Prediction of graduation rate of engineering education students using Artificial Neural Network Algorithms

Muhammad Anwar
Universitas Negeri Padang, Indonesia
*Corresponding author, e-mail: muh_anwar@ft.unp.ac.id

Abstract
The graduation rate of engineering education students on time greatly affects the quality of learning. The purpose of this study is to predict the graduation rate of engineering education students. The method uses an artificial neural network algorithm combined with particle swarm optimization and forward selection, with 234 samples. The test results with Artificial Neural Network obtained 82.61% accuracy with predictions on time 149 and not on time 62, the combination of Artificial Neural Network with Particle Swarm Optimization obtained 91.30% accuracy with predictions on time 165 and not on time 69. Furthermore, Artificial Neural Network with Particle Swarm Optimization and reduced by forward selection obtained 95.65% accuracy with predictions of the number of graduations on time 165 and not on time 69. With the combination of the three algorithms, it is able to predict the graduation rate of engineering education students with high accuracy.

Keywords: Engineering Education, Graduation, Artificial Neural Network, Particle Swarm Optimization, Forward Selection.

How to Cite: Anwar, M. (2021). Prediction of graduation rate of engineering education students using Artificial Neural Network Algorithms. International Journal of Research in Counseling and Education, 5 (1): pp. 15-23, DOI: http://doi.org/10.24036/00411za0002

Introduction
Graduates of engineering education have the competence, special skills and are able to survive in the world of work. In addition to the skill competencies possessed by engineering education graduates, they are also equipped with entrepreneurial skills (Hidayat, 2017a, 2017b; Hidayat et al., 2018a, 2018b; Hidayat et al., 2019a, 2019b, 2019c), which are expected to open up new fields after graduation work (Hidayat, & Yuliana, 2018; Hidayat et al., 2020). The need for the world of work for graduates from engineering education students is quite high due to their special skills, good competencies, and pedagogical aspects (Ganefri et al., 2017; Ganefri et al., 2018), so these graduates quickly adapt to the new world of work. The high demand for engineering education graduates from higher education is a problem for campuses because they have not been able to predict when engineering education students will graduate.

Graduation from engineering education students in higher education, whether sooner or later, of course, cannot be predicted, because graduation is influenced by several factors, including the quality of teaching lecturers, learning management, learning climate (Sari, Ganefri, & Anwar, 2020), and learning models (Hidayat, 2015; Aryanti, Anwar, & Zulwisi, 2017; Andrianis, Anwar, & Zulwisi, 2018; Anwar, 2021). Predicting the graduation rate of engineering education students is an urgent and important thing to do (Arif et al., 2017; Adekitan, & Salau, 2019), so that the graduation of engineering education students can be predicted as early as possible (Chachashvili-Bolotin, Milner-Bolotin, & Lissitsa, 2016; Mubarak, Cao, & Zhang, 2020; Naseer, Zhang, & Zhu, 2020). Being able to predict the graduation time of engineering education students will be useful for the absorption and needs of the world of work (Anwar, 2019), as well as an indicator of the quality of learning management that has been managed well (Yulastri et al., 2019). Furthermore, the percentage of engineering education students graduating on time is one of the criteria in assessing the quality of engineering education management (Mason et al., 2018; Lopez, & Jones, 2017). So if engineering education students graduate on time, it will help improve the quality of a university. On the
other hand, engineering education students will be helped and reduce the burden of spending on tuition fees (Melendez-Armenta et al., 2020).

So it is very necessary alternative solutions in order to accurately predict the graduation of engineering education students (Laugerman et al., 2015), using data mining can be an alternative solution. Data mining can find unknown or valuable data from large amounts of data (Jia, & Pang, 2018). It is a field of scientific research that integrates computer, statistics, simulation, artificial intelligence, and database technologies (Wang, 2016). So that from the data from the evaluation results of engineering education students, useful information can be obtained (Yu, 2021). There are many ways to analyze data using classification techniques from a data mining science (Imran et al., 2019).

With the advent of the big data era, significant changes have occurred in every aspect of higher education (Cao, 2021). In recent years, with the rapid development of data mining in higher education, the combination of data mining methods to analyze student behavior data has become a popular trend (Nyoman Sukajaya, Ketut Eddy Purnama, and Purnomo, 2015). It is mainly aimed at predicting student learning performance (Li, 2020). Indeed, processing the data generated by the learning environment has become a real challenge, which requires the use of big data technologies and tools to handle it (Salihoun, 2020). Engineering education curricula must take the views of industry experts to improve the quality of teaching and learning in universities (Ramamurthy, Dewitt, and Alias, 2021). The linkage of higher education and industry is considered a strategy to equip students with theoretical and practical knowledge (Birhan and Merso, 2021).

Universities are required to be able to design and implement innovative learning processes so that students can achieve optimal learning outcomes (Yustisia et al., 2021). Learned power skills to do something competently (Gambo et al., 2021). What competencies students really have at the beginning of their studies will affect their learning styles in higher education (Behrendt et al., 2015). This has an impact on the study period and timely graduation (Aldossari, 2020). The development of technology has caused many radical changes in the technical education curriculum in higher education (Bayhan and Karaca, 2020). The data mining used in this study (Moscoso-Zea, Saa, & Luján-Mora, 2019), namely the artificial neural network algorithm method combined with particle swarm optimization and forward selection algorithms. Through these three methods, the results of the prediction accuracy of graduation of engineering education students can be obtained maximum results.

**Method**

At this stage, several methods are proposed which will be combined with artificial Neural Networks, including Particle Swarm Optimization (PSO) and the Forward Selection Algorithm.

**Artificial Neural Network Algorithm**

An Artificial Neural Network (ANN) is an artificial network based on the structure of the Brain Nerves. The brain basically has the principle of learning from experience. The actual work of the brain is still not fully revealed, although its function as an extraordinary processor is known. The main components of the brain are cells, as are other parts of the body. Brain cells have the ability to remember, think and apply the experiences they have experienced (Zhang, and Jiang, 2018). An ANN generally consists of three layers, namely the input layer, hidden layer, and output layer. The input layer (input layer) consists of neurons that receive input from the external environment. The input entered is a description of a problem. The hidden layer consists of neurons that receive input from the input layer, and then carry the output to the next layer. The output layer, called output units, consists of neurons that receive output from the hidden layer and send it to the user.

**Particle Swarm Optimization (PSO) Algorithm**

Particle Swarm Optimization (PSO) has parameters such as position, maximum speed, acceleration constant, and weight of inertia. In the PSO technique, there are several ways to optimize, including increasing the attribute weight of all attributes or variables used, selecting attributes (attribute selection), and selecting features (feature selection) (Mansur, Prahasto, and Farikhin, 2014).

Each particle in the PSO is also defined by the velocity of the particles flying through the search space at a speed that is dynamically adjusted for their historical behavior. Therefore, the particles have a tendency to fly towards a better and better search area during the search process.

**Forward Selection Algorithm**

Each input unit \((x_i, i=1, \ldots, n)\) receives the input signal \(x_i\) and is forwarded to the hidden units. Each hidden unit \((z_j, z=1, \ldots, p)\) adds up the weight of the input signal and its bias

\[
z_j = \sum_{i=1}^{n} x_i \cdot v_{ij}
\]

The output of the hidden unit layer applies the arithmetic activation function:

\[
z_j = f(z_j)
\]

Each output unit \((y_k, k=1, \ldots, m)\) adds a weighted input signal by applying the arithmetic activation function:

\[
y_k = \sum_{j=1}^{p} z_j \cdot w_{jk}
\]

\[
y_k = f(y_k)
\]

(Prediction of graduation rate of engineering education students using Artificial Neural Network Algorithms)
Data collection (Data Gathering)

The data of engineering education students used in this study is data from students of engineering education of the Informatics Engineering Education Study Program class of 2013, 2014, and 2015 consisting of 37 attributes and 294 records covering course values from semester one to semester four. Data were obtained from SIA (Academic Information System) Informatics Engineering Education Study Program, Padang State University.

Initial data processing (Data Preprocessing)

Preprocessing Data in the stages of this classification method includes:

- **Original Data**
- **Data selection**
- **Data Validation**
- **Data Integration**
- **Data size reduction and discretization**

**Figure 1. Feature Selection Model**

Data Selection

At this stage, the selection of the database is carried out. Because not all of the data obtained were used, they were selected according to the attributes and variables needed in the study by selecting the data so that it became a dataset (Rolansa, Yunita, and Suheri, 2020).

To get quality data, several techniques are used as follows:

- **Data Validation**
  - Data Validation, to identify and remove odd data (outlier/noise), inconsistent data, and incomplete data (missing value).

- **Data Integration**
  - Data integration and transformation, to improve algorithm accuracy and efficiency. This data is transformed in the Rapidminer software.

- **Data size reduction and discretization**
  - Data size reduction and discretization, to obtain a data set with a number of irrelevant records such as some engineering education students' scores that have a minimum score of zero in two semesters will be deleted.

Results and Discussion

After the data preparation stage was carried out, the total data records became 234 records of course scores from semester 1 to semester 4, for engineering education students from the 2013, 2014, and 2015 batches.

Confusion Matrix results of engineering education student graduation

After testing using a combination of three algorithms, namely Artificial Neural Network, Particle Swarm Optimization (PSO), and Forward Selection to classify student graduation, the results are compared as shown in the table below:

| Algorithm                        | Accuracy | Precision | Recall | Plot AUC/ROC |
|----------------------------------|----------|-----------|--------|--------------|
| **Artificial Neural Network**    | 82.61%   | 80.00%    | 57.14% | 0.929        |
| **Particle Swarm Optimization (PSO)** | 91.30%   | 85.71%    | 85.71% | 0.973        |
| **Forward Selection**            | 95.65%   | 100.00%   | 85.71% | 0.964        |
In addition to getting the results from the Confusion Matrix, statistics on the number of graduations of engineering education students with 234 item data sets can be seen in Table 3:

**Table 2. Statistical results of graduation of engineering education students**

| Algorithm                               | On time | Not on time |
|-----------------------------------------|---------|-------------|
| Artificial Neural Network               | 149     | 62          |
| Particle Swarm Optimization (PSO)       | 165     | 69          |
| forward Selection                       | 165     | 69          |

Based on the statistical results in the table above, it can be seen that in the process using the Artificial Neural Network algorithm, 149 students of engineering education who graduated on time and who did not graduate on time were 62 people, and in the combination process of Artificial Neural Network and Particle Swarm Optimization algorithms (PSO) obtained the number of engineering education students who graduated on time as many as 165 and who did not graduate on time as many as 69 people, then in the combination process of Artificial Neural Network, Particle Swarm Optimization (PSO) and forward Selection algorithms, the number of engineering education students who graduated on time was obtained as many as 165 and who did not pass on time as many as 69 people.

**Testing the Artificial Neural Network algorithm**

The Artificial Neural Network algorithm consists of output and input layers, and a hidden layer that processes input from the input layer into something that can be accepted by the output layer.

Confusion Matrix Test results with the Artificial Neural Network algorithm are as follows:

**Table 3. Confusion Matrix Test results with Artificial Neural Network algorithm**

| Accuracy: 82.61% | On time | Not on time | Class precision |
|------------------|---------|-------------|-----------------|
| Prediction on time | 15      | 3           | 83.33%          |
| Prediction not on time | 1      | 4           | 80.00%          |
| Class recal       | 93.75%  | 57.14%      |

The Confusion Matrix from the test results using the Artificial Neural Network algorithm shows that the accuracy produced is 82.61%, with details on predictions on time of 83.33% and not on the time of 80.00%. This test curve can be seen in the image below:

**Figure 2. ROC curve with testing using the Artificial Neural Network algorithm**

The results of the Under Area Curva (AUC) test using the Artificial Neural Network algorithm model have a value of 0.929, in the best category.
Table 4. Statistics from the recapitulation results using the Artificial Neural Network algorithm

| Algorithm                      | On time | Not on time |
|--------------------------------|---------|-------------|
| Artificial Neural Network      | 149     | 62          |

Testing the Artificial Neural Network algorithm combined with Particle Swarm Optimization (PSO)

Confusion Matrix from Artificial Neural Network Algorithm Testing combined with Particle Swarm Optimization (PSO).

Table 5. Confusion Matrix test results

|                      | On time | Not on time | Class precision |
|----------------------|---------|-------------|-----------------|
| Prediction.on time   | 15      | 1           | 93.75%          |
| Prediction.not on time | 1      | 6           | 85.71%          |
| Class recal          | 93.75%  | 85.71%      |

The test results using the Artificial Neural Network algorithm combined with Particle Swarm Optimization (PSO) showed an increase in the resulting accuracy of 91.30%, with details on predictions on time of 93.75% and predictions on time of 85.71%. This test curve can be seen in the image below:

Figure 3. ROC curve of Artificial Neural Network algorithm combined with Particle Swarm Optimization

In the Under Area Curva (AUC) graph using the Artificial Neural Network algorithm model combined with Particle Swarm Optimization (PSO), it has a value of 0.973, in the best category. Furthermore, statistics are obtained from the recapitulation results for the number of engineering education students who complete their studies on time or not on time using the Artificial Neural Network algorithm combined with Particle Swarm Optimization (PSO).

Table 5. Statistics from the recapitulation results

| Algorithm                              | On time | Not on time |
|----------------------------------------|---------|-------------|
| Artificial Neural Network combined with Particle Swarm Optimization (PSO) | 165     | 69          |

From the table above, it can be seen that the number of engineering education students who graduated on time was 165 and 69 people who graduated not on time.

Testing the Artificial Neural Network algorithm combined with Particle Swarm Optimization (PSO) and reduced by the Forward Selection algorithm

Table 6. Confusion Matrix test results

|                      | On time | Not on time | Class precision |
|----------------------|---------|-------------|-----------------|
| Prediction.on time   | 16      | 1           | 94.12%          |
| Prediction.not on time | 0      | 6           | 100.00%         |
| Class recal          | 100.00% | 85.71%      |
The test results of the Artificial Neural Network algorithm combined with Particle Swarm Optimization (PSO) and reduced by the Forward Selection algorithm showed a very significant increase in accuracy, namely 95.65%, with details of on-time predictions of 94.12% and inaccurate predictions of 100.00%. This test curve can be seen in the image below:

![ROC curve](image)

**Figure 4.** ROC curve

In the Under Area Curva (AUC) graph using the Artificial Neural Network algorithm model combined with Particle Swarm Optimization (PSO) and reduced by the Forward Selection algorithm, it has a value of 0.964, in the best category.

Furthermore, statistics are obtained from the recapitulation results for the number of engineering education students who complete their studies on time or not on time using the Artificial Neural Network algorithm combined with the Particle Swarm Optimization (PSO) algorithm reduced by the Forward Selection algorithm:

| Algorithm | On time | Not on time |
|-----------|---------|-------------|
| Artificial Neural Network combined with Particle Swarm Optimization (PSO) reduced by Forward Selection algorithm | 165 | 6 |

In the table above, it can be seen the number of engineering education students who graduated on time as many as 165 and who graduated not on time 69 people using the Artificial Neural Network algorithm model combined with Particle Swarm Optimization (PSO) and reduced by the Forward Selection algorithm.

**Conclusion**

Based on the results of the research and discussion, it can be concluded that the use of data mining with the Artificial Neural Network algorithm method has a low accuracy of 82.61% with a prediction of graduation on time 149 and not on time 62. To increase the weight of the Artificial Neural Network algorithm, it is combined with Particle Swarm Optimization (PSO) algorithm and obtained 91.30% accuracy with predictions of the number of graduations on time 165 and not on time 69. Furthermore, the Artificial Neural Network algorithm combined with the Particle Swarm Optimization (PSO) algorithm is reduced again with the Forward Selection algorithm and obtained accuracy which is quite high, namely 95.65% with a prediction of the number of graduations on time 165 and not on time 69. From the results obtained using the graduation prediction method with predictions of the number of graduations on time 165 and not on time 69, it is necessary to improve the learning process and curriculum.

**References**

Adekitan, A. I., & Salau, O. (2019). The impact of engineering students’ performance in the first three years on their graduation result using educational data mining. *Heliyon, 5*(2), 01250.
Hidayat, H., Ardi, Z., Yuliana, & Herawati, S. (2019b). Exploration of the need analysis for technopreneurship scientific learning models in higher vocational education. *International Journal of Economics and Business Research*, 18(3), 356–368.

Hidayat, H., Tamin, B.Y., Herawati, S., Hidayati, A., & Muji, A.P. (2019c). Implementation of technopreneurship scientific learning for produce electronic product prototypes in engineering education. *International Journal of Innovative Technology and Exploring Engineering*, 8(11), 2842–2846. http://dx.doi.org/10.35940/ijitee.K2406.0981119

Hidayat, H., Tamin, B.Y., Herawati, S., Ardi, Z., & Muji, A. P. (2020). The Contribution of Internal Locus of Control and Self-Concept to Career Maturity in Engineering Education. *Int. J. Adv. Sci. Eng. Inf. Technol.*, 10(6), 2282–2289.

Imran, Muhammad, Shahzad Latif, Danish Mehmood, and Muhammad Saglaim Shah. (2019). Student Academic Performance Prediction Using Supervised Learning Techniques. *International Journal of Emerging Technologies in Learning*, 14(14), 92–104. doi: 10.3991/ijet.v14i14.10310.

Jia, Sujuan, and Pang, Yajing. (2018). Teaching Quality Evaluation and Scheme Prediction Model Based on Improved Decision Tree Algorithm. *International Journal of Emerging Technologies in Learning*, 13(10), 146–57. doi: 10.3991/ijet.v13i10.9460.

Laugerman, M., Rover, D. T., Shelley, M. C., & Mickelson, S. K. (2015). Determining graduation rates in engineering for community college transfer students using data mining. *International Journal of Engineering Education*, 31(6A), 1448.

Li, Na. (2020). Curriculum Data Association Organization and Knowledge Management Method for Unstructured Learning Resources. *International Journal of Emerging Technologies in Learning*, 15(6), 79–94. doi: 10.3991/ijet.15i06.13173.

Lopez, C., & Jones, S. J. (2017). Examination of factors that predict academic adjustment and success of community college transfer students in STEM at 4-year institutions. *Community College Journal of Research and Practice*, 47(3), 168–182.

Mansur, Mansur, Toni Prahasto, and Farikhin Farikhin. (2014). Particle Swarm Optimization Untuk Sistem Informasi Pendaftaran Data Di perguruan Tinggi. *Jurnal Sistem Informasi Bisnis*, 4(1), 11–19. doi: 10.21456/vol4iss1pp11-19.

Mason, C., Twomey, J., Wright, D., & Whitman, L. (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. *Research in Higher Education*, 59(3), 382–400.

Melendez-Armenta, Roberto Angel, N. Sofia Huerta-Pacheco, Luis Alberto Morales-Rosales, and Genaro Rebollo-Mendez. (2020). How Do Students Behave When Using A Tutoring System? Employing Data Mining to Identify Behavioral Patterns Associated to The Learning of Mathematics. *International Journal of Emerging Technologies in Learning*, 15(22), 39–58. doi: 10.3991/ijet.v15i22.17075.

Moscoso-Za, O., Saa, P., & Luján-Mora, S. (2019). Evaluation of algorithms to predict graduation rate in higher education institutions by applying educational data mining. *Australasian Journal of Engineering Education*, 24(1), 4–13.

Mubarak, A. A., Cao, H., & Zhang, W. (2020). Prediction of students’ early dropout based on their interaction logs in online learning environment. *Interactive Learning Environments*, 1–20.

Naseer, M., Zhang, W., & Zhu, W. (2020). Early Prediction of a Team Performance in the Initial Assessment Phases of a Software Project for Sustainable Software Engineering Education. *Sustainability*, 12(11), 4663.

Nyoman Sukajaya, I., I. Ketut Eddy Purnama, and Mauridhi Hery Purnomo. (2015). Intelligent Classification of Learner’s Cognitive Domain Using Bayes Net, Naïve Bayes, and j48 Utilizing Bloom’s Taxonomy-Based Serious Game. *International Journal of Emerging Technologies in Learning*, 10(2), 46–52. doi: 10.3991/ijet.v10i2.4451.

Rammamurthy, Viji, Dorothy Dewitt, and Norlidah Alias. (2021). The Need for Technical Communication for 21st Century Learning in Tvet Institutions: Perceptions of Industry Experts. *Journal of Technical Education and Training*, 13(1), 148–58. doi: 10.30880/jtet.2021.13.01.016.

Rolansa, Fresa, Yunita Yunita, and Suheri Suheri. (2020). Sistem Prediksi Dan Evaluasi Prestasi Akademik Mahasiswa Di Program Studi Teknik Informatika Menggunakan Data Mining. *Jurnal Pendidikan Informatika Dan Sains*, 9(1), 75. doi: 10.31571/saintek.v9i1.1696.

Salihoun, Mohammed. (2020). State of Art of Data and Learning Analytics Tools in Higher Education. *International Journal of Emerging Technologies in Learning*, 15(21), 58–76. doi: 10.3991/ijet.v15i21.16435.

Sari, P. P., Ganefi, G., & Anwar, M. (2020). The Contribution Of Principal Leadership Style, Teachers'professional Competence And School Climate On The Quality Of Learning Outcomes At Vocational High School In Padang. *Jurnal Pendidikan Teknologi Kejuruan, 3*(1), 26–30.

Wang, Lanzhong. (2016). Personalized Teaching Platform Based on Web Data Mining. *International Journal of Emerging Technologies in Learning*, 11(11), 15–20. doi: 10.3991/ijet.v11i11.6253.
Yu, Jing. (2021). Academic Performance Prediction Method of Online Education Using Random Forest Algorithm and Artificial Intelligence Methods. *International Journal of Emerging Technologies in Learning, 16*(5), 45–57. doi: 10.3991/ijet.v16i05.20297.

Yulastri, A., Hidayat, H., Ganefri, Ayu, R., & Ardi, Z. (2019). An Empirical Study on The Effects of Pedagogy Learning Tools Entrepreneurship With Product-Based Learning Approach, Learning Readiness, and Locus of Control: A Case From Engineering Education in Indonesia. *International Journal of Scientific & Technology Research, 8* (9), 1722–1727.

Yustisia, Henny, Nizwardi Jalinus, Fahmi Rizal, and Fadhillah. (2021). A New Approach of Students’ Industrial Field Experience Program in the Digital Age. *Journal of Technical Education and Training, 13*(1), 167–75. doi: 10.30880/jtet.2021.13.01.018.

Zhang, Yuan, and Wenbo Jiang. (2018). Score Prediction Model of MOOCs Learners Based on Neural Network. *International Journal of Emerging Technologies in Learning, 13*(10), 171–82. doi: 10.3991/ijet.v13i10.9461.