

Original Research Paper

Big Data in Educational Institutions using RapidMiner to Predict Learning Effectiveness

¹Evaristus Didik Madyatmadja, ²David Jumpa Malem Sembiring,
³Sinek Mehuli Br Perangin Angin, ⁴David Ferdy and ⁵Johanes Fernandes Andry

¹Department Information Systems, School of Information Systems,
Bina Nusantara University, Jakarta, Indonesia 11480, Indonesia

²Teknik Informatika, Institut Teknologi dan Bisnis Indonesia, Indonesia

³Akuntansi, Institut Teknologi dan Bisnis Indonesia, Indonesia,

^{4,5}Department of Information Systems, Faculty of Technology and Design,
Bunda Mulia University, Jakarta, 14430, Indonesia

Article history

Received: 22-02-2021

Revised: 12-04-2021

Accepted: 17-04-2021

Corresponding Author:

Evaristus Didik Madyatmadja
Information Systems
Department, School of
Information Systems, Bina
Nusantara University, Jakarta,
Indonesia 11480, Indonesia
Email: emadyatmadja@binus.edu

Abstract: Schools, universities, colleges and other educational institutions store a lot of data related to students and teachers as well as those related to education. You can analyze this data to gain insights that can improve the operational efficiency of your organization. Educational institution's needs based on student behavior, student exam results, each student's growth and changed educational regulations can be handled through statistical analysis. Big Data asphalt the way for innovative systems that students learn in exciting ways. The use of Learning Management Systems gave birth to a data explosion in higher education which became an innovation for the world of education, namely Big Data. These large amounts of digital data will provide information on what students and student behavior and involvement, their assessment, motivation and preferences, thus providing a large amount of data that can be mined for the learning experience. By utilizing the results of Big Data analysis in tertiary institutions, it can be obtained more insight into students, academics and the process in higher education so that it supports predictive analysis and increases decision making based on data which in turn can help improve the successful performance of students and institutions. This study using secondary data collection methods from journals or books. Method used is an analysis using the Decision Tree, Naive Bayes and K-Means. These results can predict the value obtained by students Learning Effectiveness in an educational institution, then a solution can be searched to increase the value of students in that institution.

Keywords: Big Data, Educational, Institutions, RapidMiner, Predict Learning, Decision Tree

Introduction

Big data transform education throughout the learning process, from basic to universities. Over time, with the development of sophisticated technology in the education system, Big Data helps instructors to realize human characters better and form new findings.

Research Background

With the advent of the Internet, technology has changed the way we are educated. Rather than taking classes in physical or face-to-face schools, many millions of students are currently enrolled in virtual schools to take classes full time or partially online (Aud and Wilkinson-Flicker, 2013).

Existence of computing currently owned by many people with a composition of 1 to 1, such as laptops, netbooks, or tablets provided by the campus to every student at school and at home are gaining momentum (Topper and Lancaster, 2013). The use of digital devices is expected to become widespread in education system schools, districts and other educational institutions (Bamiah *et al.*, 2018).

Once we digitize the education system, online activities that are carried out create a lot of digital data. The very large volume of data makes it possible to exceed the available power for computing and processing by having a single computer; therefore, several computers are needed for processing very large data (Hesse *et al.*, 2015). The data generation speed is so fast that up to 90%

of the world's data is currently generated during the 2 years this decade (IBM, 2012). The types of data are so diverse that data sources range from log files using digital devices, from web browsing data, social media data, geo-location images, to audio data and others (McHugh, 2015). Thus, large volume, speed and variation are considered the features and traits of very Big Data determinants.

Problem Identification

The problems for assessment and analysis for learning have increased enormously since the performance capacity of the new digital, user interface side and targets to be achieved by technology-powered assessments have become more complex. The increasing complexity is due in part to the role of technology in providing an interactive learning experience and diverse data collection that has led to the inclusion of data methods and techniques into scientific research to understand learning and behavior (de Freitas, 2013). This change requires new quantitative methods as well as reconceptualization of mixed methods that include existing domain experts as well as stakeholders in building knowledge of this complex system (Tashakkori and Teddlie, 2010).

The emergence of very large data is believed to be data with a large number of records starting from very different types of data, which are collected quickly for immediate action (Margetts and Sutcliffe, 2013). The need to develop literacy assessments on Instructor's, students and assessments from other audiences (Stiggins, 1995). Appraisals references are becoming more interesting and important than ever to understand how technology can influence and process judgments and especially to build confidence in making and analyzing multiple arguments from the evidence, based on the user's current of validation.

Research Questions

In this study, we will answer several questions regarding the application of this Big Data in educational institutions, namely as follows Can the analysis of the use of Big Data increase the effectiveness of student learning and can predict how much student achievement is?. and then, how is the application of Big Data in the educational institution system? We will answer that question in this journal so that readers can find out about the use of Big Data in educational institutions.

Research Benefits

From a perspective related to teaching, Big Data can provide us with some endless possibilities for improving the processes of the education system. It is also important to address issues stemming from the instructor's level of knowledge when analyzing Big Data. This is operationally acquired knowledge that some teachers

lack. It will be interesting not only to develop some training programs in various fields, but also to consider the possibility of involving some specialized staff to do this from the perspective of innovative pedagogical methods (Williamson, 2016).

West said, "Big Data can support classical education systems helping teachers analyze what students know and what techniques are most effective for each student." In this way, teachers can also learn new techniques and methods about their educational work (West, 2012). The data displayed as a Big Data model can be used as reference material for the education of schools and teachers and can evaluate the student's academic status in a timely and accurate manner to identify potential problems for the student. Predict future student performance (Li and Zhai, 2018).

The ambition is to bring more real-time data streams to our evolving digital training system. While this is still a recent attempt to use Big Data in education, the information value spent from Big Data presents some unique opportunities and promises to provide students with personalized learning that can inform the creation of systems. Once it is realized that the growing flow of Big Data can also raise concerns over data security, the need for privacy protection and access rights in accessing private digital data.

As a result, this study presents an implementation or with Big Data to predict student learning effectiveness in institutions educational. Received invitations for practitioners in the field of education, regulatory and regulatory makers and several researchers to advance the writer's understanding of Big Data in education and provide better service to students in today's digital era. It is becoming increasingly important for educators to understand the latest trends in education and Big Data analytics. Rather than relying on standardized tests to find problems, modern educators use Big Data technology to find issue areas for students.

Literature Review

Big Data

In general, there is no definition of Big Data that can be used as an academic reference (Kitchin, 2014). Scholars define Big Data differently. However, the study of Big Data has been carried out by many scholars in the context of ontology, epistemology and axiology of Big Data. Big Data can be defined as a system that integrates the real world, humans and the virtual world (Chen and Zhang, 2014). The real world is related to social reality that is reflected in the virtual world through technology and the internet of things. While humans generate is defined as large data (Valêncio *et al.*, 2020), where Big Data is produced into cyberspace through technology; computers, artificial intelligence and the mobile internet (De Mauro *et al.*, 2015).

In general, the use of data size parameters is more commonly used to answer whether Big Data is or not. The data size (volume) parameter is not the only reference. refers to the use of the following parameters: Three-V, namely Volume, Variety, Velocity, Value and Veracity (Gartner calls this a part-Three-Parts). Many definitions related to Big Data, some of them use Three-V as the emphasis (Kumar *et al.*, 2014; Hadi *et al.*, 2015).

Volume refers to the size of the numbers of data being managed (in units of MB, GB, TB, PB, ZB). The size limitations that can be called Big Data can still be said to vary. The minimum size of a Big Data system is from Terabytes (TB) to Petabytes (PB). Variety refers to the level of structural diversity in Big Data datasets, can be structured (example tables), semi-structured (XML documents) and unstructured (documents, email, text messages, audio, video, images, graphics and others). Velocity according to the manner and rate at which data is received, including whether through batch, near time, real time and stream processing (Hadi *et al.*, 2015).

Big Data requires Information Technology (IT) to combine all existing data. Information Technology or software and the systems are mainly focusing on processing data to translate information into knowledge. IT can work in every system in the world, which provides direction information for managerial activities in an education. A structure for converting data from internal and external sources into information from data and translate to communicating that information in an appropriate. Managers at all any levels in all functions to enable them to make timely and effective decisions to plan, to direct and to control the activities they carry out in the right way to be responsible (Andry *et al.*, 2018). In order to measure the progress of an organization towards strategic goal setting and organizational decision making, organizations need a reliable information system foundation. This information system provides the possibility to provide decision assistance for leaders, to reduce levels of uncertainty and to contribute to decision-making performance (Chakir *et al.*, 2020a).

Data Analytics

Some people believe that analysis data and analytical data have the same meaning. From there, sometimes several people use it interchangeably. This is technically untrue. There is a clear difference between the two. So let's discuss the unclear difference between the terms analysis and analytic because although they have the same words they have different meanings (Elgendy and Elraga, 2014).

Data analysis is a process of examining, cleaning, modifying and modeling data in order to find useful information. However, one important thing to remember is that you are analyzing things that have happened in the past. Analytics generally refers to the future, rather than describing past events. In other words, it is exploring

future potential. Analytics is basically the application of logical and computational reasoning to the component parts obtained in the analysis. In doing this analytic activity, you look for patterns in exploring what you can do in the future (Memon *et al.*, 2017).

Big data analytics is a strategic way used by some people to be able to analyze some large information that contains various information, for example, big information to reveal events that are sometimes hidden, some unclear relationships, ad flow, trends from clients and business data for other support. The results of this demonstrative test can provide proficient data, the opening of a new income, the increase and profits of some clients, the improvement of the skills of an operation and for competitor association and business payments that can be made differently (Taylor-Sakyi, 2016).

Data Mining Analytics for Predicting Student Performance

When data mining has been successfully implemented in the business world in the past decade, its use in several universities is still something new. Its use is usually to identify and extract new and potential and valuable knowledge from data. Implementation of data mining aims to develop a model that can draw conclusions about the success and academic achievement of students at universities (Osmanbegovic and Suljic, 2012).

The implementation of data mining analytics in the education system at universities can be directed to the specific needs of each student in the educational process by using Big Data. Students can then be asked to recommend additional activities, along with teaching materials and assignments that can support and enhance the learning of each student. In all educational settings, the main objective is to ensure that the learning process of students enables them to understand their own learning paths. From here, Education Data Mining (EDM) can be used, which is to provide fundamental values for educational institutions and universities and to all stakeholders who support various processes in this learning activity (Manjarres *et al.*, 2018).

In the last few years in 2000, the interest in using data mining from Big Data for educational purposes continued to increase. Because data mining is a very promising research field in the field of education and others and has special requirements that can be done in other fields as well. This very complete review of data mining in university education is one of the educational problems that is solved by data mining analysis by predicting academic performance and student performance, which aims to predict some unknown variables (such as results, grades, or scores), which describes the student (Schneider *et al.*, 2019).

Data mining analytics and Artificial Intelligence (AI) can help each other to analyze student grades and get accurate value predictions so that they can determine

student performance in an institution. There is also Artificial Intelligence (AI) which is a discipline that can be described in two ways, namely as a science that aims to discover the essence of intelligence and develop the intelligent machine itself; or as a field of science that finds techniques to be able to solve fairly complex problems that cannot be solved without applying certain other knowledge (for example, being able to make good and correct decisions based on large amounts of data and information) (Chakir *et al.*, 2020b).

Methodology

The methodology is used by the author to analyze, work on and solve the problems at hand. The theoretical framework or scientific framework is the scientific method that will be applied in conducting research. At the research stage used, namely studying the literature, collecting data, processing data, analyzing data and making reports.

Research Stages

Based on Fig. 1, it is known that this research starts from collecting data until making a report:

1. Collecting data, data collection is analyzed using secondary data from www.kaggle.com
2. Data processing, the data that has been taken is then processed by entering the dataset from Kaggle into the RapidMiner Studio application to be carried out to the next stage, namely the data analysis stage
3. Analyze data, this stage the data is analyzed in the RapidMiner Studio application. Data analysis is done by making a decision tree to display data in the form of trees that form the roots, branches and leaves of the data. Then the decision tree results are analyzed to obtain the results and discussion required in the report
4. Creating a report, this stage is carried out after the data has been analyzed. The results of the analysis are made into a research journal containing abstracts, introduction, research methods, results and discussion, conclusions and bibliography based on data that has been taken, processed, analyzed, discussed and obtained from the research

Method of Collecting Data

Subsidiary data is data collected from primary sources that can be used in current investigative studies. To collect secondary data often takes less time than collecting primary data where we as researchers have to collect every piece of information from the early stages of the research. With this data it is possible to collect more data (Kabir, 2016). Subsidiary data are often available from many sources. After the release on electronic media and the internet, the availability of secondary data has become easier to process.

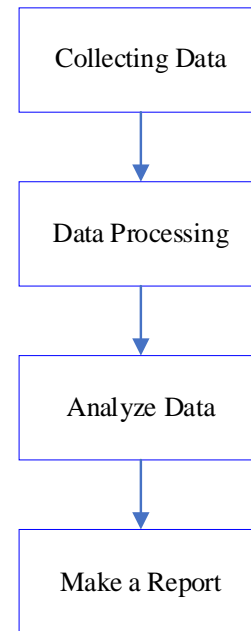


Fig. 1: Research stages (Angraeni and Rizaldi, 2020)

Data Analysis Technique

Software RapidMiner is an environment for business analytics, to predictive, data mining, to mining and machine learning. RapidMiner Studio uses predictive data analysis (AL-Ma'aitah, 2020) techniques and descriptive data analysis techniques in providing knowledge to each user so that they can make and make excellent decisions. Statistical techniques are also used in Rapid Miner. The processing of the data uses descriptive statistics or inferential statistics (Krstevski *et al.*, 2012).

Descriptive statistical analysis is carried out if the research objective is to increase the reader's knowledge, understanding and application of research. Phrases for study purposes are very important. The purpose of precise wording relates to study design and statistics. Such as describing, identifying, evaluating, checking, designing, reviewing, showing, measuring, etc. Observe emphatically that verbs associate, relate, compare, predict, predict, etc. It belongs to the field of inferential statistical analysis (Hussain, 2012).

The use of inferential statistical models includes a significance test that the authors can use to make conclusions about the existing data in their sample. This test can be divided into three basic stages depending on the intended results: Usually to evaluate differences, check for causal relationships and make certain predictions. The decision about which procedure to use can be determined, by several research questions, at the time of conducting the research design (Allua and Thompson, 2009). Model of conclusion is an interactive computer information technology that can give an alternate solution for managers (Madyatmadja, 2014).

Analysis and Discussion

This section will be discussed about Data Attributes and Description following with the explanation, Student performance data classification range, Process Implementation using software RapidMiner, Accuracy, Precision and Recall using Rapid Miner and Evaluation and Result.

The dataset in this study uses data from Kaggle entitled Student Performance Data Set. This dataset has 649 instances, with data in the form of numbers and letters. The dataset has attributes or features that describe each instance. The following is a list of attributes in this study and their explanation, Table 1.

Based on Table 3, from the G3 attribute, the data classification process will be carried out, including performance which will be grouped into 3 categories, values 0-7 will be grouped into categories (Bad), values 8-14 will be categorized (Standard) and values 15-20 will be categorized (Good). Attribute classification can be seen in Tables 1 and 2.

Implementation of the Dataset in RapidMiner

To implement Dataset using Decision Tree model on Rapid Miner, the first thing to do is Import data. The data that must be imported is the training data used in Excel and the testing data which contains all the data from the data base, namely 649 data.

When importing data, it should be noted that each table column is given the correct type of attribute. If given the wrong attribute, the final result will be wrong too.

There are several types of attributes that we will use, namely:

- a) Binominal: Data type used for data types that only have two types, such as YES/NO, or 0/1
- b) Polynomial: The type of data used for data types that have more than 2 types
- c) REAL: Used for data types that have a decimal in the number
- d) Integer: Used for data types that use numbers for data, without any decimals

Table 1: Data attributes and description

No.	Attribute	Description
1.	Universities	(Student's universities) binary: - 'RY' Ririn Yosephin - 'JF' Johanes Fernandes
2.	Gender	(Student's Gender) binary: - 'Fm' female - 'Ml' male
3.	Years	(Student's years) Numeric: from 14 to 21
4.	Mailing address	(Student's mailing address type) binary: - 'C' city - 'V' village
5.	Fam	(Family) binary: - 'LS2' less or equal to 2 - 'GR2' greater than 2
6.	FMstatus	(Father's and Mother's cohabitationstatus) binary: - 'To' living together - 'Ap' apart
7.	eduM	(education Mother's) numeric: - 4 "higher education - 3 "secondary education - 2 "th to 9th grade - 1 primary education (4th grade) - 0 none
8.	eduF	(education Father's) numeric: - 4 "higher education - 3 "secondary education - 2 "5th to 9th grade - 1 primary education (4th grade) - 0 none
9.	jobM	(job Mother's) nominal: - 'other'- 'house wife' - 'services' (e.g., administrative or police, etc.) - related 'health' care - 'college teacher'

Table 2: Attributes explanation

No.	Attribute	Description
1.	jobD	(Dady's job) nominal: 'other' 'teacher'; 'health' care related; 'services' civil); 'at_home';
2.	excuse	(excuse to choose this univercity) nominal: close to 'home'; school 'reputation'; 'course' preference; 'other'
3.	custodian	(student's custodian) nominal: - 'mother'; 'father'; 'other'
4.	trip	(home to school trip time) numeric: 1 "<10 min; 2 "10 to 20 min; 3 "20 min to 30 min; 4 ">30 min
5.	Time to learn	(weekly time to learn) numeric: 1 "<3 hours; 2 "3 to 6 hours; 3 "6 to 12 hours; 4 ">12 hours
6.	unsuccessful	(number of past class unsuccessful) numeric: from 0 to 5
7.	schoolprom	(extra educational promotion) binary: Y or N
8.	famprom	(family educational promotion) binary: Y or N
9.	payment	(extra payment classes within the course subject Math or Chinese) binary: Y or N
10.	bustle	(extra-curricular activities) binary: Y or N
11.	nursery	(attended nursery school) binary: Y or N
12.	higher	(wants to take higher education) binary: Y or N
13.	internet	(Internet access at home) binary: Y or N
14.	romance	(with a romance relationship) binary: Y or N
15.	relfam	(quality relationships of family) numeric: from 1 very bad to 5 excellent
16.	timefree	(time free after school) numeric: from 1 very low to 5 very high
17.	hangout	(hang out with friends) numeric: from 1 very low to 5 very high
18.	cDal	(workday alcohol consumption) numeric: from 1 very low to 5 very high
19.	cWal	(weekend alcohol consumption) numeric: from 1 very low to 5 very high
20.	freshness	(current freshness status) numeric: from 1 VB to 5 VG
21.	lack	(number of school lack) numeric: from 0 to 92
22.	G10	(first period grade) numeric: from 0 to 10
23.	G20	(second period grade) numeric: from 0 to 10
24.	G30	(final grade) numeric: from 0 to 10

Table 3: Student performance data classification range

G3 grade	Value
0-7	Bad
8-14	Standard
15-20	Good

Table 4: Type data

Instance	Type data
Universities, gender, mailing address, Fam, FMstatus, Schoolsup,	Binominal
Famprom, Payment, Activities, Nursery, Higher, Internet, Romance eduM, eduF, jobM, jobF, Reason, Guardian, Traveltime, Studitime, Failures, Famrel, Freetime, hangout, alcD, alcW, freshness, Performance.	Polynomial
Age, Absences, G1, G2, G3.	Integer

Based on Table 4, for each data, testing data and training data 650, following are the types of data used for training data and testing data.

At the preprocessing stage, the selected student data is confirmed to be suitable for processing. From student data, the process of data transformation and data reduction is carried out. In the transformation process, data is transformed into a form suitable for the data mining process. Furthermore, the data is reduced by removing unnecessary attributes so that the size of the database is small and only includes the attributes needed in the data mining process.

Modeling is a stage that directly involves data mining techniques, namely by selecting data mining techniques and determining the algorithm to be used. The algorithm used in this research is a Decisions Tree

algorithm. The results of the decision tree can be seen in Figs. 2 and 5.

Process Implementation using RapidMiner

Open Rapid Miner then in the repository table create a new repository then enter the student performance dataset. Select the retrieve operator then enter the student performance dataset into the retrieve operator in the rapid miner design process field, then enter the decision tree operator into the field, connect the two operators, then enter the apply model and performance operators into the design field then enter the performance dataset back into testing data then connect it with the apply model operator, the result will be like Fig. 3.

Accuracy value using RapidMiner as seen in Fig. 4, the results are as shown in Table 6.

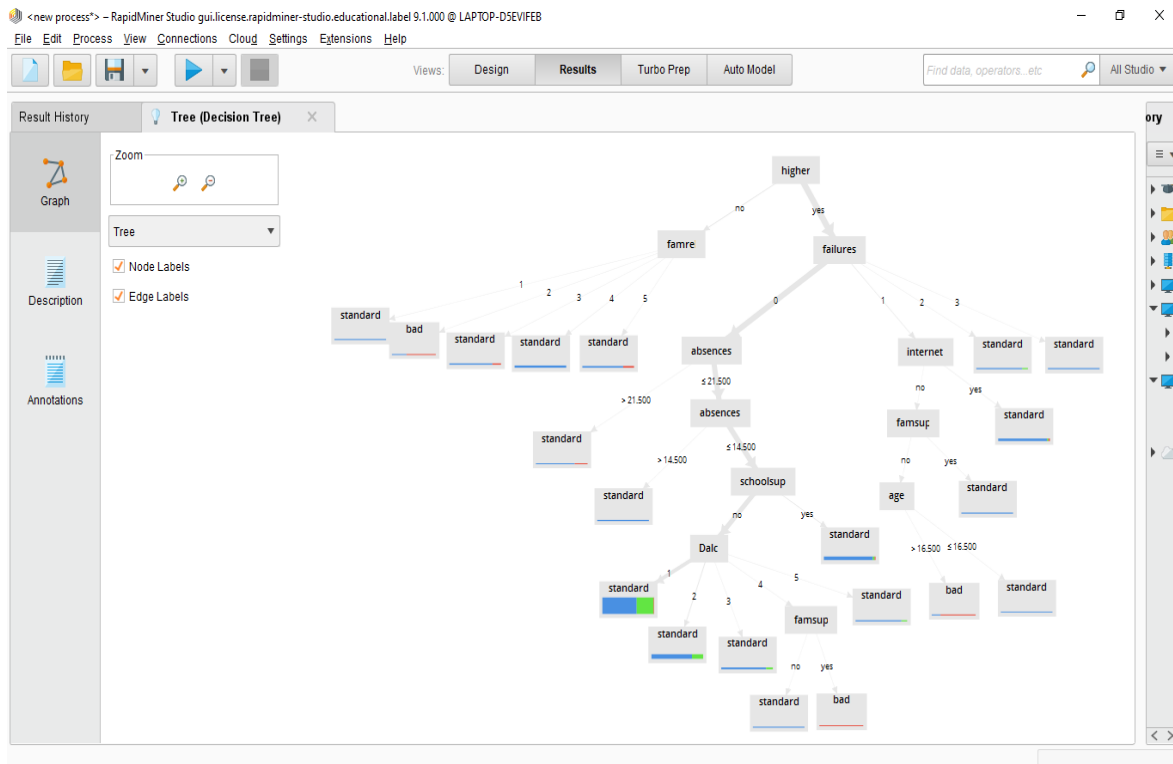


Fig. 2: The formed decision tree model using maximal depth 8

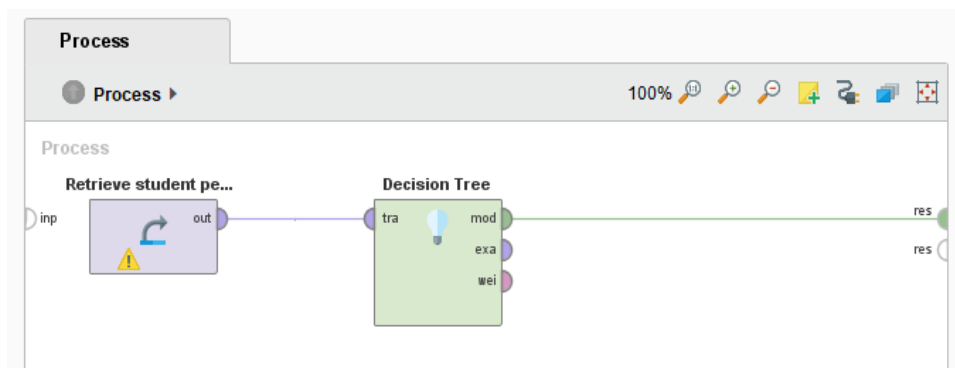


Fig. 3: Process model decision tree using RapidMiner

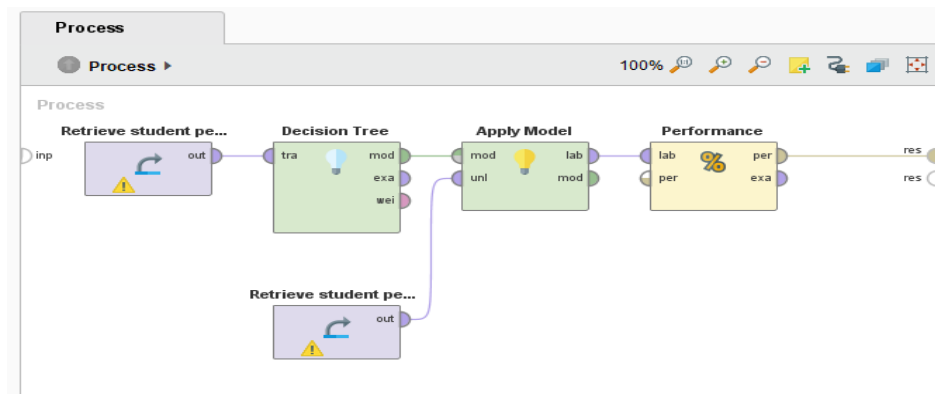


Fig. 4: Process to determine decision tree the accuracy value using RapidMiner

The resulting model with information gain criteria. The gain ratio and Gini index were tested using the Decision Tree method, it can be seen that the comparison of accuracy, precision and recall values is in Table 5, for the gain ratio criteria have the highest accuracy, precision and recall values, followed by the information gain criteria and the gini index.

results in a high enough accuracy value proving that the data used is good.

Table 5: Accuracy, Precision and Recall using Rapid Miner

Information	Gain (%)	Gain ratio (%)	Gain index (%)
Accuracy	90,00	100	85
Precision	85,71	100	79,76
Recall	93,33	100	83,33

Evaluation and Result

Comparison of accuracy values obtained from data processing using different model is as seen in Table 6.

The accuracy generated by using the Decision Tree, Naive Bayes and K-Means algorithm models in Table 3

Table 6: Accuracy result using different algorithm model

Model	Accuracy (%)
Decision Tree	100
Naive Bayes	95,07
K-Means	91,45

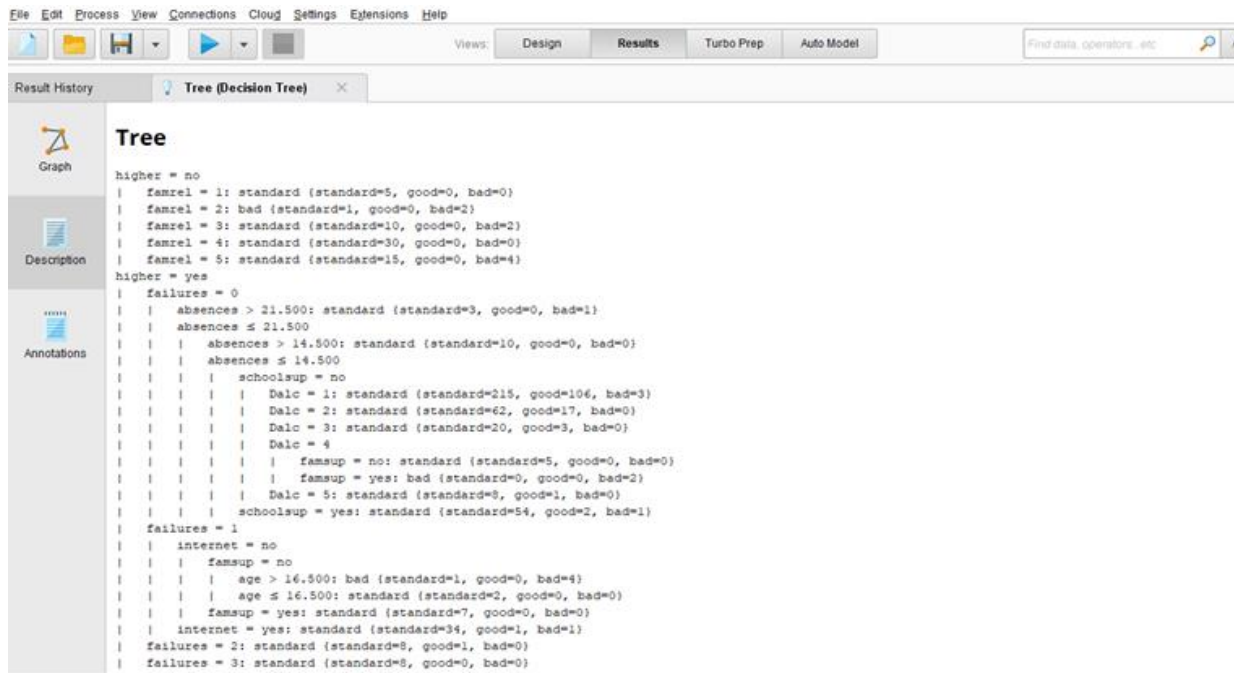


Fig. 5: Rule obtained

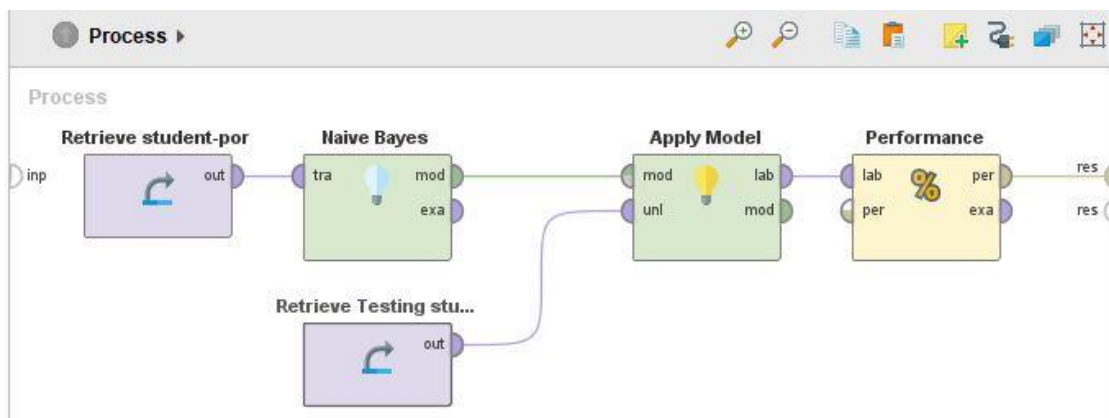


Fig. 6: Process to determines Naive Bayes accuracy value using RapidMiner

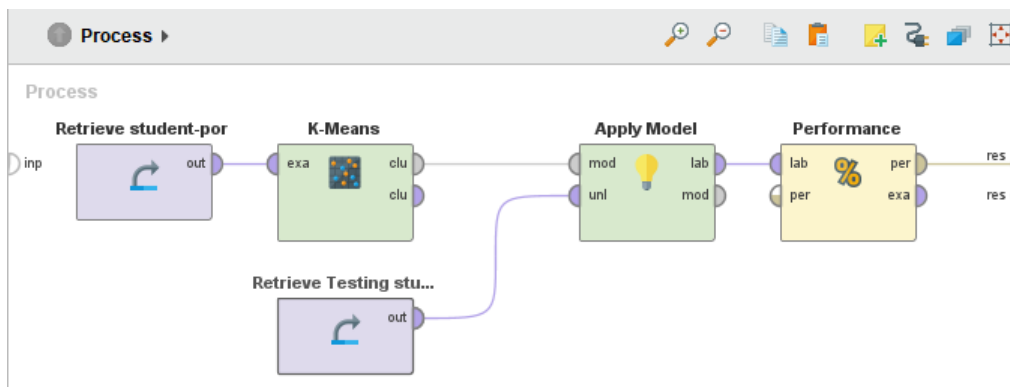


Fig. 7: Process to determines K-Means accuracy value using RapidMiner

So what's the point of high accuracy in the modeling? We can see what criteria make students score Bad, Standard, or Good. We can apply the criteria according to the instances that produce students with a Good result. This ultimately proves that data can improve student performance by imitating the way of learning methods, environment, school, health, parenting and others in accordance with existing data resulting in learning performance and get the desired results.

Figure 6, shown process determines Naive Bayes accuracy value, the results are as shown in Table 6.

Figure 7, shown process determines K-Means accuracy value, the results are as shown in Table 6.

Conclusion

In this study we conducted data mining analysis on student performance in a school using data from student performance in a school that was obtained from the Kaggle website with the name Student_Performance. This study uses the Rapid Miner application to assist us in processing student performance data so that the data processing produces output. The results obtained from this research are that in the G3 Attribute it has 3 categories of Bad, Standard and Good, if the value 0-7 will fall into the bad category, if the value 8-14 will enter the standard category, if the value 15-20 will enter the category good. The use of data mining analysis using the Decision Tree Algorithm model, Naïve Bayes and K-Means produces a fairly high accuracy value so that the data used is very good. From the three models, we can see and predict the level of student performance through existing values and the results of the analysis using Rapid Miner with good accuracy. Then we can conclude that using Big Data can help predict student scores in an institution by processing these data first so as to produce a predictable percentage value for the success rate to help increase student grades.

We are very aware in doing the research that we are doing there are still some shortcomings, therefore if this deficiency can be corrected in the next research, therefore there are some suggestions that can be used as input to do

further research. After conducting an analysis to measure student success and already getting the results, we hope that in the next development can be re-analyzed about the shortcomings in our journal.

Authors Contributions

Evaristus Didik Madyatmadja: Lead research project, coordinate developer, doing experiment, be an instructor, data analysis and writing the manuscript.

David Jumpa Malem Sembiring: Advise research project, design the experiment, data analysis and writing manuscript.

Sinek Mehuli Br Perangin Angin: Advise data analysis and writing manuscript

David Ferdy: Advise research project, design the application, data analysis, writing manuscript, proof reading.

Johanes Fernandes Andry: Advise research project, design the research methodology, data analysis, writing manuscript, proof reading.

Ethics

Authors confirm that this manuscript has not been published elsewhere and that no ethical issues are involved.

References

- Allua, S., & Thompson, C. B. (2009). Inferential statistics. *Air Medical Journal*, 28(4), 168-171. <https://doi.org/10.1016/j.amj.2009.04.013>
- AL-Ma'aitah, M. A. (2020). Utilizing of Big Data and predictive analytics capability in crisis management. <https://pdfs.semanticscholar.org/db43/03756bdeb74482e14021f25e2854b7b4cdfd.pdf>
- Andry, J. F., Suroso, J. S., & Bernanda, D. Y. (2018). Improving quality of smes information system solution with ISO 9126. *Journal of Theoretical and Applied Information Technology*, 96(14), 4610-4620. <http://www.jatit.org/volumes/Vol96No14/24Vol96No14.pdf>

- Anggraeni, D., & Rizaldi, R. (2020, July). Rapidminer's implementation in analyzing student mainstream data. In *International Conference on Social, Sciences and Information Technology* (Vol. 1, No. 1, pp. 1-6). <https://jurnal.stmikroyal.ac.id/index.php/ICoSSIT/article/view/819>
- Aud, S., & Wilkinson-Flicker, S. (2013). *The condition of education 2013*. Government Printing Office. ISBN-10: 0160920701.
- Bamiah, S. N., Brohi, S. N., & Rad, B. B. (2018). Big data technology in education: Advantages, implementations and challenges. *Journal of Engineering Science and Technology*, 13, 229-241. http://jestec.taylors.edu.my/Special%20Issue%20ICCSIT%202018/ICCSIT18_19.pdf
- Chakir, A., Chergui, M., & Andry, J. F. (2020a). A decisional smart approach for the adoption of the IT green. *Environment, Development and Sustainability*, 1-15. <https://doi.org/10.1007/s10668-020-00999-1>
- Chakir, A., Chergui, M., & Andry, J. F. (2020b). A smart updatet it governance platform based on artificial intelligence. *risk*, 8, 9. <https://doi.org/10.25046/aj050507>
- Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information sciences*, 275, 314-347. <https://doi.org/10.1016/j.ins.2014.01.015>
- de Freitas, S. (2013). *Education in computer generated environments*. Routledge. <https://doi.org/10.4324/9780203094709>
- De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is Big Data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics. <https://doi.org/10.1063/1.4907823>
- Elgendy, N., & Elragal, A. (2014, July). Big data analytics: a literature review paper. In *Industrial conference on data mining* (pp. 214-227). Springer, cham. https://doi.org/10.1007/978-3-319-08976-8_16
- Hadi, H. J., Shnain, A. H., Hadishaheed, S., & Ahmad, A. H. (2015). Big data and five V's characteristics. *International Journal of Advances in Electronics and Computer Science*, 2(1), 16-23. <https://scholar.google.com/citations?user=dfnIL0oAAA&hl=en&oi=sra>
- Hesse, B. W., Moser, R. P., & Riley, W. T. (2015). From big data to knowledge in the social sciences. *The Annals of the American Academy of Political and Social Science*, 659(1), 16-32. <https://doi.org/10.1177/0002716215570007>
- Hussain, M. (2012). Descriptive statistics--presenting your results I. *J.P.M.A. The Journal of the Pakistan Medical Association*, 62(7), 741-743. <https://pubmed.ncbi.nlm.nih.gov/23866532/>
- IBM. (2012). What is big data? - Bringing big data to the enterprise. <https://isykes.wordpress.com/2012/08/20/ibm-what-is-big-data-bringing-big-data-to-the-enterprise/>
- Kabir, S. M. (2016). *Methods of Data Collection*. In: *Basic Guidelines for Research: An Introductory Approach for All Disciplines*, (pp. 201-275), Book Zone Publication, Chittagong-4203, Bangladesh. https://www.researchgate.net/publication/325846997_Methods_of_Data_Collection.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 2053951714528481. <https://doi.org/10.1177/2053951714528481>
- Krstevski, J., Mihajlov, D., & Chorbev, I. (2012). Student Data Analysis with RapidMiner. *ICT Innovations 2011, Web Proceedings ISSN 1857-7288*, 19. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.403.6343&rep=rep1&type=pdf#page=25>
- Kumar, D., Kamesh, K., & Syed, U. (2014). A study on Big Data and its Importance. *International Journal of Applied Engineering Research*, 9(20), 7469-7479. https://www.researchgate.net/publication/267626895_A_Study_on_Big_Data_and_its_Importance
- Li, Y., & Zhai, X. (2018). Review and prospect of modern education using Big Data. *Procedia Computer Science*, 129, 341-347. <https://doi.org/10.1016/j.procs.2018.03.085>
- Madyatmadja, E. D. (2014). Decision support system model to assist management consultant in determining the physical infrastructure fund. *J. Theor. Appl. Inf. Technol*, 62(1), 269-274. <http://www.jatit.org/volumes/Vol62No1/32Vol62No1.pdf>
- Manjarres, A. V., Sandoval, L. G. M., & Suárez, M. S. (2018). Data mining techniques applied in educational environments: Literature review. *Digital Education Review*, (33), 235-266. <https://dialnet.unirioja.es/servlet/articulo?codigo=6485868>
- Margetts, H., & Sutcliffe, D. (2013). Addressing the policy challenges and opportunities of "Big data". <https://doi.org/10.1002/1944-2866.POI326>
- McHugh, D. (2015). Traffic prediction and analysis using a Big Data and visualisation approach. Department of Computer Science, Institute of Technology Blanchardstown. <http://www.dataminingmasters.com/uploads/studentProjects/TrafficAnalysis.pdf>.

- Memon, M. A., Soomro, S., Jumani, A. K., & Kartio, M. A. (2017). Big data analytics and its applications. arXiv preprint arXiv:1710.04135. <https://arxiv.org/abs/1710.04135>
- Osmanbegovic, E., & Suljic, M. (2012). Data mining approach for predicting student performance. *Economic Review: Journal of Economics and Business*, 10(1), 3-12. <https://www.econstor.eu/handle/10419/193806>
- Schneider, K., Berens, J., & Burghoff, J. (2019). Drohende Studienabbrüche durch Frühwarnsysteme erkennen: Welche Informationen sind relevant? *Zeitschrift für Erziehungswissenschaft*, 22(5), 1121-1146. <https://doi.org/10.1007/s11618-019-00912-1>
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238. <https://search.proquest.com/openview/420d00f4b01136f8d23c90af1e8936fd/1?pq-origsite=gscholar&cbl=41842>
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). Sage handbook of mixed methods in social & behavioral research. sage. <https://doi.org/10.4135/9781506335193>
- Taylor-Sakyi, K. (2016). Big data: Understanding Big Data. arXiv preprint arXiv:1601.04602. <https://arxiv.org/abs/1601.04602>
- Topper, A., & Lancaster, S. (2013). Common challenges and experiences of school districts that are implementing one-to-one computing initiatives. *Computers in the Schools*, 30(4), 346-358. <https://doi.org/10.1080/07380569.2013.844640>
- Valêncio, C. R., Silva, L. M. M., Tenório, W., Zafalon, G. F. D., Colombini, A. C., & Fortes, M. Z. (2020). Data warehouse design to support social media analysis in a Big Data environment. *Journal of Computer Science*, 126-136. <https://doi.org/10.3844/jcssp.2020.126.136>
- West, D. M. (2012). Big data for education: Data mining, data analytics and web dashboards. *Governance studies at Brookings*, 4(1), 1-10. <http://www.insidepolitics.org/brookingsreports/education%20big%20data.pdf>
- Williamson, B. (2016). Digital education governance: data visualization, predictive analytics and 'real-time' policy instruments. *Journal of Education Policy*, 31(2), 123-141. <https://doi.org/10.1080/02680939.2015.1035758>