IV. MACHINE LEARNING & DATA ANALYTICS

One can see that the field of e-learning can significantly contribute to the notion of big data as increasing number of students access educational material online and generate more data flows. It has been reported that around 58 million students have registered for online courses worldwide with nearly 7000 courses available [3]. With typical courses generating thousands of event log records per student per course, the amount of data generated is growing exponentially [3]. Thus, machine learning and data analytics also become crucial in order to make use of the growing amount of collected data generated in the field of e-learning.

Machine learning and data analytics have been proposed as possible solutions to process the increased amounts of data collected. These algorithms are the tools that can make use of the data by ``learning'' the behavior and finding interesting patterns within it.

In what follows, the different machine learning and data analytics types and algorithms are discussed briefly.

A. MACHINE LEARNING

As shown in [52], data has become more abundant and easy to obtain. However, extracting knowledge from this collected data is often expensive. With the help of computers, which can perform calculations at tremendous speeds, more complex data analysis has become available. Moreover, having computers that can ``learn'' without being told what to do is essential as this would give it a greater capacity to adapt based on new inputs. To this end, the field of machine learning has been developed. Machine learning, as per Arthur Samuel, is defined to be the ``field of study that gives computers the ability to learn without being explicitly programmed'' [5]. The learning is performed using some data or observations such as examples, direct experience, or instruction [5]. This is crucial in many cases such as when the solution changes over time or when it needs to be adapted to specific cases. Thus, ``learning'' based on experience is imperative in many future applications.

Machine learning algorithms have garnered significant attention in recent years and have been used in several applications such as pricing prediction, optical character recognition, image recognition, spam filtering, fraud detection, healthcare, transportation, and many others [5], [53] [56]. Fig. 6 shows some of the different machine learning types and algorithms that will be discussed in this work.

1) TYPES

Machine learning algorithms are divided into four main categories: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning includes having a dataset with the correct output that is used to ``train'' the system. On the other hand, unsupervised learning includes trying to find relations among the points in the dataset without having the correct results during training [57]. This means that the algorithm tries to ``cluster'' points that it believes to be highly correlated under one label based on their statistical properties only. Semi-supervised learning combines the previous two types by training the system using a dataset containing labeled and unlabeled data points. The goal is to improve the performance of the model by making use of both types of data points [58]. Last but not least, reinforcement learning in contrast uses trial-and-error to discover the set of actions that maximize some cumulative reward metric [59].

a: SUPERVISED LEARNING

Supervised learning is a branch of machine learning algorithms in which a function is inferred based on labeled training data [60]. The training data is formed of a group of training examples, each of which is a pair (x; y) where x is an input vector and y is the output value. The algorithm produces a function that can be used for mapping future unknown inputs.

Supervised learning algorithms often fall into one of two main categories: regression algorithms (output is continuous) or classification algorithms (output is discrete) [61]. Within each category, several algorithms exist which will be presented below.

1. **Regression**: Regression algorithms try to find the best fit function for the training data available. Two main algorithms are discussed below: linear regression and polynomial regression.

**Linear Regression**: One of the most common regression algorithms used in machine learning is the linear regression algorithm. This algorithm tries to find the best fit line/hyperplane for the available training data. The goal of the algorithm is to find the value of the optimal coefficient vector  such that the predictive function has a linear form. This is done by using the minimum mean squared error function.

Polynomial Regression: Another common regression algorithm used is the polynomial regression algorithm. This algorithm tries to find the best fit polynomial for the available training data. Similar to its linear counterpart, the goal of the algorithm is to find the value of the coefficient vector  such that the predictive function is a polynomial of order k.

1. **Classification**: In contrast to regression algorithms that try to find the best fit function for the training data, classification algorithms try to find the best fit class for the data by putting each input in its correct class. In such cases, the output of the predictive function is discrete with the possible values being one of the different classes available as part of the training data. Four important classification algorithms are discussed below, namely logistic regression, artificial neural networks, support vector machines, and decision trees.

**Logistic Regression**: Logistic regression is an extremely popular classification algorithm used in the literature [62] [65]. Despite its name, this algorithm is used for classification (i.e. its output is discrete) rather than being a regression algorithm. It is typically used a binary classifier where the output belongs to one of two categories only. The predictive function , also known as the hypothesis function, calculates the probability of the output being equal to 1 given a specific input. In other words,  . If this probability is greater than 0.5, the output is defined to be 1. Otherwise, the output is defined to be 0.

To determine the coefficient vector , a cost function needs to be used. However, the squared error function will not work because the sigmoid function used to determine the hypothesis function will make the output wavy and thus have many local optima. Instead, a different cost function based on the log function is used that ensures the output is convex [66]. This algorithm can also be extended for multi-class classification. In that case, instead of having just one hypothesis function , we would have multiple functions with each calculating the probability of the output being class *i* given the input available.

**Support Vector Machines**: Support vector machines (SVM) is another supervised classification algorithm. It tries to find the optimal hyperplane that separates the labeled data with the maximum margin from the closest point. It is a more powerful and restrictive classifier than the logistic regression algorithm. This algorithm replaces the sigmoid function used in logistic regression with a new function called the hinge loss function [67].

Note that the hypothesis function given by the SVM algorithm is not interpreted as the probability of the output being 1 or 0, but rather is a discrimination function that outputs either 1 or 0.

**Artificial Neural Networks**: Artificial neural networks (ANN) is a popular supervised classification algorithm. It is often used whenever we have abundant labeled training data with many features and a non-linear hypothesis function is desired [68]. ANN tries to mimic the way our brain works as it has been proven that the brain uses one ``learning algorithm'' for all its different functions [69].

Similar to neurons that act as computational units that take electrical inputs (through dendrites) and channel them towards an output (axon), the ANN algorithm adopts a model in which the features act as dendrites (nerve cell) and outputs the value of the hypothesis function. Often, one ``hidden'' layer is used that acts as an intermediate layer. This layer helps extract more information from the set of features available as part of the training data and is called the activation layer. The sigmoid function used in logistic regression is used here at each layer of the network. Since the sigmoid function is used in ANN, then the cost function used to determine the values of the coefficient vector at layer *f* is the same one used for logistic regression.

**Decision Trees**: Decision trees are another popular choice of supervised learning classification algorithms. These algorithms are often referred to as statistical classifiers since they use statistical metrics to determine the branching of the nodes [70]. To classify an instance, decision trees sort the instance down the tree from the root node to a specific leaf node. Each node within the tree represents a test of a particular feature of the instance while each branch represents a possible value of the tested feature.

There are several decision tree-based algorithms such as ID3, ASSISTANT, and random forests [70]. One of the most well-known decision trees algorithms among them is the C4.5 algorithm. The algorithm was first proposed by Quinlan [71] and is based on the notion of information entropy. Effectively, the C4.5 chooses the feature that best divides its set of samples into smaller subsets rich in one class or the other. To determine the division criterion, the normalized information gain metric (difference in entropy) is used. The feature chosen is the one with the highest information gain.

1. **Deep Learning**:

One special class of supervised machine learning algorithms is deep learning. In essence, deep learning can be thought of as a large scale neural network [72]. However, due to the fact that deep learning is also able to perform automatic unsupervised feature extraction, also commonly referred to as feature learning [73], it cannot be classified as a traditional neural network. Hence, deep learning is then considered a special case of supervised machine learning. In general, deep learning tries to model abstractions found in data using a graph with multiple processing layers [74] [76]. These processing layers contain units that apply linear and non-linear transformations on the data to extract as much useful information as possible.

Deep learning algorithms are very similar to artificial neural networks. In fact, ANN can be classified to be one of the deep neural networks learning algorithms. However, deep learning algorithms are more broad as they can be applied to both labeled and unlabeled data. Moreover, they can be applied to a much larger scale of neural networks. Andrew Ng, co-founder of Coursera and the Chief Scientist at Baidu Research, said that deep learning is just applying ANN on a large scale that can be trained with more data and have better performance because of that [74].

There are many different deep learning algorithms other than ANN. In what follows, two popular algorithms are briefly described and discussed.

**Convolutional Neural Networks**: Convolutional Neural Networks (CNN) is a version of artificial neural networks that was inspired by the connectivity patterns found in the visual cortex of animals. These connectivity patterns have been proven to be mathematically described by a convolution operation [77]. This operator replaces the sigmoid function that is typically used in artificial neural networks.

**Recursive Neural Networks**: Recursive Neural Networks (RNN) is another version of deep neural networks. RNNs are created by using the same set of weights in a recursive manner over a tree like structure [78], [79]. The tree is traversed in topological order. This algorithm is often used to process sequential data. Similar to the idea that CNNs are used to process data grouped in the form of a grid, RNNs are specialized in processing a sequence of values.

b: UNSUPERVISED LEARNING

In contrast to supervised learning, unsupervised learning is the branch of machine learning in which a function/pattern is inferred based on unlabeled training data [80]. The training data consists of only inputs  and no known outputs. Therefore, unsupervised learning algorithms aim to make sense of the training data by finding relations and patterns within it.

Unsupervised learning algorithms can be divided into 3 main categories: clustering, dimensionality reduction, and anomaly detection [80]. Several algorithms exist within each category. In what follows, some of the most popular algorithms in these categories are presented.

1. **Clustering**: One of the easiest way to make sense of a set of data points is to group/cluster them. This makes the data more understandable as it gives more structure to it by forming a finite set of groups rather than having a multitude of random data points. This is especially important in applications such as market segmentation and social network analysis. Since we do not know whether this grouping is correct or not, the term ``clustering'' is used rather than ``classification'' because the data points are not labeled to be belonging to specific classes. In what follows, a well-known clustering algorithm is discussed.

**K-Means Algorithm**: K-means is one of the most popular unsupervised clustering algorithms for automatic data grouping into coherent clusters. This algorithm tries to group the data into K clusters by finding the cluster centroid (also known as cluster mean) and group with it the data points closest to it.

To be able to properly cluster the data points into the K clusters, a cost function needs to be minimized. This cost function is dependent on 3 variables: c(j) which is the index of the cluster to which example x(j) is currently assigned to,  centroid of cluster *k*, and c(j) the centroid of the cluster to which example x(j) is currently assigned to. The algorithm aims to find the indices and centroids that will minimize the average distance of every example to its corresponding cluster centroid.

1. **Dimensionality Reduction**: Dimensionality reduction is another essential topic in the field of machine learning. The motivation behind dimensionality reduction can be summarized as follows:
2. Remove redundant data
3. Reduce the storage and computational needs
4. Simplify the visualization of data by only considering a few features.

To this end, one well-known algorithms is discussed which is the principal component analysis (PCA) algorithm.

**Principal Component Analysis**: Principal component analysis is one of the most popular dimensionality reduction algorithms in unsupervised learning. Its aim is to find the subset of features that best represents the data. For example, assuming two features x1 and x2 are given, PCA tries to find a single line that can describe both these features effectively at the same time. This is different from linear regression since the goal of PCA is to reduce the average projection error (orthogonal distance from the feature to the projection line) while linear regression tries to reduce the average error (vertical distance) to the line. Performing PCA analysis involves four steps: data preprocessing, covariance matrix computation, eigen value decomposition, and choosing first k eigen vectors.

* 1. **Anomaly Detection**:

Another important unsupervised learning algorithm is the anomaly detection algorithm. From its name, this algorithm tries to determine whether the given new example x(new) is anomalous or not. To do so, a probability function/model p(x) is calculated which gives the probability of an example not being anomalous. A threshold value, denoted by β is used as the dividing value between identifying the example as normal or anomalous.

To be able to calculate the probability function p(x), the set of features are assumed to be independent and hence the probability function p(x) becomes the product of the probabilities of the features p(). Another assumption that is made is that the features are normally distributed, i.e. p() follows the Gaussian distribution. Hence, determining whether x(new) is anomalous or not is done by comparing p(x(new)) with β: if p(x(new)) < β then the example is anomalous, otherwise it is normal.

1. SEMI-SUPERVISED LEARNING

Semi-supervised learning is a branch of machine learning techniques in which a function/patter is inferred based on partially labeled training data [58], [81]. It combines elements from supervised and unsupervised learning. The aim of semi-supervised learning is to try to make use of both the labeled and unlabeled data to get better learning models [58], [81]. Typically, these algorithms are implemented whenever the dataset has a small amount of labeled data points and an abundance of unlabeled data points.

To be able to make use of the unlabeled training data examples, some assumptions need to be made to have some structure to the data distribution. All semi-supervised learning algorithms make at least one of the following assumptions [82]:

**Smoothness Assumption**: It is assumed that points that are close to each other are more probable to share the same label.

**Cluster Assumption:** It is assumed that the data tend toform discrete clusters with points within the same cluster being more probable to share the same label.

**Manifold Assumption:** It is assumed that the data liesapproximately on a manifold of lower dimension than the input space. This is a practical assumption in cases where high-dimensional data is generated such as in voice recognition. In that case, it is know that the voice is controlled by a small number of vocal folds and thus only the space surrounding these folds is considered rather than the space of all possible acoustic waves.

Semi-supervised learning differs from transductive learning. That is because the aim of transductive learning is to predict the correct label of the set *L* using the labeled training set *M*. On the other hand, semi-supervised learning, which is inductive, aims to use both the labeled and unlabeled training sets to obtain a function that can predict the label for any ``future'' example [58], [81]. There are several algorithms that fall under the umbrella of semi-supervised learning.

In what follows, some of these algorithms are discussed.

**Self-training:** Self-training is the simplest semi-supervised learning algorithm. This algorithm first obtains a prediction function *z* using the labeled data points. It then applies this function to a subset of the unlabeled dataset and adds the outcome (*xM*+l ; *z*(*xM*+*l* )) to the labeled data. It then repeats the process to obtain a better prediction function *z* until all the unlabeled points are given a label and are used to learn a final model *z* *final* .

**Low-density Separation Models:** Another group ofsemi-supervised learning algorithms is the low-density separation models. These algorithms try to place the dividing boundaries in regions with a few data points (labeled or unlabeled). Among this class of algorithms is the transductive support vector machine algorithm (TSVM), also known as semi-supervised support vector machine algorithm (S3VM). Similar to the supervised learning SVM, TSVM aims to find the decision boundary that maximizes the margin over all the data. To do that, the unlabeled data needs to be labeled. Therefore, the TSVM algorithm enumerates all possible labeling of the unlabeled data points (2*L* possible labels) and uses SVM to find a decision boundary. After finding all the possible decision boundaries, the one with the largest margin is chosen.

1. *REINFORCEMENT LEARNING*

**Reinforcement learning** (RL) is a branch of machine learning in which the action is taken in such a manner so as to maximize a cumulative reward metric [59]. This is often done using a trial-and-error sort of way in an attempt to discover the actions with the highest rewards. The decision taken often not only affects the immediate reward, but also subsequent ones. These two features, which are trial-and-error and delayed reward, are the two most distinguishing characteristics of reinforcement learning [59].

RL problems are often modelled as a stochastic finite state machine problem with inputs being the actions taken by the agent and the output being the observations and rewards obtained by the agent [83]. The environment consists of:

* A set of states *S* describing the environment.
* A set of possible actions *A*.
* Transitioning rules between states.
* Scalar reward rules of a transition and the observation rules of an action.

RL agents take actions in discrete time steps. Hence, at time instance *t*, the agent makes observation *Ot* which includes a reward *rt*. The agent then chooses an action *at* that makes the environment move from state *St* to state *St*+1. This transition is then associated with a reward *rt+*1. The goal of the agent is to maximize the collected cumulative reward [84]. The RL model depicting the interaction between the agent and the environment is shown in Fig. 7.

There are a variety of algorithms that fall under the RL category. These algorithms have proved to be popular in a variety of applications including control theory, telecommunications, gaming (backgammon and checkers), and scheduling [59]. In what follows, some of the most popular RL algorithms are discussed [85], [86].

**Dynamic Programming:** Dynamic programming is oneof the most well known RL algorithms. This algorithm is considered a value function-based approach. Value function-based approaches try to determine a policy π, a mapping that assigns a probability distribution of the actions to all possible histories, that maximizes the reward by maintaining a set of estimates of the expected rewards for different policies. Using the recursive relationships that all value functions follow, the optimal policy can be determined using the transition probabilities between states and the expected rewards gained [86].

**Monte Carlo Methods:** Another group of algorithmsthat are used for RL is Monte Carlo Methods. The benefit of these algorithms is that they do not assume complete knowledge of the environment, but rather learn about it by experience (sample sequence of stations, actions, and rewards) [85].

This group of algorithms depend on running several iterations to evaluate the policy for the different values of the state-action pairs (*Si*; *ai*). The different iterations start at random from different states. The computed values are then averaged out to get a good estimate of the action value function of the considered policy.

Table 1 summarizes the different machine learning algorithms, the data types they can work on, and the problem types they can be applied for.

**B. DATA ANALYTICS**

Data analytics has become an increasingly important tool for users to extract knowledge and inferences from the abundant amount of collected data [52]. This analysis has become easier with the emergence of sophisticated computers and software that are able to perform exhaustive computations in lightning quick speeds. In contrast to machine learning which tries to ``teach'' computers to perform certain tasks without explicitly programming them, data analytics try to analyze and make inferences based on the raw collected data in order to take better decisions [6].

To this end, different techniques and algorithms have been developed. These algorithms have been applied in various fields such as business, marketing, health care, transportation, and education [6], [87] [89]. This has helped individuals make more informed decisions about the data available at their hands by making use of the results of such techniques.

Fig. 8 shows the different data analytics types and algorithms that will be discussed.

1) TYPES

Data analytics algorithms and techniques can be divided into three main categories: exploratory, confirmatory, and qualitative. Each category represents a different goal and approach when it comes to dealing with the data.

1. **Exploratory Data Analytics (EDA):** Exploratory DataAnalytics (EDA) is a group of techniques that tries to determine/discover new features within the data [90]. These techniques often give insights into the available data and uncover the underlying structure found within the data. The difference between EDA techniques and machine learning techniques is that EDA techniques analyze the data and try to infer the adequate model. On the other hand, machine learning techniques first try to find the model and then analyze its parameters [90]. EDA techniques can be divided into two main groups: graphical and quantitative. Some of these techniques are discussed below [90]:

Graphical:

* **Histograms**: Histograms give a graphical summary of the distribution of a univariate dataset. This can be done to one feature of the collected data.
* **Probability Plot**: Probability plots are another graphical technique that helps analysts assess whether the data follows a specific distribution or not. The data is plotted against a theoretical distribution. If the plot has a linear form, then this implies that theoretical distribution fits the data well. If the plot does not have a linear form, then the considered distribution is not representative of the data.

Quantitative:

* **Confidence intervals for the mean**: These intervals provide a range of where the true mean may lie given the dataset available. Since the data contains only a sample of the actual possible instances of a specific feature, then the calculated mean of the data is not the true mean of the feature. Hence, confidence intervals are used to indicate the lower and upper limits of the true mean.
* **Variance**: Variance is a measure of how the data varies around the mean. It gives an idea of how wide or narrow the distribution is. This can help the analyst determine how sensitive the dataset is to a variation in a specific variable.

1. **Confirmatory Data Analytics (CDA):** ConfirmatoryData Analytics (CDA) is a group of techniques that try to confirm whether a hypothesis is true or false. Such techniques assume that there is a specific hypothesis and perform tests that either confirm the validity or invalidity of the hypothesis. Two well known tests are discussed below [90].

* **Analysis of Variance (ANOVA)**: ANOVA techniques are often used to compare data collected from two or more processes. The idea is to test whether the processes have equal variances. Bartlett's test performs ANOVA by testing for what is called homogeneity of variance. The hypothesis is assumed to be that all the variances are equal. Performing this analysis answers the question of whether the variances are truly equal or not with some specific confidence level. This type of tests is often used when the dataset seems to be normally distributed and is one of a set of statistical tests called ``parametric tests'' which also include t-tests [91], [92].
* **Chi-Square Test for variance**: This test is used to see if the variance of a dataset is equal to a specific value. This is often very important in manufacturing processes because manufacturers desire that their production process's variability is greater or smaller than some threshold.

These tests output a value that allows the analyst to determine whether the original hypothesis is valid or not.

Based on this, he/she can take action and change some aspects of the considered process. This type of tests is often used when the dataset does not seem to be normally distributed, but rather follows other distributions. It is one of a set of statistical tests called ``non-parametric tests'' that include tests like Wilcoxon rank sum test and Spearman's rank correlation coefficient [91], [92].

1. **Qualitative Data Analytics (QDA):** Another type ofdata analytics is Qualitative Data Analytics (QDA). This type encompasses the set of processes and techniques that try to explain/interpret qualitative (non-numeric) data such as videos, interview transcripts, images, and documents [93]. These techniques are very prevalent in social sciences in which a lot of the data collected is non-numeric in nature. They are also prevalent in text analysis by searching for recurring words or topics.

There are some key terms that often pop up in QDA techniques. Words such as ``Theme'' that provides a general idea of the topics apparent in the data and ``Characteristic'' which represents a single item in a text often appear when performing QDA. The process of performing QDA involves having to iteratively and progressively notice, collect, and think about things [94]. This process is not linear since the cycle keeps on repeating and you do not go through it step by step, but rather do these steps simultaneously. This process is shown in Fig. 9.

To perform QDA, techniques like sorting and categorizing the collected data into groups based on theme can help simplify the analysis. After sorting the data, further analysis is performed by trying to interpret the significance of the sorting or by trying to explain the findings based on the sorted data.

These techniques have been used in different applications like text mining, marketing research, business, criminology, and sociology [95], [96].

It is worth mentioning that machine learning can be used as one method for data analysis. This is because in essence, machine learning can be thought of as a data analysis method that builds the analytical model automatically [97]. Hence, it can help identify and uncover patterns found within the data.

1. **RESEARCH OPPORTUNITIES USING MACHINE LEARNING & DATA ANALYTICS FOR E-LEARNING CHALLENGES**

As discussed before, one of the sources of data growth is the data being generated through online learning web-sites and learning management systems (LMSs) as part of e-Learning environments. Statistics show that online course websites such as Coursera, edX, and Udacity have more than 78 million students combined with around 10 thousand courses being offered by more than 600 universities [3]. Moreover, it has been estimated that there are almost one thousand event log entries per student every month and around 60,000 course visits every month for online courses [4]. This gives an idea of how big the data streams are expected to be with online courses specifically and the field of e-learning in general being a main contributor to the ``Big Data'' concept. Thus, a need to analyze and extract useful information from the collected data has risen in order to take more informed decisions and have more efficient systems. These systems can become more intelligent and responsive to users as they better cater to their needs. This can be done using ML and DA techniques such as quantitative data analysis to study students' performance and supervised classification to perform task such as sentiment analysis and students' style classification.

Based on the statistics above and the previously provided discussion on the challenges facing the field of e-learning, ML and DA techniques become promising tools to improve the e-learning platforms and processes. In what follows, a set of possible research opportunities to tackle each challenge discussed earlier is presented.

**Transmission/Delivery - Research Opportunities:** Data analytics can play an important role in tackling this challenge. For example, a case study can be conducted that studies the performance of two groups of students. The first group would take the course in a synchronous manner and the second group would take the course in an asynchronous manner. The performance of both groups can be compared based on the collected data of both groups using both quantitative exploratory and qualitative data analytics. This can help instructors determine the optimal mode of delivery for the specific course based on metrics such as students' location distribution, their interaction level on forums, and their grades throughout the course. This is important because previous work have only compared the two possible delivery mechanisms in terms of the resources needed rather than their impact on student learning. Therefore, studying the merits of each delivery mode for students should be studied.

Another possible area where qualitative data analytics can play a role is through text mining for sentiment analysis. This can be done to identify discouraged or disgruntled students based on their posts in the discussion forums. Moreover, supervised machine learning techniques such as SVM and ANN can be used to classify such posts. This will help instructors gain more insights into how the students are feeling about the material and the course. Based on this, instructor can try different ways to keep the students engaged and motivated. It can also help them make informed decisions on what to change and what to keep in the course in terms of material, assignments, topics, and activities.

**Personalization - Research Opportunities:**

As can be seen, several works have already started using machine learning and data analytics algorithms such as student clustering and content classification to tackle the challenge of personalizing the e-learning experience [23] [25], [38], [39]. However, more opportunities are available. For instance, students can be classified based on any of the previously mentioned learning styles/preferences models. This classification can be used in several ways. One way is to recommend new courses to take based on the students' learning style/preference. This is important for online programs and courses where there is an abundance of courses available to be taken. Another possible recommendation is to recommend a course format for the student taking a course. Although this might be a demanding task for instructors, it can be offset by applying this to supplemental material rather than the main material of the course. This would provide the students with extra material that caters to their preference and can help them to better understand the course they are taking.

A different opportunity is to predict students' final grade and classify them into different groups based on their performance in the first part of a course using techniques such as logistic regression and ANN. This can help instructors identify weak students that may need help in the course. This analysis can also pinpoint which concepts or learning outcomes the students seem to be struggling with. This can also help instructors adapt the content a bit or provide extra information about confusing concepts to ensure students better grasp it. Additionally, this can be used for peer student recommendation. The instructor might recommend that a weak student collaborate with a better performing student so that he/she can help him/her with the course.

Another research opportunity is to determine relationships between certain activities and the performance of students in courses. For example, case studies that investigate the statistical significance of the number of logins to an e-learning platform on the students' performance at the end of a course portion should be performed to test the validity of any assumptions or hypotheses made beforehand. This would give instructors insights into what material was accessed more and whether it helped improve students' understanding of the material or not. This analysis can be done using confirmatory data analytic techniques. Accordingly, the instructor might adapt the content throughout the course to improve the students' e-learning experience. Fig. 10 gives an overview of some of the different classifications that can be done and recommendations that can be given within the field of e-learning.