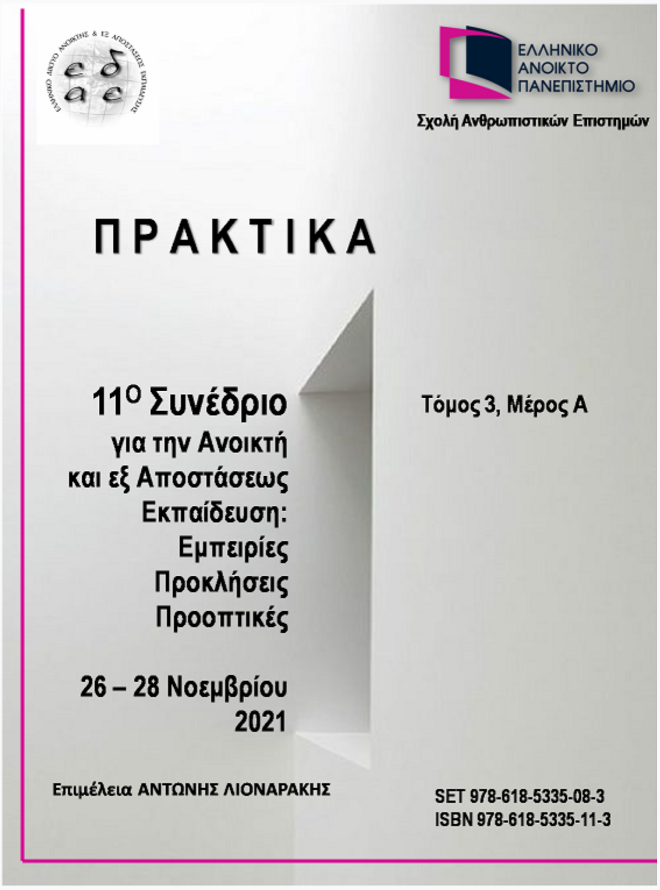
ePublishing
ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ
NATIONAL DOCUMENTATION CENTRE

Διεθνές Συνέδριο για την Ανοικτή & εξ Αποστάσεως Εκπαίδευση

Τόμ. 11, Αρ. 3A (2022)



Evaluating Contact Sessions and Assignments Grades Impact with Association Rules

Nikolaos S. S. Alachiotis, Sotiris Kotsiantis, Elias C. Stavropoulos, Vassilios S. Verykios

doi: [10.12681/icodl.3301](https://doi.org/10.12681/icodl.3301)

https://epublishing.ekt.gr | e-Εκδότης: EKT | Πρόσβαση: 21/05/2024 05:10:20

Evaluating Contact Sessions and Assignments Grades Impact with Association Rules

Nikolaos S. Alachiotis
PhD
Hellenic Open University
nalaxiot@eap.gr
https://orcid.org/0000-0002-9932-7964

Sotiris Kotsiantis
Assistant Professor
Department of Mathematics
University of Patras
kotsiantis@upatras.gr
https://orcid.org/0000-0002-2247-3082

Elias C. Stavropoulos
PhD
Hellenic Open University
estavrop@eap.gr
https://orcid.org/0000-0002-0706-781X

Vassilios S. Verykios
Professor
School of Science & Technology
Hellenic Open University
verykios@eap.gr
https://orcid.org/0000-0002-9758-0819

Abstract

The socialization of students in distance learning and the impact on their performance is a major issue. The concept of Contact Sessions tends to be substituted by videoconferences, especially in the coronavirus era. Learning Analytics (LA) is a modern scientific field that enhances the learning environments as well as the educational procedures and leads to a better comprehension of the students' learning process. The learners' activity is traced by collecting, analyzing, and reporting their data in order to redesign the course by adopting the LA new methodologies. Nowadays, there is a transition from research to practice, which formulates a new perception of learning for the students by implementing the existing LA methodologies.We utilize the opportunity of this scientific field to reveal hidden patterns for student data by employing appropriate Data Mining algorithms. Specifically, we study the impact of Contact Sessions participation on the correlation among students who either have failed or passed the final exams. For this purpose, we find the association rules with the use of the Apriori algorithm in data formulated in market basket format. This means that the data concern the participation in the Contact Sessions and also the written assignments grades. We associate these data similar to the selection of a commercial product A with commercial product B in a market basket.We use the R software environment in our experiments. In this fashion, we prove the need for the students' attendance in the Contact Sessions as a factor that influences their performance in terms of socialization by using Data Mining.

Keywords: *Learning Analytics, Contact Sessions, Distance Learning, Socialization, Association Rules*

Περίληψη

Η κοινωνικοποίηση των φοιτητών στην εξ αποστάσεως εκπαίδευση και η σημασία της στην απόδοση των φοιτητών είναι ένα μείζον ζήτημα. Στην εποχή του κορονοϊού υπάρχει

μια τάση για αντικατάσταση των Δια Ζώσης Συμβουλευτικών Συναντήσεων (ΔΖΣΣ) από βιντεοδιασκέψεις. Η ανάλυση εκπαιδευτικών δεδομένων είναι ένα σύγχρονο επιστημονικό πεδίο που βελτιώνει τα εκπαιδευτικά περιβάλλοντα και τις εκπαιδευτικές διαδικασίες. Οδηγεί σε μια καλύτερη κατανόηση της μαθησιακής διαδικασίας. Η δραστηριότητα των εκπαιδευομένων καταγράφεται με τη συλλογή, ανάλυση και αναφορά των δεδομένων αυτών. Ο σκοπός είναι ο επανασχεδιασμός του μαθήματος χρησιμοποιώντας όρους που προκύπτουν από νέες μεθοδολογίες της ανάλυσης εκπαιδευτικών δεδομένων. Στις μέρες μας υπάρχει μια μετάβαση από την έρευνα στην πράξη η οποία διαμορφώνει μια νέα αντίληψη της μάθησης. Η αντίληψη των μαθητών προκύπτει από υλοποίηση υπαρχουσών μεθοδολογιών ανάλυσης εκπαιδευτικών δεδομένων. Σε αυτή την εργασία διαπραγματευόμαστε αυτό το επιστημονικό πεδίο για να αποκαλύψουμε κρυμμένα μοτίβα δεδομένων προερχομένων από φοιτητές. Για αυτό το σκοπό χρησιμοποιούμε κατάλληλους αλγορίθμους εξόρυξης δεδομένων. Για την ακρίβεια, μελετούμε το αντίκτυπο της φυσικής παρουσίας των φοιτητών σε κάθε ΔΖΣΣ στη δημιουργία σχέσεων μεταξύ τους. Η συμμετοχή στις ΔΖΣΣ αποτελεί μορφή κοινωνικοποίησης και συνεργασίας. Επίσης αυτή τη σχέση μεταξύ φοιτητών τη συσχετίζουμε με την επιτυχία ή αποτυχία στις τελικές εξετάσεις. Για το σκοπό αυτό βρίσκουμε τους κανόνες συσχέτισης με τη χρήση του αλγορίθμου Apriori με τα δεδομένα διατυπωμένα σε μορφή καλαθιού αγοράς. Δηλαδή, ως δεδομένα έχουμε εισάγει την παρουσία ή απουσία των φοιτητών στις ΔΖΣΣ και τις βαθμολογίες στις υποβληθείσες γραπτές εργασίες. Η συσχέτιση της παρουσίας σε μια ΔΖΣΣ με τη παρουσία σε άλλες ΔΖΣΣ και με τους βαθμούς στις γραπτές εργασίες αποτελεί την ανάλυση του καλαθιού αγοράς, όπως η συσχέτιση επιλογής ενός προϊόντος με την επιλογή κάποιου άλλου προϊόντος. Με αυτό τον τρόπο αποδεικνύουμε με τη χρήση εξόρυξης δεδομένων την ανάγκη για παρακολούθηση στις ΔΖΣΣ ως ένα παράγοντα που επηρεάζει την απόδοση των φοιτητών. Χρησιμοποιούμε το R περιβάλλον λογισμικού στα πειράματα μας.

Introduction

The professional socialization of the students in distance education is an important concept. In the work of Kicherova et al. (2015) the authors mention that the students acquire advantages for their future careers by socializing. These advantages are information perception, professional communication with experts during the educational stage, and many other skills. The students achieve these goals by attending social events like conferences and forums. The most important benefit for the students when socializing is the power to persuade a target listener, by using means like presentation, attention, comprehension, and acceptance. According to Bulaev et al. (2016), the socialization parameter of a student enables him to make friends and establish harmonious relations with the teachers, the university administration, and other students. Such a student will be scientifically creative, academically active, self-confident, and stable. Independence of thought and professional fulfillment are two important competences. A student is able to get these competences during professional socialization. Moreover, other benefits are social experience and maturity.

In addition, the concept of socialization is significant in distance learning. A few not obligatory Contact Sessions are conducted during the semester or academic year.There is a dispute as regards the effect of these sessions. In the current work, we will investigate and interpret this matter. We aim to prove that the social relations which the students develop during sessions are considerable for their performance in the final exams.These relations determine the professional development, the cognitive scientific formation of the student, and many other issues to a large extent.We interpret the possible relations among students in these sessions as association rules in the terms of data mining. We use Educational Data Mining to reveal hidden patterns in the educational data to extend and enrich educational issues. Such an issue is the socialization of the students in distance learning associated with the participation in Contact Sessions. For this purpose, we employ algorithms that focus on patterns of groups of data.We have defined the data in market-basket format. This means that we assign for each student the Contact Sessions in which the student was present or absent and also the grades which the student achieved in each written assignment, similar to the products a customer has selected in a market basket.

The concept of the current work is to prove that participation in the Contact Sessions results in the development of relationships among successful students or students who fail. For this purpose, we employ the algorithm Apriori. In order to estimate socialization, we use the methodologies of Verykios et al. (2004) so as to achieve high confidence or support greater than a threshold. We use the Apriori algorithm so as to produce association rules. Our goal is to apply data mining to prove the relationships among students in Contact Sessions and the impact of these relationships on the written assignments' grades and the status of the final exam (pass or fail). We consider as itemsets the items which are included in the rules produced with data from each student's profile. In our case, the itemsets include the presence or the absence in each contact session and also the grade which a student achieved in each written assignment.

In our experiments, we sort the association rules according to two properties. As regards the students who pass the final exams,the property is the support of the rules to be greater than a threshold. As regards the students who fail in the final exams, we sort the rules according to the value of confidence.We explain the terms support and confidence below.

According to (Agrawal and Srikant, 1994) for a set of items (i.e. presence or absence in all Contact Sessions and grades in all written assignments) and a set of transactions (i.e. all students with these features as items) each transaction is a subset of the set of items. An association rule among 2 items is e.g. CS1-present→CS2-present and it exists in *D* transactions. As we observe this holds for sets of items which are subsets of the set which contains all items. The confidence of an association rule is the percentage of the transactions in *D* that e.g when the transactions contain the CS1-present they also contain CS2-present. In our example, the term confidence measures how frequently the item CS1-present exists in transactions with CS2-present. The support of an association rule is the percentage of the transactions that contain a union among two items (e.g. CS1-present ∪ CS2-present). This means that both items exist in the association rule (Suma and Shobha, 2021).

We utilize the term support to define the itemsets which overcome a support threshold. We call these items frequent itemsets according to Suma and Shobha (2021). With the term confidence, we denote the significance of a rule according to its magnitude. According to Suma and Shobha (2021), the rules with confidence and support greater than the threshold are considered as strong rules. Some of these strong rules very often a database owner desires to hide. These rules are called sensitive rules.

We focus on the participation effect of the students who attend together as a group the Contact Sessions and they pass or fail the final exams. Furthermore, in the rules, we consider the grades which the students achieve in the written assignments and the final exams.We utilize the work of Verykios et al. (2004) with another aim.The motivation of this cited work was to create association rules hiding algorithms in order to prevent data mining techniques to reveal sensitive information from databases.In the methodology of Verykios et al. (2004) the proposed algorithms perform a perturbation of parameters such as support and confidence. In the current work, we have used the values of support and confidence that produce the association rules which are determined from the most frequent itemsets.We apply the Apriori algorithm in a market-basket format of data for this purpose. We have defined this format above.This method has helped us to achieve the optimal combination of support and confidence at a higher level.
In the work of Alachiotis et al. (2017) Pivot Tables of Excel are used to visualize the attention of participants in a blended learning course. The discovery of data results is used to predict learners' performance and to optimize the course organization. Therefore, there is the opportunity for the tutor to advise the students during the learning procedure. In Alachiotis et al. (2019) the authors examine the provided ability of a data scientist to trace the learning activity of the students from Learning Management Systems (LMSs). This ability can benefit the students, faculty, and administration to enhance parameters such as the learning procedure, the course design, and the educational methodology.The purpose is to analyze learners' behavior and performance in order to make conclusions that are able to enhance the functionality of a distance learning course. In the current work, we extend these two methodologiesby using data mining terms and artificial intelligence to reveal hidden patterns among students.
A brief analysis of association rules exists in the work of (Kotsiantis and Kanellopoulos, 2006). In this work, there are some basic terms for those rules and the descriptions of algorithms like AIS for mining association rules, Apriori for pruning itemsets, FP-Tree for frequent pattern mining, TreeProjection, and other algorithms.As regards the sampling and the parallelization various algorithms are presented (e.g. Sampling Error Estimation, Fast Parallel Mining). There are also descriptions that concern the application of association rules to databases, the interestingness of an association with the term confidence. Moreover, the mining of association rules among products that either complement or conflict with each other is explored. These rules are called negative association rules. These terms are useful in market-basket analysis. The equivalence of our educational data with the market-basket analysis is the following. In a market-basket, the selection of one commercial product is associated with the selection of other commercial products (a market basket that contains A, contains also B). In our case, we use this methodology to associate the presence of each student in the CS with the presence of other students in the CS and also with the grades of the students' written assignments.Our data are formatted in this market-basket format as we describe above.

Review of Literature

The social integration and the spirit of cooperation among students are examined in Pomohaci and Sopa (2016). The relationships established within the group are also examined. The authors focus on activities such as sports. We focus on the cooperation among students who are being met in Contact Sessions because of the distant type of learning. In the current work, we try to prove this fact with the Apriori algorithm applied to quantitative data.

The promotion of social presence in face-to-face meetings of online courses is studied in (Tucker,2012). The issue is examined from a pedagogical perspective. The platform which is used for the experiments on online courses is Centra. The platform which is used for chat rooms and forums is Blackboard. Questionnaires were given to the students in order to collect data relatively with their options for both online and face-to-face classes. The surveys of quantitative data were analyzed with SPSS and the ANOVA table. The conclusions show that Centra has a positive effect on student participation and social presence in online learning. The students had the ability to discuss projects, assignments, and other courses in the Centra platform.

Ethical considerations of Learning Analytics and the experience of the LA practice by administrators, developers, researchers, teachers, learners are studied in Cerratto Pargman et al. (2021). The identification of at-risk students and the significance of attrition to the institution's reputation are presented in Gkontzis et al. (2019).The authors use Big Data mining and machine learning methods, and they also develop a prediction tool for student attrition. The viable use of data mining in education is illustrated.

In the work of Kotsiantis et al. (2013) the authors use data coming from the learning management system Moodle and data gathered from questionnaires. The goal is to predict the students' success or failure to acquire a pass course grade, by using four different methods (visualization of the variables determined by the questionnaires, C4.5 decision tree, class association rules, and *k*-means implementation). The R language and the Weka tool are used for this purpose.

In addition,Gkontzis et al.(2018)are elaborating data come from an annual course of the Hellenic Open University which takes place simultaneously in four different cities of Greece. The authors analyze the logins, replies, and quizzes in combination with the average grade of the submitted written exercises during the year, in order to predict academic performance. In this fashion, the students' learning activity and persistence in the learning procedure are revealed. The amount of data is large, and the Tableau visualization tool is used to depict the students' activity.

In the work of Ougiaroglou and Paschalis, (2012) the authors use the Apriori algorithm to find association rules regarding the interest illustrated by the students of secondary education to attend lessons. The authors have used the WEKA software for this purpose.Integer Linear Programming (ILP) is used in Suma and Shobha (2021) in order to hide sensitive association rules. The sanitization problem which concerns a database is constrained to an ILP issue. A tradeoff of support and confidence of the sensitive association rules takes place in a proposed algorithm.

The outliers students are examined in Rajeswari et al. (2014). These are the students who have passed with a borderline pass more than two courses or they have failed in more than two courses. The authors have assigned the students in itemsets with the use of the Apriori algorithm.They also have determined the minimum and the maximum support threshold. The less frequent itemsets (rare itemsets) were those that have had a smaller support value than the threshold.

In Gosch et al. (2021) the authors negotiate an ethical issue. This regards the learner autonomy by providing Learning Analytics as a service instead of intervention. The analysis which takes place deals with the issues of anonymization in combination with personal data gathering. The conclusion is that the application of LA is successful by developing a LA culture in the corresponding institution.

Various parameters are used in Zhan et al. (2020) to determine the difficulty of the course, and also to separate the grades in levels. The authors have separated the levels of the students with the use of the k-means algorithm. Furthermore, the authors employ the Apriori algorithmin order to use the validity and the difficulty of the tests to extract association rules. By following this methodology, they achieve an association of the grades among different sections of the course.

In the work of Wu et al. (2020) an investigation of the blended learning courses'significance exists. The authors have used various algorithms for this purpose. These algorithms are CART, Naïve Bayes and Apriori. The subject of the course which the authors have used in the experiments is Programming. This course has had a duration of two semesters. In this work, the authors have revealed the findings of the teachers in the exam quizzes. In addition, Wu et al. (2020) have analyzed the students' behavior. The submitted projects were offline case studies. The authors have illustrated the need for the association of the grades in the two semesters with the use of the Apriori algorithm. They also achieve the clustering of the students with the use of the k-means in data which contain the grades in the quizzes, assignments, and final exams. The authors have classified these students' features with the use of the Random Forest algorithm, according to these parameters' significance. The implemented algorithms were used for prediction purposes and the significance of each feature. Moreover, the authors have illustrated the various features of the course and the findings of the Apriori algorithm according to confidence and support.

In the perspective of the above-cited works, we employ the Apriori algorithm to reveal associations among students who participated or not in the Contact Sessions and also the grades which they have achieved in the written assignments. It is an important concept, that proves the impact of the socialization in the Contact Sessions on the written assignments' grades, and the performance in the final exams.

Data description

The Hellenic Open University (HOU) consists of eight Schools, which offer undergraduate and graduate courses to adult learners. These courses consist of modules. The students ought to submit 4-6 written assignments for a 10-month academic year. The final exams at the end of the academic year are compulsory. In addition, five not compulsory counseling Contact Sessions take place in many cities around Greece.

In this study, the data concern the "PLI30– Algorithms and Complexity" a third-year module of the undergraduate "Informatics" curricula of the School of Science and Technology in the Hellenic Open University. It is a course with also 5 not obligatory Contact Sessions. In order to successfully complete the requirements of the course, a student should submit five written assignments and a grade equal to or higher than five, in the final exams. Students can undertake the final exams of the module, only if they have successfully completed the four or five projects with a total score of twenty-five or more, on a ten-grade scale for each project. During the academic year, students have the opportunity to attend, if they wish, five Contact Sessions of four hours each. The module PLI30 covers subjects such as Algorithms and Complexity, Computational Theory, and Automatic and Standard Languages. The module "Algorithms and Complexity" contains algorithms like Greedy Algorithms, Divide and Conquer Algorithms, the algorithms Breadth-First Search, Depth First Search for traversing a graph, Asymptotic Analysis for the calculation of the complexity of Recursive Algorithms. The students are separated into groups. We have selected a sample of students and we have performed our experiments.

The data from this course consist of 5 written assignments (projects), 5 Contact Sessions, and also final exams which are repeated two times. Our sample consists of 80 students. Initially, we format the data with values of strings (words). This means that the value of an item for a student who attended e.g. the first contact session is “CS1-present”. The value of an item for a student who has received a grade e.g. in the first written assignment among 5 and 6.5 is “project1-grade1”.Specifically, the CS (Contact Sessions) items have the value “present” or “absent”, while we separate the project grades into levels grade1, grade2,grade3, and FAIL. We denote the grades among 5 and 6.5 as grade1, the grades among 6.5 and 8.4 as grade2, and the grades among 8.5 and 10 as grade3. We consider the grades below 5 as FAIL. As a next step, we consider each student as a transaction and the above values as items in each transaction. This is the market basket format, similar to commercial products selected in a market basket during a transaction.

Methodology and Results

First methodology

Initially, we try to approach the issue with the Pearson Correlation coefficient. We depict the correlation between written assignments (projects) grades in Figure 1 and the correlation between Contact Sessions in Figure 2. It is a heatmap in which we use one-hot encoding, in order to convert rules to a matrix format with numbers. We break each probable property of each categorical feature into 0-or-1 features. As a result, we are able to use the correlation with the aforementioned coefficient. This method is appropriate for measuring the relationship among pairs of variables. In the current case this relationship is positive (among 0 and 1). The advantage is the efficiency for a linear association among dependent and independent variables. In order to apply this method, the theoretical background of the data is required to understand the correlation. A detailed description of this method exists in Kumar and Chong, (2018).The following figures were created with the use of the Python language.

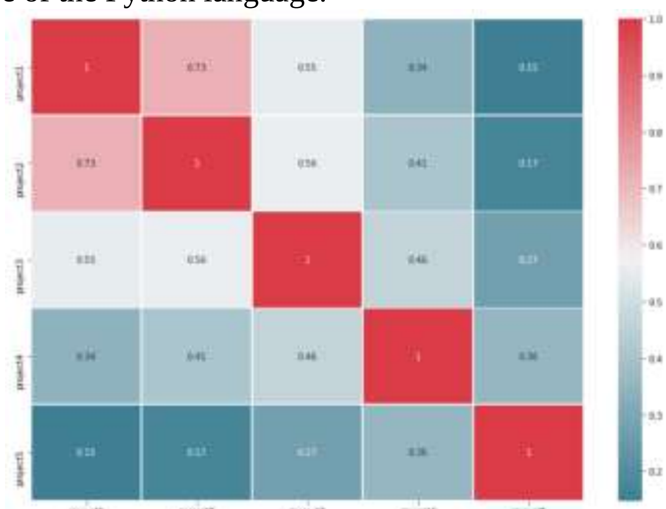


Figure 1: Correlation among grades

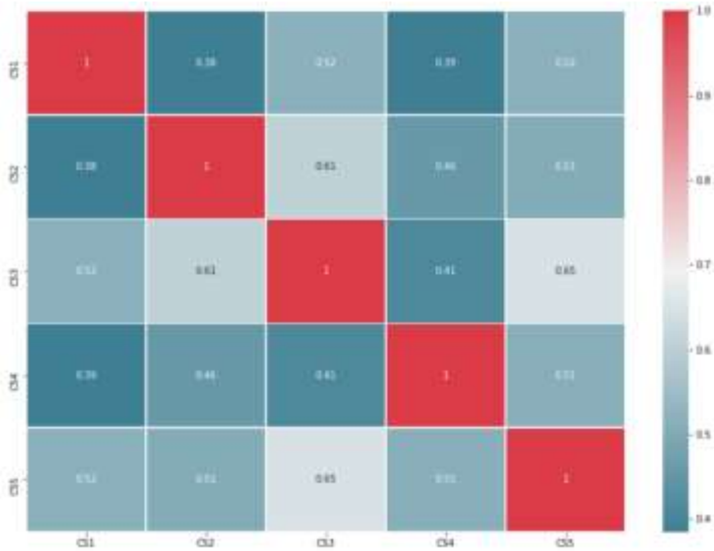


Figure 2: Correlation among Contact Sessions

The value of each block denotes the Correlation Coefficient for a pair of Contact Sessions (Figure 2) or a pair of the written assignmentsgrades (project grades in Figure 1). The values for the Contact Sessions are 0 or 1 (absence or presence respectively) and 0 to 10 for the grades of the written assignments. In Figure 1 exists a weak, moderate, and strong relationship among variables, while in Figure 2 exists a moderate relationship among variables.

In the following paragraphs, we follow a different approach than this Correlation Coefficient. More specifically, we propose below the Apriori algorithm for the association among the presence, absence in the Contact Sessions, and the grades achieved in the written assignments.

Second Methodology

In order to apply the Apriori algorithm, we use the market basket format for our data.The equivalence with the market basket analysis can be considered as a correspondence of each student to one transaction in a transactional database because items are included in the profile of each student. As it is proved Apriori algorithm is efficient in our case, where we consider our data as a transactional database. We apply the algorithm in medium-size data.

In our case, we observe that some rules even if they have high confidence, perhaps will be proved useless. Therefore, we want to hide them. In contradiction to the privacy preservation use of the Apriori algorithm, we want to hide the association rules which we don’t need and not the sensitive rules. We consider as useless association rules those rules with support and confidence below the threshold. In addition, if the left-hand side of a rule takes place after the right-hand side of this rule we determine this rule as useless. For example, for a rule which in the left-hand contains the item CS3-present and on the right-hand side contains the item CS1-present we consider it as useless.

Next, we run the Apriori algorithm so that to reveal the frequent itemsets as it is depicted in the figures below, initially for students who passed the exams and afterward for students who failed the exams. We have implemented our methodology with the R

software environment. The Apriori algorithm is appropriate because it reveals the associations among students who have attended the same Contact Sessions or they have achieved similar project grades. It also proves, the effect of these relationships on the final exam grades.
The invention of the Apriori algorithm according to Agrawal and Srikant (1994) is based to consider the candidate itemsets which were large in each previous pass of the database and also they have support greater than a threshold during the execution of the algorithm.As we have already mentioned, the itemsets with support greater than the threshold are called frequent itemsets similar to Kuswari et al. (2014) and we calculate them with the market-basket analysis.
In Figure 3 we illustratethe frequent itemsets of students who passed the exams:

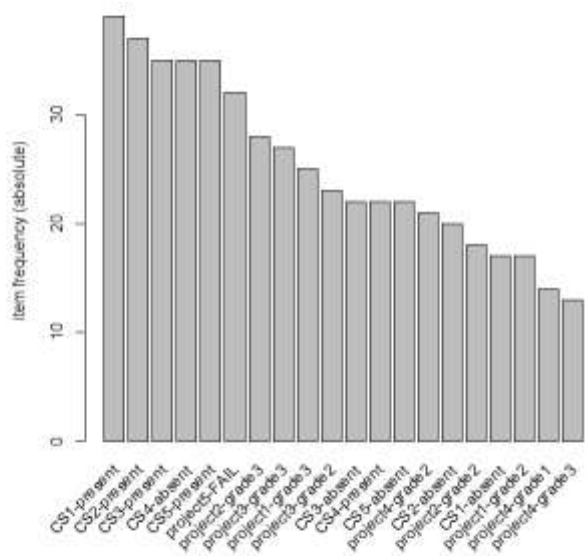


Figure 3: Frequent itemsets of the students who passed the final exams

In Figure 4 we illustrate the frequent itemsets of the students who failed the exams:

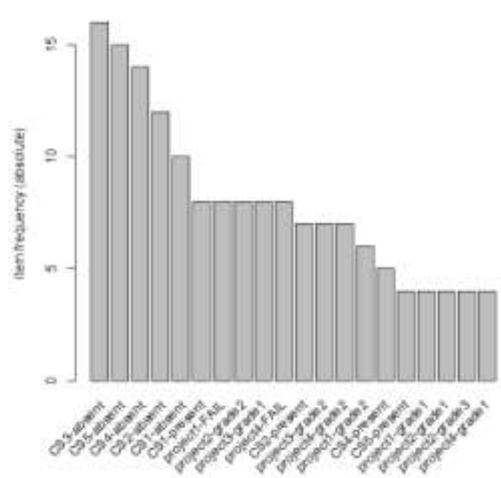


Figure 4: Frequent Itemsets of students who failed in the final exams

In Figure 5 we depict the 11 association rules for the students who have passed the final exams, while in Figure 6 we depict the top 28 association rules for the students who have failed the final exams. We consider the association rules for which the left-hand side of a rule takes place before the right-hand side of the rule. For example, on the left-hand side exists the attendance to the second contact session and a grade2 in the project4, and on the right-hand side exists the attendance to the fifth contact session. In **Figure *5*** the association rules are sorted by the support and the association rules of **Figure *6*** are sorted by confidence.

lhs		rhs	support	confidence	coverage	lift	count
{CS3-present, project4-grade2}	=>	{CS5-present}	0.2631579	1	0.2631579	1.628571	15
{CS1-present, project4-grade2}	=>	{CS5-present}	0.2631579	1	0.2631579	1.628571	15
{CS2-present, project4-grade2}	=>	{CS5-present}	0.2456140	1	0.2456140	1.628571	14
{CS2-present, project4-grade2}	=>	{CS3-present}	0.2456140	1	0.2456140	1.628571	14
{CS2-present, CS3-present, project4-grade2}	=>	{CS5-present}	0.2456140	1	0.2456140	1.628571	14
{CS1-absent, CS3-absent}	=>	{CS4-absent}	0.2280702	1	0.2280702	1.628571	13
{CS1-present, CS3-present, project4-grade2}	=>	{CS5-present}	0.2280702	1	0.2280702	1.628571	13
{CS1-present, project2-grade2}	=>	{CS2-present}	0.2105263	1	0.2105263	1.540541	12
{CS4-present, project4-grade2}	=>	{CS5-present}	0.2105263	1	0.2105263	1.628571	12
{CS1-present, CS2-present, project4-grade2}	=>	{CS5-present}	0.2105263	1	0.2105263	1.628571	12
{CS1-present, CS2-present, CS3-present, [illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Figure 5: Association rules of students who have passed the exams by support

lhs		rhs	support	confidence	coverage	lift	count
{project2-grade1}	=>	{CS2-absent}	0.2105263	1	0.210526	1.583333	4
{project2-grade1}	=>	{CS3-absent}	0.2105263	1	0.210526	1.1875	4
{project3-grade2}	=>	{CS5-absent}	0.3684211	1	0.368421	1.266667	7
{project3-grade2}	=>	{CS3-absent}	0.3684211	1	0.368421	1.1875	7
{CS1-absent}	=>	{CS5-absent}	0.5263158	1	0.526316	1.266667	10
{CS1-absent}	=>	{CS3-absent}	0.5263158	1	0.526316	1.1875	10
{CS2-absent}	=>	{CS3-absent}	0.6315789	1	0.631579	1.1875	12
{CS4-absent}	=>	{CS5-absent}	0.7368421	1	0.736842	1.266667	14
{CS1-present,CS4-present}	=>	{CS5-present}	0.2105263	1	0.210526	4.75	4
{CS4-present,project3-grade1}	=>	{CS5-present}	0.2105263	1	0.210526	4.75	4
{project1-grade2,project2-grade2}	=>	{CS4-absent}	0.2631579	1	0.263158	1.357143	5
{project1-grade2,project2-grade2}	=>	{CS5-absent}	0.2631579	1	0.263158	1.266667	5
{project1-grade2,project2-grade2}	=>	{CS3-absent}	0.2631579	1	0.263158	1.1875	5
{CS3-absent,project1-grade2}	=>	{CS4-absent}	0.2631579	1	0.263158	1.357143	5
{CS3-absent,project1-grade2}	=>	{CS5-absent}	0.2631579	1	0.263158	1.266667	5
{project2-grade2,project3-grade2}	=>	{CS4-absent}	0.2105263	1	0.210526	1.357143	4
{CS2-absent,project3-grade2}	=>	{CS4-absent}	0.2105263	1	0.210526	1.357143	4
{CS1-absent,project4-grade2}	=>	{CS4-absent}	0.2105263	1	0.210526	1.357143	4
{CS2-absent,project4-grade2}	=>	{CS4-absent}	0.2105263	1	0.210526	1.357143	4
{CS2-absent,project4-grade2}	=>	{CS5-absent}	0.2105263	1	0.210526	1.266667	4
{CS3-absent,project4-grade2}	=>	{CS4-absent}	0.2631579	1	0.263158	1.357143	5
{CS3-absent,project4-grade2}	=>	{CS5-absent}	0.2631579	1	0.263158	1.266667	5
{CS1-present,CS2-present}	=>	{project3-grade1}	0.2105263	1	0.210526	2.375	4
{CS2-present,CS3-absent}	=>	{CS5-absent}	0.2105263	1	0.210526	1.266667	4
{project2-grade2,project4-FAIL}	=>	{CS4-absent}	0.2105263	1	0.210526	1.357143	4
{project2-grade2,project4-FAIL}	=>	{CS5-absent}	0.2105263	1	0.210526	1.266667	4
{CS2-absent,project4-FAIL}	=>	{CS4-absent}	0.2105263	1	0.210526	1.357143	4

Figure 6: Top association rules of students who have failed the exams by confidence

Results

The outcomes from the association rules of the above Figure 6 for those who have failed the exams are the following:The students who were absent in the first, second and fourth Contact Sessions were absent in all Contact Sessions. The students who were present in the first and the fourth Contact Sessions were also present in the fifth Contact Session, and they have achieved a borderline pass in the third written assignment (project3). Other students were absent in all Contact Sessions and they have achieved grade2 in the first, second, third, and fourth written assignments (project1, project2, project3, project4). We interpret these indications that eventually the written assignments have been written by third persons because the students have failed the final exams. Some of the students have attended the first two Contact Sessions and they have achieved a borderline pass in project3. This means that the participation in these Contact Sessions has helped them in the attempt to write the project themselves without the help of a third person. The students who were absent in the fourth and the fifth Contact Sessions and have failed in project4 had gone to the sessions exploratory and hadn't studied alone.We sort the rules which concern the failed students by confidence.We take into account the rules for which the values of the right-hand side of a rule occur afterward the left-hand side of this rule.

The outcomes for those who have passed the final exams are the following. The students who were present in the first, second, and third Contact Sessions and have achieved a grade2 in the fourth written assignment (project4) were also present in the fifth Contact Session.The students who were present in the first Contact Session and have achieved a grade2 in the second written assignment (project2) were also present in the second Contact Session. The students who were present in the first, second,and the third Contact Sessions, and have achieved a grade2 in the second and fourth written assignments (project2 and project4), were also present in the fifth Contact Session.Hence, the students have benefitted from participation and socialization in the Contact Sessions. In addition, they have attempted to author the written assignments without the help of a third person. These students have absorbed knowledge in the Contact Sessions, and they didn't need any further help for the written assignments. The Contact Sessions were sufficient for them to pass the exams. Perhaps, socialization may have helped them to cooperate and pass.Some of the students were absent in the first, third, and fourth Contact Sessions but they have passed the final exams. Therefore, we interpret that these students preferred to study independently, and they hadn't absorbed knowledge by attending a contact session. We consider the rules for which the values of the left-hand side of a rule take place before the right-hand side of this rule, similar to the rules for the students who have failed the exams.

Conclusions

The concept of students' socialization in the Contact Sessions and the trend of the latter to be substituted by videoconferences is a major issue. We try to approach this issue in data mining terms. Specifically, we have formulated the presence of the students to the Contact Sessions and the grades which they have achieved to the written assignments and to the final exam in market basket format.This means that we use items like "CS1-present" for the presence of a student in the first contact session and "CS1-absent" for the absence in this session. In addition, we use items like project-grade1 for a grade among 5 and 6.5 in project1 and project1-grade2 for a grade among 6.5 and 8.4 in project1.We have classified the students into groups according to those who have passed and those who have failed the final exams. We try to apply a market-basket analysis with these data. This means the occurrence of some items is associated with the occurrence of other items.Afterward, we have applied the Apriori algorithm to generate the frequent itemsets. These are the itemsets that belong to an association rule with support greater than a threshold. Furthermore, we study the use of terms such as confidence and support in order to evaluate the impact of each rule.We use these terms in order to hide association rules which we don't need.In the association rules, the left-hand side values of the rules take place before the right-hand side of the rules. In addition, we have depicted graphically the correlation of the presence in pairs of Contact Sessions and the correlation of the grades in pairs of written assignments with the Pearson Correlation Coefficient as an alternative methodology. In the future, we aim to study the effect of Contact Sessions and the written assignments grades, as also their importance in the final grade prediction with machine learning algorithms. In addition, we will attempt a comparison of the algorithms' performance with educational data in order to find the effect of Contact Sessions on the students' performance prediction. Moreover, we will study the association among the variables of the current study with hidden Markov Models.

References

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. *Proc. 20th Int. Conf. Very Large Data Bases VLDB, 1215*.

Alachiotis, N. S., Stavropoulos, E. C., & Verykios, V. S. (2017, November 23-26). *Learning Analytics with Excel in a Blended Learning Course*. *Διεθνές Συνέδριο για Την Ανοικτή & Εξ Αποστάσεως Εκπαίδευση*, http://dx.doi.org/10.12681/icodl.1077 8–18.

Alachiotis, N. S., Stavropoulos, E. C., & Verykios, V. S. (2019). Analyzing learners behavior and resources effectiveness in a distance learning course: A case study of the Hellenic Open University. *Journal of Information Science Theory and Practice*.*7*(3), 6-20. https://doi.org/10.1633/JISTaP.2019.7.3.1

Bulaev, V., Koverkina, E., Soshenko, M., Shmyrev, V., Shmyrev, D . (2016). Socialization of student's youth as a factor of development and social renewal of the contemporary society: By experience of the Russian state social university. *International Review of Management and Marketing*, *6*(3), 53-64.

CerrattoPargman, T., McGrath, C., Viberg, O., Kitto, K., Knight, S., and Ferguson, R. (2021). Responsible learning analytics: creating just, ethical, and caring.*11th International Conference on Learning Analytics & Knowledge (LAK21)*.https://www.solaresearch.org/wp-content/uploads/2021/04/LAK21_CompanionProceedings.pdf

Gkontzis, A. F., Kotsiantis, S., Panagiotakopoulos, C. T., & Verykios, V. S. (2019). A predictive analytics framework as a countermeasure for attrition of students. *Interactive Learning Environments*, 1-15. https://doi.org/10.1080/10494820.2019.1674884

Gkontzis, A. F., Panagiotakopoulos, C. T., Kotsiantis, S., & Verykios, V. S. (2018). Measuring engagement to assess performance of students in distance learning. *9th International Conference on Information, Intelligence, Systems and Applications (IISA).* https://doi.org/10.1109/iisa.2018.8633607

Gosch, N., Andrews. D., Barreiros, C., Leitner, P., Staudegger, E., Ebner, M., &Lindstaedt, S., (2021, April).Learning Analytics as a Service for Empowered Learners: From Data Subjects to Controllers.*11th International Learning Analytics and Knowledge Conference April 2021, 475–481 (LAK21)* https://doi.org/10.1145/3448139.3448186

Kicherova, M. N., Efimova, G. Z., &Khvesko, T. V. (2015). Early Professional Socialization of University Students in Russia. *Procedia - Social and Behavioral Sciences*, Vol.200, 442-448https://doi.org/10.1016/j.sbspro.2015.08.093

Kotsiantis, S., &Kanellopoulos, D. (2005). Association Rules Mining: A Recent Overview. *GESTS International Transactions on Computer Science and Engineering*, *32*, 71–82.

Kotsiantis, S., Tselios, N., Filippidi, A., &Komis, V. (2013). Using learning analytics to identify successful learners in a blended learning course. *International Journal of Technology Enhanced Learning*, *5*(2), 133. https://doi.org/10.1504/ijtel.2013.059088

Kumar, S., & Chong, I. (2018). Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states. *International Journal of Environmental Research and Public Health*, *15*(12), 2907. https://doi.org/10.3390/ijerph15122907

Kuswari, H., Insani, N, &Sumarno, B. (2014). *Application Of Association Rules With Apriori Algorithm To Determine The Pattern Of The Relationship Between SBMPTN Database And Student's Grade Point Average.*International Seminar on Innovation in Mathematics and Mathematics Education. Department of Mathematics Education Faculty of Mathematics and Natural Science Yogyakarta State University.

Lin, W., & Tseng, M. (2006). Automated support specification for efficient mining of interesting association rules. *Journal of Information Science*, *32*(3), 238-250. https://doi.org/10.1177/0165551506064364

Ougiaroglou, S., & Paschalis, G. (2012). Association Rules Mining from the Educational Data of ESOG Web-Based Application. In L. Iliadis, I. Maglogiannis, H. Papadopoulos, K. Karatzas, & S. Sioutas (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 105–114). Springer Berlin Heidelberg.

Pomohaci, M., &Sopa, I. S. (2016). Study regarding socialization and social integration of students. *Scientific Bulletin*, *21*(1), 46-53. https://doi.org/10.1515/bsaft-2016-0036

Rajeswari, A.M., M Sridevi, M., C Deisy, C. (2014)- Outliers detection on educational data using fuzzy association rule mining,*International Conference on Advance in Computer Communication and Information Science (ACCIS-14), pp. 1-9*

Suma. B., S., &Shobha., G. (2021). Association rule hiding using integer linear programming. *International Journal of Electrical and Computer Engineering (IJECE)*, *11*(4), 3451-3458. https://doi.org/10.11591/ijece.v11i4.pp3451-3458

Tucker, S. (2012). Promoting socialization in distance education. *Turkish Online Journal of Distance Education*, *13*, 174–182.

Verykios, V., Elmagarmid, A., Bertino, E., Saygin, Y., &Dasseni, E. (2004). Association rule hiding. *IEEE Transactions on Knowledge and Data Engineering*, *16*(4), 434-447. https://doi.org/10.1109/tkde.2004.1269668

Wu, M., Zhao, H., Yan, X., Guo, Y., & Wang, K. (2020). Student achievement analysis and prediction based on the whole learning process. *15th International Conference on Computer Science & Education (ICCSE)*.https://doi.org/10.1109/iccse49874.2020.9201865

Zhan, J., Fan, X., & Zhao, Y. (2020, October). Automatic Rules Generation for Teaching Strategies Improvement.*ACM International Conference Proceeding Series (CSAE)*. https://doi.org/10.1145/3424978.3425079