FOCUS

A new QoE-based prediction model for evaluating virtual education systems with COVID-19 side effects using data mining

Chen Tan[1] · Jianzhong Lin[1]

Abstract

Today, emerging technologies such as 5G Internet of things (IoT), virtual reality and cloud-edge computing have enhanced and upgraded higher education environments in universities, colleagues and research centers. Computer-assisted learning systems with aggregating IoT applications and smart devices have improved the e-learning systems by enabling remote monitoring and screening of the behavioral aspects of teaching and education scores of students. On the other side, educational data mining has improved the higher education systems by predicting and analyzing the behavioral aspects of teaching and education scores of students. Due to an unexpected and huge increase in the number of patients during coronavirus (COVID-19) pandemic, all universities, campuses, schools, research centers, many scientific collaborations and meetings have closed and forced to initiate online teaching, e-learning and virtual meeting. Due to importance of behavioral aspects of teaching and education between lecturers and students, prediction of quality of experience (QoE) in virtual education systems is a critical issue. This paper presents a new prediction model to detect technical aspects of teaching and e-learning in virtual education systems using data mining. Association rules mining and supervised techniques are applied to detect efficient QoE factors on virtual education systems. The experimental results described that the suggested prediction model meets the proper accuracy, precision and recall factors for predicting the behavioral aspects of teaching and e-learning for students in virtual education systems.

Keywords Data mining · Virtual education system · E-learning · Association rules mining · Classification

1 Introduction

With advancement of new emerging technologies such as Internet of things (IoT) (Zhou et al. 2021; Failed 2020a) in educational environments, e-learning systems have used to intelligent techniques for increasing performance of learning in cloud-based computer-assisted learning applications (Zhou et al. 2019a). In higher education and e-learning systems, the coronavirus (COVID-19) epidemic conditions have caused a major troubles and technical problems to universities (Demuyakor 2020), institutions and research centers across all countries, which many

universities, institutions and research centers have led to virtual education systems such as learning management systems (LMS) (Díaz-Antón and Pérez 2006). Unexpected changes on the educational system have caused students and teachers to face many problems in how to teach and understand the course concepts in the context of virtual education. IoT-based assisted learning systems have been applied popularly in modem societies such as smart devices (Lv et al. 2021a), sensors, radio-frequency identification (RFID) and actuators. More and more data have been accumulated in cloud data centers of virtual education systems and there are many conceal and unknown patterns or knowledge in student and teacher information based on behavioral aspects of teaching and learning in virtual education systems. So, data mining as powerful technique is to search and detect valued data pattern or knowledge. Analyzing quality of experience (QoE) factors (Jiang et al. 2021) for behavioral aspects of teaching and education between lecturers and students is a critical issue based on COVID-19 side effects in virtual education systems.

✉ Jianzhong Lin
jzlin@sjtu.edu.cn

Chen Tan
764882743@qq.com

1 Shanghai Jiao Tong University, Shanghai 200040, China

Data mining is the process of abstracting unknown but latently useful information and knowledge hidden in numerous, uncompleted, noise-interweaving, blurry, stochastic and practical application data (Lou et al. 2021a). Based on data mining techniques including association rules mining, classification, clustering, sequence pattern, prediction and trend analysis, predicting QoE factors can be influenced on performance of virtual education systems for next semesters in universities and colleges (Lv et al. 2021b; Zuo et al. 2017). According to above-mentioned issues, this paper presents a prediction model based on association rules mining as one important and basic technique and supervised techniques focused on behavioral aspects of teaching and learning to evaluate the QoE factors of virtual education systems. The suggested model applies IoT-based assisted learning systems to continuously monitor the behavioral aspects of students and teachers according to existing structural attributes in virtual education systems. The presented model detects the indications of interactive negotiations between a third party framework to provide e-learning environments for students and teacher. The proposed prediction model is developed for detecting the QoE factors and identifying changes in educational levels of students and teachers with data mining methods. The main contributions of this paper are as follows:

- Proposing a prediction model for evaluating performance of virtual education systems based on COVID-19 side effects.
- Detecting behavioral aspects of teaching and learning based on QoE factors in virtual education systems.
- Applying association rules mining and classification algorithms to predict behavioral aspects of teaching and learning based on QoE factors.
- Evaluating the behavioral aspects based on QoE factors using association rules mining and supervised methods to predict the e-learning satisfactoriness in virtual education systems.

The structure of this paper is organized as follows: Sect. 2 presents a brief review and analysis of the current related works in this field. Section 3 presents the proposed prediction model in detail that includes the behavioral aspects based on QoE factors for students in the virtual education system. Section 4 demonstrates the experimental results of the suggested prediction model over statistical examination using existing supervised methods. Section 5 provides the conclusion and new research directions as future work in this field.

2 Related works

This section illustrates a review on some relevant studies that applied data mining methods on educational systems, e-learning models and teaching styles.

Ashraf, et al. (Ashraf et al. 2020) proposed a prediction approach to evaluate academic student papers in educational data mining. They used boosting algorithm with a combined synthetic minority oversampling technique and J48 classifier to compare with Naïve Bayes algorithm. The main weakness of this research is ignoring feature selection and data preprocessing on educational data set.

In other work, Isma'il et al. (2020) presented a new course recommender model to predict the expenditure of admission positions for undergraduate students. They analyzed admission procedure to some departments and schools such as health and medical science, business and economics, agriculture and management faculties. In this research, the authors did not discussed about main attributes of data set and their values. Also, there is no novelty in the prediction model. Just they have applied Naïve Bayes (NB), support vector machine (SMV), Kth-nearest neighbor (KNN) and decision tree (DT) classifiers. Finally, they have evaluated accuracy as only prediction factor in this research without any cross fold integration.

According to Troussas, et al. (Troussas et al. 2020), the authors have proposed a new prediction method for detecting interaction activities of students in learning modalities. In this method, important activities such as student traits are detected using fuzzy logic and refined important rules are analyzed as fuzzy output based on a decision-making approach. Just 40 samples as student activities were applied to detect important learning modalities. Also, there is no technical evaluation on accuracy, precision and recall factors.

Also, Aslam (2019) presented a new statistical neutrosophic testing analysis for predicting teaching test cases for students in university. In this analysis, an enhanced analysis of variance (ANOVA) algorithm was applied to evaluate important aspects of students in university. The main weakness of this research is that the authors ignored feature selection to improve the quality of the testing analysis in t-test and p-test strategies. To analyze physical aspects of education students, Zhu (2019) presented an association rules mining method based on a mutual exclusion model on the Apriori algorithm to minimize the time complexity of rule mining. The author in this research has ignored the classification method on the optimal selected features based on important extracted rules. In the same work, Qiang (2019) proposed a prediction method to enhance the correctness and increase the performance of the physical education based on stable split argument and

field programmable gate arrays (FPGA). There is no evaluation on selected features based on the FPGA method using C4.5 classifier. Also, the author just evaluates the proposed model without any comparison of other algorithms.

In prediction of e-learning systems, Azzi et al. (2019) presented a new classification model for categorizing the learning style of the teachers in e-learning environments. The authors extracted behavioral aspects of the teachers on different perspectives. Also, they applied the fuzzy C-means algorithm to cluster existing behavioral aspects of teachers. The authors just compared the number of courses in the evaluation results without comparing other clustering approaches. Daghestani et al. (2020) presented a gamified component-based learning architecture for the learning management system (LMS). The authors evaluated the proposed architecture using some classifiers such as NB, KNN and DT algorithms. In the evaluation results, some important prediction metrics such as error detection, accuracy and sensitivity have been ignored.

In other work (Assunção Flores and Gago 2020), authors have discussed advantages and weaknesses of teaching aspects on educational systems during COVID-19 pandemic in Portugal. Also, they provided a synthetic analysis on the existing remote teaching techniques for educational context as theory and practice. Also, to analyze COVID-19 side effects on online education systems (Chen et al. 2020), authors provided various aspects of students satisfaction for educational system during COVID-19 pandemic. They analyzed existing related coefficients of the effectiveness factors with applied some rule-based hypothesis. The authors just evaluated the applied data set using neural network classification. Also, some critical factors to detect teaching aspects such as accuracy and recall have been ignored.

Mitrofanova et al. (2019) provided a conceptual classification and analysis of refining teaching courses and educational methods using data mining techniques. They categorized technical aspects of educational data mining based on some important factors such as adverse student behavior, social network analysis, knowledge tracing using data mining tools. There is no experimental and evaluation results on the technical aspects in the proposed classification.

Finally, a comprehensive review was presented for evaluating educational data mining using soft computing and learning techniques in Charitopoulos et al. (2020). The authors categorized existing work in domain 2010–2018 relevant to soft computing methodologies for machine learning-based educational systems. This paper evaluates 148 relevant research studies related to soft computing techniques for educational data mining with respect to e-learning context, teaching methods and learning

management systems (LMS). The main weakness of this survey is that the authors ignored presentation of a technical taxonomy for categorization of classic and soft computing methodologies on educational data mining.

3 Proposed QoE-based prediction model

The evaluation of teaching is an important job operated by colleges frequently. We can make use of the results of evaluation, such as the teachers' proportion score, satisfactory level of e-learning system, assessment of learning-assisted equipment to comprehend the total teaching and studying circumstance of some departments or the whole university. But what factors lead to good or bad results of evaluation? In other words, which behavioral aspects are effective in the virtual education system and which learning-assisted techniques have not sufficient performance? Aiming at this question, data mining techniques can detect evaluation of behavioral aspects of teaching and learning based on QoE factors to make full advantage of original data of students' score for teachers and virtual education systems in recent semester to find out hidden rules and pattern. According to above-mentioned discussion, Fig. 1 shows a QoE-based prediction model to evaluate performance of virtual education systems using data mining techniques. This model includes four main phases: (1) virtual education system platform, (2) data preparation, (3) rule mining and (4) prediction model.

In virtual education platform, students and teacher login to the system with respected username and password. Students can use a categorization of e-learning courses and teacher can choose his/her specified course to teach. Teacher can use existing computer-assisted learning methods such as presenting PowerPoint and PDF files, screen capturing, audio capturing, video capturing, file sharing, e-handwriting and many design skills. The collection of these learning methods in the LMS platform as virtual education system can be considered for student to achieve best knowledge for each course section. In other side, the reaction of each of e-learning tools to student can be specified as behavioral aspects in QoE factors from students after finalizing e-learning courses. After completion of each course, students complete specified questionnaire based on behavioral aspects of virtual education system platform. Each filled questionnaire gives a conceptual QoE-based behavioral aspects from e-learning satisfactory level for student in virtual education system.

In data preparation phase, all QoE-based information are gathered to feature collection database. To achieve optimal performance of the prediction method, a data preprocessing is required to clean the existing QoE factors from the virtual education system. Furthermore, the normalization
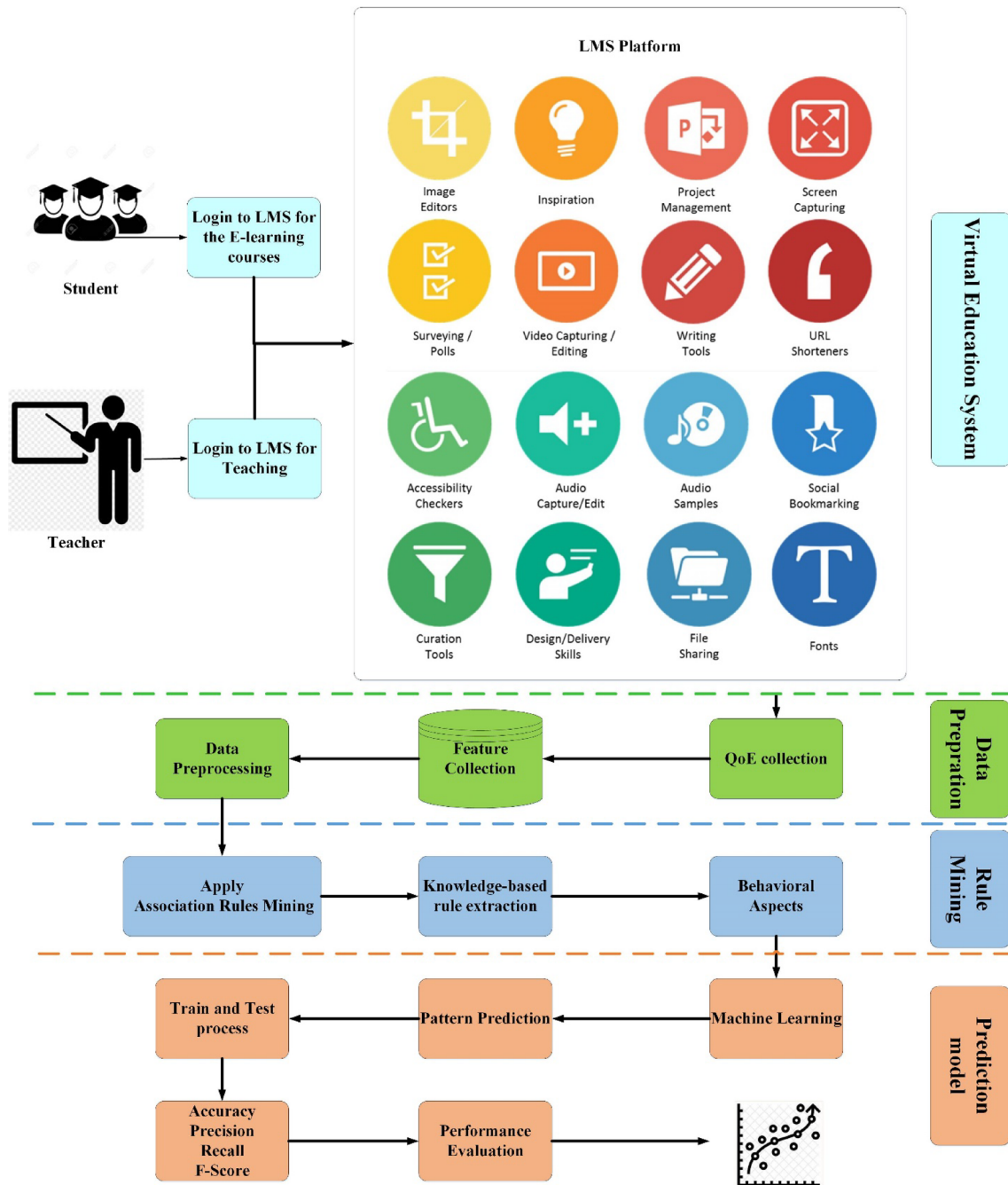
Fig. 1 QoE-based prediction model for virtual education system

procedure is used for reducing the dimension of data for easier classification processes in the proposed prediction model. Some QoE attributes have nominal domain and others have numerical domain. Based on feature normalization method, all QoE attributes are reconfigured to same domain.

Rule mining phase illustrates a knowledge-based rule extraction according to association rules mining methods. Then, the behavioral aspects of the QoE-based virtual

education system are specified. Based on the above QoE factors as behavioral features in virtual education system, we can refine important association rules in the data set to achieve a sufficient pattern for predicting effective behavioral aspects of QoE-based virtual education system.

Finally, in prediction phase, machine learning is provided to detect unknown patterns of the proposed data set with train and test procedures. Existing prediction factors

such as accuracy, precision, recall and f-measure are evaluated to examine performance of prediction model.

3.1 Data collection and preprocessing

According to the proposed prediction model on QoE factors, we used the LMS platform based on virtual education system for gathering data and instances of electronic courses.[1] Table 1 shows the content of each QoE factor which is made e-learning platform including all features of behavioral aspects for teaching and learning in English education systems. The existing QoE factors were collected using evaluation of virtual education system by students in 32 features.

In the QoE factors, five basic features are illustrated for each student including gender, age, academic degree, academic field and employment status. In addition, six academic fields are including computer, electrical and mechanical engineering, economic, social sciences and accounting. Also, 25 questions have been provided to specify evaluation of virtual education system. The main concept of these 25 questions is related to evaluation of behavioral aspects of students based on the existing computer-assisted learning methods in virtual education systems. Some of the existing questions are related to technological aspects of e-learning and rate of accessibility of students to virtual education infrastructures. We illustrated description of each question in Table 1. Finally, two features as the GPA with rang of [0–20] and status of each student have presented to illustrate degree of satisfactory each student based on e-learning in virtual education system.

After illustrating data collection, the data preprocessing phase is required to clean the existing QoE factors from the virtual education system. Furthermore, the normalization procedure is used for reducing the dimension of data for easier classification processes in the proposed prediction model. We finalized 543 samples as student's information for the proposed prediction model based on the QoE factors in applied virtual education system.

3.2 Association rules mining method

Association rules mining is to search interesting relationships, which determine that what will happen coincidently, among data (Wu et al. 2018; Zhu et al. 2020). Based on the above QoE factors as behavioral features in virtual education system, we can refine important association rules in the data set to achieve a sufficient pattern for predicting effective behavioral aspects of QoE-based virtual education system. For example, a computer engineering (CE)

student who was satisfied based on behavioral aspects of virtual education system with "*Very Good*" (VG) grade and achieved *A* score in total average probably gain a succeed grade in final status, then we can achieve to an association rule between existing behavioral aspects based on QoE factors $CE => VG => A => succeed$. This is an instance of association rules mining. The formal description of association rules mining is as follows: $I = (i_1, i_2, i_3 … i_k)$ is a set consisted of k different elements, each element is called an item, the set is called items set. The set which includes k items is called k items set (Lou et al. 2021b; Bai et al. 2021).

Postulated a transmit data, each transmit T in which is a subset of the data item I, that is $T \subseteq I$, and only if $X \subseteq T$, we say transmit T includes items set X. So the mining association rules is the implication such as "$X => Y$", in which $X \subseteq I$, $Y \subseteq I$, $X \cap Y = \Phi$. The mining association rules $X => Y$ are well-founded in the transmit database and have support degree and confidence degree (Abedini and Zhang 2021; Huang et al. 2020). Support degree expresses that s% transmits of D include $X \cup Y$. (This sign presents a set which supports both X and Y.) Confidence degree expresses that c% transmits of those which includes X also include Y, the description is: confidence ($X => Y$): the number of transmits which includes both X and Y/the number of transmits which includes X (Zhang and Wang 2020). The user will give a MinSup and a MinConf, for items sets X and Y, if the confidence degree of the rule $X => Y$ is not lower than MinCon, we called this rule as a mining association rules. From semantic angle, the confidence degree expresses the correct degree of the rule (Liu et al. 2020); support degree expresses what percent's objects we can induce from this rule, vise the impatience of this rule to all data. For example, among 200 student grade records there are 30 records which expresses that capability of learning is A, and among these 30 records there are 15 records which expresses the grade of extracurricular activities is C. So the rule that the learning grade is A implies that the extracurricular activities grade is C has the confidence degree $C = 15/30 = 0.5$ and the support degree $S = 15/200 = 0.075$. Figure 2 shows a conceptual Apriori algorithm in association rules mining (Shi et al. 2020a).

After bring forth the association rules mining, bring out the core algorithm of this miring-Apriori algorithm. This algorithm can be seen as the classic algorithm of association rules mining, the core thought is as follows (Failed 2020b): (1) to find out all the frequent item sets in the transmit database. The so-called frequent item sets is the item sets whose support degree is higher or equal to MinSup. Now we simply descript the first step of this algorithm as follows (chart 1): This algorithm uses the recursion method. First, it generate the set of frequent 1 item sets L_1, then the set of frequent 2 item sets L_2, when

Table 1 QoE factors as behavioral features in virtual education system

Feature	Description	Value
Gender	Male (M), Female (F)	M, F
Age	(Lower than 20) A/(21–30) B/(31–40) C/(41–50) D	A, B, C, D
Academic degree	Bachelor (B), Master (M), PhD (P)	B, M, P
Academic field	Computer Engineering (CE), Electrical Engineering (EE), Mechanical Engineering (ME), Economic (EC), Social Sciences (SS) and Accounting (AC)	CE, EE, ME, EC, SS, AC
Employment status	Employed (EM), Not-Employed (NE)	EM, NE
Q1	The rate of achievement of educational goals appropriate to each lesson?	VG, G, Av, L, VL
Q2	Satisfaction rate of e-learning?	VG, G, Av, L, VL
Q3	The relevance of the content to the level of information and abilities of the student?	VG, G, Av, L, VL
Q4	Proportion of content volume to the number of courses?	VG, G, Av, L, VL
Q5	Provide content in a logical order and sequence?	VG, G, Av, L, VL
Q6	Enjoyable use of e-learning system?	VG, G, Av, L, VL
Q7	Appropriate time of use and learning time of the system?	VG, G, Av, L, VL
Q8	Ability to review previous information by the system?	VG, G, Av, L, VL
Q9	Backup ability for saving educational information by the system?	VG, G, Av, L, VL
Q10	Using PPT and PDF files as educational instruments?	VG, G, Av, L, VL
Q11	Creating group discussion and cooperation?	VG, G, Av, L, VL
Q12	The level of student participation in teaching?	VG, G, Av, L, VL
Q13	The content of interactive lessons?	VG, G, Av, L, VL
Q14	The content of the lessons in audio form?	VG, G, Av, L, VL
Q15	The content of the lessons in video form?	VG, G, Av, L, VL
Q16	The content of the lessons in share-screen form?	VG, G, Av, L, VL
Q17	Satisfaction with e-learning social interactions?	VG, G, Av, L, VL
Q18	Ability to use online text conversations?	VG, G, Av, L, VL
Q19	Ability to use video conferencing to satisfy the e-learning environment?	VG, G, Av, L, VL
Q20	Satisfaction with technical facilities and e-learning support?	VG, G, Av, L, VL
Q21	The amount of guidance and roadmap in the system?	VG, G, Av, L, VL
Q22	The effect of Internet speed and bandwidth on e-learning system?	VG, G, Av, L, VL
Q23	Use of communication tools by the teacher such as chat and voice?	VG, G, Av, L, VL

Table 1 (continued)

Feature	Description	Value
Q24	The amount of access to the teacher to solve content-related problems?	VG, G, Av, L, VL
Q25	Attendance meetings to solve academic problems?	VG, G, Av, L, VL
GPA	GPA of existing semester	(0–20)
Status	Degree of student learning satisfactory on the virtual education system?	Succeed, Average, Damaged

```
Apriori(T, ε)
    L_1 ← {large 1 − itemsets}
    k ← 2
    while L_{k−1} ≠ ∅
        C_k ← {c = a ∪ {b} | a ∈ L_{k−1} ∧ b ∉ a, {s ⊆ c | |s| = k − 1} ⊆ L_{k−1}}
        for transactions t ∈ T
            D_t ← {c ∈ C_k | c ⊆ t}
            for candidates c ∈ D_t
                count[c] ← count[c] + 1
        L_k ← {c ∈ C_k | count[c] ≥ ε}
        k ← k + 1
    return ⋃_k L_k
```

Fig. 2 Apriori algorithm in association rules mining (Wu et al. 2009; Lv 2021)

some r make L_r empty, the algorithm ceases (Zhang et al. 2019). Here, in the k round circle, it first generates the set of all candidate k item sets, then scans the database to compute the support degree of each item sets and preserves Lk consisted of the k item sets whose support degree is higher or equal to MinSup. The generation of C_k needs to suffice such demands: (1) relation: each item sets in C_k is generated by the connection of two item sets of $L_{k−1}$. These two item sets have the same first $k−2$ items and the different last item ($k−1$ item). (2) Pruning: Because of any non-empty frequent item sets of frequent item sets is a frequent item sets, so after the connection of the former step, it needs to delete all that item sets which has non-frequent subset. This is called pruning. The application of pruning lowers the cost of the calculation of all the candidate item sets' support degree and improves the performance of the algorithm (Yin et al. 1045). To apply association rules mining, we used Weka 3.9 software installed on a system with operation system window-64-bit, 8 GB (RAM), with an intel ® Core ™ i5-6200U CPU.

Figure 3 illustrates extracted association rules using Apriori algorithm for QoE factors of student in virtual education system according to the proposed data set. We show ten important rules with high confidence score in existing QoE mining. Each confidence score as a context-based threshold illustrates percentage of acceptance for matching dependencies between extracted rules. Based on existing confidence scores, the maximum confidence scores are related to rules 4 and 9 with score "1". Also, the minimum confidence score is related to rule 7 with score "0.84".

3.3 Classification approach

In classification process, existing classification techniques are applied to evaluate the proposed prediction model on the collected QoE instances. We applied machine learning algorithms such as multilayer perceptron (MLP) (Yu et al. 2021), C4.5 (J48) (Zhang et al. 2021), KStar classifier (Gholipour et al. 2020), sequential minimal optimization (SMO) (Wang et al. 2021), K-nearest neighbors (IBk)(Xu et al. 2018), Random Forest (Chen et al. 2021), Naïve Bayes (NB) (Li et al. 2021) and hybrid J48 + BinarySplites algorithm (Zuo et al. 2015). In the collected QoE data set, 70% of the instances are divided as train and 30% of them are applied as testing data set. Also, the k-fold cross-validation technique is employed in which the data set is separated into k mutual-exclusive folds of close equal size in a random way for k times training and testing of the classification process. The evaluation factors of prediction are considered as the total number of right classifications, which is divided by the instances in the data set in cross-validation method.

4 Experimental results

To examine the efficiency of the proposed prediction model, five main predictive factors including accuracy, precision, recall, F-measure and execution time are considered. Tables 2 and 3 show the mentioned performance evaluation attributes and prediction factors with equations

Fig. 3 Extracted association rules using Apriori algorithm for QoE factors

```
Apriori
=======
Best rules found:

 1. Q8=VL Q19=L 283 ==> Q9=VL Q25=VL 280    conf:(0.99) < lift:(1.87)> lev:(0.24) [129] conv:(33.23)

 2. Q9=VL Q25=VL 288 ==> Q8=VL Q19=L 280    conf:(0.97) < lift:(1.87)> lev:(0.24) [129] conv:(15.32)

 3. Q5=G Q7=G 283 ==> Q1=G Q17=G 280    conf:(0.99) < lift:(1.71)> lev:(0.21) [116] conv:(29.84)

 4. Q5=G Q16=G Status=Succeed 314 ==> Q1=G Q2=G 280    conf:(1) < lift:(1.71)> lev:(0.21) [116]

 5. Q16=G Q15=G Q14=G 283 ==> Q17=VG Status=Succeed 280    conf:(0.99) < lift:(1.71)> lev:(0.21) [116]

 6. Q17=VL Q6=L Q4=L 314 ==> Status=Damaged conf:(0.89) < lift:(1.71)> lev:(0.21) [116] conv:(4.3)

 7. Q9=VL Q19=L 335 ==> Q8=VL Q25=VL 280    conf:(0.84) < lift:(1.62)> lev:(0.2) [106] conv:(2.89)

 8. Q8=VL Q25=VL 281 ==> Q9=VL Q19=L 280    conf:(0.99) < lift:(1.62)> lev:(0.2) [106] conv:(53.82)

 9. Academic Field =EC Q6=VL 348 ==> Q1=VL Q2=VL 298    conf:(1) < lift:(1.56)> lev:(0.2) [107]
    conv:(3.08)

10. Q3=VL Q4=VL Q2=VL 298 ==> Status=Damaged Employment status=EM 298    conf:(0.88) < lift:(1.56)>
    lev:(0.2) [107]
```

Table 2 Performance evaluation attributes to detect prediction parameters (Zhou et al. 2019b; Alam et al. 2021)

Feature	Description
True Positive (TP)	Shows the cases that have been correctly categorized
True Negative (TN)	Defines the QoE factors that have been wrongly classified as normal samples
False Positive (FP)	Indicates the normal instances that have been grouped as QoE factors
False Negative (FN)	Specifies the QoE instances that have been detected as normal

Table 3 The prediction's parameters of the supervised method (Xie et al. 2018; Chao et al. 2020)

Prediction parameters	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F-measure	$\frac{2\times Precision\times Recall}{Precision+Recall}$

concerning the confusion metrics which are usually employed for measuring the performance of machine learning classifiers.

K-fold cross-validation method is employed for reducing the bias of random choice of instances in the training step for assessing the accuracy of different classification methods. Typically k-fold cross-validation technique is a common procedure in which the folds are provided from the same portion of labels for creating dissimilar data sets (Shi et al. 2020b). Here, for performance evaluation of the applied classifiers, stratified k-fold cross-validation is employed with values between 5, 10, 15 and 20 for k-folds.

The results from different classification methods are presented in Fig. 4 to Fig. 8. The figures show the evaluation of the performance of the mentioned classification algorithms based on some different cross folds values. As exposed in figures, all the methods of classifications have reached approximately close and satisfactory outcomes according to the accuracy, precision, recall and F-measure values.

According to Fig. 4, the J48 + BinarySplits classifier showed an accuracy of 95.9% than other classifier algorithms. Likewise, hybrid J48 + BinarySplits algorithm has the highest quality regarding the results. Hence, it can result that in the QoE-based prediction model, the J48 + BinarySplits classifier has the maximum performance among the others.

Figure 5 illustrates the evaluation of precision factor to show how many selected QoE factors in all features of data set are relevant to detect satisfactory status of students. We observed that the hybrid J48 + BinarySplits classifier has the maximum level of the precision factor with 95.8% to detect relevant selected features for predicting satisfactory status of each student in evaluation of virtual education system. Also, SMO and IBK classifiers have 94.2% and

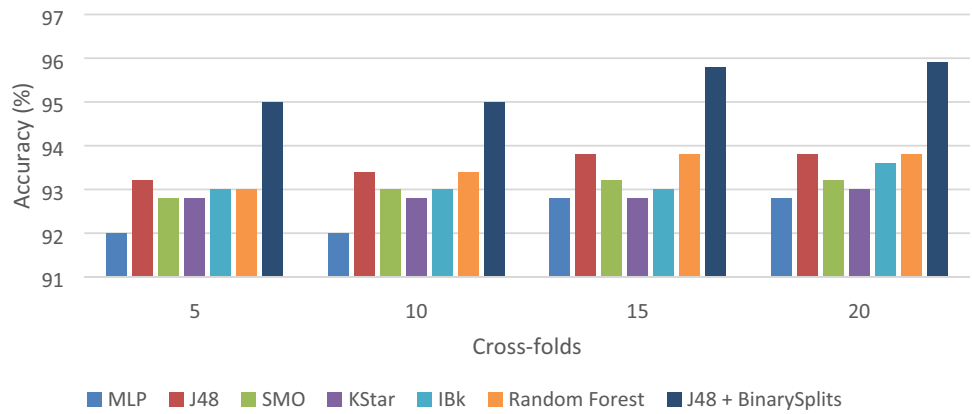**Fig. 4** The accuracy of each classifier algorithm with 20 cross-validation fold

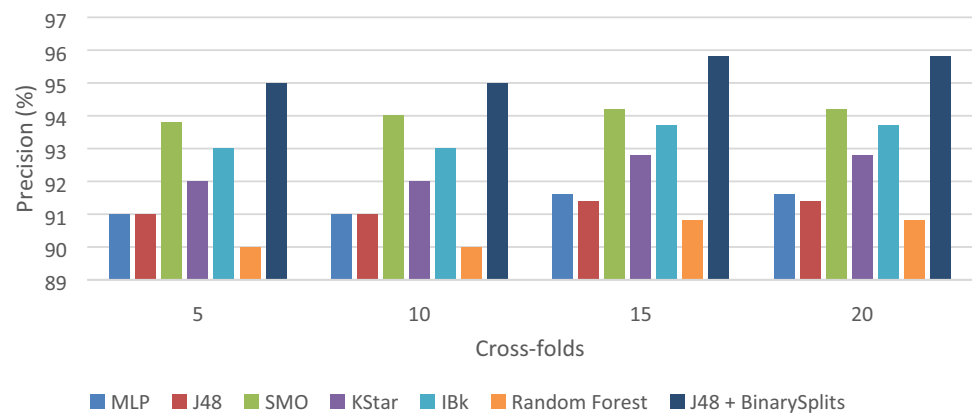**Fig. 5** The precision factor of each algorithm with 20 cross-validation folds

Fig. 6 The recall factor of each classifier algorithm based on QoE data set

93.7% for achieving maximum precision factor based on QoE data set.

To describe the performance of recall factor in Fig. 6, we conclude that the hybrid J48 + BinarySplits classifier has the maximum level of the recall metric with 96.8% to

select relevant features for predicting satisfactory status of each student in evaluation of virtual education system.

According to Fig. 7, by comparing the f-measure factor of the existing classifier algorithms, it is found that the hybrid J48 + BinarySplits classifier has the highest recognition f-measure, which can reach 96.2% than other

Fig. 7 The F-measure of each classifier with existing cross-validation folds
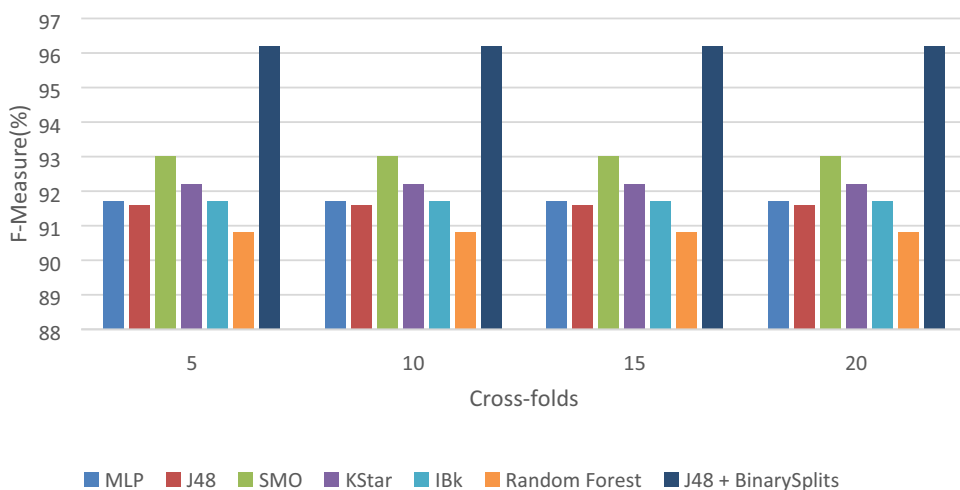
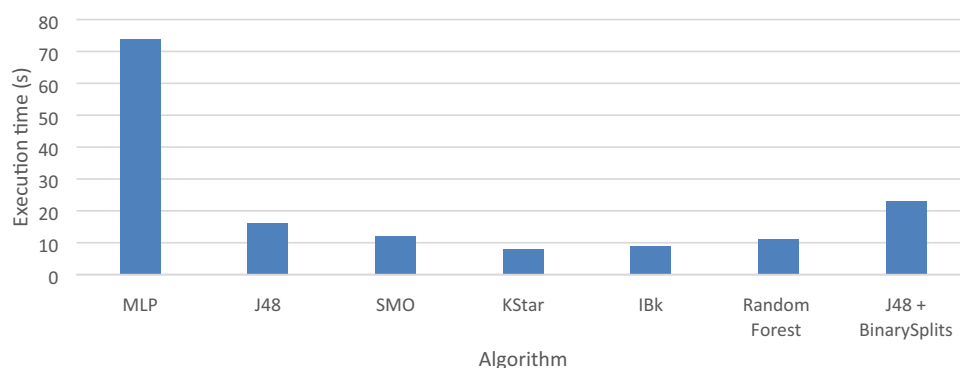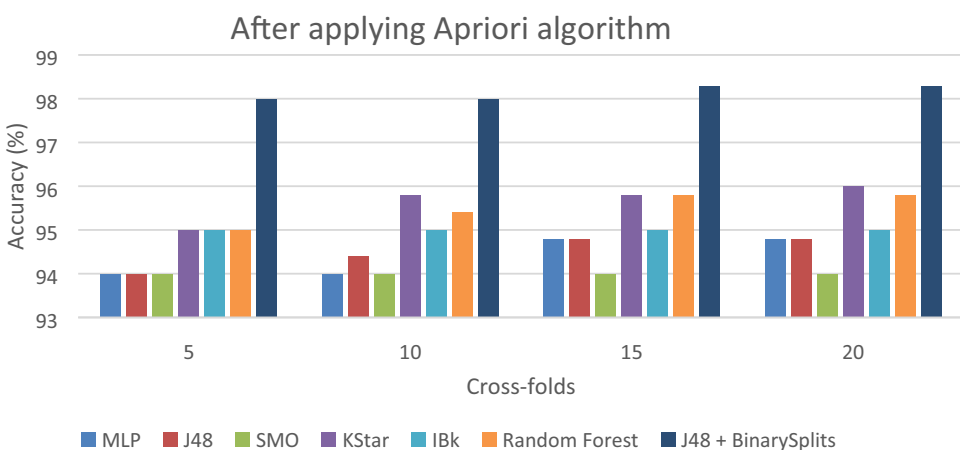**Fig. 8** Average execution time for applied classifier algorithms

Fig. 9 The accuracy of each classifier algorithm after applying Apriori algorithm

algorithms. Also, SMO classifier has 93% for evaluating f-measure factor.

Nevertheless, since the execution time of the employed classifiers is an important assessment factor, Fig. 8 shows that the KStar classifier has achieved the minimum execution time with a major difference comparing to the other

algorithms. The MLP has highest execution time than other classifiers to achieve the prediction results.

After applying Apriori algorithm on the data set to select high confident features, we have executed existing classification algorithms, respectively. Based on Fig. 9, the J48 + BinarySplits classifier showed an accuracy of 98.3%

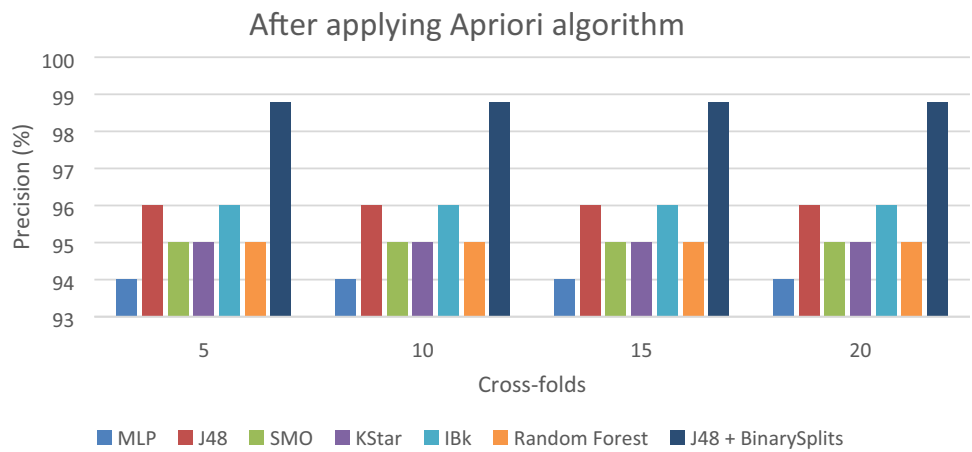**Fig. 10** The precision factor of each algorithm after applying Apriori algorithm

After applying Apriori algorithm

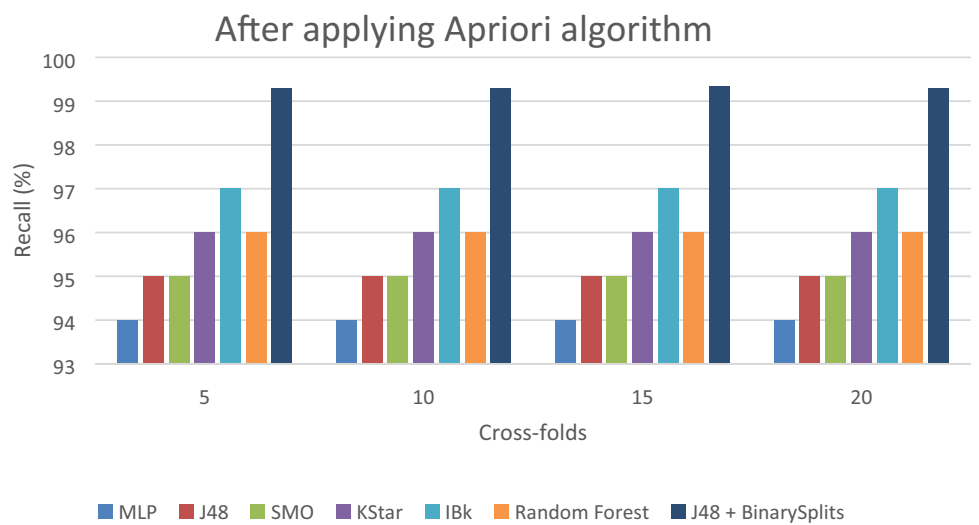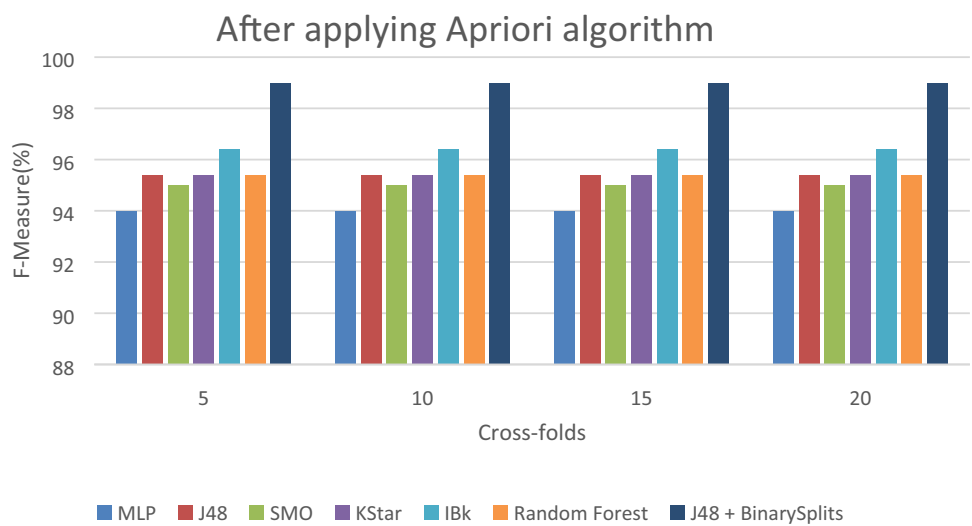**Fig. 11** The recall factor of each classifier algorithm after applying Apriori algorithm

After applying Apriori algorithm

Fig. 12 The F-measure of each classifier after applying Apriori algorithm
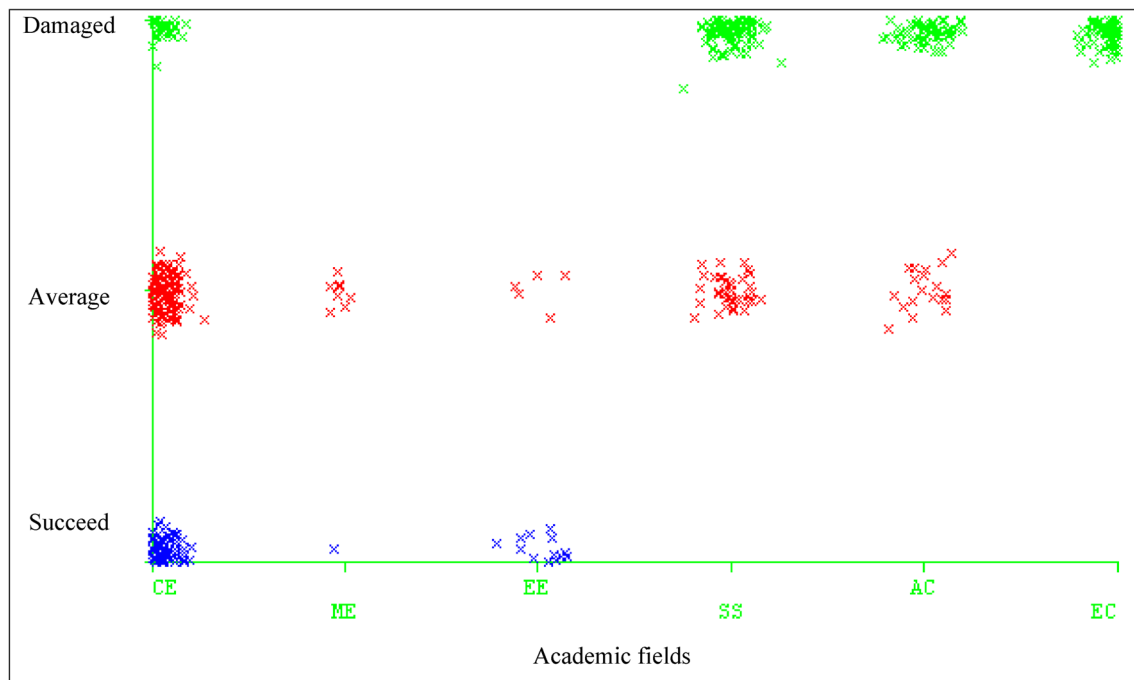

After applying Apriori algorithm

Fig. 13 Classification of existing academic fields based on level of satisfactory scores
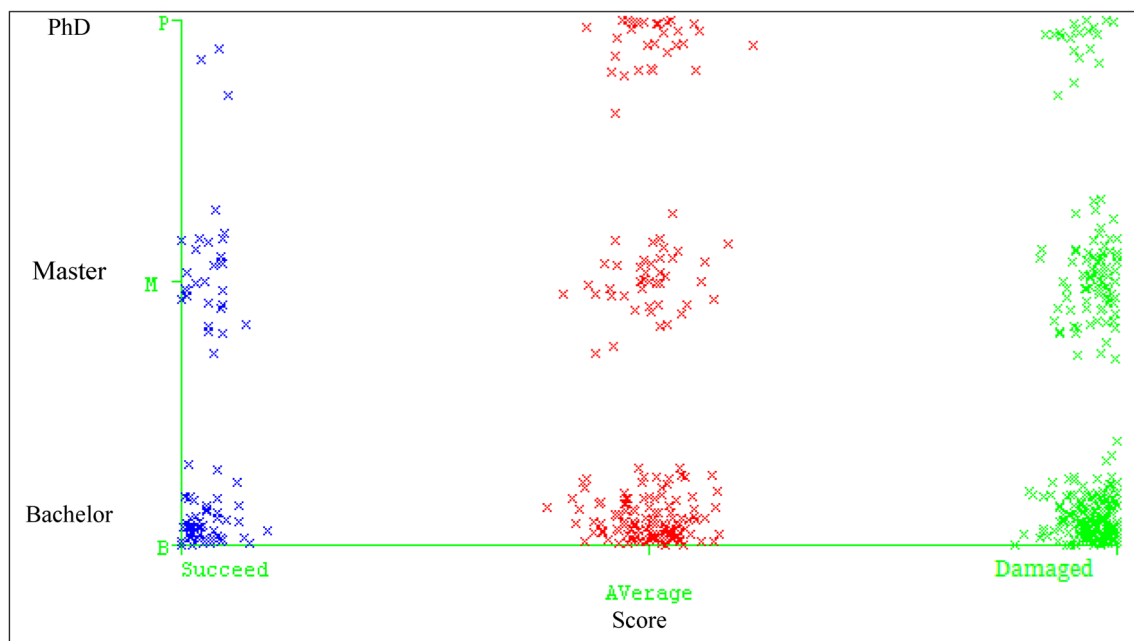

Fig. 14 Classification of existing academic degrees based on level of satisfactory scores

than other classifier algorithms when Apriori algorithm was applied to feature selection. Also, hybrid J48 + BinarySplits algorithm has the highest accuracy in the QoE-based prediction model.

Figure 10 illustrates the evaluation of precision factor to show how many selected QoE factors in all features of data set are relevant to detect satisfactory status of students. After applying Apriori algorithm, the hybrid J48 + BinarySplits classifier achieves to the highest level of the
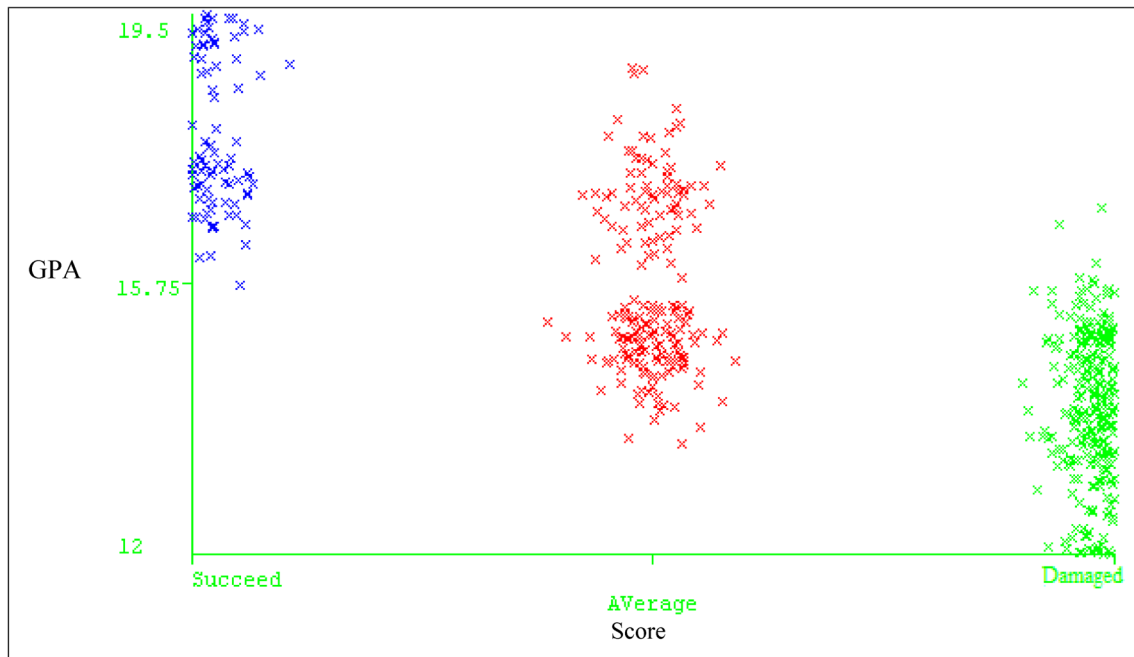
Fig. 15 Classification of GPA based on level of satisfactory scores

precision factor with 98.8% to detect relevant selected features for predicting satisfactory status of each student in evaluation of virtual education system. Also, J48 and IBK classifiers have 96% for achieving maximum precision factor based on QoE data set, respectively.

To describe the performance of recall factor in Fig. 11, we conclude that the hybrid J48 + BinarySplits classifier after performing Apriori algorithm has the maximum level of the recall metric with 99.3% to select relevant features for predicting satisfactory status of each student in evaluation of virtual education system.

According to Fig. 12, by comparing the f-measure factor of the existing classifier algorithms, it is found that the hybrid J48 + BinarySplits classifier with feature selection of Apriori algorithm has the highest recognition f-measure, which can reach 99% than other algorithms. Also, IBK classifier has 97% for evaluating f-measure factor.

According to the extracted association rules based on QoE factors in Sect. 3.2, we show statistical analysis on classified existing QoE factors. Figure 13 shows classification of existing academic fields based on level of satisfactory scores. Blue color describes class of succeed satisfaction, red color shows average satisfaction and green color depicts class of damaged satisfaction score. According to this classification, we observed that maximum number of succeed satisfaction related to computer engineering students. Little number of electrical engineering students have succeed satisfaction score in learning procedure of virtual education system. It is observable that

all economics students have damaged in their learning procedure of virtual education system. Also, social sciences and accounting students have negative performance in their learning procedure of virtual education system.

Figure 14 illustrates classification of existing academic degrees based on level of satisfactory scores. We have three academic degrees bachelor, master and PhD that have different density between satisfactory statuses including succeed, average and damaged. According to the classified academic degrees, bachelor students have damaged on virtual education system in COVID-19 side effects. Also, PhD students have little damage than master student on the e-learning environment based on QoE factors in virtual education system. Finally, Fig. 15 shows categorization of GPA scores based on level of satisfactory status we observed that some students with the GPA scores more than 15.75 have succeed on the virtual education system. Of course, many students have average satisfactory status for the virtual education system. Finally, students with the GPA scores lower than 15.75 have damaged using the virtual education system.

5 Conclusion

In this paper, a new QoE-based prediction model was proposed for evaluating performance of virtual education systems in COVID-19 pandemic. The proposed model used association rules mining and classification algorithms to

capture the direct and indirect behavioral aspects of virtual education system on satisfactory level of students in e-learning. The Apriori analysis evaluated important rules in QoE factors to specify student sentiments in the virtual education system. The positive relationship established between important QoE factors and satisfactory status of students that valuable for the improving virtual education system architecture in next semester. Then, some important classification algorithms have been applied to predict important metrics of behavioral aspects of students based on QoE factors in the virtual education system. The experimental results showed that our hybrid J48 + BinarySplits classifier after applying Apriori algorithm has an accuracy of 98.3%, precision of 98.8%, recall of 99.3% and f-measure of 99% than other classifier algorithms to predict maximum number of succeed satisfaction related to behavioral aspects of students based on QoE factors. As limitations of this research, we can suggest applying a meta-heuristic algorithm for feature selection of educational data preprocessing. In the future work, some meta-heuristic algorithms can applied to improve the feature selection strategy based on finding important behavioral aspects of QoE factors in virtual education system.

Declarations

Conflict of interest We declare that there is no conflict of interest between authors in the content of the manuscript.

References

Abedini M, Zhang C (2020) Performance assessment of concrete and steel material models in ls-dyna for enhanced numerical simulation, a state of the art review. Arch Comput Methods Eng 28:2921–2942. https://doi.org/10.1007/s11831-020-09483-5

Alam Z, Sun L, Zhang C, Su Z, Samali B (2021) Experimental and numerical investigation on the complex behaviour of the localised seismic response in a multi-storey plan-asymmetric structure. Struct Infrastruct Eng 17:86–102

Ashraf M, Zaman M, Ahmed M (2020) An intelligent prediction system for educational data mining based on ensemble and filtering approaches. Procedia Comput Sci 167:1471–1483

Aslam M (2019) Neutrosophic analysis of variance: application to university students. Complex Intell Syst 5:403–407

Assunção Flores M, Gago M (2020) Teacher education in times of COVID-19 pandemic in Portugal: national, institutional and pedagogical responses. J Educ Teach 46(4):507–516

Azzi I, Jeghal A, Radouane A, Yahyaouy A, Tairi H (2019) A robust classification to predict learning styles in adaptive E-learning systems. Educ Inf Technol 25:437–448

Bai Y, Wang S, Mou B, Wang Y, Skalomenos KA (2021) Bi-directional seismic behavior of steel beam-column connections with outer annular stiffener. Eng Struct 227:111443

Chao M, Kai C, Zhiwei Z (2020) Research on tobacco foreign body detection device based on machine vision. Trans Inst Meas Control 42(15):2857–2871

Charitopoulos A, Rangoussi M, Koulouriotis D (2020) On the use of soft computing methods in educational data mining and learning analytics research: a review of years 2010–2018. Int J Artif Intell Educ 30:371–430

Chen T, Peng L, Yin X, Rong J, Yang J, Cong G (2020) Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic. In: Healthcare, 200

Chen Y, Zheng W, Li W, Huang Y (2021) Large group Activity security risk assessment and risk early warning based on random forest algorithm. Pattern Recogn Lett 144:1–5

Daghestani LF, Ibrahim LF, Al-Towirgi RS, Salman HA (2020) Adapting gamified learning systems using educational data mining techniques. Comput Appl Eng Educ 28:568–589

Demuyakor J (2020) Coronavirus (COVID-19) and online learning in higher institutions of education: a survey of the perceptions of ghanaian international students in China. Online J Commun Media Technol 10:e202018

Díaz-Antón G, Pérez MA (2006) Towards an ontology of LMS-A conceptual framework. In: International conference on enterprise information systems, 161–164

Gholipour G, Zhang C, Mousavi AA (2020) Nonlinear numerical analysis and progressive damage assessment of a cable-stayed bridge pier subjected to ship collision. Marine Struct 69:102662

Huang B, Li C, Zhang Y, Ding W, Yang M, Yang Y, Zhai H, Xu X, Wang D, Debnath S, Jamil M, Li HN, Ali HM, Gupta MK, Said Z (2021) Advances in fabrication of ceramic corundum abrasives based on sol–gel process. Chin J Aeronaut 34(6):1–17

Isma'il M, Haruna U, Aliyu G, Abdulmumin I, Adamu S (2020) An autonomous courses recommender system for undergraduate using machine learning techniques. In: 2020 international conference in mathematics, computer engineering and computer science (ICMCECS), pp 1–6

Jiang D, Wang F, Lv Z, Mumtaz S, Al-Rubaye S, Tsourdos A et al (2021) QoE-aware efficient content distribution scheme for satellite-terrestrial networks. IEEE Trans Mobile Comput. https://doi.org/10.1109/TMC.2021.3074917

Li Y, Qiao L, Lv Z (2021) An optimized byzantine fault tolerance algorithm for consortium blockchain. Peer-to-Peer Network Appl 1–14

Liu J, Yi Y, Wang X (2020) Exploring factors influencing construction waste reduction: a structural equation modeling approach. J Clean Prod 276:123185

Lou R, Lv Z, Dang S, Su T, Li X (2021) Application of machine learning in ocean data. Multimed Syst. https://doi.org/10.1007/s00530-020-00733-x

Lou R, Wang W, Li X, Zheng Y, Lv Z (2021) Prediction of ocean wave height suitable for ship autopilot. IEEE Trans Intell Transp Syst. https://doi.org/10.1109/TITS.2021.3067040

Lv Z, Chen D, Li J (2021) Novel system design and implementation for the smart city vertical market. IEEE Commun Mag 59(4):126–131

Lv Z, Lou R, Li J, Singh AK, Song H (2021a) Big data analytics for 6G-enabled massive internet of things. IEEE Internet Things J 8(7):5350–5359

Lv Z, Chen D, Lou R, Alazab A (2021b) Artificial intelligence for securing industrial-based cyber–physical systems. Futur Gener Comput Syst 117:291–298

Mitrofanova YS, Sherstobitova AA, Filippova OA, Modeling smart learning processes based on educational data mining tools. In: Smart education and e-learning 2019, Singapore, pp 561–571

Qiang S (2019) Formulation of physical education and training program based on multidimensional education data mining. Cluster Comput 22:5017–5023

Shi K, Zhong S, Tang Y, Cheng J (2020a) Hybrid-driven finite-time $H\infty$ sampling synchronization control for coupling memory

complex networks with stochastic cyber attacks. Neurocomputing 387:241–254

Shi K, Wang J, Tang Y, Zhong S (2020b) Reliable asynchronous sampled-data filtering of T-S fuzzy uncertain delayed neural networks with stochastic switched topologies. Fuzzy Sets Syst 381:1–25

Troussas C, Krouska A, Sgouropoulou C (2020) Dynamic detection of learning modalities using fuzzy logic in students' interaction activities. In: International conference on intelligent tutoring systems, pp 205–213

Wang L, Peng Y, Xie Y, Chen B, Du Y (2021) A new iteration regularization method for dynamic load identification of stochastic structures. Mech Syst Signal Process 156:107586

Wu H, Lu Z, Pan L, Xu R, Jiang W (2009) An improved apriori-based algorithm for association rules mining. In: 2009 sixth international conference on fuzzy systems and knowledge discovery, pp 51–55

Wu Z, Cao J, Wang Y, Wang Y, Zhang L, Wu J (2018) hPSD: a hybrid PU-learning-based spammer detection model for product reviews. IEEE Trans Cybern 50:1595–1606

Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J (2018) Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR mHealth uHealth 6:e94

Xu M, Li C, Chen Z, Wang Z, Guan Z (2018) Assessing visual quality of omnidirectional videos. IEEE Trans Circuits Syst Video Technol 29:3516–3530

Yin Q, Li C, Dong L, Bai X, Zhang Y, Yang M et al (2021) Effects of physicochemical properties of different base oils on friction coefficient and surface roughness in MQL milling AISI 1045. Int J Precis Eng Manuf-Green Technol. https://doi.org/10.1007/s40684-021-00318-7

Yu Z, Amin SU, Alhussein M, Lv Z (2021) Research on disease prediction based on improved DeepFM and IoMT. IEEE Access 9:39043–39054

Zhang C, Wang H (2020) Swing vibration control of suspended structures using the active rotary inertia driver system: theoretical modeling and experimental verification. Struct Control Health Monitor 27:e2543

Zhang C, Gholipour G, Mousavi AA (2019) Nonlinear dynamic behavior of simply-supported RC beams subjected to combined impact-blast loading. Eng Struct 181:124–142

Zhang J, Wang M, Tang Y, Ding Q, Wang C, Huang X et al (2021) Angular velocity measurement with improved scale factor based on a wideband-tunable optoelectronic oscillator. IEEE Trans Instrum Meas 70:1–9

Zheng L (2020) Research on E-commerce potential client mining applied to apriori association rule algorithm. In: 2020 international conference on intelligent transportation, big data & smart city (ICITBS), pp 667–670

Zhou M, Wang Y, Liu Y, Tian Z (2019a) An information-theoretic view of WLAN localization error bound in GPS-denied environment. IEEE Trans Veh Technol 68:4089–4093

Zhou Y, Tian L, Zhu C, Jin X, Sun Y (2019b) Video coding optimization for virtual reality 360-degree source. IEEE J Sel Topics Signal Process 14:118–129

Zhou M, Li X, Wang Y, Li S, Ding Y, Nie W (2020) 6G multi-source information fusion based indoor positioning via gaussian kernel density estimation. IEEE Internet Things J

Zhou M, Li Y, Tahir MJ, Geng X, Wang Y, He W (2021) Integrated statistical test of signal distributions and access point contributions for Wi-Fi indoor localization. IEEE Trans Veh Technol. https://doi.org/10.1109/TVT.2021.3076269

Zhu S (2019) Research on data mining of education technical ability training for physical education students based on Apriori algorithm. Cluster Comput 22:14811–14818

Zhu L, Kong L, Zhang C (2020) Numerical study on hysteretic behaviour of horizontal-connection and energy-dissipation structures developed for prefabricated shear walls. Appl Sci 10:1240

Zuo C, Chen Q, Tian L, Waller L, Asundi A (2015) Transport of intensity phase retrieval and computational imaging for partially coherent fields: the phase space perspective. Opt Lasers Eng 71:20–32

Zuo C, Sun J, Li J, Zhang J, Asundi A, Chen Q (2017) High-resolution transport-of-intensity quantitative phase microscopy with annular illumination. Sci Rep 7:1–22