International Journal of Scientific Engineering and Research (IJSER)
ISSN (Online): 2347-3878
Impact Factor (2022): 7.741

Data Mining and Its Applications in Higher Education

Sankalp Kohli

Education & Research in Computer Networking India, Ministry of Electronics and Information Technology Government of India

Abstract: ***Data mining is an arena of intersection of computer science and statistics intended to ascertain patterns in the information bank. The main purpose of the data mining process is to extract the useful information from the dossier of data and mould it into an understandable structure for future use. There are various process and techniques used to carry out data mining successfully.***

Keywords: Analytics, Data Mining, Higher Education, Statistics

1. Introduction

A. What Is Data Mining?

Data mining is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Data mining depends on effective data collection, warehousing, and computer processing.

B. Data Mining Techniques

Data mining uses algorithms and various techniques to convert large collections of data into useful output. The most popular types of data mining techniques include:

- Association rules, also referred to as market basket analysis, searches for relationships between variables. This relationship in itself creates additional value within the data set as it strives to link pieces of data. For example, association rules would search a company's sales history to see which products are most commonly purchased together; with this information, stores can plan, promote, and forecast accordingly.
- Classification uses predefined classes to assign to objects. These classes describe characteristics of items or represent what the data points have in common with each. This data mining technique allows the underlying data to be more neatly categorized and summarized across similar features or product lines.
- Clustering is similar to classification. However, clustering identified similarities between objects, then groups those items based on what makes them different from other items. While classification may result in groups such as "shampoo", "conditioner", "soap", and "toothpaste", clustering may identify groups such as "hair care" and "dental health".
- Decision trees are used to classify or predict an outcome based on a set list of criteria or decisions. A decision tree is used to ask for input of a series of cascading questions that sort the dataset based on responses given. Sometimes depicted as a tree-like visual, a decision tree allows for specific direction and user input when drilling deeper into the data.
- K-Nearest Neighbor (KNN) is an algorithm that classifies data based on its proximity to other data. The basis for KNN is rooted in the assumption that data points that are close to each are more similar to each other than other bits of data. This non-parametric, supervised technique is used to predict features of a group based on individual data points.
- Neural networks process data through the use of nodes. These nodes are comprised of inputs, weights, and an output. Data is mapped through supervised learning (similar to how the human brain is interconnected). This model can be fit to give threshold values to determine a model's accuracy.
- Predictive analysis strives to leverage historical information to build graphical or mathematical models to forecast future outcomes. Overlapping with regression analysis, this data mining technique aims at supporting an unknown figure in the future based on current data on hand.

C. The Data Mining Process

To be most effective, data analysts generally follow a certain flow of tasks along the data mining process. Without this structure, an analyst may encounter an issue in the middle of their analysis that could have easily been prevented had they prepared for it earlier. The data mining process is usually broken into the following steps.

Step 1: Understand the Business

Before any data is touched, extracted, cleaned, or analyzed, it is important to understand the underlying entity and the project at hand. What are the goals the company is trying to achieve by mining data? What is their current business situation? What are the findings of a SWOT analysis? Before looking at any data, the mining process starts by understanding what will define success at the end of the process.

Step 2: Understand the Data

Once the business problem has been clearly defined, it's time to start thinking about data. This includes what sources are available, how it will be secured stored, how information will be gathered, and what the final outcome or analysis may look like. This step also critically thinks about what limits there are to data, storage, security, and collection and assesses how these constraints will impact the data mining process.

Step 3: Prepare the Data

It's now time to get our hands on information. Data is gathered, uploaded, extracted, or calculated. It is then cleaned, standardized, scrubbed for outliers, assessed for mistakes, and checked for reasonableness. During this stage of data mining, the data may also be checked for size as an overbearing collection of information may unnecessarily slow computations and analysis.

Step 4: Build the Model

With our clean data set in hand, it's time to crunch the numbers. Data scientists use the types of data mining above to search for relationships, trends, associations, or sequential patterns. The data may also be fed into predictive models to assess how previous bits of information may translate into future outcomes.

Step 5: Evaluate the Results

The data-centered aspect of data mining concludes by assessing the findings of the data model (s). The outcomes from the analysis may be aggregated, interpreted, and presented to decision-makers that have largely be excluded from the data mining process to this point. In this step, organizations can choose to make decisions based on the findings.

Step 6: Implement Change and Monitor

The data mining process concludes with management taking steps in response to the findings of the analysis. The company may decide the information was not strong enough or the findings were not relevant to change course. Alternatively, the company may strategically pivot based on findings. In either case, management reviews the ultimate impacts of the business and re-creates future data mining loops by identifying new business problems or opportunities.

Different data mining processing models will have different steps, though the general process is usually pretty similar. For example, the Knowledge Discovery Databases model has nine steps, the CRISP-DM model has six steps, and the SEMMA process model has five steps.

D. How Data Mining Works

Data mining involves exploring and analysing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users.

The data mining process breaks down into five steps. First, organizations collect data and load it into their data warehouses. Next, they store and manage the data, either on in-house servers or the cloud. Business analysts, management teams, and information technology professionals access the data and determine how they want to organize it. Then, application software sorts the data based on the user's results, and finally, the end-user presents the data in an easy-to-share format, such as a graph or table.

Data mining programs analyse relationships and patterns in data based on what user’s request. For example, a company can use data mining software to create classes of information. To illustrate, imagine a restaurant wants to use data mining to determine when it should offer certain specials. It looks at the information it has collected and creates classes based on when customers visit and what they order. In other cases, data miners find clusters of information based on logical relationships or look at associations and sequential patterns to draw conclusions about trends in consumer behaviour.

Warehousing is an important aspect of data mining. Warehousing is when companies centralize their data into one database or program. With a data warehouse, an organization may spin off segments of the data for specific users to analyse and use. However, in other cases, analysts may start with the data they want and create a data warehouse based on those specs. Cloud data warehouse solutions use space and power of a cloud provider to store data from data sources. This allows smaller companies to leverage digital solutions for storage, security, and analytics.

E. Applications of Data Mining

In today's age of information, it seems like almost every department, industry, sector, and company can make use of data mining. Data mining is a vague process that has many different applications as long as there is a body of data to analyze.

Sales

The ultimate goal of a company is to make money, and data mining encourages smarter, more efficient use of capital to drive revenue growth. Consider the point-of-sale register at your favorite local coffee shop. For every sale, that coffeehouse collects the time a purchase was made, what products were sold together, and what baked goods are most popular. Using this information, the shop can strategically craft its product line.

Marketing

Once the coffeehouse above knows it’s ideal line-up, it's time to implement the changes. However, to make its marketing efforts more effective, the store can use data mining to understand where its clients see ads, what demographics to target, where to place digital ads, and what marketing strategies most resonate with customers. This includes aligning marketing campaigns, promotional offers, cross-sell offers, and programs to findings of data mining.

Manufacturing

For companies that produce their own goods, data mining plays an integral part in analyzing how much each raw material costs, what materials are being used most efficiently, how time is spent along the manufacturing process, and what bottlenecks negatively impact the process. Data mining helps ensure the flow of goods is uninterrupted and least costly.

Fraud Detection

The heart of data mining is finding patterns, trends, and correlations that link data points together. Therefore, a company can use data mining to identify outliers or correlations that should not exist. For example, a company may analyze its cash flow and find a reoccurring transaction to an unknown account. If this is unexpected, the company may wish to investigate should funds be potentially mismanaged.

Human Resources

Human resources often has a wide range of data available for processing including data on retention, promotions, salary ranges, company benefits and utilization of those benefits, and employee satisfaction surveys. Data mining can correlate this data to get a better understanding of why employees leave and what entices recruits to join.

Customer Service

Customer satisfaction may be caused (or destroyed) for a variety of reasons. Imagine a company that ships goods. A customer may become unhappy with ship time, shipping quality, or communication on shipment expectations. That same customer may become frustrated with long telephone wait times or slow e-mail responses. Data mining gathers operational information about customer interactions and summarizes findings to determine weak points as well as highlights of what the company is doing right.

F. Benefits of Data Mining

Data mining ensures a company is collecting and analyzing reliable data. It is often a more rigid, structured process that formally identifies a problem, gathers data related to the problem, and strives to formulate a solution. Therefore, data mining helps a business become more profitable, efficient, or operationally stronger.

Data mining can look very different across applications, but the overall process can be used with almost any new or legacy application. Essentially any type of data can be gathered and analyzed, and almost every business problem that relies on qualifiable evidence can be tackled using data mining.

The end goal of data mining is to take raw bits of information and determine if there is cohesion or correlation among the data. This benefit of data mining allows a company to create value with the information they have on hand that would otherwise not be overly apparent. Though data models can be complex, they can also yield fascinating results, unearth hidden trends, and suggest unique strategies.

G. Limitations of Data Mining

This complexity of data mining is one of the largest disadvantages to the process. Data analytics often requires technical skill sets and certain software tools. Some smaller companies may find this to be a barrier of entry too difficult to overcome.

Data mining doesn't always guarantee results. A company may perform statistical analysis, make conclusions based on strong data, implement changes, and not reap any benefits. Through inaccurate findings, market changes, model errors, or inappropriate data populations, data mining can only guide decisions and not ensure outcomes.

There is also a cost component to data mining. Data tools may require on-going costly subscriptions, and some bits of data may be expensive to obtain. Security and privacy concerns can be pacified, though additional IT infrastructure may be costly as well. Data mining may also be most effective when using huge data sets; however, these data sets must be stored and require heavy computational power to analyze.

Even large companies or government agencies have challenges with data mining. Consider the FDA's white paper on data mining that outlines the challenges of bad information, duplicate data, underreporting, or over reporting.

H. Examples of Data Mining

Data mining can be used for good, or it can be used illicitly. Here is an example of both.

eBay and e-Commerce

eBay collects countless bits of information every day, ranging from listings, sales, buyers, and sellers. eBay uses data mining to attribute relationships between products, assess desired price ranges, analyze prior purchase patterns, and forms product categories. eBay outlines the recommendation process as:

1) Raw item metadata and user historical data is aggregated.
2) Scripts are run on a trained model to generate and predict the item and user.
3) A KNN search is performed.
4) The results are written to a database.
5) The real-time recommendation takes the user ID, calls the database results, and displays them to the user.

I. What Are the Types of Data Mining?

Data mining is broken into two basic aspects: predictive data mining and descriptive data mining. Predictive data mining is a type of analysis that extracts data that may be helpful in determining an outcome. Description data mining is a type of analysis that informs users of that data of a given outcome.

J. How Is Data Mining Done?

Data mining relies on big data and advanced computing processes including machine learning and other forms of artificial intelligence (AI). The goal is to find patterns that can lead to inferences or predictions from otherwise unstructured or large data sets.

What Is Another Term for Data Mining?

Data mining also goes by the less-used term knowledge discover in data, or KDD.

K. Where Is Data Mining Used?

Data mining applications range from the financial sector to look for patterns in the markets to governments trying to identify potential security threats. Corporations, and especially online and social media companies, use data mining on their users to create profitable advertising and marketing campaigns that target specific sets of users.

L. The Bottom Line

Modern businesses have the ability to gather information on customers, products, manufacturing lines, employees, and storefronts. These random pieces of information may not tell a story, but the use of data mining techniques, applications, and tools helps pieces together information to drive value. The ultimate goal of the data mining process is to compile data, analyze the results, and execute operational strategies based on data mining results.

M. Key Takeaways

- Data mining is the process of analyzing a large batch of information to discern trends and patterns.
- Data mining can be used by corporations for everything from learning about what customers are interested in or want to buy to fraud detection and spam filtering.
- Data mining programs break down patterns and connections in data based on what information users request or provide.
- Social media companies use data mining techniques to commodify their users in order to generate profit.
- This use of data mining has come under criticism lately as users are often unaware of the data mining happening with their personal information, especially when it is used to influence preferences.

Facebook-Cambridge Analytica Scandal

Another cautionary example of data mining includes the Facebook-Cambridge Analytica data scandal. During the 2010s, the British consulting firm Cambridge Analytical collected personal data belong to millions of Facebook users. This information was later analyzed to assist the 2016 presidential campaigns of Ted Cruz and Donald Trump. It is also suspected that Cambridge Analytica interfered with other notable events such as the Brexit referendum.

In slight of inappropriate data mining and misuse of user data, Facebook agreed to pay $100 million for misleading investors about the use of consumer data. The Securities and Exchange Commission claimed Facebook discovered the misuse in 2015 but did not correct disclosures for more than two years.

Research Elaborations:

The following three case studies illustrate key applications of data mining in higher education.

Case study one: Creating meaningful learning outcome typologies

Challenge

"What do institutions know about their students?" If the answer is a recital of enrollment percentages or other basic counts, institutions do not know their students as well as they could. This case study demonstrates how suburban community colleges can establish learning outcome typologies for students using unsupervised data mining.

A typical suburban community college with an enrollment of 15, 000 traditionally identifies its students as "transfer oriented," "vocational education directed," or "basic skill upgraders." These identifications, however, are based on students' initial declarations of educational goals at enrollment. While these are inclusive classifications, they don't help to illustrate the differences between each student type.

Solution

To establish appropriate typologies for the 15, 000 students, researchers used both TwoStep and K-means, two powerful clustering algorithms. They first applied the algorithms to the general groupings identified above, with mixed results. The boundaries among clusters were unclear and dispersed, and even after repeated testing on holdout datasets, as well as the removal of suspected outliers (cases that do not appear to belong to any group), the results did not improve significantly. It's possible that the students' initial declaration of goals did not dictate their academic behavior.

The researchers then used a replacement method that looked at educational outcomes in combination with lengths of study. Defining educational outcomes is easier said than done. Enough time must pass to conclude that a student has reached a certain milestone. Dropping out is also an outcome by itself. Further work was conducted to determine length of study, which required decisions on how to deal with "stopouts," students who left school and later returned.

All of these situations test the data miner's domain knowledge. There are no absolutely right or wrong typologies. In essence, a typology is a good one if it serves a particular research objective.

After either removing the outliers or adding them to a particular cluster, the TwoStep algorithm produced the following clusters: “Transfers,” “vocational students,” “basic skills students,” “students with mixed outcomes,” and “dropouts.” K-means validated these clusters. Introducing the length-of-study element gave new dimensions to each cluster. Some transfer students completed their studies quickly; some vocational students took longer; and other students appeared to simply take one or two courses at a time.

Results

Data mining, combined with student demographics and other information, enabled the college to improve its understanding of its student types. Certain older students, for example, tended to take their time, while younger students with more privileged socioeconomic backgrounds often took high credit courses and graduated quickly. One of the most interesting steps in classification is naming the typologies. The college used the term “transfer speeders,” for example, to describe students who quickly accumulated units, while those who took classes for a considerable length of time were “college historians.”

Other student clusters were “fence sitters,” “skill upgraders,” etc.

Typologies are important because they go beyond conventional student profiling to identify homogenous groups of students, thus increasing the accuracy of predictive modeling algorithms. Even if a data mining project ends with the discovery of appropriate typologies, the newly discovered patterns and relationships help educators and administrators better meet the needs of varied student groups.

Case study two: Academic planning and interventions-transfer prediction

Challenge

This case study showcases a solution to a vexing higher education problem: How to accurately predict academic outcomes in order to facilitate timely academic intervention. When institutions use data mining to predict which students are most at risk, institutions can prevent a student from failing before the student is even aware that he or she is at risk.

More than half of community college students identify transferring to four-year universities as their goal. Due to academic difficulties, however, many either take a long time to transfer or never transfer at all. While it has traditionally been difficult to discover which students transfer, the National Student Clearing House now allows community colleges and universities to match their data. This means that data miners and decision makers can link the academic behavior of community college students to their transfer outcomes.

Solution

Building an effective data mining model with this data involves a combination of typologies and domain knowledge. Transfer education domain knowledge emphasizes that the most effective means of increasing student transfers is to identify transfer-directed students as early as possible. Grooming those who are most likely to transfer is far more meaningful than counting the number of students who have accumulated enough units to transfer.

Using the transfer outcome data, analysts built a dataset containing students who fell under the general transfer clusters of “speeders” and “laggards.” The dataset was split into a test dataset and a validation dataset, using a proprietary randomization method. The outcome variable was transfer. Other variables, such as demographics, courses taken, units accumulated, and financial aid, were predictors to be analyzed without stepwise testing for significance. Data mining is very tolerant of variable interactions and non-linear relationships in data. Supervised data mining was the obvious and appropriate method; therefore, the analysts ran neural network and rule induction algorithms simultaneously in order to contrast and compare the prediction accuracy.

Results

Data mining enabled the college to accurately identify good transfer candidates. After extensive machine learning, the neural network algorithm, Neural Net, had a prediction accuracy of 72 percent, and the rule induction algorithms, C5.0 and C & RT, had a prediction accuracy of 80 percent. The models then ran against the test dataset and produced similar results, indicating their grasp of the patterns within the data.
Case study three: Predicting alumni pledges

Challenge

For a typical urban university of 25, 000, the alumni population can be as much as ten times its enrollment. Most universities send mailings to alumni on a regular basis, even when alumni fail to respond. These mailings typically cost more than

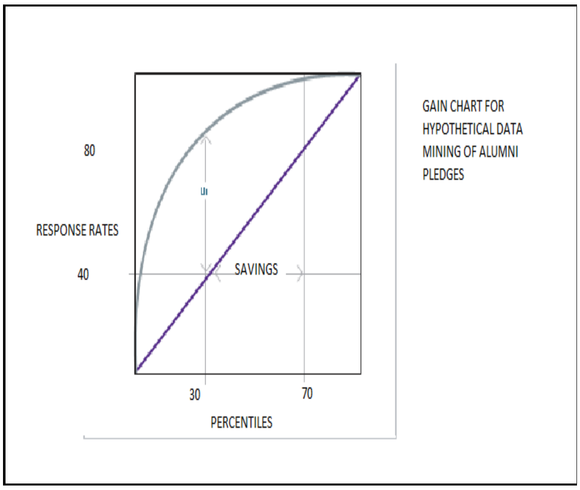
$100, 000 a year. This case study shows how data mining helps universities focus on the alumni most likely to make pledges.

Solution

It’s often difficult to determine whether mailings directly affect the volume and value of alumni pledges. Given the same type of mailing, one alumnus may contribute regularly while another may not. Adding to the confusion is the presence of outliers, such as alumni who unexpectedly contribute large sums. How do institutions identify and cultivate relationships with “outlier” alumni?

In Figure 2, on the next page, the chart shows the benefit of using data mining to determine alumni mail recipients versus simply mailing to all alumni. The curved line is the optimal return rate (alumni contributions) as predicted by data mining. The straight, 45-degree line is the predicted result if the entire population received the mailing. In this case, the chart indicates that when the mailing reached the 30th percentile of the population predicted by data mining to be responsive, 80 percent would respond with a pledge. If the entire population received the mailing, only 40 percent would respond. If every percentage point = $2, 500, savings = (70% * $2, 500) - (30% * $2, 500) = $175, 000-$75, 000 = $100, 000). Without data mining, therefore, it would cost $100, 000 more to reach all 80 percent.

Figure 2, below, shows the benefits that data mining may bring to alumni pledge campaigns.



Results

Using data mining, the college discovered a way to make its mailing more effective and increase alumni pledges, while reducing mailing costs. This is best described using a concept called "lift." If 20 percent of alumni respond to a pledge request, the college should concentrate on those 20 percent. If data mining can quickly identify potential donors by a ratio of two to four (correctly predicting two out of four who will donate), then the university can achieve results by mailing only to the indicated 40 percent of the alumni population, thus saving considerable time and money.

Conclusion

Data mining is a powerful analytical tool that enables educational institutions to better allocate resources and staff, proactively manage student outcomes, and improve the effectiveness of alumni development. With the ability to uncover hidden patterns in large databases, community colleges and universities can build models that predict-with a high degree of accuracy-the behavior of population clusters. By acting on these predictive models, educational institutions can effectively address issues ranging from transfers and retention, to marketing and alumni relations.

Table 2.1: Classification of Data Mining Tasks and Tools

Tasks	Supervised	Unsupervised
Classification	Memory based reasoning, genetic algorithm, C&RT, link analysis, C5.0, ANN	Kohonen nets
Estimation	ANN, C&RT	
Segmentation	Market basket, analysis, memory based reasoning, link analysis, rule induction	Cluster detection, K-means, generalized rule induction, APRIORI
Description	Rule induction, market basket analysis	Spatial visualization

Name of Tool and Developer	Source (Commercial)	Function/Features	Techniques/Tools	Environments
MSSQL Server 2005 (Microsoft) An	Commercial	Provides DM functions both in relational db system and Data Warehouse (DWH) system environment.	Integrates the algorithms developed by third party vendors and application users.	Windows, Linux
Enterprise Miner (SAS Institute)	Commercial	Provides variety of Statistical analysis tools	Association Mining, Classification, Regression, Time series analysis, Statistical analysis, Clustering	Windows, Solaris, Linux Insightful Miner
Oracle Data Mining (Oracle Corporation)	Commercial	Provides an embedded DWH Infrastructure for multi dimensional data Analysis	Association Mining, Classification, Prediction, Regression, Clustering, Sequence Similarity search and analysis	Windows, Mac, Linux
DBMiner (DBMiner technology Inc)	Commercial	Provides multiple mining algorithms, Data-cube-based on-line analytical mining,	Discovery-driven OLAP analysis, association, classification,	Windows, Linux

		frequent-pattern mining functions and integrated visual classification methods	and clustering	
SPSS Clementine	Commercial	Provides an integrated Data mining	Association Mining, Clustering,	Windows, Solaris, Linux
(IBM)		Development environment for end users and developers.	Classification, Prediction and Visualization tools	

Tool	Admission	Training	Research	Social Environment and Support	Administration	Quality Assurance	Management
AWE Benchmarking		XX	X				X
Suite Blackboard		X	X		X	X	X
Analytics Canvas C4S Data Hero Degree	X	XXX	XX X			X	
Compass Echo360 Eduten Playground		XX	X				X
Faculty			X				
Insight GISMO GLASS JISC	XXXX			XX		XX	X
LATch LATqe	X			X			
Learning Catalytics	XX			XX			
Mobile APP Mobile	X			X			
Teacher App Moodle Learning Analytics API MyUni Next-Lab	XXXXXX			XXXXXX	X	X XX	X
OLAP Open LIPASS Pyramid Analytics SEAtS SIGNALS SNAP Teacher Ease Uni Vu Zoola 5Lab	XXXXX X			XXXXX X		XX	XX

Each of the 35 data analysis tools was examined in terms of supported processes typically taking place in higher education institutions. The results from the examination clearly show that none of the considered tools offers support for processes in all seven areas.

Result Mining: Analysis of Data Mining Techniques in education

2. Summarization of Results

Table Head	Application of Various Techniques and Algorithms		
	Technique	**Algorithm**	**Application**
1	Classification	Decision Tree	Predicting Failure Predicting Dropout Alumini Fund Raising Classify the students as Bad Average Very Good and Excellent Prediction of Placements Student Retention
		Bayesian	Classify the students as Bad Average Very Good and Excellent Academics & Recruitment
2	Association	Aproiri	Finding Adept Teacher Dealing with Students

			Failure pattern extraction
3	Clustering	K-Means	Persistent and Non-Persistent student Comprehension Grouping Similar Students
4	Outlier Detection		Rare Events Analysis and understanding Understanding irregular events

EDM Applications	Techniques	The highest Prediction Accuracy
Predicting student performance	Na¨ıve Bayes	93.6%
	Bayesian Network	98.08%
	Decision Tree	94.23%
	Rule Based	71.3%
	Neural Network	90.2%
	K-Nearest Neighbor (KNN)	90.91%
	Multilayer Perception	75%
	REPTree	76.7334%
	OneR	76.7334%
	Iterative Dichotomiser 3 (ID3)	88%
	Random Forest	99%
	PART	74.33%
	Logistic Regression	89.15%
Detecting undesirable student bahaviors	Social Network Analysis (SNA)	92.89%
	Na¨ıve Bayesian	87%
	JRip	77.30%

Grouping students	Clustering	78.8%
	Neural Network	76%
Student modeling	Social Network Analysis (SNA)	81%
	Decision Trees	80%

TABLE I: Educational data mining methedology categories.

Category	objectives	Key applications
prediction	Develop a model to predict some variables base on other variables. The predictor variables can be constant or extract from the data set.	Identify at-risk students. Understand student educational outcomes
Clustering	Group specific amount of data to different clusters based on the characteristics of the data. The number of clusters can be different based on the model and the objectives of the clustering process.	Find similarities and differences between students or schools. Categorized new student behavior
Relationship Mining	Extract the relationship between two or more variables in the data set.	Find the relationship between parent education level and students drooping out from school. Discovery of curricular associations in course sequences; Discovering which pedagogical strategies lead to more effective/robust learning
Discovery with Models	It aims to develop a model of a phenomenon using clustering, prediction, or knowledge engineering, as a component in more comprehensive model of prediction or relationship mining.	Discovery of relationships between student behaviours, and student characteristics or contextual variables; Analysis of research question across wide variety of contexts
Distillation of Data for Human Judgement	The main aim of this model to find a new way to enable researchers to identify or classify features in the data easily.	Human identification of patterns in student learning, behaviour, or collaboration; Labelling data for use in later development of prediction model

Applications in Learning

Big Data techniques can be used in a variety of ways in learning analytics as listed below:

- Performance Prediction
  - Student's performance can be predicted by analyzing student's interaction in a learning environment with other students and teachers
- Attrition Risk Detection
  - By analyzing the student's behavior, risk of students dropping out from courses can be detected and measures can be implemented in the beginning of the course to retain students.
- Data Visualization
  - Reports on educational data become more and more complex as educational data grow in size. Data can be visualized using data visualization techniques to easily identify the trends and relations in the data just by looking on the visual reports.
- Intelligent feedback

Learning systems can provide intelligent and immediate feedback to students in response to their inputs which will improve student interaction and performance.

- Course Recommendation

  - New courses can be recommended to students based on the interests of the students identified by analyzing their activities. That will ensure that students are not misguided in choosing fields in which they may not have interest.

Student skill estimation

  - Estimation of the skills acquired by the student

- Behavior Detection

  - Detection of student behaviors in community based activities or games which help in developing a student model

- Grouping & collaboration of students
- Social network analysis
- Developing concept maps
- Constructing courseware
- Planning and scheduling

How Data Analytics Can Improve the Student Journey

Calculate the Student Success

Predict a success of a student learning about any course or overall information about your institution.

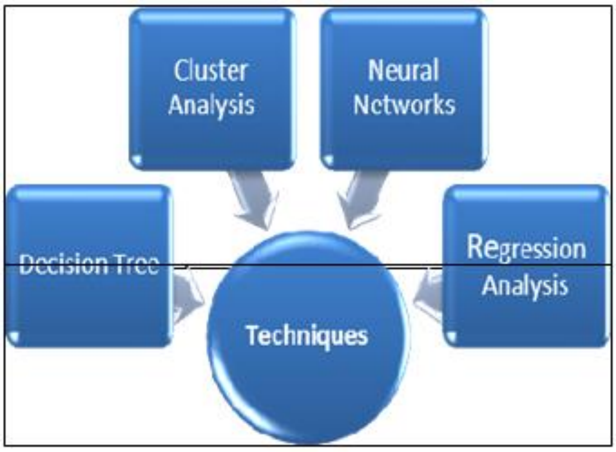
Improve Learning Outcomes with Customized Modules Data Analytics can Enhance the learning and development of the students with customized Curricula.

Helps Teachers Understand their Students

With the help of data gathered from the behaviour analysis and other assessments, teachers can better understand a student's interest and can work on improving the learning experience for them.

With the appropriate data gathered, gain insights to analyse which learning format is most efficient and productive students. Additionally, this information can provide awareness on

How to improve students' overall academic experience



References

[1] A. V. Manjarres, L. G. M. Sandoval, and M. S. Sua´rez, "Data mining techniques applied in educational environments: Literature review," Digital Education Review, pp.235-266, 2018.

[2] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol.40, no.6, pp.601-618, 2010.

[3] B. Bakhshinategh, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," Education and Information Technologies, vol.23, no.1, pp.537-553, 2018.

[4] I. F. Siddiqui, Q. A. Arain, et al., "Analyzing students' academic performance through educational data mining.," 3C Tecnologia, 2019.

[5] A. Algarni, "Data mining in education," International Journal of Advanced Computer Science and Applications, vol.7, no.6, pp.456- 461, 2016.

[6] K. Karthikeyan and K. PALANIAPPAN, "On improving student performance prediction in education systems using enhanced data mining techniques," International Journal of Advanced Research in Computer Science and Software Engineering, vol.7, no.5, 2017.

[7] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," Applied Mathematical Sciences, vol.9, no.129, pp.6415-6426, 2015.

[8] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," Procedia Computer Science, vol.57, pp.500-508, 2015.

[9] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and predicting students' academic performance using data mining techniques," Inter-national Journal of Modern Education and Computer Science, vol.8, no.11, p.36, 2016.

[10] H. Almarabeh, "Analysis of students' performance by using different data mining classifiers," International Journal of Modern Education and Computer Science, vol.9, no.8, p.9, 2017.

[11] H. Mousa and A. Y. Maghari, "School students' performance prediction using data mining

classification," School Students' Performance Predication Using Data Mining Classification, vol.6, no.8, 2017.

[12] B. Kapur, N. Ahluwalia, and R. Sathyaraj, "Comparative study on marks prediction using data mining and classification algorithms," International Journal of Advanced Research in Computer Science, vol.8, no.3, 2017.

[13] A. U. Khasanah et al., "A comparative study to predict student's performance using educational data mining techniques," in IOP Conference Series: Materials Science and Engineering, vol.215, p.012036, IOP Publishing, 2017.

[14] M. MAKHTAR, H. NAWANG, and S. N. WAN SHAMSUDDIN, "Analysis on students performance using na¨ıvebayes classifier," Jour-nal of Theoretical & Applied Information Technology, vol.95, no.16, 2017.

[15] I. A. A. Amra and A. Y. Maghari, "Students performance prediction using knn and na¨ıvebayesian," in 2017 8th International Conference on Information Technology (ICIT), pp.909-913, IEEE, 2017.

[16] Y. Salal, S. Abdullaev, and M. Kumar, "Educational data mining: Student performance prediction in academic," IJ of Engineering and Advanced Tech, vol.8, no.4C, pp.54-59, 2019.

[17] K. Agarwal, E. Maheshwari, C. Roy, M. Pandey, and S. S. Rautray, "Analyzing student performance in engineering placement using data mining," in Proceedings of International Conference on Computational Intelligence and Data Engineering, pp.171-181, Springer, 2019.

[18] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," Heliyon, vol.5, no.2, p. e01250, 2019.

[19] S. Ray and M. Saeed, "Applications of educational data mining and learning analytics tools in handling big data in higher education," in Applications of Big Data Analytics, pp.135-160, Springer, 2018.

[20] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using weka," Indonesian Journal of Electrical Engineering and Computer Science, vol.9, no.2, pp.447-459, 2018