Teacher Quality Affects On Graduation Of Study Programming Data Approach There With CRISP-DM Method

Mawaddah Harahap[1]*, Namira Hidayati[2], Sumiati Panjaitan[3], Enjelyna Tambunan[4], Juniati Sihombing[5]
[1,2,3,4,5]Universitas Prima Indonesia, Indonesia
[1]mawaddah@unprimdn.ac.id, [2]namirahidayati8@gmail.com,
[3]sumiatipanjaitan2019@gmail.com, [4]enjelynatambunan@gmail.com,
[5]juniatisihombing36@gmail.com

Abstract: Each student's graduation is influential to the teacher in every subject that can be predicted based on the pattern of habits of the teacher who presents the subject. Web Proggramming is the subject of study that must be completed by every student. If this course is not completed, it is not allowed for the student to take other courses related to it. The custom patterns of teachers in this study were taken from 300 student respondents. An analysis is done to compare the results of questionnaire scores with the assessment of college admissions teachers. From the results of the comparison, it is possible to predict the graduation rate of students to the web programming course. The results of the experiment were that 72% of the students received highly influential predictions, 12% Influential, 7% Sufficient, 5% Influential and 4% Highly Influential.

Keywords: Teacher Quality Affects, *Fuzzy Logic algorithms, CRISP-DM, Data Science*

INTRODUCTION

Programming has become one of the most interesting learning for many students. If the lecturer's performance is good then the results of the teaching process will also be good, this significantly affecting the quality of graduates produced in an educational institution. Lecturer's performance is important in achieving optimal learning goals. Programming one of the courses required to be taken on the program of the study of informatics engineering. In programming courses, the material discussed in the learning process relates to instructions for execution by computers. Programming is a course that students must complete. If the programming course is not completed (failed) then it is not allowed for students to take other related courses. *(M. Sungkar et al., 2021).*

Data science is used for data grouping to extract knowledge or information from data. Data science processes must be measurable, reliable and meet agreed standards. CRISP-DM is the standardization of data science used in data projects to produce optimal results that can answer questions of a problem that is to be solved. There are the various stages in the CRISP-DM method, namely Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment. *(Shedriko et al., 2022).* With a relatively large number of students, the presence of a tool about student graduation estimates in a course strongly supports the teaching process. So that research is conducted on students who attend Web Proggramming courses with several classes taken as samples. In this homogeny data, a total of 300 students who will be observed by other assessment parameters are then considered as training data. *(P. Rohmawan, 2022)*

Create prediction programs about graduation of 1st to 4th graders based on web programming lecture

*name of corresponding author

taking using CRISP-DM method. Meanwhile, the benefit of this study is helping the programming experts of Prima University Indonesia to predict the graduation rate of students taking web programming courses. As a comparative evaluation of various other approaches.

LITERATURE REVIEW

The data science discussed here is the CRISP-DM method. Data science deals with CRISP-DM frameworks that the project develops from a process perspective and manages the overall flow and various steps in data mining and data science According to John Rollins, CRISP-DM is a data science related to the project-developed framework from a process perspective that manages the overall flow and various steps in data mining and data science. According to Hubber (2018) in a study entitled The Influence of Teacher Quality on Graduation of Programming College through Data Science Approach Using CRISP-DM Method. based on the results of tests that researchers have conducted on questionnaires demonstrated that the method tested using the CRISP-DM method. According to Yerik Afrianto Singgaen. (2023) the algorithm performance evaluation results show that DT, k-NN and SVM algorithms can be used as models in CRISP-DM framework. Altrhythm DT when using SMOTE operators where operating values are 93.25%, precision is 88.74%, recall is 99.10% and f-measure is 93.62%. According to Agung Pambudi, Zaenal Abidin, Permata (2022) The implementation of CRISP-DM modeling results using multiple linear regression and K-Fold got R2 value of 100% with MAPE value of 0.0013%. The obtained model has a ROC-AUC value of 1,000 and an RMSE value of 0.0548 which is an exceptional regression. The maximum value obtained from the index_individual feature is 0.18388841. (John Rollins, 2022)

METHOD

The data in this study were obtained from the documentation of data collection results through questionnaires that had attended 300 respondents' web programming lectures *(Suhanda, I. Kurniati, 2020)*

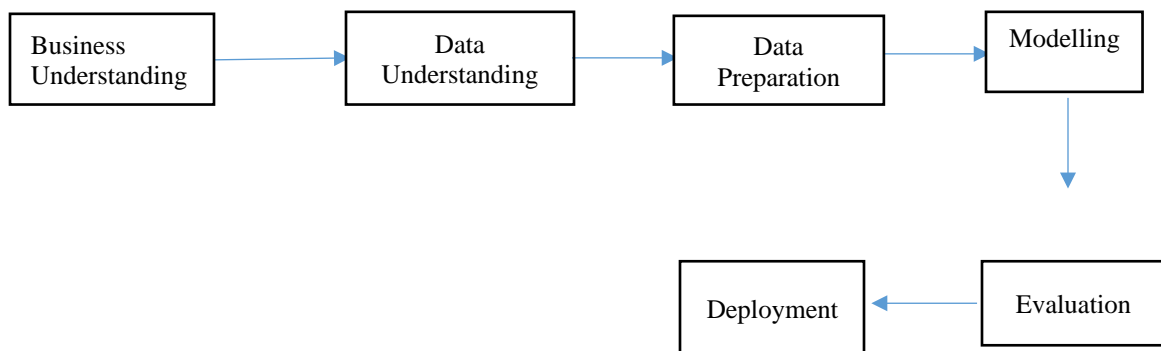The steps in this methodology are:

Figure 1. Work Prosedure of CRISP-DM Methods

Business Understanding

Between entering and graduating students, there are no statistics noteworthy. This is evident from the currently accessible accreditation data, where it is known that the percentage of scholars who have graduated on time in the past two years has not exceeded 50%. Using the CRISP-DM approach, this study investigated the impact of student graduation on the caliber of instructors in web programming courses.

Data Understanding

Researchers analyze students' academic information for 2020–2022 received from PRIMA University Indonesia to find out the impact of students' graduation from teaching quality on web programming

*name of corresponding author

courses. Two Microsoft Excel documents containing the collected information are displayed. A total of 4205 rows and 23 columns of data, including major_choice, class_year, nim, name, gender, province, ipsmt_1 to ipsmt_14, GPA, and status, became the first document of student information for 2020–2022. A total of 4205 rows of data and 9 columns, including the year_class, major, choice_type, name, gender, status, school_type, and principal_school, form the student data for the 2020–22 class, which is the second document.

Data Preparation

This study could not directly use 4205 rows and 32 columns of student academic data from the Data and Information Center (PUSTIPANDA) PRIMA Indonesia University for the 2012–2014 academic year.

Modelling

Different modeling methods are selected and applied and the parameters calibrated to the optimal values are performed at this stage.

Evaluation

At this point, a model that looks very credible based on data analysis was created. Before adopting a model, it is essential to thoroughly evaluate and examine all stages of development to determine whether the model actually achieves business objectives.

Deployment

At this point, user testing is performed to ensure that the results match the user's needs and expectations.

At this stage, the process of collecting data on qualitative and quantitative methods to observed and analyze the graduation scores of students taking web programming courses was carried out. From the results of this test, the total value of each respondent will be determined.

The data in this study were obtained from the documentation of data collection results through questionnaires that had attended 300 respondents' web programming lectures, Data Collection, At this stage, a data collection process for qualitative and quantitative methods to observe and analyze graduation scores of students taking web programming courses was carried out. Problem Analysis, At the problem analysis stage, the author analyzed the data obtained from the research done at the data collection stage to find out the problems that occurred when predicting the graduation scores of students who took web programming courses at University Prima Indonesia. Analysis of Applying Methodology, At this stage, the author implements the CRISP-DM method by deciphering the steps taken to complete CRISP-DM in learning more effectively based on student course scores. Testing,

The test is the stage that the author performs to test the data obtained on the prediction of the results of the CRISP-DM method as well as its flaws in the research to be developed in subsequent studies. Documentation, Documentation is the stage in which the author prepares this research report according to the research topic.

The testing process will use the data from the interview results of each student's statements of their respective academic conditions. From the results of this test, the total value of each respondent will be known. *(Suhanda, I. Kurniati, 2020)*

In addition to the *CRISP-DM* method, The author also uses *Fuzzy algorithms.* A fuzzy system is a system contructed with a clear definition, way of working and description based on fuzzy logic theory. The steps in Fuzzy logic methods are as follows :

Fuzzy Set Formation, both input and output variables are divided into one or more corresponding fuzzy sets. Establishing An Implication Function theimplication function used is minus. Composition Of Fuzzy Rules For Inference, The rule of inference on fuzzy methods is the max method, the maximum value of the rule, then using it to modify the fuzzy region and apply it to output by using a union operator.

Fuzzy logic is an appropriate way to map an input space. Mapping or mapping of input and output relationship of a system based on input – output data. Fuzzy logic is a fuzzy inference method for rules reprented in the form of IF – THEN, where the system's output is not a fuzzy set, but rather a constant. The prediction of a teacher's graduation to a programming course is important because it provides information about the teaching level of one of the lecturers in a programming course. This is very useful for teachers to improve teaching quality. Graduation predictions are also useful to provide recommendations for students who want to find information about the teacher. According to Agung Pambudi, Zaenal Abidin, Permata (2022) implementasi of CRISP-DM modeling results using *multiple*

*name of corresponding author

linear regression and K-FOLD was obtained an R2 value of 100% with a MAPE value of 0.0013%. The obtained ROC-AUC value is 1,000 and RMSE value is 0,0548 an exceptional regression *(excellent regression).*

The maximum value obtained from the index_individual feature is 0.18388841. *(Agung Pambudi, Zaenal Abidin, Permata et al 2022).* CRISP-DM is the standardization of data science used in data projects to produce optimal results that can answer questions of a problem that is to be solved. *(Shedriko and M. Firdaus et al 2022)*

Fuzzy logic similar to the concept of human thinking can present human knowledge into mathematical form by more closely resembling human thingking. Fuzzy logic is a methodology of counting words (linguistic variable). (Arman dan Defiariany et all 2020)

RESULT

Suppose a student fills out a questionnaire for a lecturer with a total lecturer score of 4500. It is possible to calculate the degree of membership of each set by means of total fuzzyness of the lecturer's scores.

The fuzzy set generated by the composition process serves as an input to the defuzzification process, while the value (crisp) serves as an output. In the equation RU(k) = IF x1 is A1k and then... and xn is Ank, the fuzzy IF- THEN THENY rule is Bk, where i = 1, 2,..., n and x = (x1, x2,..., xn) U and y V are respectively input and output variables (characteristics) of the fuzzy system. A1k and Bk, respectively, are fuzzy sets in Ui R (U and V are physical domains). The defuzzifier in the above equation is described as a mapping from the firm point y*V to the fuzzy set Bk on VR (which is the result of fuzzy inference). Defuzzification in the Fuzzy Logic approach is defined as a mapping from the fuzzy set Bk in VR (which is the output of fuzzy inference) to the point crispy * V. In the Fuzzy logic method the defuzzification is performed by calculating the Weight Average (WA):

$$WA = \frac{\alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 \dots + \alpha_n z_n}{\alpha_1 + \alpha_2 + \alpha_3 + \dots \alpha_n}$$

With α_n: rule predicate value to- n

z_n : output value index (default) to- n

Table 1. Lecture's Grade

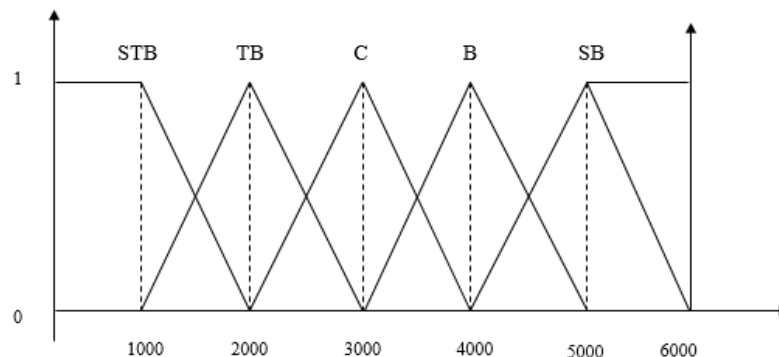| Number	Value Range	Information	Code
1	5000-6000	Highly Efecctive	SB
2	4000-4999	Impact	B
3	3000-3999	Enound	C
4	2000-2999	Happening	TB
5	0-1999	Very Not Effect	STB

Figure 1. Satisfaction Value Variable Input

Lecturer's value variable with weight (x)= 4500:
The total value of lecturer variable membership function

*name of corresponding author

$$\mu_{STB}[x] = \begin{cases} 1; & x \le 1000 \\ \dfrac{2000-x}{1000}; & 1000 \le x \le 2000 \\ 0; & x \ge 2000 \end{cases}$$

$$\mu_{STB}[4500] = \begin{cases} 1; & x \le 1000 \\ \dfrac{2000-x}{1000}; & 30 \le x \le 2000 \\ 0; & x \ge 2000 \end{cases}$$

$$\mu_{STB}[4500] = 0$$

$$\mu_{TB}[x] = \begin{cases} 0; & x \le 2000 \\ \dfrac{x-2000}{3000-2000}; & 2000 \le x \le 3000 \\ \dfrac{3000-x}{3000-2000}; & 2000 \le x \le 3000 \end{cases}$$

$$\mu_{TB}[4500] = \begin{cases} 0; & x \le 2000 \\ \dfrac{x-2000}{3000-2000}; & 2000 \le x \le 3000 \\ \dfrac{3000-x}{3000-2000}; & 2000 \le x \le 3000 \end{cases}$$

$$\mu_{TB}[4500] = 0$$

$$\mu_{c}[x] = \begin{cases} 0; & x \le 4000 \ or \ x \ge 3000 \\ \dfrac{x-4000}{4000-3000}; & 3000 \le x \le 4000 \\ \dfrac{5000-x}{5000-3000}; & 4000 \le x \le 5000 \end{cases}$$

$$\mu_{c}[4500] = \begin{cases} 0; & x \le 4000 \ or \ x \ge 3000 \\ \dfrac{x-4000}{4000-3000}; & 3000 \le x \le 4000 \\ \dfrac{5000-x}{5000-3000}; & 4000 \le x \le 5000 \end{cases}$$

$$= (5000\text{-}4500)/(5000\text{-}3000)$$

$$= 0.25$$

$$\mu_{B}[x] = \begin{cases} 0; & x \le 5000 \ or \ x \ge 4000 \\ \dfrac{x-5000}{6000-5000}; & 5000 \le x \le 6000 \\ \dfrac{7000-x}{7000-6000}; & 6000 \le x \le 7000 \end{cases}$$

$$\mu_{B}[4500] = \begin{cases} 0; & x \le 5000 \ or \ x \ge 4000 \\ \dfrac{x-5000}{5000-4000}; & 5000 \le x \le 6000 \\ \dfrac{7000-x}{7000-6000}; & 6000 \le x \le 7000 \end{cases}$$

$$= 0$$

$$0; \quad x \le 6000 \ or \ x \ge 5000$$

*name of corresponding author

$\mu_{SB}[x]$ $= \{ \dfrac{x-5000}{7000-5000} ; 5000 \le x \le 6000$

$1; 6000 \le x \le 7000$

$0; x \le 6000 \; or \; x \ge 5000$

$\mu_{SB}[4500]$ $= \{ \dfrac{x-5000}{6000-5000} ; 5000 \le x \le 6000$

$0; 6000 \le x$

$= 0$

Value α Predicate adalah:

α1 = min (0.25)

α1 = 0

$z1 = \dfrac{(0.25)}{0.25} x 1000$

z1 = 1000

Therefore, the effect of student graduation on web programming courses can be seen using Table 2 which is the level of "Very Influential".

Figure 2. Questionnaire Data Display

In this study, questionnaire display data were divided into 5 ranks as shown in the previous figure, with the highest number of ranks at 100 and the lowest rank at 0. However, making predictions with 5 ratings proved difficult and there was also a lack of data at low ratings.

Figure 3. Interview Data Input Display

*name of corresponding author

In this study, the interview data which are CRISP-DM variables, were divided into two categories. Very influential for 0-29, no influence fo 30-49, sufficient for 50-59, influential for 60-69 and very influential 70-100. From this data, 300 respondents were

Figure 4. Processing Data Display

This the prediction result of the CRISP-DM and Fuzzy Logic methods in this study.

Figure 5. Prediction Result Display

Result of this methods is, Improve the accuracy of the ranking by further developing the criteria for evaluating lecturers. From the results of the comparison, it is possible to predict the graduation rate of students to the web programming course. The results of the experiment were that 72% of the students received highly influential predictions, 12% Influential, 7% Sufficient, 5% Influential and 4% Highly Influential

DISCUSSIONS

Based on the results of the crispdm method, it can be said that the crispdm model has the advantage of producing accurate respondent data. This is evident from respondents' results and model validation accuracy. attributes. It produces selected characteristics that produce ideal values for predicting graduation. Department, choice, class, gender, province, school type, school major, ipsmt_1 to ipsmt_3, and this property. After creating a dataset of 1442 records with 11 selected attributes, testing and changing the data back to evaluate which transform resulted in the most accuracy results. 70% of training data (1009 records) and 30% of testing data (433 records) are the percentage of gaps used for all experiments. WEKA version 3.8 with 11 characteristics was used for research. Some of the experiments and data changes made were to produce the best.

*name of corresponding author

CONCLUSION

In this study, the conclusion can be made.

The application of student graduation effects from teacher quality to lecture web programming using the CRISP- DM method can display prediction information, namely teacher quality influence information on student graduation.Can provide student questionnaire data in the form of lecturer's behavior that is capable of each lecture. Can implement Fuzzy Logic algorithms on the process of determining the effect of faculty quality on Web Programming course graduation. The results of 300 respondents' tests were obtained, 72% of the students received Very Influential predictions, 12% Influential, 7% Sufficient, 5% Uninfluential and 4% Very Uninfluential.

REFERENCES

Arora, S. *et al.* (2022) 'Reasoning over Public and Private Data in Retrieval-Based Systems', pp. 1–27. Available at: http://arxiv.org/abs/2203.11027.

Bayu Baskoro, B., Susanto, I. and Khomsah, S. (2021) 'Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR)', *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 3(2), pp. 21–029. doi: 10.20895/INISTA.V3.

Clark, K. *et al.* (2020) 'ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators', pp. 1–

18. Available at: http://arxiv.org/abs/2003.10555.

Divya, K. *et al.* (2020) 'An Interpretation of Lemmatization and Stemming in Natural Language Processing', *Journal of University of Shanghai for Science and Technology*, 22(10), pp. 350–357. Available at: https://www.researchgate.net/publication/348306833.

HaCohen-Kerner, Y., Miller, D. and Yigal, Y. (2020) 'The influence of preprocessing on text classification using a bag-of-words representation', *PLoS ONE*, 15(5), pp. 1–22. doi: 10.1371/journal.pone.0232525.

Jayanto, R., Kusumaningrum, R. and Wibowo, A. (2022) 'Aspect-based sentiment analysis for hotel reviews using an improved model of long short-term memory', *International Journal of Advances in Intelligent Informatics*, 8(3), pp. 391–403. doi: 10.26555/ijain.v8i3.691.

Kusumaningrum, R. *et al.* (2021) 'Sentiment analysis of Indonesian hotel reviews: from classical machine learning to deep learning', *International Journal of Advances in Intelligent Informatics*, 7(3), pp. 292–303. doi: 10.26555/ijain.v7i3.737.

Moch Sambas *et al.* (2022) 'Analysis of Lodging and Competition on the Island of Bali during Covid-19 with Big Data', *International Journal of Travel, Hospitality and Events*, 1(3), pp. 214–228. doi: 10.56743/ijothe.v1i3.172.

Morama, H. C., Ratnawati, D. E. and Arwani, I. (2022) 'Analisis Sentimen berbasis Aspek terhadap Ulasan Hotel Tentrem Yogyakarta menggunakan Algoritma Random Forest Classifier', *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 6(4), pp. 1702–1708.

Priyantina, R. A. and Sarno, R. (2019) 'Sentiment analysis of hotel reviews using Latent Dirichlet Allocation, semantic similarity and LSTM', *International Journal of Intelligent Engineering and Systems*, 12(4), pp. 142–155. doi: 10.22266/ijies2019.0831.14.

Utami, T. A. (2021) 'Sentiment Analysis of Hotel User Review using RNN Algorithm', *International Journal of Informatics and Computation*, 3(1), p. 30. doi: 10.35842/ijicom.v3i1.34.

Yao, L. (2022) 'Sentiment analysis based on CNN - LSTM hotel reviews', *Journal of Physics: Conference Series*, 2330(1). doi: 10.1088/1742-6596/2330/1/012018

*name of corresponding author