Application of Big Data for Students' Behavior Prediction in Education Industry

Meltem Yontar Aksoy
Ph.D. student in Industrial Engineering Program
İstanbul Technical University
Istanbul, Turkey
yontarm@itu.edu.tr

Abstract

Big data is a highly innovative technology that provides economic and social benefits for all segments of society. Big data analysis offers significant savings and new possibilities by moving the decision support of many critical areas of knowledge to an upper dimension in many different industries such as banks, energy companies, the pharmaceutical industry, health services, public services, etc. Even though there are few studies about big data applications in education, big data will revolutionize the learning industry in the coming years. Therefore, this study deals with big data applications in the education sector, especially in predicting student behavior. This paper investigated the existing literature for big data usage for predicting students' behavior in the education industry. Then it proposed a specific methodology for determining students who are likely to drop out early from the university. Because the increasing graduation rate is the strategic goal of any public or private educational institution, it is believed that with the proposed model, they can take the right action at the right time to reduce the drop-out rate and increase the graduation rate.

Keywords
Big data, learning analytics, behavior prediction, performance prediction, education

1. Introduction
Today, big data is crucial for our lives, and in modern life, it is emerging as one of the essential technologies. The use and adoption of big data are valuable and provide efficiency in cost and productivity. Because of that, more and more companies focus on big data. Nowadays, big data technologies are used in many fields such as; health services, retail sector, medical area, public services, education, oil and gas, automotive industry, high technology, media, travel, and transport sector, etc. In the future, it is thought that big data will revolutionize the learning industry. For that reason, this study deals with big data applications in education.

Big Data is a novel technology designed to deliver value efficiently by providing high-speed capture and analysis from highly variable, high-volume data clusters (Gantz and Reinsel 2011). According to IBM (2012), "big data is any structured and unstructured data such as activity data, text data, sensor data, audio, video, click streams data, log files data, the internet of things data and more". Big Data contains large volume, high speed, and expandable data types from different sources. Big data applications are suitable for raw structured data (relational data), semi-structured data (XML data), and unstructured data (Word, PDF, Text, Media Logs).

Big data formation has five components: volume, velocity, variety, veracity, and value. The most significant property of big data is the volume representing the size of the big data set. Velocity indicates the rate at which new data are generated and the speed at which they move around the data. Variety means that the obtained data is obtained from different sources. Analyzing unnecessary and altered big data can lead to a severe waste of time and erroneous results. For this reason, the veracity of the data is one of the most sensitive criteria. The definitions of volume, velocity, variety, veracity describe Big Data. After producing and processing the data, these analyses must make a 'value' to reach a helpful result. The process of turning big data into a value is shown in Figure 1. So, V of Value is added to the other 4V.

As seen in Figure 1, the input of the process is called the "Datafication" of our world. These are data sources of big data. These data sources can be listed as activities, conservation, words, social media, browser logs, photos, videos, etc. Some examples of five data sources categories are presented below.

- Activities: This category includes website tracking information, application logs, sensor data such as check-in and other location-tracking generated content, and the data generated by the processors found within vehicles, video games, cable boxes, or household appliances.
- Conversations: Widespread conservation sources are e-mails, web clickstreams, CRM information, supply chain tools, etc.
- Social Media: This category includes Twitter, Facebook, LinkedIn, Yahoo, Instagram, Snapchat, Tumblr, Pinterest, Google, and specific interest social or travel sites to gather individuals' profiles and demographic information.
- Browser Logs: User logs bring a unique understanding of their online user behaviors, such as what they did on website, things they clicked on, etc., during the buyer's journey. Analyzing raw user logs allows more control, accuracy, and transparency into user activities beyond statistics provided by standard web analytics services like Google Analytics or Omniture.
- Videos: A massive amount of content is created and shared continuously across multiple networks and in video format. YouTube is the top video-sharing site online, with one billion users, four billion videos viewed daily, and more than six billion hours of video viewed every month. Other examples are Vimeo and Daily Motion.

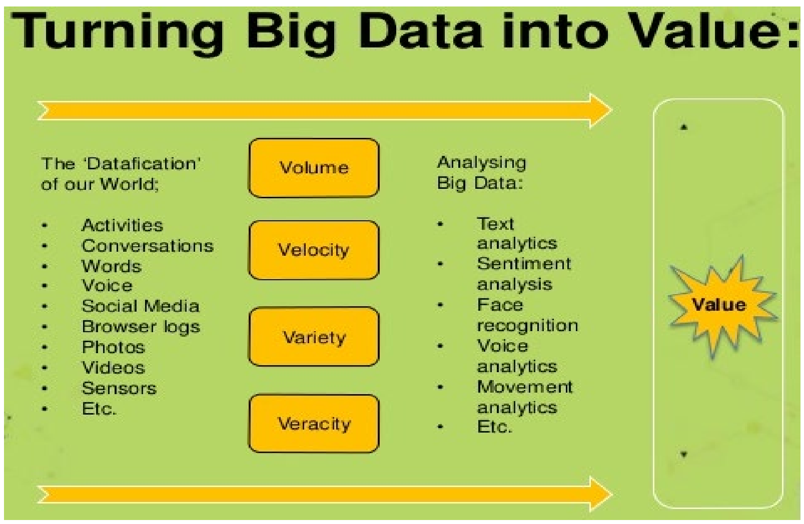


Figure 1. Turning big data into value (Marr 2016)

After determining the data sources, three crucial challenges need to be handled: capturing, storage, and analyzing the data. Several technologies exist that help collect, store, and analyze big data, such as MongoDB, Hadoop, MapReduce, Orange. Hadoop is an open-source library used to process large data on ordinary servers. It provides massive storage, very high processing power, and the ability to manage an almost unlimited number of simultaneous tasks for all kinds of data. It enables efficient management and processing of large data in a distributed computing environment. MapReduce is a system that allows easy analysis of very large data on a distributed architecture. MongoDB is, by the simplest definition, an open-source NoSQL database application. It is a document-based and scalable application. Due to its fast and open-source, MongoDB, which is more preferred in structures where traditional relational databases cannot keep up and remains cumbersome, has gained popularity worldwide. Orange is a python-based tool for processing and mining big data. It has an easy-to-use interface with drag & drop functionalities and a variety of add-ons.

Analyze stage includes applying various analysis techniques to the data to draw meaningful conclusions from the data obtained in different formats from different sources, in other words, to produce value. There are many big data analysis methods, but the paper mentioned most commonly used globally. It can be listed as data mining, machine learning, text analysis, in-memory analytics, predictive analysis, sentiment analysis, face recognition, voice analysis, movement analytic, etc. There is not a single technology that covers big data analysis. Of course, advanced examinations can be applied to big data. Still, in reality, several types of technology are used together to help get the most value from your information.

Machine learning: Machine learning, a precise subclass of Artificial Intelligence that trains a machine on how to learn, makes it possible to rapidly and automatically produce models that can analyze extensive and complex data, deliver faster and more precise results even on a vast scale. By creating precise models, an organization can identify profitable opportunities or avoid unknown risks.

Data management: To be reliably analyzed, the data must be high quality and well managed. With data constantly entering and exiting an organization, it's essential to create repeatable processes to create and maintain data quality standards. Once the data is trusted, organizations should create a master data management program that brings the entire business together on the same page.

Data mining: Data mining supports inspecting at big amounts of data to notice patterns in it. With data mining software, all the chaotic and repetitive data can be reviewed, determined what is relevant, used this information to evaluate possible outcomes, and then increase the speed at making informed decisions.

In-memory analytics: By analyzing data in system memory, it can get instant insights from data and quickly respond to it. This technology eliminates data preparation and analytical processing delays to test new scenarios and build models.

Predictive analytics: Predictive analytics technology uses statistical algorithms, and machine learning techniques to determine the probability of forthcoming outcomes based on past data. It's all about providing the best assessment of what will happen in the future. Thus, organizations can be more confident that they make the best possible business decision.

Text mining: Text mining uses natural language processing and/or machine learning technology to scan documents to analyze large amounts of information and discover new topics and term relationships.

Sentiment Analysis: Sentiment analysis is a method to extract, identify, or characterize the sentiment content of a text unit by using Neuro-Linguistic Programming, statistics, or machine learning methods. Sentiment Analysis is applied in many social media applications, ranging from marketing to customer service.

Face recognition: Face recognition is an application that can identify or verify a person from a video frame, a digital image or video source. It has become popular as a commercial identity and marketing tool.

Voice analytics: Voice analytics (or speech analytics) gathers customer information by analyzing the recorded calls to increase communication and interaction. It enables the recorded conversations between customer representatives and customers in the call centers to be converted into correspondence and the relevant statistical and emotional analyses.

1.1 Objectives

Nowadays, students constitute an essential part of all levels of education in the management of education, and schools have a lot of data about the students in their large databases. The increasing number of universities and colleges aware of these data have already used big data to improve the quality of education. One of the most general usages of big data in the education industry is monitoring success and identifying students at risk. In brief, data is crucial for institutions to produce better results for students by serving teachers appreciate the cause.

This study focuses on learning analytics in education by using big data techniques. Learning analytics is the procedure of collecting, analyzing, and reporting students' data and their contexts to understand and optimize the environments they are learning and experiencing. Considering the advantages of the learning analytics and big data, the transaction process must be mainly electronic rather than manual. It is vital to collect instructional data when they occur to move to more comprehensive and time-sensitive applications of learning analytics applications. Classical face-to-face teaching may support traditional data-driven decision-making processes. Due to the Covid -19 pandemic, face-to-face education has largely left its place in online education. In this way, it has become much easier to store training data.
Most importantly, it has become more important to analyze this data and use it in decision-making processes. Like the data that a person creates when shopping on the internet, a student who receives training online systems unwittingly creates their data. Examples of this data include the user's demographic characteristics, the training they receive, the

time they spend in training, and which activity they have completed or not completed. Higher education institutes have started to use these analytics to improve their services and increase students' continuity with graduation.

Big Data techniques can be used in various ways in learning analytics such as students' performance prediction, skill estimation, behavior detection, course recommendation (Manolis et al. 2013). A student's performance can be estimated by analyzing a student's demographic features, historical performance data, student's interaction with teachers and other students in a learning environment. By analyzing the students' behaviors, the risk of the students being separated from the lessons can be perceived, and precautions can be applied at the beginning of the course to protect the students. Learning systems can provide instant and intelligent feedback to students in response to their input, which will increase their interaction and performance. New courses can be recommended according to their interests by analyzing students' activities. As the number of educational data increases, the reports on educational data become increasingly complex. The data can be visualized using data visualization techniques to identify trends and associations based solely on visual reports quickly.

Low graduation rates have social, individual, and economic impacts on all higher education stakeholders. This ratio is the most critical component of performance-dependent funding. Moreover, students consider the effects of low graduation rates when making university selections. Big Data Analysis can help to increase the graduation rate with different aspects. This study focused on predicting and taking precautions for drop-out students to increase the graduation rate.

The primary purpose of this paper is to predict when students are likely to drop out early in their university education. Also, it examines how to identify the academic and demographic character of the students by using big data and how to determine students likely to drop out.

2. Literature Review

In the 21st century, new learning and teaching technologies have begun to emerge. Online learning is one of the most prominent examples of this. Online learning is a form of simultaneous or asynchronous learning through the internet or a computer network, which is realized by the self-learning of the individual. The spread of online education to the world offers new data collection, analysis, and reporting opportunities. Online learning makes it possible to gather real-time data in detail from each student. This can be done to determine the best way of learning, identify the points that students are challenged with, and test teaching and learning principles. Traditional universities and other educational institutions have also greatly expanded their learning opportunities in this direction.

While big data analyses have been applied in various sectors, little interest has been in the educational field. More recently, work has begun in Educational Data Mining (EDM), a new discipline that focuses on applying data mining tools and techniques to educational practices. Especially with the spread of online education, big data produced in education grows day by day. Therefore, it is now possible for some educational purposes to produce meaningful results from these produced data. The significant data obtained are being implemented to achieve objectives such as improving student learning processes, providing personalized learning systems to students, creating a course management system, and making predictions for the future. To achieve these objectives, educational data mining researchers use technical methods such as classification, regression, density estimation, and clustering have also been applied to the analysis in the academic field. Using these techniques, educational data mining researchers are developing various models to conduct their research.

The most crucial technique is predictive analytics, a data analysis topic that tries to predict future or unknown data using existing knowledge with modeling, machine learning, statistics, and data mining techniques. Predictive models have been successfully applied to the field of educational big data to predict the previous unknown students' performance, to model students' failure and drop-out, and student other behaviors regards to education. Predicting student performance models fall into two categories: the feature-driven predictive model and the generative predictive model. Identifying critical factors such as past grades, student behavior, and student psychological factors is crucial to predicting student performance. The feature-driven predictive model mainly focuses on feature selection and model selection. Identifying the critical features is very important for predictive models. In addition, choosing a suitable model can obtain good predictive performance. The feature-driven predictive models are used supervised learning approaches, such as linear regression, -nearest neighbor (KNN), support vector machine (SVM), linear discriminant analysis, random forest, XGBoost, etc. In contrast, generative models are hard to get but offer better interpretability.

The generative predictive models usually use Naive Bayesian, Bayesian networks, Markov, etc. This paper focus on feature-driven predictive model.

Classical data mining approaches such as classification, regression, clustering analysis, one of the multivariate statistical techniques is used to classify unknown and ungrouped data according to their similarity. Besides, social network analysis is used graph theory to understand, develop, and conduct quantitative and qualitative analyses of relationships in social networks. Existing literature for big data usage for predicting students' behavior in the education industry is summarized in Table 1. As seen in Table 1, the studies in the literature are examined and summarized according to the specific point of the studies, techniques, selected features and datasets used in the studies.

Table 1. Literature review

Author(s)	**Specific Point of the Study**	**Techniques**	**Features used**	**Dataset**
Iatrellis et al. (2020)	Prediction outcomes of students	K-means, random forest algorithm	GPA, specialization field, capstone project grade, time to complete basic cycle, first year performance in core subjects, high school, rank-in-class, student mobility program, internship, awards, grades in high school, grades in entrance exams	1100 records over six academic years, from 2007 to 2012
Xu et al. (2019)	Prediction of student's learning performance associated with internet usage behaviors	Decision trees, support vector machines, artificial neural networks	Final grades of undergraduate students' compulsory courses, online time, offline time, Internet download volume, Internet upload volume, terminal device type	20 million records from 4000 sample students
Zhang et al. (2018)	Prediction student performance based on student behavior pattern	Regularized multi-task	Behavior set, including breakfast, lunch, dinner, shopping, exercise, treatment, buying materials, library entrance, card serve, school bus, dorm access	Student's performance on 12 courses from an Asia university
Ge et al. (2018)	Student behavior warning	Hadoop, principal component analysis	Sex, major, grade, dormitory, age, domicile of origin, hobbies, classroom attendance information, online learning situation, library, dormitory homing rate, entrance guard system, campus id	More than 10000 students
Ashraf et al. (2018)	Student performance prediction	Decision tree, alternative decision tree, multilayered perceptron algorithm, logical regression, support vector machine, k-nearest neighbor, and Naive Base	Age, gender, region, residence, guardian info, cleared certificates, scholarships and results, grades, social network details, extra-curricular activities, and psychometric factors	Different datasets
Shingari et al. (2017)	Student performance prediction	Decision tree (ID3, C4.5), Artificial neural network	Student's grade, medium of teaching, hometown, place of stay, family income status, father's qualification level, mother's qualification level, newspapers reading, novels reading, the internet usage for studying, waking up early in the morning, daily exercise, mode of commutation	NA
Uddin and Lee (2017)	Prediction drop-outs, low retention, poor student performances, lack of motivation, and unnecessary change of study majors and re-admissions	Stochastic probability	GPA, previous grade, language, age, previous degree level, previous degree area, IQ, previous housing, future housing, family size, financial status, scholarship, full-time status, major changing, reading skills, writing skills, oral skills, returning student, drop out, marital status, Facebook membership, twitter membership, number of Facebook friends, number of Facebook post daily, number of tweets daily, consistency, friendliness, writing activity, liking activity, re-tweeting, types of tweets, blog writer	Academic data from 17 Universities around the world for 8–10 years of records for undergraduate and graduate's students; An online survey was designed and used to collect data from students; Twitter API and Facebook API
Xu et al. (2017)	Prediction of student performance	Ensemble learning, Support Vector Machine (SVM)	GPA, the SAT scores, the average grade, and total credits	367 anonymized students at UCLA Mechanical and Aerospace Engineering Department

Author(s)	Specific Point of the Study	Techniques	Features used	Dataset
Jose et al. (2016)	Analyze the relationship between age and academic performance, Student performance prediction	Correlation Analysis, regression Analysis	Historical score, age	1037 undergraduate student at Middle East College datasets
Boysen (2016)	Behavior Warnings in Abnormal Psychology	Descriptive statistics	A survey of Abnormal Psychology instructors	Data is collected through a survey of Abnormal Psychology instructors from 131 students
Sweeney et al. (2015)	Next term student grade prediction	Factorization machine	Historical grade	Historical grade data of summer, fall, and spring terms of a public university (George Mason University) from 2009 to 2014
Meier et al. (2015)	Grade prediction	Grade prediction algorithm	Homework assignments, quizzes, or midterm exams	The scores of approximately 700 students
Iam-On and Boongoen (2015)	Early detection of students at risk of drop-out	Ensemble clustering	Gender, province, department, overall grade and average grades across four subject groups	811 students at Mae Fah Luang University, Chiang Rai, Thailand
Tamhane et al. (2014)	Prediction student risk of poor academic performance	Decision tree, decision table, logistic regression	Scores, gender, ethnicity, free meal, gifted and special education, number of absent days, number of suspensions, and number of discipline incidents	NA
Sobecki (2014)	Course recommendation	Ant colony optimization, particle swarm optimization, intelligent weed optimization, bee colony optimization, and bat algorithm	Semester, studies specialization, studies type, recruitment grade and recruitment number, student identification number, course identification number and course grade	33000 students at Wroclaw University of Technology
Baker (2010)	Investigating the psychological state during their use of three different computer-based learning environments	ANOVA	Psychological state (incidence, persistence, and impact of boredom, frustration, confusion, engaged concentration, delight, and surprise affective states)	There are different datasets from three different computer-based learning environments: 1. AutoTutor:28 undergraduate students from a university in the mid-south of the USA, 2. The Incredible Machine: 36 students in a private high school in Quezon City, 3. Aplusix: 140 high school students from four schools within Metro Manila and one school in Cavite
O'Connor et al. (2009)	Influence of behavioral approach system and behavioral inhibition system on university students' drinking, smoking, and gambling behavior	Correlation Analysis, Regression Analysis	Sex, age, reward responsiveness, drive, fun-seeking, alcohol use, cigarette use, gambling	533 undergraduate students enrolled at an Eastern Canadian university
Vialardi et al. (2009)	Course recommendation	Classification tree (C4.5)	Courses enrolled in, name of course, accumulative GPA, grade	58871 students' records at the School of System Engineering at Universidad de Lima

1. Methodology

Thanks to the rise of online learning environments in recent years, gathering and analyzing real-time data in detail for each student are much more manageable. Almost every university has an automation system and a database of this system. This database contains data about the academic and demographic characteristics of the students. Universities also offer students web applications (online program services), like e-learning environments. A relationship can be found between the tendency of drop-out and the characteristics variable gathered from the automation system and the online program services.

This paper proposes a strategy to structure reference models of students who do and do not drop out to handle this problem. Reference models collect various patterns that are used to predict whether a student will drop out during an academic term. With this proposed structure, teachers could recognize students who are likely to drop out and take action to prevent this situation. The overall architecture of the drop-out or not drop-out prediction system is shown in Figure 2.

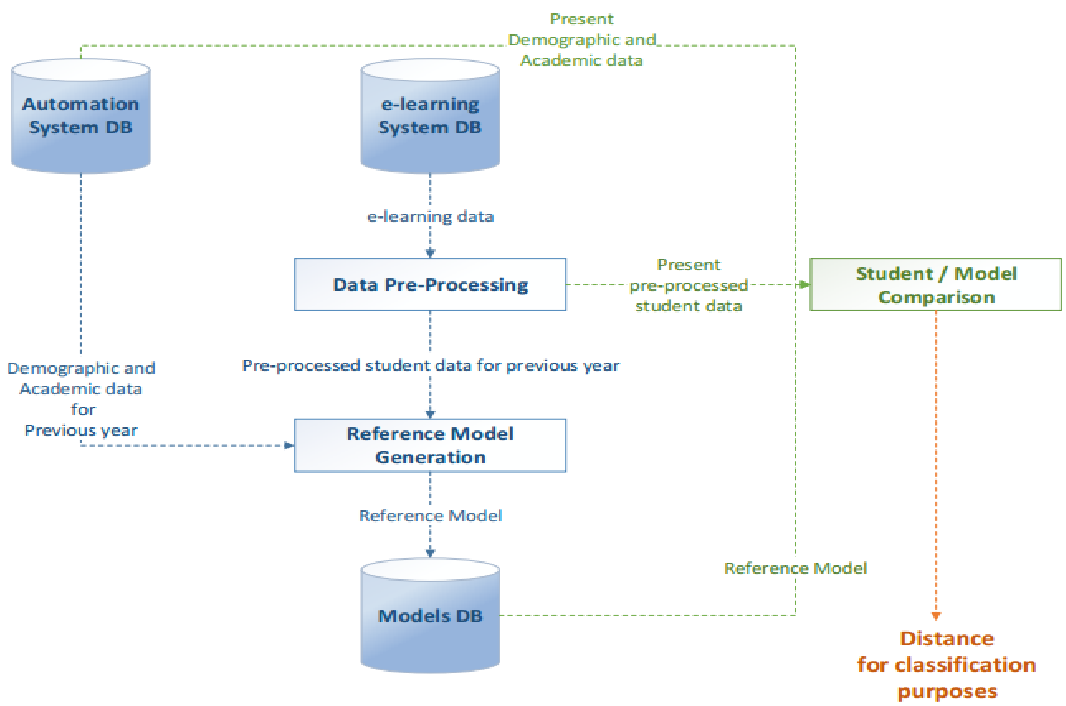


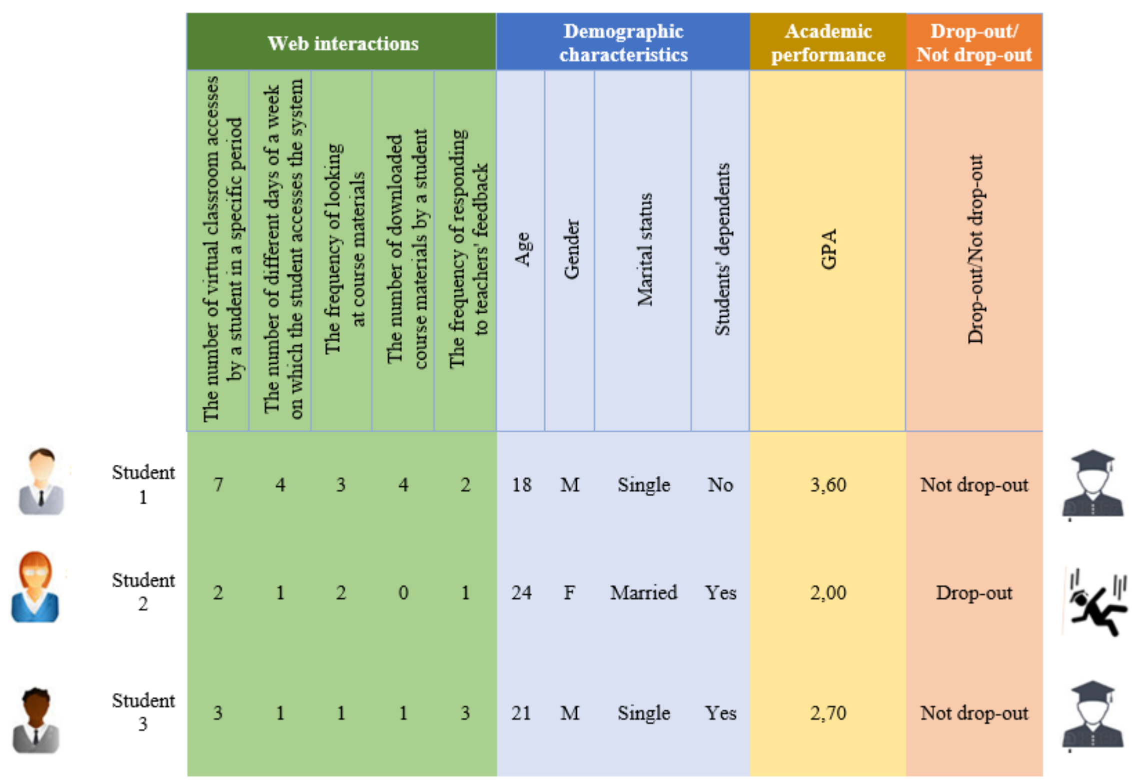
Figure 2. The overall architecture of the drop-out/not drop-out prediction system

As seen in Figure 2, the proposed system included three different databases: automation system database, e-learning system database, and models database. The automation database consists of demographic and academic data of previous years. In this study, age, gender (Female, male), marital status (married, single, divorced), and student's dependents (is there anybody a student has to look after? Yes, no) are chosen as demographic features. As academic performance features, students' GPAs are selected. From e-learning system database, the number of virtual classrooms accessed by a student in a specific period, the number of different days of a week on which students access the systems, the frequency of looking at the course material, and the frequency of responding to teacher feedback are extracted as web interaction data. The online learning systems store record for each action executed by a student. The probable actions on a learning system are shown in Table 2. It is understood that the proposed model focuses on visualizing and adding actions. This activity data should be pre-proceeded by a specialized module called data pre-processing.

Table 2. Action recorded in e-learning system

Actions Recorded in e-Learning System	
Action	**Description**
Visualize	Student opens and views the resource
Add	Student adds a new entry to the resource
Update	Student changes a formerly added entry
Delete	Student removes a formerly added entry

The demographic and academic data and pre-processed e-learning data are used to generate reference models. Reference model data are stored in a special-purpose database called Models database. Also, there is a comparison instrument that can measure the distance of a student to each of the previously produced drop-out and not drop-out reference models. A student can be classified into the nearest groups relevant to this distance. The current student is compared to these reference models, and the closeness to the two clusters is measured. The proximity values of current students to this cluster may be a warning for teachers. This tracking system can warn teachers by classifying students according to their web interactions, demographic variables, and academic performance.
An example of the final state of the pre-processed web interactions data is presented in Figure 3. Figure 3 also shows selected features of students' demographic and performance characteristics. The purpose is to classify the students according to these features into two classes: drop out and not drop out.



	Web interactions					Demographic characteristics				Academic performance	Drop-out/ Not drop-out
	The number of virtual classroom accesses by a student in a specific period	The number of different days of a week on which the student accesses the system	The frequency of looking at course materials	The number of downloaded course materials by a student	The frequency of responding to teachers' feedback	Age	Gender	Marital status	Students' dependents	GPA	Drop-out/Not drop-out
Student 1	7	4	3	4	2	18	M	Single	No	3,60	Not drop-out
Student 2	2	1	2	0	1	24	F	Married	Yes	2,00	Drop-out
Student 3	3	1	1	1	3	21	M	Single	Yes	2,70	Not drop-out

Figure 3. Example of data for the reference model

2. Conclusion and Future Studies

Big data is a highly innovative technology that provides economic and social benefits for all segments of society. Big data analysis offers significant savings and new possibilities by moving the decision support of many critical areas of knowledge to an upper dimension in many different industries such as banks, energy companies, the pharmaceutical industry, health services, public services, etc. Even though there are few studies about big data applications in education, big data will revolutionize the learning industry in the coming years.

Low graduation rates have social, individual, and economic impacts on all higher education stakeholders. This ratio is the most critical component of performance-dependent funding. Moreover, students consider the effects of low graduation rates when making university selections. Big Data Analysis can help to increase the graduation rate with different aspects. To solve the low graduation rate problem, this study proposed a model to predict which students are likely to drop out early in their university education. It focused on predicting and taking precautions for drop-out students to increase the graduation rate.

This paper investigated the existing literature for big data usage for predicting students' behavior in the education industry. It proposed a specific methodology for determining students who are likely to drop out early from the university using big data. It is planned to assess the proposed model with real data from different universities in future studies. In addition, future studies can focus on different demographic, academic, and action features that can affect prediction accuracy. In addition, other than the two classes that are tried to be predicted, the different situations or behaviors of the students can be predicted by the model presented.

References

Ashraf, A., Anwer, S., Khan, M.G., A comparative study of predicting student's performance by use of data mining techniques, American Academic Scientific Research Journal for Engineering, Technology, and Sciences, Vol. 44 no. 1, 122–136, 2018.

Baker, R. S. J., D'Mello, S.K., Rodrigo, M.M.T., and Graesser, A.C., Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive affective states during interactions with three different computer-based learning environments, International Journal of Human-Computer Studies, vol. 68, no. 4, pp. 223–241, 2010.

Boysen, G.A., Wells, A.M., and Dawson, K.J., Instructors' use of trigger warnings and behavior warnings in abnormal psychology, Teaching of Psychology, vol. 43, no. 4, pp. 267–275, 2016.

Gantz, J., and Reinsel, D., Extracting value from chaos, *IDC review*, vol. 1142, pp. 9-10, 2011.

Ge, S., Bai, C., Wan, Q., Hadoop-based college student behavior warning decision system, IEEE International Conference on Big Data Analysis, pp. 217–221, 2018.

IBM, what is big data? Available: http://www.01.ibm.com/software/data/bigdata, 2012.

Iam-On, N., Boongoen, T., Improved student dropout prediction in Thai university using ensemble of mixed-type data clusterings, International Journal of Machine Learning and Cybernetics, vol. 8, no. 2, pp. 1–14, 2015.

Iatrellis, O., Savvas, I. K., Fitsilis, P., and Gerogiannis, V. C., A two-phase machine learning approach for predicting student outcomes. *Education and Information Technologies*, 2020.

Jose, M., Kurian, P., and Biju, V., Progression Analysis of Students in A Higher Education Institution Using Big Data Open-Source Predictive Modeling Tool, *IEEE 3rd MEC International Conference on Big Data and Smart City*, 2016.

Manolis M., Patricia C., and Demetra K., The Potential of Learning Analytics and Big Data, Available: http://www.ariadne.ac.uk/issue71/charlton-et-al, 2013.

Marr, B., *Big Data in Practice*, United Kingdom: Wiley, 2016.

Meier, Y., Xu, J., Atan, O., and Schaar, M.V.D., Personalized grade prediction: a data mining approach, IEEE International Conference on Data Mining, pp. 907–912, 2016.

O'Connor, R.M., Stewart, S.H., and Watt M.C., Distinguishing bas risk for university students' drinking, smoking, and gambling behaviors, Personality and Individual Differences, vol. 46, no. 4, pp. 514–519, 2019.

Shingari, I., Kumar, D., Khetan, M., A review of applications of data mining techniques for prediction of students' performance in higher education, Journal of Statistics and Management Systems, vol. 20, no. 4, pp. 713–722, 2018.

Sobecki J., Comparison of Selected Swarm Intelligence Algorithms in Student Courses Recommendation Application, International Journal of Software Engineering and Knowledge Engineering, vol. 24, no.1, pp. 91-109, 2014.

Sweeney, M., Lester, J., Rangwala, H., Next-term student grade prediction, IEEE International Conference on Big Data, pp. 970–975, 2015.

Tamhane, A., İkbal, S., Sengupta, B., Duggirala, M., and Appleton, J., Predicting student risks through longitudinal analysis, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1544–1552, 2014.

Uddin, M.F., and Lee, J., Proposing stochastic probability-based math model and algorithms utilizing social networking and academic data for good fit students prediction, Social Network Analysis and Mining, vol. 7, no. 1, pp. 29, 2017.

Vialardi, J. Braver, J., Shaftr, L., and Ortiaosa, A., Recommendation in higher education using data mining techniques. 2nd International Conference on Educational Data Mining, pp. 190-199, 2009.
Xu, J., Han, Y., Marcu, D., and Schaar, M., Progressive prediction of student performance in college programs, The Thirty-First AAAI Conference of Artificial Intelligence. AAAI, pp. 1604–1610, 2017.
Xu, X., Wang, J., Peng, H., & Wu, R, Prediction of academic performance associated with Internet usage behaviors using machine learning algorithms, *Computers in Human Behavior*, 2019.
Zhang, X., Sun, G., Pan, Y., Sun, H., He, Y., and Tan, J., Students performance modeling based on behavior pattern, Journal of Ambient Intelligence and Humanized Computing, vol. 3, no. 3, pp. 1–12, 2018.

Biography

Meltem Yontar Aksoy, MSc, Ph.D. Candidate is a Ph.D. student in Industrial Engineering Program at Istanbul Technical University. She received her master's degree in Industrial Engineering from Istanbul Technical University in 2017. She got her bachelor's degree in Industrial Engineering from Gaziantep University in 2012. Parallel to academic life, she has worked as a project auditing expert in the Support Unit of Istanbul Development Agency since October 2013.