1. **Introducere**- poate conține: definirea domeniului, alte articole de tip studiu bibliografic (related work), întrebările de cercetare pentru care se caută răspuns, scurtă descriere pentru următoarele secțiuni

[1]

Educational institutions, ranging from primary schools to univer sities, were forced to convert to distance teaching during the covid-19 pandemic. Based on (UNESCO (2021)), half of the world ’s students are still affected by the pandemic, and over 100 million additional kids will have difficulties on their learning journey as a result of the health crisis. Online learning became a popular surrogate to overcome this crisis with the major drawback that most educators lost direct and nonverbal communication with the students in the classroom, which was the main channel for the teachers to read their students ’ needs and allow them for early intervention. The main problems with online learning are low students achievements and higher drop-out rates (Park and Choi (2009); Lee and Choi (2011); Rovira, Puertas, and Igual (2017); Xing, Tang, and Pei (2019)). One can say that such problems are also present in traditional learning systems. Still, the studies show that the failing and withdrawn student rates in online learning are much more significant than in traditional learning (Levy (2007); Parr (2013); Hone and El Said (2016)). Nevertheless, studies show that instructors/teachers play a critical role in leveling up the students ’ academic achievements and preventing/lowering the drop-out rates (Onah, Sinclair, and Boyatt (2014); Tang, Xing, and Pei (2019); Carvalho, Sana, and Yan (2020) . Unfortunately, this role is facing many new challenges once the learning platform switches from physical to online classes (Gunnarsson and Alterman (2012); Xing, Pei, Li, Chen, and Xie (2019)).

Artificial intelligence in education (AIEd) has been shown to provide support for teachers to gain more and betterinsights into their students’ performance on online platforms (Luckin, Holmes, Griffiths, and Forcier (2016)). Another promising area that offers support to educators is learning analytics (LA) (Ferguson (2012)). LA is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs, as defined back in 2011 for the first LAK conference (Long, Siemens, Conole, and Gasevic (2011); Lang, Wise, Merceron, Gas ˇevic ́, and Siemens (2022)). Students’ performance prediction models give the educators insight into how well/poorly their students are doing in the course (Xu, Yuan, and Liu (2020) ). Early intervention based on the prediction of students’ academic performance showed promising results as described in (Burgos et al. (2018)) where the drop-out rate was lowered by 14%. Another study with a control group and an experimentalgroup showed that the groupof students who received an individualized intervention -the experimental group-had a significantly better academic performance (Zhang et al. (2020)).

[2]

Due to the digitization and use of technology in the educational field, there is a large amount of educational data. Educational Data Mining helps to analyze and extract useful information, such as selecting the factors that affect the students‟ performance, predicting students‟ performance, etc., from a large amount of educational data. As students or youths are the future of any nation, predicting the success rate of students in their academic area is a very important and beneficial task. This may be achieved with the help of educational data mining, which utilizes various machine learning techniques.

Among the applications of EDM, detecting student failure at an early stage has been an appealing research topic for researchers due to its social impact. The prediction of the students at risk of being dropouts from an institute or school becomes difficult due to the large number of factors that may influence the academic performance of the students. Thus, it is quite important to predict low-performing students at an early stage with higher accuracy, along with the important factors that may affect their performance.

[3]

Educational systems contain massive data about students’ behavior, enrollment of students, results, and attendance that could be analyzed to improve outcomes of the educational process [1]. Therefore, Educational Data Mining (EDM) has become a necessity to discover knowledge that helps decision makers to improve the educational process [2].EDM can reduce the drop-out rate by providing the academic institutions with knowledge to be able to develop appropriate strategies. It can also provide a timely decision to help students who are vulnerable to failure. Data mining deals with the educational field to identify and evaluate several important learning indicators from the data [3]. Students’ performance prediction is one of the most challenging and interesting topics of EDM research [4]. It helps instructors to track their students’ performance to identify those at-risk [12]. Research on students’ performance prediction is useful to identify the features, behaviors and hidden relations that affect the students’ performance [6][11].

There are two research perspectives, the first one is to find the features that affect the performance of students. The second one is to find an effective methodology to predict students’ performance [4]. It is very crucial to apply the feature selection to discover high influence features that need to be improved the dropout rate and enhance student performance [5]. Machine learning is applied to discover hidden patterns and the relations between the features in addition to the prediction of at-risk students [20].

The massive growth of the educational data gives the education institutes the opportunity to apply data mining techniques to extract useful and hidden information for predicting students’ academic performance effectively [22].

Predicting learning outcomes and especially students’ performance became very important in the overall learning and educational process.

One of the essential objectives of educational data mining is to accurately predict students who are vulnerable to dropwww.ijacsa.thesai.org 882 | P a g

One of the essential objectives of educational data mining is to accurately predict students who are vulnerable to dropout for providing them with more support and the suitable intervention. EDM helps the academic institutions to make an appropriate intervention to assist those students enhance their performance.

[4]

Information and knowledge discovery through data mining is known as knowledge discovery in databases. Scientists working in this field are looking for information that may aid educational institutions in better managing their student populations. It is believed that this will help students in better managing their studies and producing finished work. It is largely concerned with deriving precise rules, classifications, and predictions to assist students in their future education by comprehending and analyzing educational data that signals their future educational achievement.

Classifying data and forecasting values are the most popular methods for data mining. Using Educational Data Mining (EDM) to analyze and categorize survey data to predict and classify students' future semester performance is another notable example of this approach.

Classification is a key component of data mining and knowledge discovery. Every strategy has its advantages and disadvantages.

Many new online methods have been introduced in education, and student digital data has become big data. Processing educational data with data mining techniques allow us to generate students' rules and predictions. Information on the socioeconomic, learning, and course environment can be used for prediction, significantly impacting a student's success or failure. Predicting pupils' academic progress is a major focus of educational data mining. As technology has progressed, educational technology investments have grown. E-Learning systems, including web-based online learning and multimedia technologies, have arisen along with technological breakthroughs, reducing learning costs while eliminating time and space constraints.

Because of digitization, we now have access to a profusion of knowledge on virtually any subject. It is critical to understand how to use vast amounts of data. Several machine learning techniques are used in data mining to extract information from large amounts of data. Making connections between data and making solid future predictions is made possible through this process. Data mining in education, which uses machine learning techniques, allows us to make predictions by analyzing the data obtained thus far in education. Data mining methodologies include classification, clustering, and association rule mining, to name a few. Data mining methods can differ depending on the topic of the study and the sort of data we have available.

It is critical to choose the suitable characteristics for a model to succeed.

[5]

The advancement of computing power and information processing techniques has facilitated the development and application of artificial intelligence (AI). AI-enhanced technology is playing an increasingly prominent role in modernsociety. In the educationfield, AI applications has been widely used in various contexts, and it enables instructors to employ diverse teaching approaches. For example, numerous instructors now use a learning management system (LMS) in teaching to support various learning activities. Researchers can analyze the data recorded in the LMS database by using educational data mining models and learning analytic approaches and can provide relevant insights to instructors; for example, academic success can be predicted, and at-risk students can be identified (Baker et al., 2015 ; Smith et al., 2012 ).

The increasing adoption of AI applications in educational contexts significantly impacts the learning environment and learning approach. Students’ learning performance, motivation, and engagement in the context of blending learning, online learning, and ubiquitous learning have gained researchers’ attention in recent years (Chen et al., 2020b , 2020c ). One advantage of using AI-enhanced applications is collecting data on students’ behaviors during their learning, which enables behavioral analysis. As learning behaviors often reflect students’ cognitive skills, analyzing these data provides a comprehensive understanding of student learning, and individualized guidance can be provided to students.

[6]

The study of the factors that affect the dropout of students in educational institutions is one of the most frequent topics in the educational field.

In recent years, advanced data mining techniques [2–4] and the use of classification models [5], [6] have become one of the main tools for predicting academic performance [7–9] and student dropout statistics at all educational levels.

[7]

Online learning has become a conventional mode of learning in higher education (You, 2016); its major advantages include time flexibility, accessibility, and visibility (Bouhnik & Marcus, 2006). However, it also has several disadvantages. One that has received growing attention in the scientific literature is the high drop-out rate (Arora, Singhal, & Bansal, 2014; Shea & Bidjerano, 2019; Soffer, Kahan, & Livne, 2017), which varies according to the setting in which the course is delivered: academic, open online courses, or courses for professional development (Onah, Sinclair, & Boyatt, 2014).

[8]

Student dropout is considered the most complex and significant issue in the education system (Kim & Kim, 2018). This problem causes economic, social, academic, political, and financial damage to the main agents of education, i.e., from the students to the governmental and promotional agencies for effective and efficient strategies to minimize the indexes of school dropout, since the measures adopted up to now did not have the positive effects on this problem (Martinho et al., 2013).

Previous research has established various definitions of student dropout. The most common definition focuses on whether students will continue to be active until the end of the week or if the current week is the last week in which students are active.

Early identification of students at risk of dropping out is critical to reducing the problem and allowing for the necessary conditions to be targeted. As a result, timing considerations are important for the dropout problem. Some studies have found out that 75% of dropouts occur in the first few weeks (Moreno-Marcos et al., 2018). Dropout prediction is frequently regarded as a time series prediction problem or a sequence labeling problem (Mubarak et al., 2020 ). These can correspond to students’ final status (Jin, 2020). On the other hand, the time dimension can be indirectly included in the prediction of dropout by using the input features available in a specific time window, which allows for the selection of a suitable form of intervention (Drlik et al., 2021). Specifically, students’ dropout causes educational deficiencies which can severely affect the social and economic well-being of current and future generations (Rumberger, 1987). Besides, several losses to the society may be encountered because the productive capacity of a nation can be challenged by lack of skilled labor force, and the issues of dropout can lead to poor standards of living, unemployment issues, and disruptive behaviors in the society (Catterall, 1987). Based on different demerits of students’ dropouts, this issue is considered as a huge hindrance to educational development by researchers, policymakers, and educators (Kim & Kim, 2018).

To address this issue, a warning to dropout can enable schools to identify the behaviors that can accelerate the risks of dropping out and take precautionary proactive measures to solve the issue before occurrence (Balfanz et al., 2007). When students intend to drop out, they apply their decision without carefully taking into consideration the dangers that can arise, either asking the guidance from their parents, relatives, experts, or successful colleagues (Sara et al., 2015). The kind intervention followed by the dropout early warning system can help those who are at the risk of dropping out to remain focused on their studies until they graduate and prepare for a better future (Dynarski & Gleason, 2002).

Several governments have developed and implemented dropout early warning systems to deal with this issue. To list a few measures, the department of education and early childhood development in the state of Victoria in Australia developed the student mapping tool to help private and public schools to analyze students at the risk of disengagement and dropout (Lamb & Rice, 2008). Similarly, the state of Wisconsin in the United States developed the same system to predict students’ dropouts (Knowles, 2015). Furthermore, dropout early warning systems were implemented in the United States during 2014–2015 for more than half of public high schools (Sullivan, 2017). Machine learning is a promising tool for building a predictive model for student dropout and offers early warning to responsible authorities to take alternative measures to students at the risk of dropping out the school (Del Bonifro et al., 2020).

In the last decades, artificial intelligence (AI) has shown the ability to change many aspects of our society and our lives because it offers the technological basis for new services and tools that help decision-making in daily life (Nagy & Molontay, 2018). Education is not immune to this revolution, indeed, AI and ML algorithms can play a significant role in improving several aspects of the learning process (Lykourentzou et al., 2009). The main reason for emphasizing these technologies is the possibility of building new predictive systems which can be utilized to help students to plan for their future and improve their academic careers (Del Bonifro et al., 2020).

One of the most frequently researched topics in educational data mining (EDM) and learning analytics (LA) disciplines is predicting a student’s learning success or failure (Prenkaj et al., 2020 ).

EDM is concerned with the analysis of study-related data to comprehend student behavior. These techniques are typically used to provide more effective learning environments by revealing useful information for modifying course structure or to aid in the prediction of student performance and behavior (Baker & Inventado, 2014).

LA, on the other hand, is concerned with the measurement, collection, analysis, and reporting of student data and backgrounds to understand and improve learning and the environments in which it occurs (Siemens & Baker, 2012).

EDM and LA methods are typically at the heart of current prediction approaches. Predicting the likelihood of students completing or failing a course, particularly in the early weeks (Alamri et al., 2019), has been one of the hottest research topics in learning analytics, as has EDM (Romero et al., 2010).

Once a reliable performance prediction is available, it can be used to identify weak students and provide feedback to students, as well as predict students’ failure (Skalka & Drlik, 2020).

al., 2017). According to (Lang et al., 2017 ), classification and regression methods, neural networks, Bayesian networks, support vector machines, logistic regression, and linear regression could be used to solve the student performance prediction problem. These models are frequently referred to as black-box models because they are difficult to understand and interpret. They are all heavily reliant on feature extraction. Feature extraction, also known as attribute selection, is the compilation of a subset of unique predictive features for the predictive problem in modeling. The process aids in identifying relevant attributes in the dataset that contribute to the accuracy of the prediction model, such as the student’s most recent activity on the correspondingcourse to predict a dropout (Queiroga et al., 2020). Strategies that work well for one type of dataset may not work well for another. In this case, it is frequently necessary to manually develop new feature extraction strategies (Li, 2018).

[9]

D ROPOUT in higher education is a difficult topic to explain; for this reason, [1] proposes to analyze the phenomenon from different perspectives: student, institutional, and state or national. From the student's perspective, there are expectations, goals, intellectual capacities, and socio-economic origin. In the institutional aspect, those options offered to students to include them in university life follow them up and retain them. At the state or national level, desertion must be considered the interruption of studies in any modality, and the policies that promote retention must be analyzed.

In Mexico, 38% of the people who can access higher education do not graduate, which is why the OECD (Organization for Economic Cooperation and Development) places Mexico and Turkey (with the same percentage) as the countries with a severe problem in terms of school dropouts. This percentage contrasts with Germany and Finland, where they have 4.03% and 0.45%, respectively [2]. Additionally, the effects of dropping out are reflected in labor and social inequality because the probability of finding better jobs with greater privileges is permeated by this problem [3]. According to OECD data (2018), 85% of people with higher education (25 to 64 years old) are employed, compared to 75.2% of those with only an average higher education, which shows that there are more significant job opportunities for those who advance and complete their academic studies. In Mexico’s case, this proportion is 80% for those with higher education and 70.6% for those with lower education, which is below the OECD average [4].

On the other hand, according to the report on Higher Education in Mexico (2018), by the OECD, the hiring of young people between 25-34 years of age with higher education is 80.7%, which is less than 84.1% of the average of the other member countries. Thus, [3] indicates that reducing student dropout in higher education impacts positively by promoting a society better prepared to meet global challenges in which we are immersed and improve people’s quality of life, get better jobs, wages, opportunities for intellectual growth, among others.

Recently, however, artificial intelligence, data mining, and computational learning methods have gained great importance in predicting college dropouts. For example, commonly used methods for predicting college dropout are: neural networks [12]-[16], K-nearest neighbors and logistic regression [14],[17], random forest [14],[17], Bayesian networks [18], decision trees [16],[19],[20], support vector machines [21], [29], statistical methods [22]-[27] and finally, deep learning in [28].

Educational Data Mining (EDM) is an area of computer knowledge focused on creating methods to examine the unique types of data that come from large volumes of data in educational environments to provide answers to educational questions or improve educational or administrative processes automated manner.

EDM methods are drawn from various areas, including data mining, computer learning, psychometrics, statistics, information visualization, and computer modeling [31]. Currently, some algorithms have been applied in various real-world contexts to provide solutions with high precision. These algorithms include decision trees, k-means, vector support machines, artificial neural networks, Bayesian learning, instance based methods, and Bayesian models. In this section, decision trees and artificial neural networks are briefly described, and the metrics used to evaluate their performance.

Decision tree algorithms are supervised learning techniques, easy to implement and very useful, composed of a single initial node and underneath other independent trees that indicate the predictive attributes [32]. The decision trees are located within a branch of automatic learning called symbolic learning, in which there are also decision rule models closely related to the trees.

Artificial neural networks (ANN) are computer models that try to mimic the neurons in the human brain and solve complex learning problems. They are composed of algorithms that process a set of data to find non-linear relationships. They can learn and improve their functioning [32]. The simplest type of ANN is the so-called perceptron, which takes a vector of real values as input, calculates a linear combination of these inputs, and produces a value (usually 0 or 1) according to a function. A typical ANN is formed by interconnected neurons arranged in three layers (this may vary). The data entry through the input layer passes through the hidden layer and exits through the output layer. It is worth mentioning that the hidden layer can create several layers. In other words, neurons organize in layers (monolayer and multilayer), and the output of some neurons are the inputs of other neurons, then they are forward (feedforward) if they have connections backward, then they are feedback [30].

[10]

In recent years, education has changed as a result of technological advances available that are directed to the instrumentation of the educational sector, both in software aimed at teaching and in the digital administration of academic records by institution managers. Data Mining is an area of study whose objective was to extract patterns and relevant knowledge from the data. One of the ways to extract these patterns is to use techniques of Learning from Machines, a sub-area of Artificial Intelligence, whose main objective is to develop computational techniques capable of acquiring knowledge automatically. These patterns can be useful to build predictive models to assist in decision-making processes associated with student dropout and their performance analysis. The mining area of educational data is constantly growing.

[11]

An overabundance of data relating to students in digital format permits the educational institutes to analyze the patterns for decision making. The acceptance of technology allows us to process and transform the data that will assist educators, administrators, and policymakers in bettering education quality [9].

Students' academic performance depends on socioeconomic, personal, and other environmental variables. Meanwhile, understanding such aspects and their influence on the performance hep them to make an early decision.

Evaluating students' information to categorize them in making good decisions or developing their performance is essential to research that focuses primarily on interpreting and analyzing educational data. Later, this data can be used for Learning Analytics (LA), which is provided to stakeholders via data mining methods known as Educational Data Mining (EDM).