

## Development of Graduation Prediction Model for Industrial Engineering Students Using Decision Tree

Rifki Muhendra

Program Studi Teknik Industri Universitas Bhayangkara Jakarta Raya  
Jl. Perjuangan No.81, RT.003/RW.002, Marga Mulya, Kec. Bekasi Utara,  
Kota Bekasi, Jawa Barat, Indonesia  
E-mail [rifki.muhendra@dsn.ubaharajaya.ac.id](mailto:rifki.muhendra@dsn.ubaharajaya.ac.id)

*Article received : December 2021; revised : January 2022 ; accepted : March 2022*

**DOI : 10.17977/um025v6i12021p28**

---

**Abstract:** Data mining is one of the rapidly growing data processing sciences. One application of data mining is to build a model of student graduation criteria. In this study, a prediction model for student graduation criteria was developed using a decision tree (DT). Several factors that influence the graduation criteria of students studied in this study include GPA, field of study, age, number of credits completed, and so on. The development of this model uses the open source Rapidminer software which is proven to have ease in processing but is very good at producing models. There are 3 prediction models produced, namely the DT model using the Gini Index method, Information gain and gain ratio. The resulting model has a fairly large root distribution in the predicate is very satisfactory. This means this predicate in the process does quite a lot of iterations. These three models can be used to predict student graduation because they have an accuracy and Kappa value greater than 80%. This shows that this model has a high level of confidence and can describe what is happening. The Gini Index model has the highest accuracy and kappa value compared to the information gain and gain ratio models with accuracy and kappa values of 0.963 and 0.932, respectively. This shows that the Gini Index model is superior for processing large data.

**Keywords:** Data Mining; Prediction Model; Student Graduation Criteria; Decision Tree

graduating on time is every student's dream. On-time graduation is usually defined as the length of time for completion of studies at a university in 8 semesters or 4 years. If a student can finish faster than 8 regular semesters, it is classified as a faster graduate. However, if a student completes study for more than 8 semesters, it is usually classified as late. Graduating on time is considered an ideal graduation criterion because students are considered to have met the initial target before the lecture, have sufficient experience during the lecture period, and of course, meet the financial budget prepared before entering the lecture(Hendley, 2021).

With the development of science, especially data processing, statistics, and data mining, timely graduation can be calculated earlier or modeled. Some of the advantages of this graduation model include being able to predict the type of graduation criteria for a student, as a guide for new students to complete lectures and for tertiary institutions, being able to give early warning to students when the student is undergoing studies that are not as expected. The student graduation prediction model usually utilizes previous student graduation data which is then processed using a statistical method to produce a new rule or pattern of knowledge(Bassi et al., 2019; Qin and Phillips, 2019; Lagman et al., 2020).

Prediction models for timely graduation can be built using several methods, including Naïve Bayes, Artificial Neural Network, and Decision tree. The method or Naive Bayes algorithm is known

as a classification algorithm that can predict the class membership of data based on probability calculations (Lagman et al., 2019; Sutoyo and Almaarif, 2020; Marzuqi, Laksitowening and Asror, 2021; Sembiring and Tambunan, 2021). An artificial Neural Network (ANN) or artificial neural network is a network of a group of small processing units that are modeled based on the behavior of human neural networks (Arsad, Buniyamin and Manan, 2014; Bassi et al., 2019; Olalekan, Egwuiche and Olatunji, 2020). Decision Tree is the algorithm that is most often used to develop a classification model for a set of data. Decision Tree abbreviated as DT is considered to have a predictive model with accuracy, ease of interpretation, and stability (N.Undavia, M. Dolia and P. Shah, 2013; Undavia, Dolia and Patel, 2016; Romadhona, Suprapedi and Himawan, 2017; Mulia and Muanas, 2021). In addition, DT is also effective in adjusting non-linear relationships because it can solve data adjustment challenges, such as regression and classification.

In this study, a prediction model for on-time graduation of Industrial Engineering students from Bhayangkara University, Greater Jakarta was developed using DT. Industrial Engineering has a relatively large number of students compared to other study programs at the University. This predictive model is expected to increase students' timely graduation and also as a guide for improving lecture infrastructure at the University. The newness of this study compared to previous research is the criteria for data collection, ease of data processing, and the expected level of accuracy of more than 90%. This means that the developed model can represent the actual situation.

## METHODS

This type of research method is a structured experimental research type as shown in Figure 2. This research starts from the study of literature on theories that affect student graduation. This theory is then used as the basis for developing research instruments. Then data was collected on student graduation at the Faculty of Engineering, Industrial Engineering Study Program, Bhayangkara University, Greater Jakarta. The number of student graduation data collected must be greater than 255 data. We assume that the more data collected the better the resulting model. Each student's graduation data consists of data on identity no, gender, age of entering college, number of credits completed during the education period, the field of research, GPA, having a certificate of expertise or not, and also graduation criteria as shown in Figure 4.

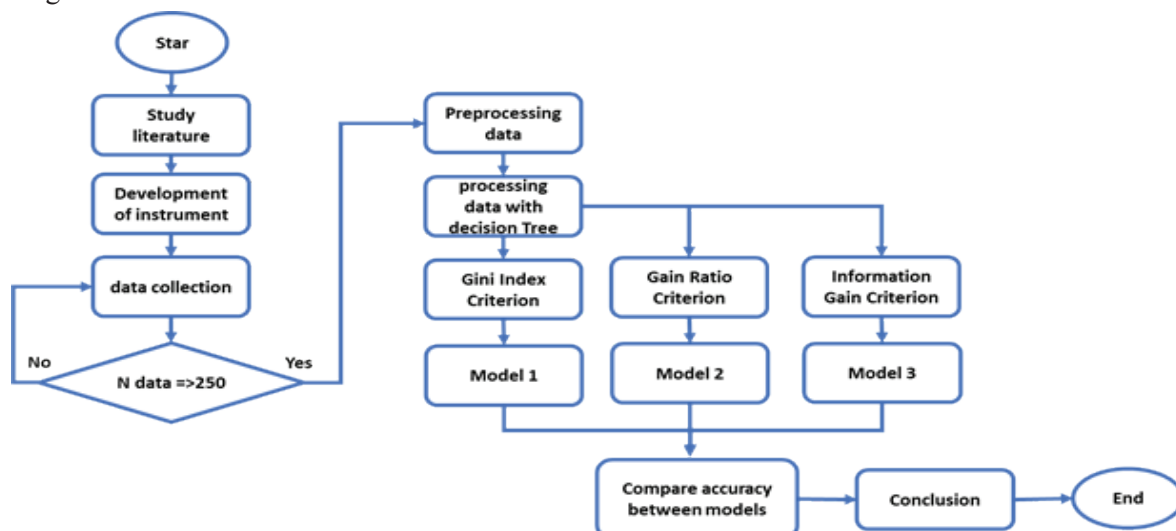


Figure 2. Research flow

### Factors that affect student graduation

College can be a stressful time for many students. Stress can come from many sources such as course load, academic requirements, or adjustment to living away from home. With the high cost of education and living costs, many students are now faced with the financial burdens and pressures that

come with paying for college. In reaction to financial stress, students respond in a number of ways: some find jobs or seek help from their parents, and in more drastic circumstances, students have turned to crime and even suicide (Johnson, 2005).

On student engagement, their individual practices and experiences also influence their studies. Sidhu et al. (2017) briefly showing students as key players in successfully completing postgraduate studies required to keep in mind several aspects. First and foremost, they must be aware of their role and responsibilities as a graduate student. In addition, they need to learn to work towards developing interesting, meaningful, and positive working relationships with their superiors. Finally, they need to ensure their readiness and competence required to carry out a postgraduate degree as Jeyaraj (2018) emphasizes. Their study also revealed that while students considered themselves to be moderately prepared in terms of their critical reading, writing, research and conceptual skills, their supervisors felt otherwise. Supervisors highlight that the majority of graduate students have limited readiness and competencies required to handle postgraduate studies.

In this context, it is more important to determine students' readiness to carry out research work which in turn motivates them throughout the process. Students may be in a different emotional state when they enroll as graduate students. Students with high awareness of the perception of challenges can develop more skills (Picken, 2015). So, their emotional state needs to be taken care of. Placing too high expectations on a challenge may be worse at developing the skills they need in writing.

A student's willingness and commission in research also depend on attitude. Their attitude should be built among postgraduate students early in their studies (Papanastasiou, 2005; Siamian et al., 2016). Similarly, sharing and caring for their emotions, beliefs, behaviors and interactions will determine a positive attitude in doing research work (Zan & Martino, 2007). In short, graduate students need to work in a communicative society. The role of the community aims to promote useful research in terms of its contribution to life and relevance in addition to reducing anxiety facing the difficulties of conducting research. The most important thing is to experience emotional freedom in the learning environment (Varvogli & Darviri, 2011). Thus, students need more communication and support between each other as well as with the institution.

In general, there are internal and external factors that affect a student's graduation (Espinoza *et al.*, 2019; Srinadi and Nilakusmawati, 2020; Marzuqi, Laksitowening and Asror, 2021). Internal factors are factors that come from within the students themselves, such as gender, age, GPA, and so on. External factors are factors that come from outside of students, such as finances, learning environment, and the number of outside lectures that are followed. In addition to the two factors above, other factors affect student graduation. This factor is usually a student's choice of a situation. One example is the field of research. The field of research is a choice or decision taken by students. Sometimes the university frees students to take a topic with a preferred field of research, sometimes directly determines, and also sometimes there is communication between students and the campus before determining the research field.

## **Data Collection**

In this study, the data collected came from data from graduates of the Bachelor of Industrial Engineering, Bhayangkara University, Greater Jakarta for 4 graduation periods. This data consists of personal data of students who have graduated such as gender, age, number of credits taken and others in part. However, this data is also complemented by other data that does not include graduate data, including research fields, ownership of professional certificates and so on. The number of industrial engineering graduate student data collected which is used as research data is as many as 255 students.

## **Preprocessing Data**

Data preprocessing is a technique used to convert raw data into useful and efficient formats. This initiative is necessary because raw data is often incomplete and has inconsistent formats. The quality of the data itself has a direct correlation with the success of any project that involves data analysis. Preprocessing itself involves data validation and imputation. The purpose of validation is to assess the

level of completeness and accuracy of the filtered data. On the other hand, the purpose of imputation is to correct errors and enter missing values, either manually or automatically through a business process automation (BPA) program. Data preprocessing is usually used in rule-based and database-based applications.

Several stages of data preprocessing include:

### Data cleaning

In this work stage, the data is cleaned through several processes such as filling in missing values, smoothing noisy data, and resolving inconsistencies found. Data can also be cleaned by dividing into segments of similar size and then binning. You can also adjust it with linear or multiple regression functions (regression), or by grouping them into groups of similar data (grouping).

### Data integration

The next work stage in the data preprocessing process is data integration. Here, data with different representations are pooled and all internal conflicts in them are resolved. This work stage is an advanced process of data cleansing with the aim of making the data smoother.

### Data transformation

Data transformation is the next stage of work in the data preprocessing process. At this stage, the data will be normalized and generalized. Normalization itself is a process where the company ensures that there is no redundant data. All data will be stored in one place and all dependencies must be logical. This step is also taken to transform the data into a form suitable for the mining process.

### Data reduction

The final work stage in the data preprocessing work process is data reduction. This technique is in the form of increasing storage efficiency and reducing data representation in the data warehouse.

N	Jenis Kelami	Kriteria Jumlah SKS Lulus	PREDIKAT	Kriteria Usia Masuk Kulial	Bidang Penelitian	Sertifika	Kriteria Lulus
1	L	LEBIH	Memuaskan	20-25 tahun	managemen industri	Tidak	Tidak Tepat Waktu
2	L	PAS	Sangat Memuaskan	Kecil dari 20 tahun	manufaktur	Tidak	Tidak Tepat Waktu
3	L	LEBIH	Memuaskan	Kecil dari 20 tahun	managemen industri	Tidak	Tidak Tepat Waktu
4	L	LEBIH	Memuaskan	Kecil dari 20 tahun	managemen industri	Tidak	Tidak Tepat Waktu
5	L	LEBIH	Sangat Memuaskan	Kecil dari 20 tahun	manufaktur	Tidak	Tidak Tepat Waktu
6	L	PAS	Sangat Memuaskan	20-25 tahun	managemen industri	Tidak	Tidak Tepat Waktu
7	L	LEBIH	Sangat Memuaskan	20-25 tahun	managemen industri	Tidak	Tidak Tepat Waktu
8	L	LEBIH	Sangat Memuaskan	20-25 tahun	managemen industri	Tidak	Tidak Tepat Waktu
9	L	LEBIH	Sangat Memuaskan	20-25 tahun	manufaktur	Tidak	Tidak Tepat Waktu
10	L	LEBIH	Memuaskan	20-25 tahun	managemen industri	Tidak	Tidak Tepat Waktu
11	L	LEBIH	Sangat Memuaskan	20-25 tahun	manufaktur	ada	Tidak Tepat Waktu

Figure 4. Form of data collected

After the data has been collected, preprocessing the data is carried out. This process is a process of cleaning data into data that is more structured, neat, and easy to process further. These data are then processed using DT to produce a classification model of student graduation criteria. There are three types of DT processing methods used, namely Gini index, gain ratio, and information gain. In each of these methods, accuracy and kappa measurements were carried out. Accuracy can be interpreted as a measure of the accuracy of the model to the actual data. Kappa is a statistical method that can measure the degree of agreement of two raters in classifying objects. Both accuracy and kappa have a range from 0 to 1, where if the accuracy and kappa data are close to 1, it can be concluded that the level of confidence in the model is getting higher. Then, from the comparison of the 3 measurements of accuracy and kappa of the resulting models, one model that has the best value will be selected and concluded.

## Data Mining

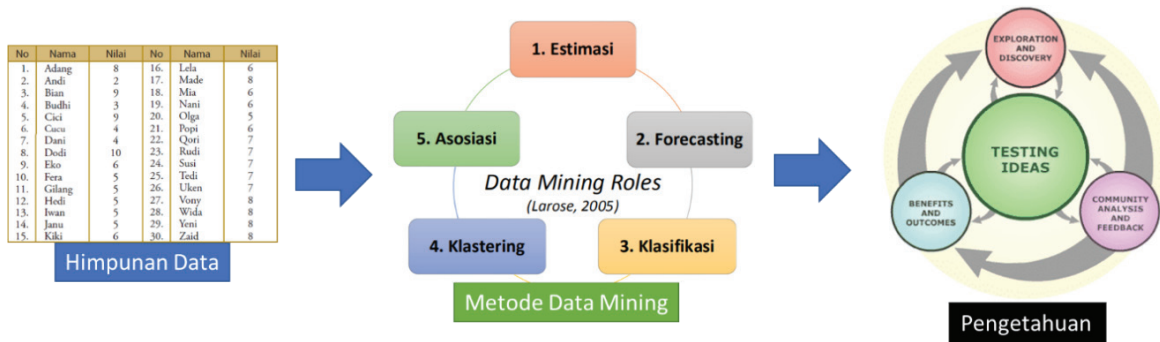


Figure 5. Data Mining Concept

Rapidminer is one of the data mining software that is widely used by scientists and processions. Some of the advantages of RapidMiner compared to other data mining software are that it is easy to use, has many data processing methods, attractive visualizations, and so on. In addition, Rapidminer is also available in free form for education, community, and development services (Aprilla Dennis, 2013; Ristoski, Bizer and Paulheim, 2015).

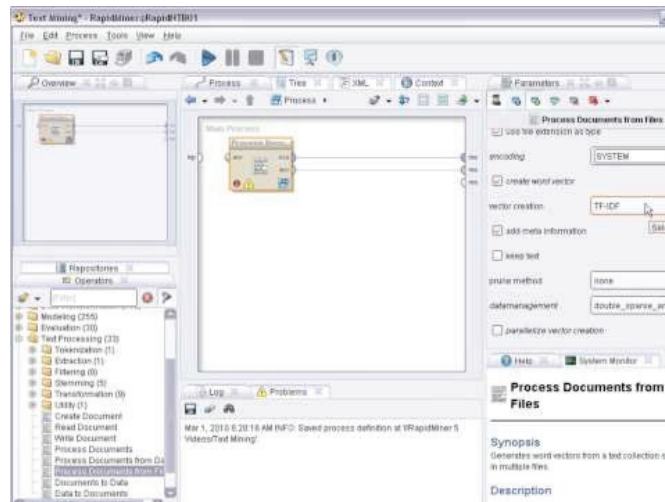


Figure 6. Interface RapidMiner

## Decision Tree

The decision tree is the most widely used algorithm for classification problems. A decision tree consists of several nodes, namely the tree's root, internal nodes, and leaves (N. Undavia, M. Dolia and P. Shah, 2013). The concept of entropy is used to determine which attributes a tree will split (split). The higher the entropy of a sample, the more impure the sample. The results of the decision quality obtained from the decision tree method are very dependent on how the tree is designed. So if the decision tree made is less than optimal, it will affect the quality of the decisions obtained. The application of feature selection in this study is to calculate the information gained on each attribute. Information gain of an attribute, obtained from the entropy value before the separation minus the entropy after the separation [13]

$$Entropy (S) = \sum_{i=1}^k (P_i) \log_2 (P_i)$$

P<sub>i</sub> is the proportion of data S with class I, and k is the number of classes at the output S. The entropy value after splitting is as follows:



$$Entropy(S, A) = \sum_{i=1}^V \frac{|Sv|}{S} \times Entropy(Sv)$$

Creating a decision tree is selecting the attributes that should be tested at each node. This process is called Information Gain, which is useful for determining which attributes are used for each node. Information gain can be calculated using a unit called Entropy

$$Gain(S, A) = Entropy(S) - Entropy(S, A)$$

In the Decision tree method, several attributes will be found of numeric (continuous) and discrete types. To determine the type of attribute can use the Gain Ratio. Gain Ratio can give a slightly more specific value than information gain. To perform the Gain Ratio calculation, first, calculate the Split Information value.

$$Split\ Information = - \sum_{i=1}^c \frac{Si}{S} \log_2 \frac{Si}{S}$$

Si to Sc is the sub-subject c resulting from solving attribute A which has many C values. So that the determination of the Gain Ratio(Roy *et al.*, 2019) can use the equation:

$$Gain\ Ratio = \frac{Gain(S, A)}{Split\ information(S, A)}$$

The Gini Index is also commonly used to provide a more accurate model shape. The Gini index can measure the degree of probability of certain variables being misclassified when randomly selected(Arabshahi and Fazlollahtabar, 2018). It is suitable for a larger number of partitions or attributes.

$$Gini\ Index = 1 - \sum_{i=1}^n (Pi)^2$$

where pi is the probability of an object being classified to a particular class.

## RESULTS

At this stage, a model of student graduation criteria has been produced in the Industrial Engineering study program, Bhayangkara University, Greater Jakarta as shown in Figure 5. In this model, there are 3 types of student graduation criteria, namely not on time, on time, and faster. The parameter that greatly affects the graduation criteria for this student is the student's GPA. While the parameter that has little contribution in determining the criteria is the age of entering college. Some of the conclusions generated by this model include: If a student has a satisfactory predicate or a GPA less than 3.00 then the graduation criteria are not on time. If the student has a Praise predicate, then the graduation criteria are on time. If the student has a very satisfactory graduation predicate, the field of research will be reviewed first, the age of entering college, certificate ownership, the number of credits completed, and others in part to determine the graduation criteria.

```

Tree
PREDIKAT = Memuaskan: Tidak Tepat Waktu {Tidak Tepat Waktu=40, Lebih Cepat=0, Tepat Waktu=0}
PREDIKAT = Pujian: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=18}
PREDIKAT = Sangat Memuaskan
  Bidang Penelitian = K3 & Ergonomi
    Kriteria Usia Masuk Kuliah = 20-25 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=17, Lebih Cepat=0, Tepat Waktu=2}
    Kriteria Usia Masuk Kuliah = Besar dari 25 tahun: Lebih Cepat {Tidak Tepat Waktu=1, Lebih Cepat=2, Tepat Waktu=0}
    Kriteria Usia Masuk Kuliah = Kecil dari 20 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=17, Lebih Cepat=0, Tepat Waktu=3}
  Bidang Penelitian = Manajemen industri|
    Sertifikat = Tidak
      Jenis Kelamin = L
        Kriteria Jumlah SKS Lulus = LEBIH: Tepat Waktu {Tidak Tepat Waktu=4, Lebih Cepat=0, Tepat Waktu=5}
        Kriteria Jumlah SKS Lulus = PAS
          Kriteria Usia Masuk Kuliah = 20-25 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=1, Lebih Cepat=0, Tepat Waktu=1}
          Kriteria Usia Masuk Kuliah = Kecil dari 20 tahun: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=2}
        Jenis Kelamin = P: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=1}
      Sertifikat = ada: Tidak Tepat Waktu {Tidak Tepat Waktu=2, Lebih Cepat=1, Tepat Waktu=1}
    Bidang Penelitian = Manufaktur: Tidak Tepat Waktu {Tidak Tepat Waktu=15, Lebih Cepat=0, Tepat Waktu=0}
  Bidang Penelitian = Rantai pasok & Logistik
    Kriteria Jumlah SKS Lulus = LEBIH
      Kriteria Usia Masuk Kuliah = 20-25 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=8, Lebih Cepat=0, Tepat Waktu=0}
      Kriteria Usia Masuk Kuliah = Kecil dari 20 tahun: Tepat Waktu {Tidak Tepat Waktu=1, Lebih Cepat=0, Tepat Waktu=2}
    Kriteria Jumlah SKS Lulus = PAS: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=1}
  Bidang Penelitian = Rantai pasok & logistik
    Sertifikat = Tidak: Tidak Tepat Waktu {Tidak Tepat Waktu=4, Lebih Cepat=0, Tepat Waktu=0}
    Sertifikat = ada: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=1}
  Bidang Penelitian = manajemen Industri
    Kriteria Jumlah SKS Lulus = LEBIH
      Kriteria Usia Masuk Kuliah = 20-25 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=4, Lebih Cepat=0, Tepat Waktu=2}
      Kriteria Usia Masuk Kuliah = Besar dari 25 tahun: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=1}
      Kriteria Usia Masuk Kuliah = Kecil dari 20 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=4, Lebih Cepat=0, Tepat Waktu=3}
    Kriteria Jumlah SKS Lulus = PAS: Tidak Tepat Waktu {Tidak Tepat Waktu=3, Lebih Cepat=0, Tepat Waktu=0}
  Bidang Penelitian = manajemen industri: Tidak Tepat Waktu {Tidak Tepat Waktu=85, Lebih Cepat=1, Tepat Waktu=0}
  Bidang Penelitian = manufaktur
    Kriteria Jumlah SKS Lulus = LEBIH: Tidak Tepat Waktu {Tidak Tepat Waktu=21, Lebih Cepat=0, Tepat Waktu=3}
    Kriteria Jumlah SKS Lulus = PAS
      Jenis Kelamin = L
        Sertifikat = Tidak
          Kriteria Usia Masuk Kuliah = 20-25 tahun: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=3}
          Kriteria Usia Masuk Kuliah = Besar dari 25 tahun: Tepat Waktu {Tidak Tepat Waktu=0, Lebih Cepat=0, Tepat Waktu=1}
          Kriteria Usia Masuk Kuliah = Kecil dari 20 tahun: Tidak Tepat Waktu {Tidak Tepat Waktu=2, Lebih Cepat=0, Tepat Waktu=1}
        Sertifikat = ada: Tidak Tepat Waktu {Tidak Tepat Waktu=2, Lebih Cepat=0, Tepat Waktu=0}
      Jenis Kelamin = P: Tidak Tepat Waktu {Tidak Tepat Waktu=2, Lebih Cepat=0, Tepat Waktu=0}
  
```

Figure 5. Model of graduation criteria for Industrial Engineering students, Bhayangkara University, Greater Jakarta

From the resulting DT model, a model was developed using the Gini Index, Information Gain, and Gain Ratio methods. Then, each model has measured the value of accuracy and kappa. A comparison of the accuracy and kappa values of each model can be seen in Figure 6.

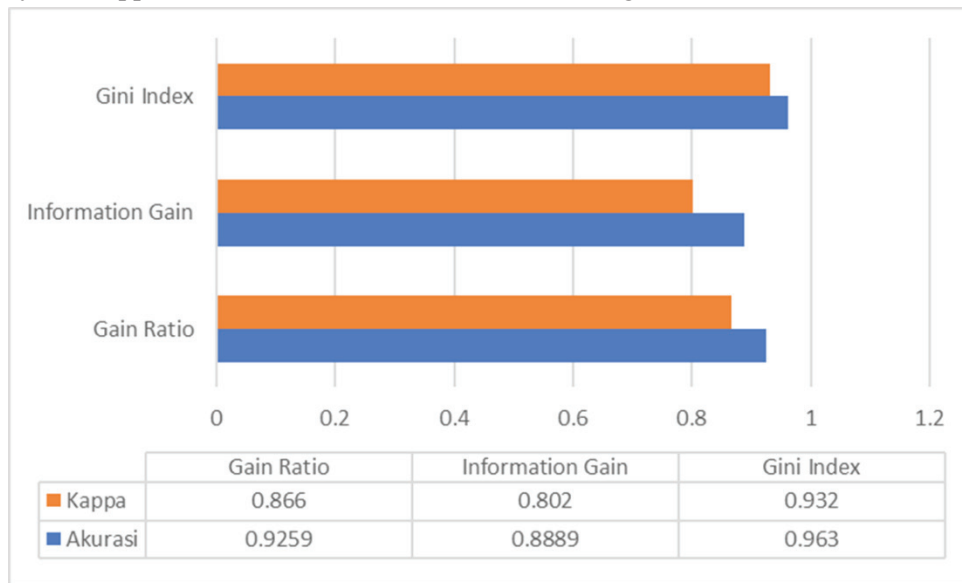


Figure 6. Comparison graph of accurate measurement and Kappa models of student graduation criteria

Based on Figure 6, the DT models of student graduation criteria have high accuracy and kappa values, which are greater than 80% or 0.85. This shows that all three models are reliable. The DT model using the Gini Index has the highest accuracy and kappa, namely 0.963 and 0.932. The DT model using Information Gain has the smallest accuracy and kappa, namely 0.88 and 0.80.

Figure 7 shows a comparison of student graduation criteria with student graduation criteria resulting from the DT model using the Gini Index.

Row No.	Kriteria Lulus	prediction(Kriteria ...	Row No.	Kriteria Lulus	prediction(Kriteria ...
1	Tidak Tepat Waktu	Tidak Tepat Waktu	14	Tepat Waktu	Tepat Waktu
2	Tidak Tepat Waktu	Tidak Tepat Waktu	15	Tepat Waktu	Tepat Waktu
3	Lebih Cepat	Lebih Cepat	16	Tepat Waktu	Tepat Waktu
4	Lebih Cepat	Lebih Cepat	17	Tepat Waktu	Tepat Waktu
5	Tidak Tepat Waktu	Tidak Tepat Waktu	18	Tepat Waktu	Tepat Waktu
6	Tidak Tepat Waktu	Tidak Tepat Waktu	19	Tepat Waktu	Tidak Tepat Waktu
7	Tidak Tepat Waktu	Tidak Tepat Waktu	20	Tepat Waktu	Tepat Waktu
8	Tidak Tepat Waktu	Tidak Tepat Waktu	21	Tepat Waktu	Tepat Waktu
9	Tidak Tepat Waktu	Tidak Tepat Waktu	22	Tepat Waktu	Tepat Waktu
10	Tidak Tepat Waktu	Tidak Tepat Waktu	23	Tepat Waktu	Tepat Waktu
11	Tidak Tepat Waktu	Tidak Tepat Waktu	24	Tepat Waktu	Tepat Waktu
12	Tepat Waktu	Tepat Waktu	25	Tepat Waktu	Tepat Waktu
13	Tepat Waktu	Tepat Waktu	26	Tepat Waktu	Tepat Waktu
14	Tepat Waktu	Tepat Waktu	27	Tepat Waktu	Tepat Waktu

Figure 7. Comparison of student graduation criteria data with the results of the DT Gini Index model

Figure 7 shows 27 comparisons of the original student criteria data with the results of the DT Gini Index model. Of the 27 data, 1 data is not suitable, namely, the 19th data where the original criteria are shown are on time while the model results are not on time. This test shows that the DT Gini Index model is the best produced and has a small error rate.

## DISCUSSION

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, machine learning to extract and identify useful information and related knowledge from large databases (Turban et al. 2005). An easy-to-use open source in the development of predictive models is Rapid Miner. Rapid Miner is a solution for analyzing data mining, text mining and predictive analysis. Rapid Miner uses a variety of descriptive and predictive techniques to provide users with insights so they can make the best decisions.

The data classification process uses the decision tree method to determine the attributes that are very influential in the prediction process. This method is useful for exploring data, finding hidden relationships between a number of potential input variables and a target variable. Because this method combines data exploration and modeling, it is excellent as a first step in the modeling process even when used as a final model for some other techniques. And the results obtained on average are more stable. If you look at the resulting model, the predicate is very satisfying to be a predicate that has quite a lot of distribution. This indicates that students who graduate with very satisfactory predicate more than other predicate. Because of this large distribution, of course, there will be more iterations of model development. And here the role of Rapid Miner becomes more important, namely the process of developing the model to have a high speed without reducing the quality of the resulting model.

There are three models that have been successfully developed. The three models can be accepted in predicting the graduation of Industrial Engineering students at Bhayangkara University, Jakarta Raya because they have an accuracy value above 80%. This means that these models can represent the actual situation. The Gini index model occupies the value with the highest accuracy, which is 96.3%. The Gini index facilitates a larger distribution making it easy to implement. The Gini Index method is used by the CART algorithm. CART was developed to perform classification analysis on response variables,



either nominal, ordinal, or continuous. The principle of this classification method is to separate all observations into two groups of observations and re-segregate the group of observations into the next two groups of observations, until the minimum number of observations is obtained in each subsequent group of observations. The Gini index operates on categorical target variables in terms of “success” or “failure” and only performs binary splitting.

## CONCLUSION

A model of student graduation criteria has been produced in the Industrial Engineering study program, Bhayangkara University, Greater Jakarta. The parameter that greatly affects the graduation criteria for this student is the student's GPA. Some of the conclusions generated by this model include if a student has a satisfactory predicate or a GPA less than 3.00 and not on time. DT models of student graduation criteria have high accuracy and kappa values, which are greater than 80% or 0.85. The DT model using the Gini Index has the highest accuracy and kappa, namely 0.963 and 0.932. based on resulting model, the predicate is very satisfying to be a predicate that has quite a lot of distribution. This indicates that students who graduate with very satisfactory predicate more than other predicate. Because of this large distribution, of course, there will be more iterations of model development. And here the role of rapidminer becomes more important, namely the process of developing the model to have a high speed without reducing the quality of the resulting model.

## REFERENCES

- Aprilla Dennis (2013) ‘Belajar Data Mining Dengan Rapidminer’, *Innovation And Knowledge Management In Business Globalization: Theory & Practice, Vols 1 And 2*, 5(4).
- Arabshahi, H. And Fazlollahtabar, H. (2018) ‘Classifying Innovative Activities Using Decision Tree And Gini Index’, *International Journal Of Innovation And Technology Management*, 15(3). Doi: 10.1142/S0219877018500256.
- Arsad, P. M., Buniyamin, N. And Manan, J. A. B. (2014) ‘Students’ English Language Proficiency And Its Impact On The Overall Student’s Academic Performance : An Analysis And Prediction Using Neural Network Model’, *WSEAS Transactions On Advances In Engineering Education*, 11.
- Bassi, J. S. *Et Al.* (2019) ‘Students Graduation On Time Prediction Model Using Artificial Neural Network’, *IOSR Journal Of Computer Engineering (IOSR-JCE)*, 21(3).
- Espinoza, O. *Et Al.* (2019) ‘Factors That Affect Post-Graduation Satisfaction Of Chilean University Students’, *Studies In Higher Education*, 44(6). Doi: 10.1080/03075079.2017.1407306.
- Han, J., Kamber, M. And Pei, J. (2012) *Data Mining: Concepts And Techniques, Data Mining: Concepts And Techniques*. Doi: 10.1016/C2009-0-61819-5.
- Hand, D. J. (2008) ‘Data Mining: Methods And Models By D. T. Larose’, *Biometrics*, 64(1). Doi: 10.1111/J.1541-0420.2008.00962\_9.X.
- Hendley, K. (2021) ‘Rethinking The Role Of Personal Connections In The Russian Labor Market: Getting A Job As A Law Graduate In Russia’, *Post-Soviet Affairs*, 37(3). Doi: 10.1080/1060586X.2021.1874768.
- Lagman, A. C. *Et Al.* (2019) ‘Embedding Naïve Bayes Algorithm Data Model In Predicting Student Graduation’, *In Pervasivehealth: Pervasive Computing Technologies For Healthcare*. Doi: 10.1145/3369555.3369570.
- Lagman, A. C. *Et Al.* (2020) ‘Classification Algorithm Accuracy Improvement For Student Graduation Prediction Using Ensemble Model’, *International Journal Of Information And Education Technology*, 10(10). Doi: 10.18178/Ijiet.2020.10.10.1449.
- Marzuqi, A., Laksitowening, K. A. And Asror, I. (2021) ‘Temporal Prediction On Students’ Graduation Using Naïve Bayes And K-Nearest Neighbor Algorithm’, *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 5(2). Doi: 10.30865/Mib.V5i2.2919.
- Mulia, I. And Muanas, M. (2021) ‘Model Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree C4.5 Dan Software Weka’, *JAS-PT (Jurnal Analisis Sistem Pendidikan Tinggi Indonesia)*, 5(1). Doi: 10.36339/Jaspt.V5i1.417.
- N.Undavia, J., M. Dolia, P. And P. Shah, N. (2013) ‘Prediction Of Graduate Students For Master Degree Based On Their Past Performance Using Decision Tree In Weka Environment’, *International Journal Of Computer Applications*, 74(11). Doi: 10.5120/12930-9877.

- Olalekan, A. M., Egwuche, O. S. And Olatunji, S. O. (2020) 'Performance Evaluation Of Machine Learning Techniques For Prediction Of Graduating Students In Tertiary Institution', In *2020 International Conference In Mathematics, Computer Engineering And Computer Science, ICMCECS 2020*. Doi: 10.1109/ICMCECS47690.2020.240888.
- Qin, L. And Phillips, G. A. (2019) 'The Best Three Years Of Your Life: A Prediction Of Three-Year Graduation With Diagnostic Classification Model', *International Journal Of Higher Education*, 8(6). Doi: 10.5430/Ijhe.V8n6p231.
- Ristoski, P., Bizer, C. And Paulheim, H. (2015) 'Mining The Web Of Linked Data With Rapidminer', *Journal Of Web Semantics*, 35. Doi: 10.1016/J.Websem.2015.06.004.
- Romadhona, A., Suprapedi And Himawan, H. (2017) 'Prediksi Kelulusan Mahasiswa Tepat Waktu Berdasarkan Usia, Jenis Kelamin, Dan Indeks Prestasi Menggunakan Algoritma Decision Tree', *Jurnal Teknologi Informasi*, 13.
- Roy, S. *Et Al.* (2019) 'Dispersion Ratio Based Decision Tree Model For Classification', *Expert Systems With Applications*, 116. Doi: 10.1016/J.Eswa.2018.08.039.
- Sembiring, M. T. And Tambunan, R. H. (2021) 'Analysis Of Graduation Prediction On Time Based On Student Academic Performance Using The Naïve Bayes Algorithm With Data Mining Implementation (Case Study: Department Of Industrial Engineering USU)', *IOP Conference Series: Materials Science And Engineering*, 1122(1). Doi: 10.1088/1757-899x/1122/1/012069.
- Srinadi, I. G. A. M. And Nilakusmawati, D. P. E. (2020) 'Analisis Waktu Kelulusan Mahasiswa Fmipa Universitas Udayana Dan Faktor-Faktor Yang Memengaruhinya', *E-Jurnal Matematika*, 9(3). Doi: 10.24843/Mtk.2020.V09.I03.P300.
- Sutoyo, E. And Almaarif, A. (2020) 'Educational Data Mining For Predicting Student Graduation Using The Naïve Bayes Classifier Algorithm', *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(1). Doi: 10.29207/Resti.V4i1.1502.
- Undavia, J. N., Dolia, P. And Patel, A. (2016) 'Customized Prediction Model To Predict Post-Graduation Course For Graduating Students Using Decision Tree Classifier', *Indian Journal Of Science And Technology*, 9(12). Doi: 10.17485/Ijst/2016/V9i12/83335.