

PROCEEDINGS OF

2017 International Conference on Data and Software Engineering (ICoDSE)

 **NOVEMBER**
1st - 2nd, 2017

 **Aryaduta Palembang Hotel**
Palembang, Indonesia



"Delivering Data-Intensive Software System"

Organized by:



Technical co-Sponsored by:



Sponsored by:



ISBN: 978-1-5386-1448-8

Proceedings of

**2017 International Conference
on Data and Software Engineering (ICoDSE)**

Aryaduta Palembang Hotel, Palembang, Indonesia
November 1st - 2nd, 2017

General Chair's Message

Welcome to ICoDSE 2017,

It is with a great pleasure that we extend our warm welcome to all the participants of the 4th International Conference on Data and Software Engineering 2017. This is an annual conference which started in 2014. In the beginning the conference was organized by Institut Teknologi Bandung. In 2015, the conference is held by Gadjah Mada University. Last year, the conference is held by Udayana University. This year, the Universitas Sriwijaya together with Knowledge and Software Engineering Research Group (from School of Electrical Engineering and Informatics, Institut Teknologi Bandung) is organizing this conference and technically co-sponsored by IEEE Indonesia Section.

The theme of this year conference is ***"Delivering Data-Intensive Software System"***. The ICoDSE 2017 aims to bridge the knowledge between Academic, Industry and Community. This is a forum for researchers, scientists and engineers from all over the world to exchange ideas and discuss the latest progress in their fields. The two-day conference highlights recent and significant advances in research and development in the field of Data/Knowledge and Software Engineering. The conference welcomes contributions from two tracks, namely Research Track and Industrial Track.

This year, we have received 110 submissions from 6 countries around the world, namely Indonesia, Malaysia, P.R. China, South Korea, United Arab Emirates, and United States of America. All submissions were peer-reviewed (blind) by at least 2 reviewers drawn from external reviewers and the committees, and 49 papers are accepted for presentations.

Finally, as the General Chair of the Conference, I would like to express my deep appreciation to all members of the Steering Committee, Technical Programme Committee, Organizing Committee and Reviewers who have devoted their time and energy for the success of the event. We also would like to thank our sponsor Intens, Lintasarta, BukitAsam, Teluu, and BukaLapak.

For all participants, I wish you an enjoyable conference.

Prof. Saparudin, Ph.D
General Chair of 2017 International Conference on Data and Software Engineering (ICoDSE)

Committee

General Chair

Saparudin Universitas Sriwijaya – Indonesia

Co–Chair

Reza Firsandaya Malik Universitas Sriwijaya – Indonesia
Yudistira D. W. Asnar Institut Teknologi Bandung – Indonesia

Steering Committee

A Min Tjoa Vienna University of Technology – Austria
Benhard Sitohang Institut Teknologi Bandung – Indonesia
Ford Lumban Gaol IEEE – Indonesia
Iping Supriana Institut Teknologi Bandung – Indonesia
Richard Lai La Trobe University – Australia
Siti Nurmaini Universitas Sriwijaya – Indonesia
Zainuddin Nawawi Universitas Sriwijaya – Indonesia

Technical Program Committee

Chair :

Muhammad Zuhri Catur C. Institut Teknologi Bandung – Indonesia

Member :

Agung Trisetyarso Bina Nusantara University – Indonesia
Ayu Purwarianti IEEE Indonesia Education Activity Committee
Bayu Hendradjaya Institut Teknologi Bandung – Indonesia
Bernaridho Hutabarat UPI YAI – Indonesia
David Taniar Monash University – Australia
Dwi H. Widiyantoro Institut Teknologi Bandung – Indonesia
Fazat Nur Azizah Institut Teknologi Bandung – Indonesia
GA Putri Saptawati Institut Teknologi Bandung – Indonesia
I Ketut Gede Darma Putra Udayana University – Indonesia
I Putu Agung Bayupati Udayana University – Indonesia
Iwan Pahendra Universitas Sriwijaya – Indonesia
Jamaiah Yahaya The National University of Malaysia – Malaysia
Khabib Mustofa Universitas Gajah Mada – Indonesia
Linawati Udayana University – Indonesia
Made Sudiana Mahendra Udayana University – Indonesia
Maman Fathurrohman Sultan Ageng Tirtayasa University – Indonesia
MM Inggriani Liem Institut Teknologi Bandung – Indonesia
Muhammad Asfand-e-yar Bahria University – Pakistan
Noor Maizura Mohamad Noor University Malaysia Terengganu – Malaysia
Robert P. Biuk–Aghai University of Macau – China
Saiful Akbar Institut Teknologi Bandung – Indonesia
Soon Chung Wright State University – USA
Wikan D. Sunindyo Institut Teknologi Bandung – Indonesia
Yudistira D. W. Asnar Institut Teknologi Bandung – Indonesia

Organizing Committee

Deris Stiawan Universitas Sriwijaya – Indonesia
Erwin Universitas Sriwijaya – Indonesia
Fazat Nur Azizah Institut Teknologi Bandung – Indonesia
Fitra Arifiansyah Institut Teknologi Bandung – Indonesia

Hari Purnama
Hastie Audytra
Latifa Dwiyanti
M. Fachrurrozi

Institut Teknologi Bandung – Indonesia
Universitas Sriwijaya – Indonesia
Institut Teknologi Bandung – Indonesia
Universitas Sriwijaya – Indonesia

Reviewers

<i>Adi Mulyanto</i>	Institut Teknologi Bandung – Indonesia
<i>Ayu Purwarianti</i>	Institut Teknologi Bandung – Indonesia
<i>Bayu Hendradjaya</i>	Institut Teknologi Bandung – Indonesia
<i>Benhard Sitohang</i>	Institut Teknologi Bandung –Indonesia
<i>Bernaridho Hutabarat</i>	UPI YAI – Indonesia
<i>Dade Nurjanah</i>	Telkom University – Indonesia
<i>Dessi Lestari</i>	Institut Teknologi Bandung – Indonesia
<i>Dwi Widyantoro</i>	Institut Teknologi Bandung – Indonesia
<i>Ermatita</i>	Universitas Sriwijaya – Indonesia
<i>Erwin</i>	Universitas Sriwijaya – Indonesia
<i>Fajar Ekaputra</i>	TU Wien – Austria
<i>Fazat Nur Azizah</i>	Institut Teknologi Bandung – Indonesia
<i>Firdaus</i>	Universitas Sriwijaya – Indonesia
<i>Hira Laksmiwati</i>	Institut Teknologi Bandung – Indonesia
<i>Inggriani Liem</i>	Institut Teknologi Bandung – Indonesia
<i>Jamaiah Yahaya</i>	The National University of Malaysia – Malaysia
<i>M. Fachrurrozi</i>	Universitas Sriwijaya – Indonesia
<i>Masayu Leylia Khodra</i>	Institut Teknologi Bandung – Indonesia
<i>Mewati Ayub</i>	Maranatha Christian University – Indonesia
<i>Muhammad Asfand-e-yar</i>	Bahria University – Pakistan
<i>Muhammad Zuhri Catur C.</i>	Institut Teknologi Bandung – Indonesia
<i>Putri Saptawati</i>	Institut Teknologi Bandung – Indonesia
<i>Reza Firsandaya Malik</i>	Universitas Sriwijaya – Indonesia
<i>Richard Lai</i>	La Trobe University – Australia
<i>Riza Satria Perdana</i>	Institut Teknologi Bandung – Indonesia
<i>Robert Biuk-Aghai</i>	University of Macau – Macao
<i>Saiful Akbar</i>	Institut Teknologi Bandung – Indonesia
<i>Soon Chung</i>	Wright State University – USA
<i>Sukrisno Mardiyanto</i>	Institut Teknologi Bandung – Indonesia
<i>Tricya E. Widagdo</i>	Institut Teknologi Bandung – Indonesia
<i>Wikan Sunindyo</i>	Institut Teknologi Bandung – Indonesia
<i>Yani Widayani</i>	Institut Teknologi Bandung – Indonesia
<i>Yoppy Sazaki</i>	Universitas Sriwijaya – Indonesia
<i>Yudistira Asnar</i>	Institut Teknologi Bandung – Indonesia

Table of Content

Data and Information Engineering Track

A Classification of Sequential Patterns for Numerical and Time Series Multiple Source Data - A Preliminary Application on Extreme Weather Prediction

Regina Yulia Yasmin, Andi Eka Sakya and Untung Merdijanto

A Statistical and Rule-Based Spelling and Grammar Checker for Indonesian Text

Asanilta Fahda and Ayu Purwarianti

An Effective Random Forest Model of Intrusion Detection System in IoT Network

Rifkie Primartha and Bayu Adhi Tama

An Extensible Tool for Spatial Clustering

Venny Larasati Ayudiani and Saiful Akbar

Analysis of Weakness of Data Validation from Social CRM

Ali Ibrahim, Ermatita, Saparudin and Zefta Adetya

Aspect-Sentiment Classification in Opinion Mining using the Combination of Rule-Based and Machine Learning

Zulva Fachrina and Dwi H. Widyantoro

Cells Identification of Acute Myeloid Leukemia AML M0 and AML M1 using k-Nearest Neighbour Based on Morphological Images

Esti Suryani, Wiharto, Sarngadi Palgunadi and Yudha Rizki Putra

Classification and Clustering to Identify Spoken Dialects in Indonesian

Jacqueline Ibrahim and Dr.Eng. Dessi Puji Lestari, S.T., M.T.

Comparison of Optimal Path Finding Techniques for Minimal Diagnosis in Mapping Repair

Inne Gartina Husein, Saiful Akbar, Benhard Sitohang and Fazat Azizah

Comparison of Similarity Measures in HSV Quantization for CBIR

Jasman Pardede, Benhard Sitohang, Saiful Akbar and Masayu Leylia Khodra

Content Based Image Retrieval for Multi-Objects Fruits Recognition using k-Means and k-Nearest Neighbor

Erwin, M. Fachrurrozi, Ahmad Fiqih, Bahardiansyah Rua Saputra, Rachmad Algani and Anggina Primanita

Content-based Clustering and Visualization of Social Media Text Messages

Sydney A. Barnard, Soon M. Chung and Vincent A. Schmidt

Density Level Detection of The Use of Transportation Mode Based on GPS Data With Data Mining Technology

Irrevaldy and Gusti Ayu Putri Saptawati

Enhancing Clustering Quality of Fuzzy Geographically Weighted Clustering using Ant Colony Optimization

Arie Wahyu Wijayanto, Siti Mariyah and Ayu Purwarianti

Extensible Analysis Tool for Trajectory Pattern Mining

Vanya Deasy Safrina and Saiful Akbar

FVEC-SVM for Opinion Mining on Indonesian Comments of YouTube Video

Ekki Rinaldi and Aina Musdholifah

Graph Analysis on ATCS Data in Road Network for Congestion Detection

Apip Ramdlani, G.A. Putri Saptawati and Yudistira Dwi Wardhana Asnar

Graph Clustering using Dirichlet Process Mixture Model (DPMM)

Imelda Atastina

Implementation of LANDMARC Method with Adaptive K-NN Algorithm on Distance Determination Program in UHF RFID System

Ahmad Fali Oklilas, Fithri Halim Ahmad and Dr.Reza Firsandaya Malik

Imputation of Missing Value Using Dynamic Bayesian Network for Multivariate Time Series Data

Steffi Pauli Susanti and Fazat Nur Azizah

Rapid Data Stream Application Development Framework

Wilhelmus Andrian Tanujaya, Muhammad Zuhri Catur Candra and Saiful Akbar

Scheme Mapping for Relational Database Transformation to Ontology: A Survey

Paramita Mayadewi, Benhard Sitohang and Fazat N. Azizah

Strategic Intelligence Model in Supporting Brand Equity Assessment

Agung Aldhiyat and Masayu Leylia Khodra

The Grouping Of Facial Images Using Agglomerative Hierarchical Clustering To Improve The CBIR Based Face Recognition System

Muhammad Fachrurrozi, Saparudin, Erwin, Clara Fin Badillah, Junia Erlina, Mardiana and Auzan Lazuardi

Traffic Speed Prediction From GPS Data of Taxi Trip Using Support Vector Regression

Dwina Satrinia and G.A. Putri Saptawati

Software Engineering Track

A Generic Tool For Modelling And Simulation Of Fire Propagation Using Cellular Automata

Taufiqurrahman and Saiful Akbar

Arduviz, A Visual Programming IDE for Arduino

Adin Baskoro Pratomo and Riza Satria Perdana

Designing Dashboard Visualization for Heterogeneous Stakeholders (Case Study: ITB Central Library)

Tjan Marco Orlando and Dr. techn Wikan Danar Sunindyo

Evaluation of Greedy Perimeter Stateless Routing Protocol On Vehicular Ad Hoc Network in Palembang City

Reza Firsandaya Malik, Muhammad Sulkhan Nurfatih, Huda Ubaya, Rido Zulfahmi and Erizal Sodikin

Hybrid Attribute and Personality based Recommender System for Book Recommendation

'Adli Ihsan Hariadi and Dade Nurjanah

Hybrid Recommender System Using Random Walk with Restart for Social Tagging System

Arif Wijonarko, Dade Nurjanah and Dana Sulistyio Kusumo

Identification Process Relationship of Process Model Discovery based on Workflow-Net

Ferdi Rahmadi and Gusti Ayu Putri Saptawati

Implementation of Regular Expression (Regex) on Knowledge Management System

Ken Ditha Tania and Bayu Adhi Tama

Interaction Perspective in Mobile Banking Adoption: The Role of Usability and Compatibility

Hotna Marina Sitorus, Rajesri Govindaraju, Iwan I. Wiratmadja and Iman Sudirman

Interaction Quality and The Influence on Offshore IT Outsourcing Success

Yogi Yusuf Wibisono, Rajesri Govindaraju, Dradjad Irianto and Iman Sudirman

Minutia Cylinder Code-based Fingerprint Matching Optimization using GPU

Muhamad Visat Sutarno and Achmad Imam Kistijantoro

Modelling Online Assessment in Management Subjects Through Educational Data Mining

Mewati Ayub, Hapnes Toba, Maresha Caroline Wijanto and Steven Yong

Moodle Plugins for Quiz Generation Using Genetic Algorithm

Muhammad Rian Fakhruy and Yani Widayani, S.T., M.T.

On the Implementation of Search Based Approach to Mutation Testing

Mohamad Tuloli, Benhard Sitohang and Bayu Hendradjaya

Predicting Defect Resolution Time using Cosine Similarity

Pranjal Ambardekar, Anagha Jamthe and Mandar Chincholkar

Profile Hidden Markov Model for Malware Classification

Ramandika Pranamulia, Yudistira Asnar and Riza Satria Perdana

Rule Generator for IPS by Using Honeypot

Daniel Silalahi, Yudistira Asnar and Riza Satria Perdana

Semi-Automated Data Publishing Tool for Advancing The Indonesian Open Government Data Maturity Level (Case Study: Badan Pusat Statistik Indonesia)

Chairuni Aulia Nusapati and Wikan Danar Sunindyo

Software Quality Requirement Analysis of Context Aware "Family Tracking Mobile Application" on Cross Platform with Hybrid Approach

Anggy Trisnadoli and Indah Lestari

The Development of Data Collection Tool on Spreadsheet Format

Feryandi Nurdiantoro, Yudistira Asnar and Tricya Esterina Widagdo

The Effectiveness of Using Software Development Methods Analysis by The Project Timeline in an Indonesian Media Company

Putri Sanggabuaana Setiawan, Muhammad Ikhwan Jambak and Muhammad Ihsan Jambak

Two-Step Graph-Based Collaborative Filtering Using User and Item Similarity: Case Study of e-Commerce Recommender Systems

Aghny Arisya Putra, Rahmad Mahendra, Indra Budi and Qorib Munajat

Utility Function Based-Mixed Integer Nonlinear Programming (MINLP) Problem Model of Information Service Pricing Schemes

Robinson Sitepu, Fitri Maya Puspita and Shintya Apriliyani

Web Application Fuzz Testing

Ivan Andrianto, M. M. Inggriani Liem and Yudistira Dwi Wardhana Asnar

Modelling Online Assessment in Management Subjects through Educational Data Mining

Mewati Ayub¹, Hapnes Toba², Maresha Caroline Wijanto³, Steven Yong⁴

Department of Informatics Engineering
Faculty of Information Technology, Maranatha Christian University
Bandung, Indonesia

mewati.ayub@it.maranatha.edu¹, hapnestoba@it.maranatha.edu², maresha.cw@it.maranatha.edu³, steven.yong@outlook.com⁴

Abstract—Educational data mining (EDM) has been used widely to investigate data that come from a learning process, including blended learning. This study explores educational data from a Learning Course Management System (LMS) and academic data in two courses of Management Study Program, Faculty of Economics at Maranatha Christian University, which are Change Management (CM) in undergraduate program and Creative Leadership (CL) in master degree program as case studies. The main aim of this research is to provide feedback for the learning process through the LMS in order to improve students' achievement. EDM methods used are association rule mining and J48 classification. The results of association rule mining are two sets of interesting rules for the CM course and three sets of rules for CL course. Using J48 classification, two J48 pruned trees are obtained for each course. Based on those results, some suggestions are proposed to enhance the LMS and to encourage students' involvement in blended learning.

Keywords— *blended learning; educational data mining; association rules; J48 classification*

I. INTRODUCTION

The growth of Internet utilization in education in recent years has encouraged the development of blended learning, a variant of e-learning, which merges face-to-face instructions with technology-mediated instruction [1], [2]. Using blended learning, students have opportunities to obtain knowledge in a more convenient manner. Interaction of student and lecturer in blended learning could generate a huge amount of data, that recorded in a learning course management system [3]. Educational data mining (EDM) could be used to explore such data to acquire hidden knowledge. The knowledge can further be used to improve the system [4], [5].

This study explores EDM from a learning course management system (LMS) and academic data in the courses of Management Study Program, Faculty of Economics at Maranatha Christian University. The courses adopt a blended learning system with full face-to-face instruction. The main objective of this research is to provide feedback for the learning process through the LMS in order to improve students' achievement. As case studies, this research investigates data from two courses of human resources management subjects. The first course is Change Management

(CM) in undergraduate program and the second is Creative Leadership (CL) in master degree program. The findings from EDM of these data would be proposed to enhance the system.

This research is an extension of previous study [6], the objective of which is to analyze students' activity in programming subjects through blended learning in order to enhance students' achievement in their study. There are some differences between this work and [6]. Data set of the prior work derived from one academic year, while data set of this work came from three academic years. Subject of the prior work was programming subject, while subject of this work was social subject. Students online activities in the prior work were viewing resource, doing exercise, and attempting quiz, while students' online activities in this work were attempting quiz and attempting examination. EDM techniques used in this study are not only association rules, but also decision tree classification.

Based on [11], there are some critical factors affecting students in learning using the LMS, such as: learner computer anxiety, instructor attitude toward e-learning, e-learning course flexibility, e-learning course quality, perceive usefulness, perceived ease of use, and diversity in assessment. In [12], the critical factors include the following aspects: learner's characteristics, instructor's characteristics, institution and service quality, infrastructure and system quality, course and information quality, and extrinsic motivation.

We are focusing our study in the following research question: what kind of students' activities need to be emphasized to improve engagement in a blended learning environment? We hypothesize that our existing LMS can only attracts students in limited number of (obligated) activities which lead them to achieve a certain level of final grade, but not really contributes in the learning process itself.

Online assessment data, which is extracted from the LMS, will be analyzed by using association rules and classification techniques. Association rules mining is consider as a good tool to obtain general rules which indicate significant students' activities and their achievement in a learning process. Further, classification technique, based on tree induction, is use to analyze the robustness of a generic rule, which would contribute to the LMS' features enhancement. The overall

results will be used to generate several enhancement features which could improve students' involvement and enthusiasm in the learning process.

II. LITERATURE STUDY

In Romero [7] and Baker [8], there are some data mining techniques in educational data mining work, namely prediction, clustering, relationship mining, distillation of data for human judgment, and discovery with models. In this study, the data mining methods used for EDM are association rule mining and classification for prediction.

A. Association Rule

An association rule is written as an implication of the form $A \Rightarrow B$ where $A \subset X, B \subset X, A \neq \emptyset, B \neq \emptyset, A \cap B = \emptyset$, and $X = \{x_1, x_2, \dots, x_n\}$ is an item set. D is a set of transactions T where $T \subseteq X, T \neq \emptyset$. The rule $A \Rightarrow B$, has support (s), which defined as a probability that a transaction T contains $A \cup B$. Confidence (c) of the rule is defined as a probability that a transaction having A also contains B . A strong association rule is a rule that satisfies a minimum support threshold and a minimum confidence threshold [9].

The generation of rules in association rules mining consists of two step process [9]. The first process is finding all frequent item sets based on minimum support threshold. The next process is producing strong association rules from frequent item sets that complied minimum support and minimum confidence threshold.

Lift is a correlation value to measure importance of a rule. For rule $A \Rightarrow B$, lift is defined as [9]:

$$lift(A, B) = \frac{P(A \cup B)}{P(A)P(B)} \quad (1)$$

A lift value less than 1 indicates that the occurrence of A is negatively correlated with the occurrence of B . A lift value greater than 1 indicates that A and B is positively correlated. However, a lift value equal to 1 indicates that there is no correlation between A and B [9].

B. Classification

The classification technique is one of predictive data mining tasks utilized to classify hidden data [9], [10]. The classification model is generated in a learning step to analyze a training data. The next step is to apply the model to predict a new data class. This study used a decision tree induction as a classification method, that is J48 classification. The J48 is Weka's implementation of C4.5 decision tree learner as an improved version, that called C4.5 revisions 8 [10].

III. METHODOLOGY

The research methodology used in this study consists of following steps :

A. Data preparation

Data set was extracted from two courses of human resources management subjects, which are: CM course in undergraduate program and CL course in magister program. Both courses are conducted as blended learning courses. The courses combine face-to-face instructions with a learning management system (LMS). The LMS is used to perform online quizzes and online examination. The data set was extracted from even semester in three academic year, 2014/2015, 2015/2016, and 2016/2017 for both courses.

B. Data set of CM Course

For CM course, the data set has been extracted from 234 students. Each piece of student data consists of personal data, online quizzes scores, online examination scores, mid semester final scores, end semester final scores, course final score, and activity level. The activity level is counted as the frequency of students participations in quizzes. Each online quiz was conducted weekly from beginning until end semester, so there were 14 online quizzes. Online quizzes and online examinations combined with written examinations would determine final scores for mid semester and end semester. Course final grade would be ascertained by mid semester final grade, end semester final grade, and final assignment grade. In this case, final assignment grade is not involved in the data set, because online activities are not included in assessing assignment.

As shown in Table I, a group of attributes has been selected for EDM. These attributes are :

- online quizzes grades, divided into quizzes grade before mid semester and quizzes grade before end semester
- online examination grades, consisted of online mid semester exam and online end semester exam
- examination final grades, consisted of mid semester final grade and end semester final grade
- activity level, transformed from participation frequency in quizzes
- course final grade

C. Data set of CL Course

For CL course, the data set has been extracted from 180 students. Each piece of student data consists of personal data, online quizzes scores, online examination scores, mid semester final scores, end semester final scores, assignments final scores, course final score, and activity level. The activity level is counted as the frequency of students' participations in quizzes. Quizzes were conducted as many as three until nine quizzes in the semester. Course final grade would be ascertained by mid semester final grade, end semester final grade, and final assignment grade. In this case, final assignment grade is involved in the data set, because online activities are included in assessing assignment.

As shown in Table II, a group of attributes has been selected for EDM. These attributes are:

- online quizzes grades, transformed from the average of quizzes scores

- online examination grades, consisted of online mid semester exam and online end semester exam
- examination grades, consisted of mid semester exam grade and final exam grade
- assignment final grades, calculated from online quizzes scores and assignments scores
- activity level, transformed from participation frequency in quizzes
- course final grade

TABLE I.
STUDENT'S DATA SET FOR CM COURSE

Attribute Name	Description	Possible Values
GradePreMid	Online quizzes grade before mid semester	[Excellent, Good, Fair, BelowAvg, Poor]
GradeMidOL	Online mid exam grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeMidF	Mid semester final grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradePreFinal	Online quizzes grade before end semester	[Excellent, Good, Fair, BelowAvg, Poor]
GradeFinalOL	Online final exam grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeFinal	End semester final grade	[Excellent, Good, Fair, BelowAvg, Poor]
ActivityQ	Activity level	[Low, Medium, High]
GradeC	Course final grade	[Excellent, Good, Fair, BelowAvg, Poor]

TABLE II.
STUDENT'S DATA SET FOR CL COURSE

Attribute Name	Description	Possible Values
GradeMidOL	Online mid exam grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeMidF	Mid semester exam grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeFinalOL	Online final exam grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeFinal	End semester Final grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeQ	Online quiz grade	[Excellent, Good, Fair, BelowAvg, Poor]
GradeAss	Assignment final grade	[Excellent, Good, Fair, BelowAvg, Poor]
ActivityQ	Activity level	[Low, Medium, High]
GradeC	Course final grade	[Excellent, Good, Fair, BelowAvg, Poor]

D. Data mining techniques

In this study, investigation on the students' data set was conducted to discover whether the LMS could be used effectively in learning process to enhance the students' academic achievement. The experiments were performed twice, one for the students' data set of the CM course and the

other for the students' data set of the CL course. For each data set, exploration utilized association rule mining and tree classification mining.

IV. RESULT AND DISCUSSION

The data mining tools used during the experiments is WEKA version 3.8.1. To determine whether a rule is interesting in association rule mining, three parameters are considered, those are: support, confidence, and lift. The minimum support is set as 0.1, the minimum confidence is 0.75, and the lift must be greater than 1.0.

A. Result of Experiments of CM data set

Fig. 1 shows a histogram of distributions of grades in the CM course. Grade distributions for online quizzes before end semester (QPreFinal) is better than online quizzes before mid semester (QPreMid). Likewise, grade distributions for online final exam (FinalOL) is better than online mid exam (MidOL).

The results of association rule mining against the students' data set of CM course are shown in Table III and Table IV. In Table III, a set of interesting rules indicates associations between activity level, grade of online quizzes before mid semester, grade of online mid exam, and final grade of mid semester. The students that obtained fair and good for mid semester final grade, had high activity level. Similarly, the students that gained excellent and good for online mid exam, also had high activity level. However, activity level did not distinguish online quizzes grade before mid semester. The last rule in Table III, shows that students with medium activity level obtained poor grade for online quizzes before mid semester.

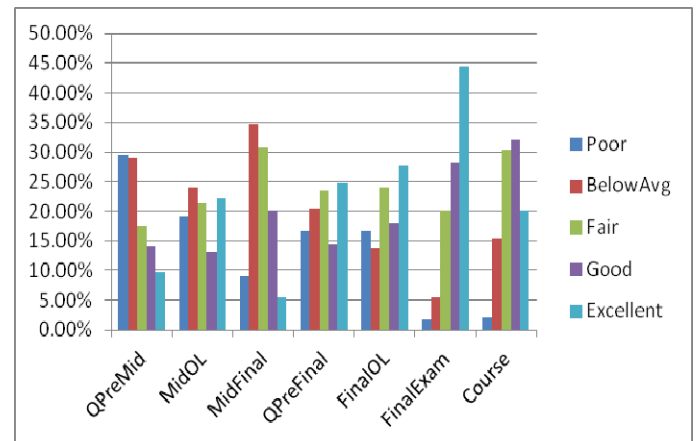


Fig. 1. Histogram of distributions of grades in CM course

In Table IV, a set of interesting rules indicates associations between activity level, grade of online quizzes before end semester, grade of online final exam, and final grade of end semester. The rules in table 3 dominates by students that had excellent grade for end semester final grade, online final exam, and online quizzes before end semester. All the students had high activity level.

From rules in Table III and Table IV, we might conclude that the students studied more seriously towards end semester compared to pre-mid semester, especially from online quizzes.

TABLE III.
RULES OF STUDENTS' ACTIVITY OF THE CM COURSE BEFORE MID SEMESTER

No	Association Rule	Sup.	Conf.	Lift
1.	GradeMidF=Fair ==> ActivityQ=High	0.26	0.83	1.13
2.	GradeMidF=Good ==> ActivityQ=High	0.19	0.96	1.3
3.	GradeMid=Excellent ==> ActivityQ=High	0.2	0.9	1.22
4.	GradeMid=Good ==> ActivityQ=High	0.12	0.94	1.27
5.	GradePreMid=BelowAvg ==> ActivityQ=High	0.25	0.85	1.15
6.	GradePreMid=Fair ==> ActivityQ=High	0.16	0.93	1.25
7.	GradePreMid=Good ==> ActivityQ=High	0.14	1	1.35
8.	GradePreMid=Excellent ==> ActivityQ=High	0.1	1	1.35
9.	ActivityQ=Medium ==> GradePreMid=Poor	0.17	0.76	2.59

This study further explored the relationships between course final grade with the group attributes using the J48 classification with ten folds cross validation. The classification is used to derive general rules from data set to indicate which students' activities that affect students' final grade in the course.

The result of the classification for CM course students' data is shown in Fig. 3 in the form of a J48 pruned tree with 60.25% accuracy. There are 141 correctly classified instances and 93 incorrectly classified instances. As shown in Fig. 3, there are two numbers (n/m) for each leaf in the tree, which mean that n instances reach the leaf, but m instances are classified incorrectly [10].

TABLE IV.
RULES OF STUDENTS' ACTIVITY OF THE CM COURSE AFTER MID SEMESTER

No	Rule	Sup.	Conf.	Lift
1.	GradeFinal=Excellent ==> ActivityQ=High	0.42	0.94	1.27
2.	GradeOnlineFinal=Excellent ==> ActivityQ=High	0.26	0.92	1.25
3.	GradePreFinal=Excellent ==> ActivityQ=High	0.25	1	1.35
4.	GradePreFinal=Fair ==> ActivityQ=High	0.18	0.78	1.06

The most affective attribute in predicting course final grade is end semester final grade (GradeFinal). If GradeFinal is fair, then course final grade would be determined by online final exam grade. If GradeFinal is good or below average, then course final grade would be determined by online mid exam grade. If GradeFinal is excellent or poor, then course final grade would be determined by mid semester final grade.

In the subtree of fair GradeFinal, the course final grade are below average or fair. In the subtree of good GradeFinal, the course final grade are good or fair. In the subtree of excellent GradeFinal, the course final grade are good or excellent, but there is an instance that has below average for the course final grade. In the subtree of poor GradeFinal, the course final grade are below poor or fair. In the subtree of below average GradeFinal, the course final grade are poor, below average or fair. In this case, online exam contributes in determining course final grade, especially for the students that have good, fair, or below average GradeFinal.

B. Result of Experiments of CL data set

Fig. 2 shows a histogram of distributions of grades in the CL course. Grade distributions for grade distributions for online final exam (FinalOL) is better than online mid exam (MidOL). However, grade distributions for online quizzes (Quiz) is worse than online exam (MidOL or FinalOL).

The results of association rule mining against the students' data set of CL course are shown in Table V, Table VI, and Table VII. In Table V, a set of interesting rules indicates associations between activity level, grade of online mid exam, and final grade of mid semester. The students that obtained excellent for mid semester final grade, had high activity level. Similarly, the students that gained excellent and good for online mid exam, also had high activity level.

In Table VI, a set of interesting rules indicates associations between activity level, grade of online final exam, and final grade of end semester. The students that obtained excellent for end semester final grade, had high activity level. Similarly, the students that gained excellent and fair for online final exam, also had high activity level.

In Table VII, a set of interesting rules indicates associations between activity level, grade of online quizzes, and final grade of assignments. The students that obtained excellent for assignment final grade, had high activity level. However, activity level did not distinguish online quizzes grade. From the rules listed in Table V, Table VI, and Table VII, we might conclude that the students performed online exam more seriously compared to online quizzes.

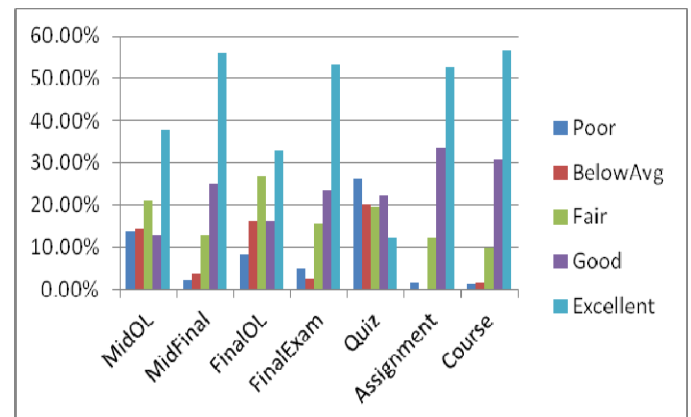


Fig. 2. Histogram of distributions of grades in CL course

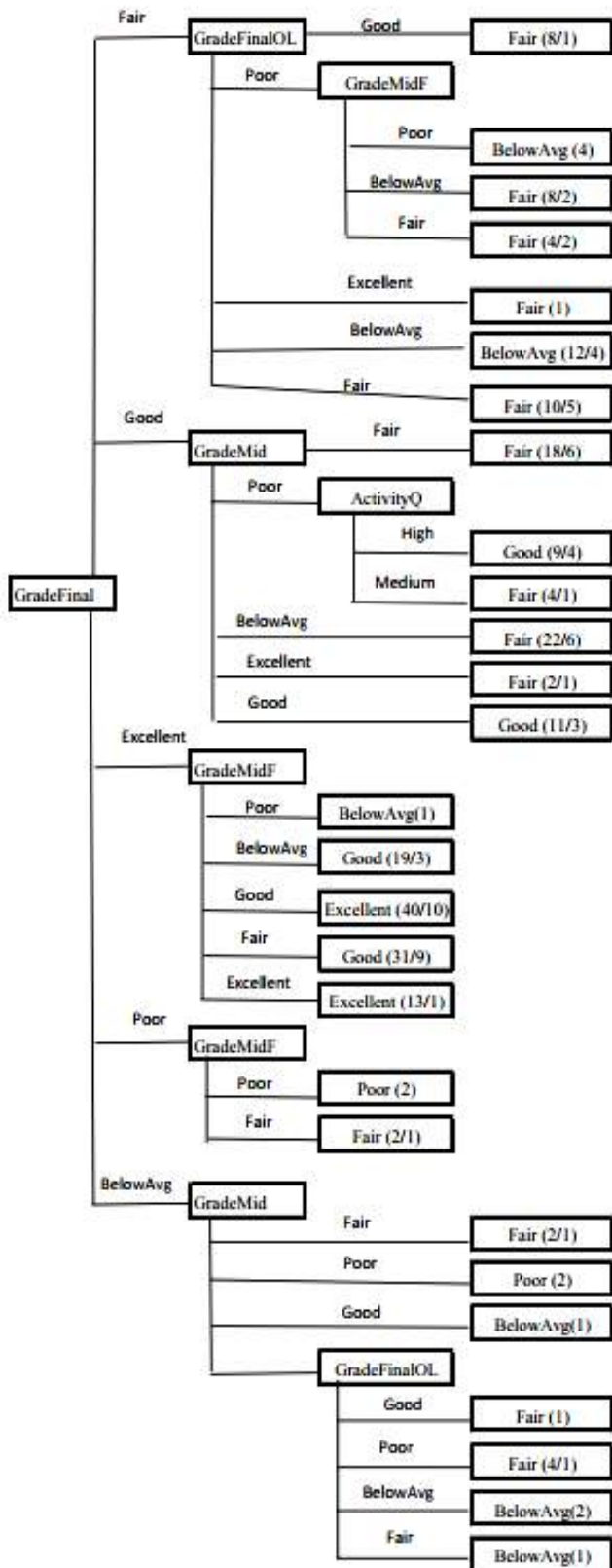


Fig. 3. J48 Decision Tree for CM course

The result of the classification for CL course students' data is shown in Fig. 4 in the form of a J48 pruned tree with 78.89% accuracy. There are 142 correctly classified instances and 38 incorrectly classified instances. The most affective attribute in predicting course final grade is final assignment grade (GradeAss). If GradeAss is excellent or good, then course final grade would be determined by end semester final grade. However, if GradeAss is fair or poor, course final grade is determined directly by GradeAss. In the subtree of excellent GradeAss, the course final grade are good or excellent. In the subtree of good GradeAss, the course final grade are excellent, good or fair. In this case, although online quizzes contribute in determining GradeAss, however online activities do not explicitly appear in J48 pruned tree.

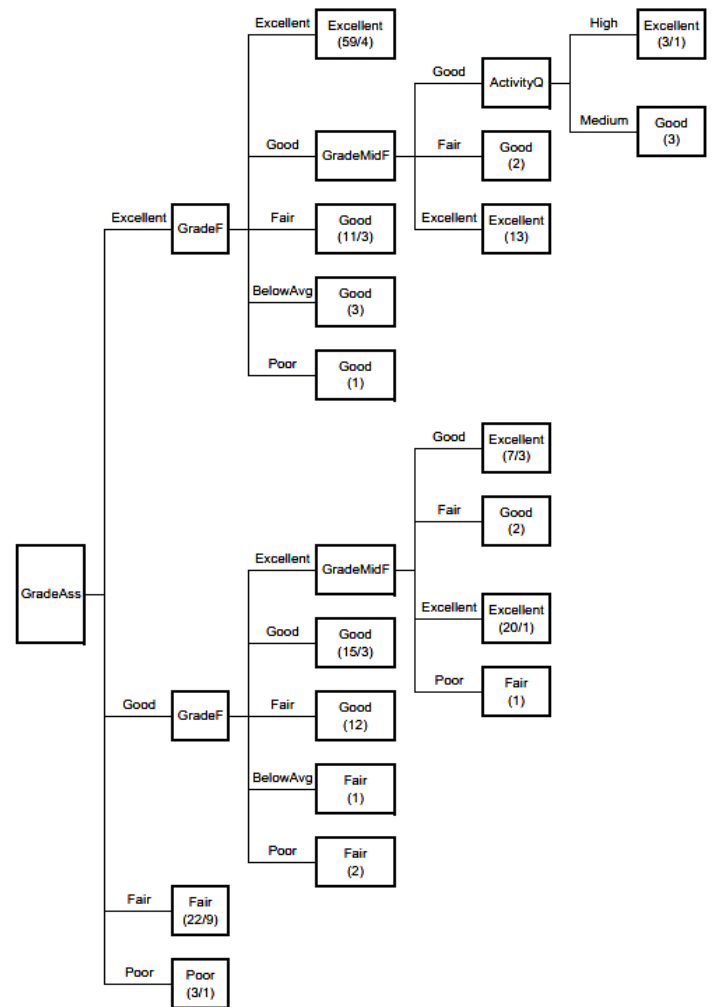


Fig. 4. J48 Decision Tree for CL course

Based on the results have been described, some suggestions are proposed to improve students' involvement and enthusiasm in learning process. The LMS should be facilitated with a recommender to give feedback to the students personally based on their activities [7], such as their achievement, a notification of next tasks, and assistance to adapt learning contents. To encourage students' motivation in

learning, some features such as a leader board, gamification, or a tournament could be provided as proposed in [6].

ACKNOWLEDGMENT

TABLE V.
RULES OF STUDENTS' ACTIVITY OF THE CL COURSE BEFORE MID SEMESTER

No	Rule	Sup.	Conf.	Lift
1.	GradeMidF=Excellent ==> ActivityQ=High	0.45	0.8	1.14
2.	GradeMidOL=Excellent ==> ActivityQ=High	0.30	0.79	1.13
3.	GradeMidOL=Good ==> ActivityQ=High	0.11	0.87	1.23

TABLE VI.
RULES OF STUDENTS' ACTIVITY OF THE CL COURSE AFTER MID SEMESTER

No	Rule	Sup.	Conf.	Lift
1.	GradeF=Excellent ==> ActivityQ=High	0.43	0.81	1.15
2.	GradeFOL=Excellent ==> ActivityQ=High	0.27	0.83	1.18
3.	GradeFOL=Fair ==> ActivityQ=High	0.21	0.77	1.09

TABLE VII.
RULES OF STUDENTS' ACTIVITY OF THE CL COURSE IN ASSIGNMENTS

TABLE	Rule	Sup.	Conf.	Lift
1.	GradeAss=Excellent ==> ActivityQ=High	0.46	0.87	1.24
2.	GradeQ=Good ==> ActivityQ=High	0.22	1	1.42
3.	GradeQ=Fair ==> ActivityQ=High	0.17	0.86	1.21
4.	GradeQ=BelowAvg ==> ActivityQ=High	0.15	0.75	1.06
5.	GradeQ=Excellent ==> ActivityQ=High	0.12	1	1.42
6.	ActivityQ=Low ==> GradeQ=Poor	0.12	1	3.83

V. CONCLUSION

This research have investigated students' activities data using EDM techniques to discover interesting rules and patterns in a blended learning system. This study reveals that there are strong correlation between students' activities in the form of online assessment and examination with their final grade. To improve students achievement and engagement in blended learning, the LMS should be facilitated with a recommender feature and some features to increase students motivation during their learning process.

The authors would like to acknowledge the financial support provided by the Directorate General of Research and Development Strengthening, Ministry of Research, Technology and Higher Education of the Republic of Indonesia, under the Research Grant number 1598/K4/KM/2017. The authors would also like to express their sincerest thanks to Mr Sunjoyo, lecturer of Change Management course and Creative Leadership course for his support and providing the data.

REFERENCES

- [1] C. Graham, "Blended Learning Models," in *Encyclopedia of Information Science and Technology*, Hershey: PA: Idea Group Inc., 2009, pp. 375-383.
- [2] J. Pankin, J. Roberts and M. Savio, "Blended learning at MIT," Massachusetts Institute of Technology Repository, 2015.
- [3] C.Romero, S.Ventura and E.Garcia, "Data mining in course management systems: Moodle case study and tutorial," *Computers and Education*, vol. 51, no. 1, p. 368-384, 2008.
- [4] C.Romero and S.Ventura, "Educational Data Mining : A survey from 1995 to 2005," *Expert System with Applications*, vol. 33, no. 1, pp. 135-146, 2007.
- [5] C. Romero and S. Ventura, "Data Mining in Education," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 3, no. 1, pp. 12-27, 2013.
- [6] M. Ayub, H. Toba, M. C. Wijanto and S. Yong, "Modelling students' activities in programming subjects through educational data mining," *Global Journal of Engineering Education*, in press.
- [7] C.Romero, "Educational Data Mining : A Review of the State-of-the-Art," *IEEE Transaction on Systems, Man, and Cybernetics, Part C(Applications and Reviews)*, vol. 40, no. 6, pp. 601-618, 2010.
- [8] R. Baker and K.Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3 - 16, 2009.
- [9] J. Han, M. Kamber and J. Pei, *Data Mining Concepts and Techniques*, Waltham: Elsevier, Inc., 2012.
- [10] I.H.Witten, E.Frank and M.A.Hall, *Data Mining Practical Machine Learning Tools and Techniques*, Burlington: Elsevier Inc., 2011.
- [11] P. C. Sun, R. J.Tsai, G. Finger, Y. Y. Chen and D. Yeh, "What drives a successfull e-learning? An empirical investigation of the critical factors influencing learner satisfaction.," *Computers and Education*, vol. 50, no. 4, pp. 1183-1202, 2008.
- [12] W. Bhuasiri, O. Xaymoungkhoun, H. Zo, J. J. Rho and A. P. Ciganek, "Critical success factor for e-learning in developing countries: A comparative analysis between ICT experts and faculty," *Computers and Education*, vol. 58, no. 2, pp. 843-855, 2012.