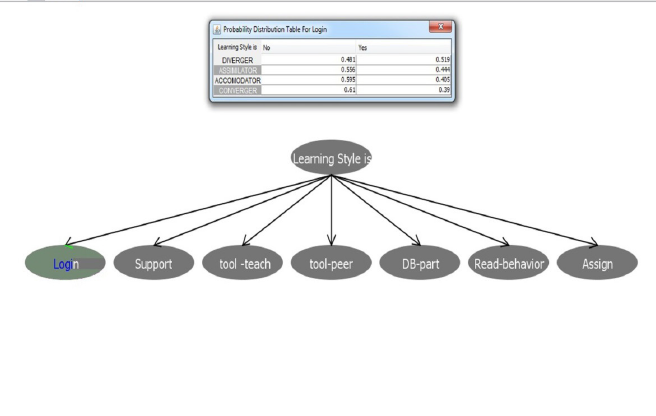


Fig7: CPT student behavior

The values obtained of the probabilities in CPT are displayed above the averaged value of the students log records were matched with known learning styles. The threshold values of the Bayesian network that were already known were then matched with the probabilities found for every learning styles.

The authors concluded that their study used a “*data mining technique BN for detection of learning styles on LMS, the acquired CPT values were evaluated using an experiment and found promising results. However, the results can be better in future by improving the technique. The technique used can be incorporated in personalized E-learning systems to know about the students’ learning preferences to teach them accordingly.*”

[45] 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 it’s record compares it’s 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:*

*1. The average learning time for each topic.*

*2. The average learning time for each course which consists of fifteen lectures*

*3. The number of times the learner accessed the e-learning system.*

*4. 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)*

*5. The number of topics which the learner has learned.*

*6. The average number of times the learner has completed each topic. (This implies the time the learner repeated each topic.)*

*7. The average learning time for each lecture, which consists of several types of contents and runs 90 minutes*

*8. The number of times which the learner has posted opinions or comments to the discussion board.*

*9. The average learning time for each course which consists of fifteen lectures*

*10. 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.)*

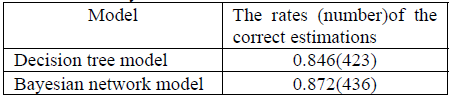


Fig8: Model accuracy

The figure given above shows the performance accuracies of two models that is the Decision tree and the Bayesian network. It was found out that the proposed model performs better than the decision tree model. “*Furthermore, the system was evaluated by comparing a class that used the proposed agent system with one that did not for one semester and one that used 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 authors concluded that *“the proposed method and the agent using the decision tree show that the proposed method has better prediction performances and effective to decrease the number of students withdrew from classes.*”

[46] Carmano, Castillo and Millan in their paper on designing a dynamic Bayesian network for modelling students learning styles. 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.

The authors 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 made a selection of 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 [46]. The changes in students’ preferences have been accounted for and after validation can include more objects for the students.

[49] Francesco and Moreno in their paper on Using Bayesian Networks in the Global Adaptive E-learning Process 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.

People hardly track all their progress individually and it’s hard to get feedback all by yourself. The learning metrics can generate a paradigm shift in the teaching and e-learning systems to significantly to improve the research system.

[49] Francesco and Moreno believe that *“*t*he extended learning and personalized learning opportunities that Information and Communications Technologies (ICT) brings enable learners to take much greater personal responsibility for learning. It is this,*

*combined with the different mechanisms through which ICT adds value to teaching and learning, which is a major factor in producing the great improvements in learning, than are seen in schools and colleges that have fully embedded ICT use. To enable students to take greater personal responsibility for their learning - whether this is a matter of slightly increased concentration in class, or fully autonomous learning in and out of class - learning needs to become more explicit. Students need to be able to see their progress. Teachers need to be able to use information on students' approaches to learning and progress, to help them. Though ICT enables much more work to be done on computers, and at a distance, which makes it more difficult for teachers to see what their students are doing, it can also capture information on the learning process in ways that have not been possible before.”*

We now have access to a lot of data and it’s up to us as to how we use it to capture and analyze in a wide variety of systems for the purpose of learning. According to the student’s profile and the conduct form that he has along with his previous knowledge will determine the learning style of the student. The conduct form gets extracted through the learning modules usage by the students on the LMS Platform which would then undergo pre-processing and generate the learner profile. The entire process is shown in the figure below.

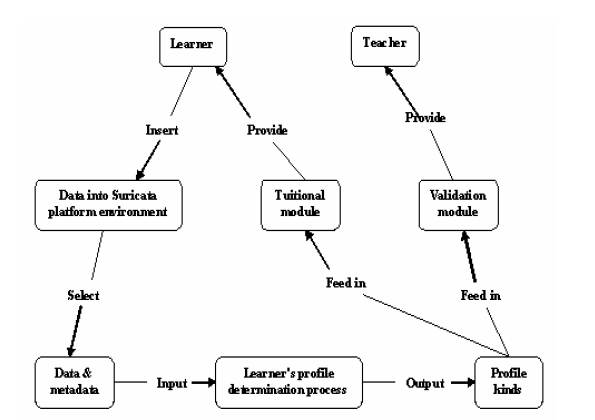


Fig9: Student’s profile determination

Francesco and Moreno concluded that a learner’s profile is created through the activities that he selects and realizes using this it’s 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 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.*”

[48] In 2010 Colace and De Santo researched about developing an ontology for E-Learning through a Bayesian Approach. They found that E-Learning is gaining a lot of popularity among researchers in the last decade with a constant research going on to develop tools and approaches that can adapt to the special needs of an individual learner.

Introduction of new services to improve the process quality was discussed through the launch of a formalized ontology. A significant improvement through the introduction can be observed through ontologies in the knowledge domain of a course.

The authors concluded that the “*By the automatic analysis of students’ learning performances, the proposed approach can analyze the courses’ ontology and propose corrective actions. In this way, teachers can redesign their courses and better understand the requirements of their students. This approach also provides an ontological basis for determining learning paths to personalize learning. An integrated method tool used for matching between the ontology and the Bayesian network allowed effective tutoring and a better adaptation of the learning process to the demands of students.*” This proposed ontological idea proposed would in turn help us get an in-depth analysis of the students resulting in more accurate model.

[47] Carmona , Castillo and Milan in their paper discussed a more accurate learning model in order to get better results to discover the student preferences.

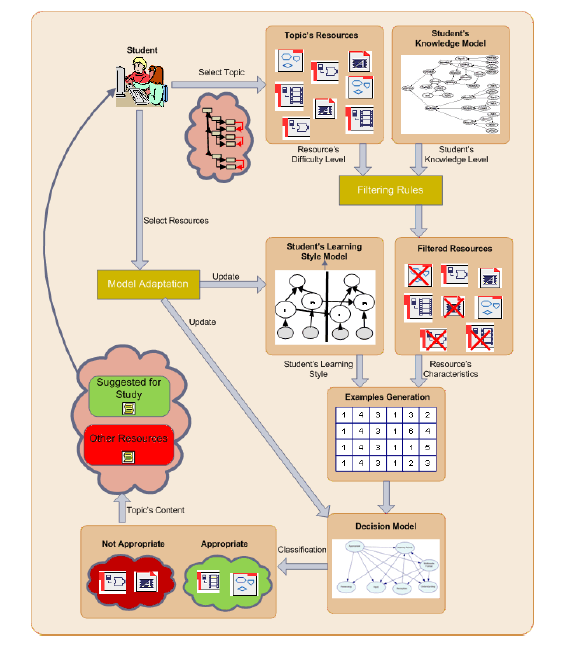


Fig10: Learning Resource task selection

They proposed a whole sort and filter types process to determine a learner’s process. The entire process described by them is mentioned in the diagram above and performed in the following steps:

* *Filtering****:*** *when a student selects a topic we apply some deterministic filtering rules to obtain the learning resources for this topic. This filtering process is performed according to the matches between the resource’s difficulty level and the student’s knowledge level.*
* *Prediction****:*** *using the current decision model, each filtered resource is classified as ‘appropriate’ or ‘not appropriate’ for the student. With this purpose, examples including the learning style features (obtained from the learning style model) and the resource’s characteristics are automatically generated and classified by the decision model. As a result the set of available resources is partitioned into these two classes. Decision****:*** *since the classifier returns probabilities, all the resources of the same class can be ranked. Then, a document is sent to the student including two separated ranked lists: Resources suggested for study (those classified as appropriate) and Other resources for study (those classified as not appropriate).*
* *Adaptation: when the student selects a resource in one of the two lists we assume that this resource is interesting to the student not by its content (since all the shown resources must explain the same concept), but by the learning activity and the multimedia format that this resource represents (each learning resource implements a learning activity in a multimedia format). Moreover, the user can explicitly rank a resource in order to obtain some confidence levels about how much does she/he like it. This way we can obtain a new labelled example that can be used to adapt both the learning style model and the decision model, accordingly.*

The authors concluded this paper with a model that discover the student preferences of educational materials over a period. The proposed model is cable 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.”*

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

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

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

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

Table

Description automatically generated

Fig11: An Uncleaned Dataset Table

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

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

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

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

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

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

The figure given below gives a rough idea of the optimal categorical that can be used for your model depending upon the usage of the Machine Learning algorithm type whether it is a classification or regression. The Data Type available also plays an important role in determining the data category in use. The approach can also be then found out. There are several other techniques that are present which can be used for the same application and getting similar results. For instance, we can use both Fact Extraction and Semi-supervised + Active Learning for classification of text types data’s. Nevertheless, the table is just a guide there are several other factors also that play a role in determining the optimum accuracy. The table is an appropriate guide for determining the data type approach that we need to start with for the model training.

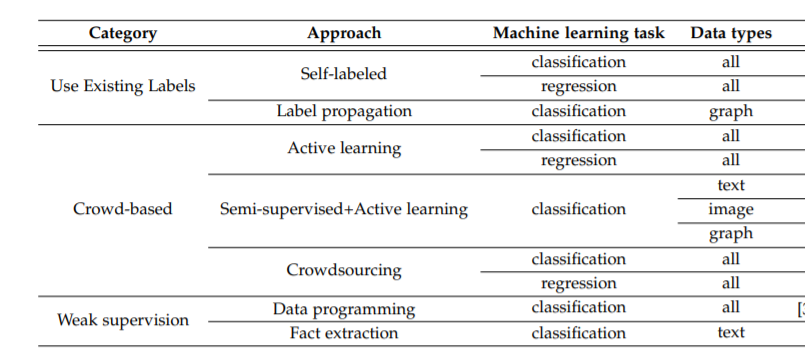


Fig12: Classification of Data Labelling Techniques

[35]In another paper published by Roh, Heo and Wang in 2019 discuss the popularity that machine learning is gaining and the importance to acquire large amount of data to feed into your model for training your model and generating an optimal output In this particular research the authors conducted a survey to check the landscape of research being currently performed in these areas and the optimal technique to be used and the way in which these techniques can complement each other to get the outcome. They concluded their research with discussion about the integration of machine learning deep learning techniques in the process of data collection to make the entire process of data collection simpler.

According to Roh and Heo “*An interesting observation is that the data collection techniques come not only from the machine learning community (including natural language processing and computer vision, which traditionally use machine learning heavily), but have also been studied for decades by the data management community, mainly under the names of data science and data analytics.*” The research overview of the topics that have contributed towards data management community is highlighted in the figure given below.

Roh and Heo further stated that “*Traditionally, labeling data has been a natural focus of research for machine learning tasks. For example, semi-supervised learning is a classical problem where model training is done on a small amount of labeled data and a larger amount of unlabeled data. However, as machine learning needs to be performed on large amounts of training data, data management issues including how to acquire large datasets, how to perform data labeling at scale, and how to improve the quality of large amounts of existing data become more relevant. Hence, to fully understand the research landscape of data collection, one needs to understand the literature from both the machine learning and data management communities.*”

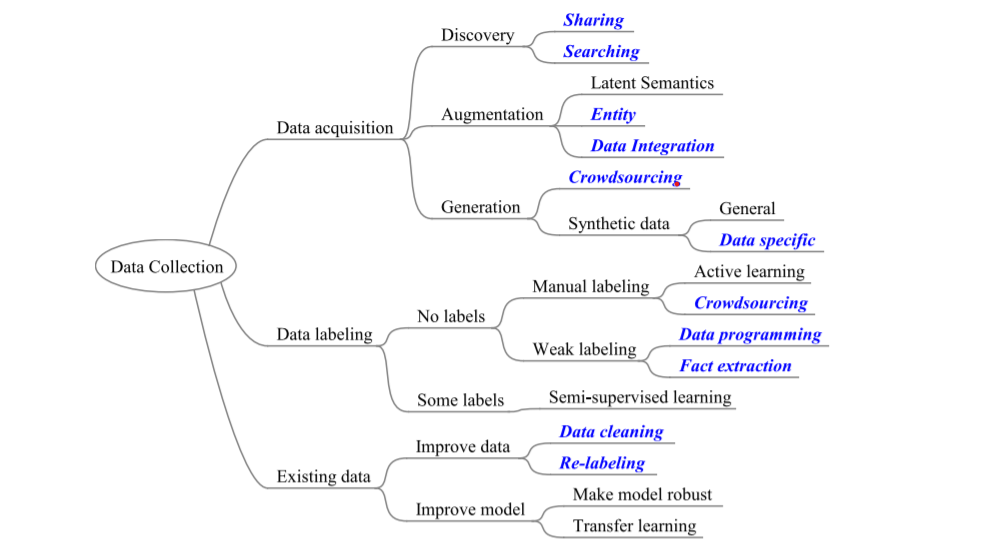


Fig13-:A high level research landscape of data collection for machine learning. The topics that are at least partially contributed by the data management community are highlighted using blue italic text. Hence, to fully understand the research landscape, one needs to look at the literature from the viewpoints of both the machine learning and data management communities

The Extraction of data and its cleaning is one of the most important steps in the framework of training a machine learning model to get the most optimal solution. That being stated, the utmost importance is to maintain integrity while extracting the data to make sure that the process followed is ethical. David Hand in 2018 in his paper on “*Aspects of Data Ethics in a Changing World: Where Are We Now?*” discussed the constraints and overviews of ethics in data. The data is very complex and there is not a hard and fast rigid framework that will ensure that the ethics are not violated. The nature of data along with the usage of it with the privacy of the data provider that is making his data available to us the author gives us the point in form of a checklist to maintain and keep in mind to always ensure integrity.

The structure of Ethical Guidelines that are mentioned and discussed can not generalized to everyone and the availability of everything on an open-source platform can lead to unethical behavior even tough the research was ethical one carried out in an ethical way. So, it is also important to think of the future ways in which there can be any breach of invasion on the privacy policies or the ethical guidelines.

|  |
| --- |
| **A Case Study** |
| *An instance involving the 1991 U.S. Federal Policy for the Protection of Human Subjects, the so-called Common Rule.6,7 This is a U.S. ethical structure for biomedical and behavioral research on human subjects, based largely on the Belmont report. The Common Rule,7 Section 46.101(b), gives a number of exemptions from these ethical guidelines. Among these exemptions is ‘‘Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.’’ Now, one of the characteristics of modern data work is that very often analyses are made on existing publicly available data sets (often characterized as ‘‘open data’’), and it is one of the particular strengths of these new technologies that discoveries can be made by linking and merging them. Such discoveries can be of scientific and medical value (e.g., in epidemiology), of economic value (e.g., in estimates of gross domestic product or inflation), and they can also be highly sensitive. There are, for example, classic cases of individuals being reidentified by linking public domain anonymized data sets. At the very least, this involves invasion of privacy, and it can lead to worse.*  *Source:* *https://www.liebertpub.com/doi/pdfplus/10.1089/big.2018.0083* |

Table1: A case Study on privacy invasion

A basic aspect of the study is also that while using or extracting the data one does not know how that data will be used in the near future. Due to this issue, it is very difficult to be able to classify data sets as public or private. It is a common belief that it is the data which raises the ethical issues rather it is the way in which they are used and the analysis they are subjected to for the same.

[36] The Ethical issues to make the technology distinct using data are as follows:

|  |
| --- |
| * *The pervasiveness of modern data technology means we might legitimately regard it as an aspect of societal infrastructure, in the same way that mathematics, language, transport, and so on are infrastructural.* * *The interconnectedness of data. Data on travel may be used for discovering things far beyond mere travel patterns; data on purchases are not solely relevant to purchases; and so on.* * *The dynamic nature of data. Modern data sets often evolve and accumulate over time so that they may permit discoveries in the future that they do not permit today.* * *Real-time and online analysis and decision making, as data arrive* * *Synergistic analysis through merging and combination of data sets.* * *Lack of space, time, and social context limitation on scope of data (data may describe and be used regardless of where, when, and for what purpose they were collected).* * *Ability to use for unexpected purposes and to reveal unexpected information (this is the core purpose of data mining).* * *Risk of exceptional intrusiveness since it is impossible to avoid having data about individuals stored in multiple databases.* * *Potential for misuse, privacy breach, blackmail, and other crimes.* * *Subtle ownership issues (‘‘my’’ data might also be your data; I can sell ‘‘my’’ data while retaining them, and so on, as discussed in detail below).* |

Table2: Checklist Ethical Issues

Hand discussed the Examples in the data world including “*the original Code of Practice of the UK Statistics Authority which are based on eight principles, the report of the joint British Academy/Royal Society on Data Management and Use: Governance in the 21st Century based on five principles, the Accenture Universal Principles of Data Ethics that had twelve principles, the ACM Code of Ethics and Professional Conduct beginning with seven general moral principles (which ‘‘a computing professional should .’’), and others. At the highest level, the principles include such things as integrity, honesty, objectivity, responsibility, trustworthiness, impartiality, nondiscrimination, transparency, accountability, fairness, robustness, resilience, usability, efficiency, and independence. All good and desirable characteristics. These are then refined into lower, but still high-level principles.*”

van Asbroeck et al. [38] reviewed the legal aspects in the EU, commenting that ‘‘*the current legal framework relating to data ownership is not satisfactory. No specific ownership right subsists in data and the existing data-related rights do not respond sufficiently or adequately to the needs of the actors in the data value cycle.*’’ Further as expected for the lack of the harmonization of legal aspects, he also pointed out that that the ‘‘*the issue of data ownership is even more complicated by the data value cycle, which can be rather complex and involves numerous stakeholders. This increases the difficulties in determining who could or would be entitled to claim ownership in data. Many of such stakeholders may attempt claiming ownership in data because, for instance, they create or generate data, or because they use, compile, select, structure, reformat, enrich, analyze purchase of, take a license on, or add value to the data.*’’ So it's not merely the trouble of aggregation the information within the initial place which could confer the possession, however the trouble of manipulating in addition. in additional detail, among the conclusions reached by van Asbroeck et al.19 are as follows: case law at EU level doesn't expressly acknowledge Associate in Nursing possession right in knowledge; that non-public data aren't essentially owned by the individual, in order that Associate in Nursing possession right for knowledge controllers or processors (as outlined within the GDPR) cannot be excluded, however that this may be subject to the individual’s management over his or her personal data; the principles regarding possession of physical entities aren't invariably relevant to data; there area unit several laws which will impact ‘‘*a company’s management of, the access to, or the rights in knowledge*’’; and none of belongings rights or trade secrets provides ‘‘adequate protection of (ownership in) data.’’ Suggesting that resolution the problems with written agreement agreements would be onerous and possibly not possible to control with legal certainty, they are going on to propose an answer via ‘‘the creation of a nonexclusive, flexible, and protrusible possession right in data(sets), with a knowledge traceability obligation as a safeguard.’’ This notion of traceability looks to American state to be elementary to easing several of the moral and untechnical knowledge challenges. Consent and Purpose The notions of consent or consent have long been necessary in analysis, particularly in tending analysis. therein domain, the thought is that consent ought to be obtained before Associate in Nursing intervention, which the intervention ought to be supported a sound understanding of its implications and attainable consequences. However, it's questionable what quantity this is often relevant to or practicable within the fashionable knowledge world, for reasons as well as the 2 points noted above: that the terribly essence of the promise of recent huge knowledge is that future applications area unit one and unknown (indeed, unknowable), which the info usually exist already in databases in order that studies area unit largely noninterventional.

Hand briefly discussed about the data and it’s nature of information, particularly personal knowledge, whether or not the notion of possession is important, consent and purpose, trustiness of information yet as of algorithms and of these mistreatment the information, and matters of privacy and confidentiality. None of those topics is simple in a very knowledge context: all have nuances and even apparent contradictions. A primary supply of the complexities is that the undeniable fact that the employment to that knowledge may be place is essentially unknowable—but this is often the supply of the terribly power of information science, that some marvelous new and maybe even unknown insight would possibly emerge from the perspicacious analysis of information. In pains to strike an acceptable balance between profit and risk, we want to be clear regarding World Health Organization advantages and World Health Organization incurs the chance. If these square measures borne by totally different actors, imbalances will occur—and so moral disasters may end up.

There are different concepts that are adopted by the different bodies as to the concept of data ethics. The Data Governance Working Group defines human budding as the ‘‘*the overarching principle that should guide the development of systems of data governance*.’’ [39] In spite of this we can say that the ethical principles are not universally acceptable. In many countries, there are laws that are defined about what can be done with data and its possibilities. [40] Based on these general considerations, the HLEG derived four major ethical principles for AI:

*1.Respect for human autonomy*

*•*

* *Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process.*
* *AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills.*
* *•*
* *The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight over work processes in AI systems.*

*•*

*2. Prevention of harm*

* *This entails the protection of human dignity as well as mental and physical integrity.*
* *•*
* *AI systems and the environments in which they operate must be safe and secure. They must be technically robust and it should be ensured that they are not open to malicious use.*
* *•*
* *Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems.*
* *•*
* *Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens.*
* *•*
* *Preventing harm also entails consideration of the natural environment and all living beings.*

*3. Fairness*

*3. Fairness*

*There are both substantive and procedural dimensions of fairness:*

* + *Substantive dimension of fairness:*
* *Ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. If unfair biases can be avoided, AI systems could even increase societal fairness. Equal opportunity in terms of access to education, goods, services and technology should also be fostered. Moreover, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice.* 
  + *Procedural dimension of fairness:*
* *The ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them. In order to do so, the entity accountable for the decision must be identifiable, and the decision-making processes should be explicable*

*4. Explicability*

* *Explicability is crucial for building and maintaining users’ trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested.*

The ethical principles discussed have been transformed into 7 key requirements for achieving an AI system that is ethical and trustworthy. The 7 key requirements are stated in the figure below. The AI developers as well as the designers need to keep in mind these requirements before implementing them in the system. All the buyers and the developers who are deploying it on their system need to make sure that they follow all the guidelines in their product and are ethical.



Figure14: 7 Key requirements HLEG (Source: European Commission High Level Expert Group on Artificial Intelligence (April 2019, p. 14))

In the present generation everything is available open-source, and it is very easy to just scrape any website without looking at the privacy policy of the website and asking for permission to scrape. Wel and Royakkers in 2004 discussed about the ethical issue in data mining through web. According to them “*Web mining refers to the whole of data mining and related techniques that are used to automatically discover and extract information from web documents and services. When used in a business context and applied to some type of personal data, it helps companies to build detailed customer profiles, and gain marketing intelligence. Web mining does, however, pose a threat to some important ethical values like privacy and individuality. Web mining makes it difficult for an individual to autonomously control the unveiling and dissemination of data about his/her private life. Web content and structure mining is a cause for concern when data published on the web in a certain context is mined and combined with other data for use in a totally different context. Web usage mining raises privacy concerns when web users are traced, and their actions are analysed without their knowledge.*”

The individuality and privacy is supposed to be protected as well as respected to ensure that people are treated fairly and protected. There should be aware ness about unethical practices and dangers that these practices can possess. To maintain integrity all the members of the value chain should make sure that they maintain integrity and take responsibility in their action and in whatever they do.

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

Developments In E-Learning- E-learning has been gaining a lot of popularity in recent times all around the world and has pushed a significant number of researchers into exploring and coming up with new idea in this domain. In disciplines like IT, Distance Learning, Mass Communication and Education there is a lot of research that is in progress. Along with this a lot of research projects are having E-learning as a part or domain included in them.

Many researchers and scholars have been pursuing M.Phil. and doctoral research in the fields of E-Learning. There are also a lot of Journal which are completely related to E-Learning.

[31] In 2015 Gaur mentioned that E-Learning can be divided into 2 categories that are Complete Online Learning- “*This type of learning depends completely on e-learning tools. Delivery of course materials, discussions, assignment evaluation, examination and other evaluations are performed only on learning platform. It provides maximum flexibility to the learners regarding place and time of learning.”* Whereas the Blended Learning is “*used to increase the effectiveness of conventional face-to-face methods as on add on tools. Sometimes these tools are used for reducing face-to-face contact time. Some part of learning activities is occurred in classroom and rest of the part at e-learning platform.*”

Further the findings of the qualitative study preformed by Gaur in 2015 stresses on the need for technology in education. She concluded that “*There is a possibility to make education interesting, easy to understand and broader with the help of technology. E-learning tools have potential to change the whole educational scenario but few limitations are also there. These are psychological and infrastructure related. As far as research trends in e-learning is concerned we can say without fail that sufficient research works have been done and e-learning is still a popular area of research for the researchers of various disciplines.*”

[32] Gunasekaran, McNeil and Shau identified that “*Corporate and campus agendas have started to recognize e-learning as having the power to really transform the performance, knowledge and skills landscape, so much so that the International Data Corporation estimates that the corporate spend on e-learning alone will increase from $1bn in 1999 to over $11bn in 2003 (Henry, 2001)[34]. Education and training is poised to become ne of the largest sectors in the world.*”

They proposed a framework which had the following steps included for the success of E-Learning:

1. *Develop an interactive and online resource model that should consider the stakeholders at various levels including lecturers, students and tutors in the distance process.*
2. *Develop significant and active learning by stimulating student participation in the use of the different resources.*
3. *Improve and ensure the procurement of the most relevant information and the establishment of the most significant communications for each user type by providing quick, efficient personalized access.*
4. *Promote new ways of communication that facilitate the establishment of working groups of students and lecturers with common interests by increasing the flow of information between all the protagonists participating in the process.*

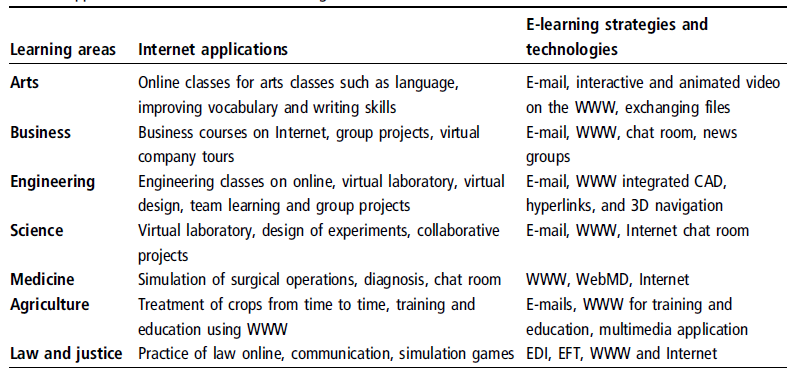
*(5) Stimulate the use of technological resources available among users.*

*The platform specifications should include clear and structured documentation on the*

*site's possibilities, search service, updated documentation on the students, and*

*interaction spaces (shared workspace, monitoring people, online annotations, news*

*groups, mailing lists, adaptive interface, communication of events, glossary of terms, electronic bulletin, software repository, documentation). It has been emphasized that it is possible to support all aspects of the educational process to at least some degree within an online distance learning scenario.[32]*

**

*Table3: Application of Internet in E-Learning*

Within a few years the technology of computer would get outdated. Bradband availability is only going to increase the learning options provided in the online mode by making use of advanced techniques like Deep Learning , Augmented Reality, Virtual Reality, Animation Techniques and Holograms with virtual Laboratories. The Learning Opportunities are going to only extend futher with the digitalization of the world and the industry. The fact that E-Learning is pocket friendly to the masses, along with it’s ease of access, flexibility and repeatibilty.

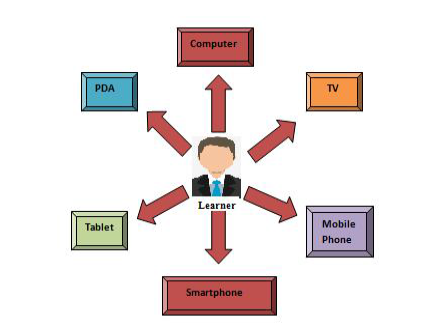
Gunasekaran, McNeil and Dennis Shauconcluded that *“The objectives of e-learning are dependent on the quality of the teaching process and the effectiveness of online access. Several script languages and Internet- oriented languages have been used to embed interactivity into Web pages.”*

Makara and Satish in 2017 analyzed the emerging trends of E-Learning in India. According to them “*India is a home of many latest e-learning trends in education that are being used by the developed countries from a very*

*long period. Some of the emerging trends of learning in India are:*

* *Distance education - Postal, Radio, TV*
* *E-Learning*
* *Open Educational Resources (OER)*
* *Cloud based E-Learning*
* *Big Data in E-Learning*
* *Automated Course Authoring*
* *Responsive LMS*
* *Ubiquitous Learning*
* *Massive Open Online Courses*

Online Learning provides people to access the same amount of facilities which others are getting thus eliminating the hierarchical concept of education for the privileged and making it accessible to all and everyone at a minimal cost which a lot of underprivileged students may not be able to achieve due to personal or financial constraints. And in a large country like India where there is a lot of poverty E-learning is the ultimate source for the abolishment of poverty and for giving rise to a new tomorrow for the country where everyone gets an equal opportunity irrespective of where they are born.

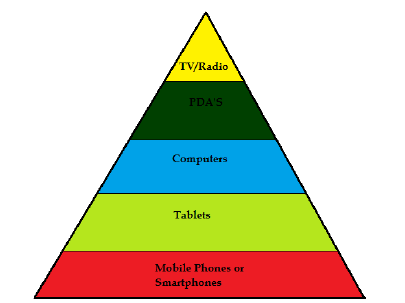
**

*Fig15: Devices used for E-Learning*

Further analyzed the Technological Trend present in E-Learning. Few decades ago all the education was on hard papers but after the advent of computers the entire game changes and the hard copies slowly started revolutionizing into softcopies. And after internet was introduced a new variety of distance courses were launched which enabled students to learn subjects and skills of his choice at a very low cost.

Diving further into the facts it was found out by Makara and Satish in 2017 that “*There was time since 1840’s, first Isaac Pitman taught his pupils shorthand via correspondence. Shorthand is a symbolic writing was designed to improve writing speed and was popular amongst secretaries, and other individuals who did a great deal of note taking or writing. In 1924, the first testing machine was invented. This device allowed students to test themselves. Then, in 1954, BF Skinner, a Harvard Professor, invented the teaching machine, which enabled schools to administer programmed instruction to their students. It was not until 1960 however that the first computer-based training program was introduced to the world. This computer-based training program (or CBT program) was known as PLATO-Programmed Logic for Automated teaching Operations. It was originally designed for students attending the University of Illinois but ended up being used in schools throughout the area. Because of evolution of computer and internet in the late 20th century, e-learning tools, and delivery methods also got expanded. The first MAC in the year 1980’s enabled individuals to have computers in their homes, making it easier for them to learn about subjects and develop certain skill sets. Then, in the following decade, virtual learning environments began to truly thrive, with people gaining access to a wealth of online information and e-learning opportunities. Presently in 2000’s, businesses began using e-learning to train their employees. New and experienced workers alike now had the opportunity to improve upon their industry knowledge base and expand their skill sets. At home individuals were granted access to programs that offered them the ability to earn online degrees and enrich their lives through expanded knowledge. Today, e-learning is more popular than ever, with countless individuals realizing the benefits that online learning can offer.”*

In today’s century going to a college was very popular and a college degree was a necessity for almost anyone but with the advent of E-Learning a lot has changes and quality information is accessible and available to everyone at a much lower costs and the privatization of education is starting to become a boon where monopoly has been abolished and a lot more players are venturing out into the field of E-Learning. The online degrees are now proving to be equally important as the college counterparts and are being provided based on the learners understanding and skills. There are a lot of devices depending on the accessibility and its usage to the students and common people. The figure given below shows the usage of these devices in E-Learning



*Fig16: Pyramid of E-Devices used for E-Learning*

There were 10 different technological trend that were identified by Satish and Makara in 2017. The trend identified were as follows-

*1. Mobile Learning:*

*Mobile learning is the ability of an individual to obtain or provide educational content on personal pocket devices such as PDAs, Smartphone's and mobile phones. These devices with access to internet connection or with the availability of the resources on itself will be a great source of e-learning. Today everyone has smart mobile phones on which huge memory and faster internet availability can be achieved and this makes the user to learn anything he desire at anytime and in any place. These devices are so portable that as if they walk with the user and never be a hurdle to carry them from one place to another. Mobile learning apps are also great source of this kind of learning. Mobile learning has a great role to accomplish the success of e-learning in a country not only like India but also in many developing countries of the world. Mobile phones are considered to be best platform for e-learning because:*

* Potential to reach masses.*

* Carried all the time.*

* Easy to use.*

* Cheap.*

* Just in time learning.*

*2. Micro Learning:*

*Micro learning is a way of teaching and delivering content to learners in small, very specific bursts. The learners are in control of what they are learning and when they are learning.Typically designed and delivered in rich media formats, it is a learner-centric approach that provides just-in-time training that is available on multiple devices. All these aspects ensure that it can be easily accessed, quickly completed, and easily applied by the learners. The main benefits of this kind of learning are:*

* Accessible.*

* Rich media.*

* Less time consuming.*

* Just-in-time*

* Learner-centric*

* Less time consuming.*

*Micro Learning also provides huge benefits to*

*business environments like:*

* Affordable and agile.*

* Shorter development cycle.*

* Easy to update.*

* Wider application.*

* High impact.*

*3. Internet of Things (IoT):*

*It refers to the ever growing network of physical things or objects around us which hold IP address for internet connectivity, and the communication that occurs between these connected objects and other internet enabled devices and systems. It include not only the traditional things like desktop, laptops, smartphones, tablets etc., but also all other things that utilize embedded technology to communicate and interact with the external environment via the internet. Here IoT can be a great tool for the learners instant learning for like daily study exercises, daily news or any information study the learner instantly plans to learn. In this kind of technology the learner will be assumed like an object in the whole system, here the system connected will detect the new updated and specific learner IP and gives him all the*

*updates he desires from the whole network of physical devices or objects in which he is connected.*

*Some major benefits of IoT in case of e-learning can be listed as follows:*

* Learner is part of the learning system as an*

*entity.*

* Continuous tracking of learner by the system.*

* Continuous updates to learners.*

*4. Cloud based E-Learning:*

*This type of e-learning is creating ripples in the field of education and business. Thesee-learning systems are hosted on the internet and can be easily accessed by logging into a service provider's site. Rather than*

*installing all the software and course on user's or learner's computer, the instructional designers will simply use their internet browsers to upload course content, create new courses, and communicate with*

*learners and users directly. This is all done by learner management system, which also gives the designer the ability to store information on the cloud, which can be remotely accessed by other, approved users.*

*here are some notable advantages of using this type of e-learning methodology. They are:*

*• Faster deployment.*

*• Cost predictability.*

*• Easier to maintain.*

*• More storage space.*

*• Fully customizable and scalable.*

*• Learner oriented service request.*

*5. Gamification:*

*In e-learning, gamification is the hot topic today with lots of good reasons. It has proven to be useful in helping learners further comprehend and apply new information they want. This type of e-learning is through games and it really depends on the program and the audience desires. Especially children can benefit more from this kind of e-learning because it create interest in them and make them do again and again for long time. Gamification not only helps online learners acquire knowledge and skills more effectively but also it allows them to retain the information and commit it to long term memory for the future use. Some of the important benefits of the Gamification in e-learning are:*

*• Better learning experience.*

*• Better learning environment.*

*• Takes to Mastery level.*

*• Instant feedback.*

*• Helps to remember for long time.*

*6. Adaptive E-Learning:*

*Adaptive e-learning uses computers as an interactive teaching devices. These methodology arrange the allocation of human and mediated resources according to the unique learning needs of each learner. This is also known as intelligent tutoring and it has its origin from artificial intelligence and started gaining its popularity in recent decades. Adaptive learning system can be implemented on the internet for use in distance learning and group collaboration. The field of distance learning is greatly incorporating the aspects of adaptive learning.Adaptive learning has been implemented in several kinds of educational systems such as adaptive hypermedia, intelligent tutoring systems, computerized adaptive testing, and computer-based pedagogical agents. Some major benefits of this technological trend in case of e-learning can be:*

*• Very rich study resources.*

*• Tracking of learner by Tutor or human allocated.*

*• Saves lots of time of learners.*

*7. Augmented Reality:*

*This technology superimposes a computer-generated image on a user view of real world.It is related to a more general concept called mediated reality. It is really a great boon technology for the students or learners in general. Whenever the learner wants to know more of the things, he is seeing in the real world, using a device like mobile phone on which the augmented reality software in enabled, the learner can get all the information regarding the object. This technology needs device, internet, and software of augmented reality. This technology has good future in e-learning. This technology has long way to go for making the learners learn the things just by projecting the device. This technology has just introduced in 1962. Google glass is a very good example for augmented reality. Some benefits of augmented reality in case of e-learning can be as follows:*

*• Easy instant deep learning of things.*

*• Huge collection of information.*

*8. Video E-Learning:*

*This kind of learning helps the learner to grasp the content by watching the videos. When a learner wants to get an idea on some specific topic in details, he visits the youtube.com for sure. This kind of e-learning gives a very quick idea and helps to understand the things with multimedia affects. Youtube.com is a rich source of video e-learning content and in the same way TV, CD's and storage devices with educational videos paves the way to this kind of e-learning and these days this kind of learning is on full swing as it saves lots of time of the learner compared to reading line by line full stuff on computer or any other electronic device and also it might take little concentration of the learner than to reading stuff. This serves as very effective medium of e-learning. Some benefits of video e-learning can be as follows:*

*• Video explanation.*

*• More information in less time.*

*• Best learning experience.*

*9. Beacon E-Learning:*

*This is one more boon technological trend to the elearning. This beacon e-learning or beacon technology is a wireless device that transmit signals to other nearby devices via low-energy Bluetooth connections. This is used as an Indoor Positioning System (IPS). These IPS beacons can wirelessly locate people and objects within a specific range and then trigger an action on a nearby wireless device. Most importantly this is safe and secure, only accessible through paired applications andeasy to download and use. In these modern times almost 99% college students have smartphone, and they can use their phones for effective e-learning. Some great benefits of this technological trend is:*

*• Easier Campus Navigation.*

*• Better Accessibility.*

*• More Powerful Communication.*

*• Increased Intellectual Discovery.*

*• Insightful Data.*

*• Improved In-Class Experiences.*

*10. Artificial Intelligence:*

*This technological trend can also be a great shaper to e-learning world. Artificial Intelligence and robots are not same things. It is something like intelligent software which is designed to take some intelligent*

*actions reading entire environment around it. This can produce a very good instructor which helps in making each student a good expert in their own field. Some latest outcomes in artificial intelligence*

*learning instructors are like SIRI and VIV, these are software which will answer simple queries to much complex queries of the learners. The involvement of artificial intelligence in e-learning helps the learner to*

*take very wise decisions and quality resource, which in turn helps the learner to excel and make his achievements at faster phase. Some major benefits of AI in case of e-learning can be as follows:*

*• Provides expert tutors for learners.*

*• Automated teaching.*

*• Hugh and rich information.*

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

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

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

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

A close up of text on a white background

Description automatically generated

*Figure17: Graphical model for a Bayesian Network*

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

For this study we picked 1 Electrical Electronics Department course ADVD for the study to be conducted upon. For the students the data was categorized and collected through floating a google form which had 4 main attributes as shown in the figurebelow 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

Fig18: The Collected Gform Data

The collected data about the students and the platform attributes would be refined using several methods of cleaning, encoding and normalizations,

According to Zhuang and Zhang 3 types of models were proposed that could be trained by the Bayesian Networks the first one being data centric model, second one being an efficiency centric model and the third one being an expert centric model.

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

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

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

Description automatically generatedTable

Description automatically generatedpredictions.

Fig19:The precprocessed Data Table

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

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

We then Created a Gaussian Classifier with-model = GaussianNB(), then trained the model using the training sets and the model created above.

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

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

Text

Description automatically generated

Fig20: 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

*Fig21: Correlation Matrix*

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

*Columns = {' Comm\_Platform\_Satisfaction(1-Yes, 0-No)':'A', 'Pref\_Live\_Lectures':'B', 'Pref\_Audio\_Lectures':'C', 'Pref\_Recorded\_Content':'D', 'Pref\_PowerPoint\_presentation':'E', 'Pref\_Interactive\_Sessions':'F', 'PercentileAbleTo\_Score':'G', 'Question\_Difficulty':'H', 'Curr\_Live\_Lectures':'I', 'Curr\_Audio\_Lectures':'J', 'Curr\_Recorded\_Content':'K', 'Curr\_PowerPoint\_presentation':'L', 'Curr\_Interactive\_Sessions':'M', 'PlatformAccessEase(1-Easy, 2-Medium, 3-Hard)':'N', 'platform\_UI\_intuitive \_easy-to-use':'O', 'E-Learning\_feature\_incorporate':'P', 'THA':'Q', 'PracticeExercises':'R', 'ReadingMaterials':'S', 'QuizSolutions':'T', 'TextBooks':'U', 'Research Papers':'V', 'Materials\_provision\_Platform':'W', 'Tests\_Assignments\_in\_course':'X', 'Result\_Show\_Time':'Y', 'AvgDur\_Tests':'Z', 'WeeklyHours\_Browse\_Platform':'AA', 'Portal\_Login\_Freq':'AB', 'CourseProgress\_Satisfaction\_x':'AC', 'Prim\_Comm\_Platform':'AD' }*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Chart, pie chart

Description automatically generated

Fig22: Pie Chart - Course Progression

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

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

References-

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

***30.*** *Iqbal, Khalid & xu, Yin & Hao, Hong-Wei & Ilyas, Qazi & Ali, Hazrat. (2015). An Overview of Bayesian Network Applications in Uncertain Domains. International Journal of Computer Theory and Engineering. 7. 416-427. 10.7763/IJCTE.2015.V7.996.*

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

***32.*** *Gunasekaran, Angappa & McNeil, Ronald & Shaul, Dennis. (2002). E-learning: Research and applications. Industrial and Commercial Training. 34. 44-53. 10.1108/00197850210417528.*

***33.*** *Makara, Vivekananda & R U V N, Satish. (2017). EMERGING TRENDS OF E-LEARNING IN INDIA. 4.*

***34****. Henry, P. (2001), ``E-learning technology, content and services'', Education + Training, Vol. 43 No. 4, pp. 249-55*

***35.*** *Roh, Y., Heo, G., & Whang, S. E. (2019). A Survey on Data Collection for Machine Learning: A Big Data -- AI Integration Perspective. ArXiv:1811.03402 [Cs, Stat].* [*http://arxiv.org/abs/1811.03402*](http://arxiv.org/abs/1811.03402)

***36.*** *Hand, D. J. (2018). Aspects of Data Ethics in a Changing World: Where Are We Now? Big Data, 6(3), 176–190.* [*https://doi.org/10.1089/big.2018.0083*](https://doi.org/10.1089/big.2018.0083)

***37.*** *van Wel, L., & Royakkers, L. (2004). Ethical issues in web data mining. Ethics and Information Technology, 6(2), 129–140.* [*https://doi.org/10.1023/B:ETIN.0000047476.05912.3d*](https://doi.org/10.1023/B:ETIN.0000047476.05912.3d)

***38.*** *van Asbroeck B, Debussche J, Ce´sar J. 2017. Building the European Data Economy: Data ownership. White Paper, Bird and Bird. Available online at https://sites-twobirds.vuture.net/1/773/uploads/white-paperownership-of-data-(final).PDF (last accessed January 16, 2018).*

***39.*** *DGWG. 2017. Data management and use: Governance in the 21st century. A joint report by the British Academy and the Royal Society. Available online at www.britac.ac.uk/sites/default/files/Data%20management%20 and%20use%20-%20Governance%20in%20the%2021st%20century.pdf (last accessed March 26, 2018).*

***40.*** *Lenz, Rainer. (2019). Big Data: Ethics and Law. SSRN Electronic Journal. 10.2139/ssrn.3459004.*

***41.*** *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)

***42.*** *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)

***43.*** *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)

***44.*** *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)

***45.*** *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)

***46.*** *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)

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

***48.*** Colace, F., & Santo, M. D. (2010). Ontology for E-Learning: A Bayesian Approach. *IEEE Transactions on Education*, *53*(2), 223–233. <https://doi.org/10.1109/TE.2009.2012537>

***49.*** 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