

Utilizing Genetic Algorithms to Enhance Student Graduation Prediction with Neural Networks

Witriana Endah Pangesti^{1)*}, Indah Ariyati²⁾, Priyono³⁾, Sugiono⁴⁾, Rachmat Suryadithia⁵⁾

¹⁾ Universitas Nusa Mandiri, Indonesia

^{2,3,4,5)} Universitas Bina Sarana Informatika, Indonesia

¹⁾ witriana.weg@nusamandiri.ac.id, ²⁾ indah.ayi@bsi.ac.id, ³⁾ priyono.pyo@bsi.ac.id, ⁴⁾ sugiono.sgx@bsi.ac.id, ⁵⁾ rachmat.rcs@bsi.ac.id

Submitted : Nov 8, 2023 | **Accepted** : Nov 22, 2023 | **Published** : Jan 1, 2024

Abstract: The prediction of student graduation plays a crucial role in improving higher education efficiency and assisting students in graduating on time. Neural networks have been used for predicting student graduation; however, the performance of neural network models can still be enhanced to make predictions more accurate. Genetic algorithms are optimization methods used to improve the performance of neural network models by optimizing their parameters. The problem at hand is the suboptimal performance of neural networks in predicting student graduation. Thus, the objective is to leverage genetic algorithms to improve the accuracy of student graduation predictions, measure the improvements obtained, and compare the accuracy results between the genetic algorithm-optimized neural network model and the neural network model without optimization. The training process of the neural network model is conducted using training data obtained through experiments, and the accuracy results of the neural network model with and without genetic algorithm optimization are compared. The research findings indicate that by harnessing genetic algorithms to optimize the parameters of the neural network model, the accuracy of student graduation predictions increased by 2.78%. Furthermore, the Area Under the Curve (AUC) also improved by 0.037%. These results demonstrate that integrating genetic algorithms into the neural network model can significantly enhance prediction performance. Thus, this study successfully utilized genetic algorithms to improve student graduation predictions using a neural network model. Experimental results show that prediction accuracy and AUC values significantly increased after optimizing the neural network model's parameters with genetic algorithms. Therefore, the use of genetic algorithms can be considered an effective approach to improving student graduation predictions, thereby assisting educational institutions in improving efficiency and helping students graduate on time.

Keywords: Genetic algorithm, Graduation data mining, Graduation prediction, Neural network, Student graduation.

INTRODUCTION

Higher education institutions face challenges in effectively predicting student graduation. Student graduation issues are of paramount importance in the field of education, as they can impact funding efficiency and the quality of the institution's education. Especially when dealing with student graduation data that has not been consistently observed, graduation can become uncontrolled, with some students graduating on time and others exceeding the maximum allowed study duration. Therefore, the

*name of corresponding author



development of accurate and advanced graduation prediction models becomes a priority for educational institutions. Student graduation prediction has been extensively researched recently, utilizing various data sources and algorithms, including classification, clustering, and prediction methods.

Some of the methods used for student graduation prediction include Decision Trees (DT), Support Vector Machine (SVM), Neural Networks (NN)(Riyanto, Hamid, & Ridwansyah, 2019), K-Nearest Neighbor (Widaningsih, 2019), Naive Bayes (Salmu & Solichin, 2017). Researchers have further developed these methods by incorporating optimization techniques, such as SVM with Particle Swarm Optimization (PSO) (Suhardjono, Wijaya, & Hamid, 2019), SVM with Genetic Algorithm (GA) (Ridwansyah, Wijaya, & Purnama, 2020), DT with PSO (Hendra, Azis, & Suhardjono, 2020), and NN with PSO (Nurdin, Sartini, Sumarna, Maulana, & Riyanto, 2023).

Given the existing research with various methods, the current research aims to maximize accuracy to create an efficient tool for predicting student graduation while harnessing the power of GA combined with NN in the context of student graduation prediction. This requires a testing method capable of accurately predicting student graduation. In this regard, testing student graduation data using NN with GA optimization, a method not previously explored, is considered. NN has become a popular tool in data analysis and prediction modeling due to its ability to handle large datasets like student graduation data. Despite its widespread use, NN still has limitations in terms of learning and generalization capabilities. Genetic algorithms, on the other hand, are powerful methods for optimizing model parameters, which can be used to enhance NN performance.

With the stated objectives, the hope is to create a more accurate prediction model that can assist higher education institutions in identifying factors contributing to student graduation and providing better insights to improve graduation rates and educational efficiency.

Consequently, this research can contribute to the field by offering an innovative approach to predicting student graduation by combining GA and NN to enhance graduation prediction accuracy, enabling higher education institutions to take more effective measures for improving student graduation rates, curriculum enhancements, and educational programs

LITERATURE REVIEW

Neural Network

Neural networks offer greater flexibility as they are not bound by previous constraints, as applied in conventional statistical modeling approaches. This results in neural networks having a tendency to produce more accurate predictions (Moro & Laureano, 2011). They have been employed in predicting fields such as banking (Ridwansyah & Purwaningsih, 2018), education (Ariyati, Ridwansyah, & Suhardjono, 2018), diagnostics, medicine, pattern recognition, and data classification (K & M, 2012). One of the advantages of neural networks is their ability to operate effectively with data that has a non-normal distribution (Satapathy, Chittineni, Krishna, Murthy, & Reddy, 2012). They are also highly useful for solving complex and large datasets (Nazari & Pacheco Torgal, 2013) that are nonlinear in nature (Chen, Zhang, Xu, Chen, & Zhang, 2012). With learning capabilities, neural networks can adapt and change their structure based on the information received from the dataset, as long as the data is well-received by the system (S & N, 2016).

Genetic Algorithm

Genetic algorithms are one of the potential candidate methods for optimizing neural network weights (K & M, 2012). Their ability to tackle complex problems through processes involving selection, recombination, and mutation to generate new solutions makes them suitable for this purpose. In the beginning, in the search for a solution using a GA (Genetic Algorithm), the first step is to determine the initial population. This initial population consists of a group of solutions generated randomly. Each individual in this population is called a chromosome and serves as a representation of various possible solutions. Next, each chromosome is evaluated for its fitness level using a predefined function. After that, natural selection occurs between two chromosomes (parents) with the aim of producing new chromosomes with higher fitness levels as the next generation (offspring). This process is called a generation, and each generation of chromosomes is evaluated based on their fitness values. After going through several generations, the genetic algorithm will successfully find the best chromosome representing the optimal solution.

*name of corresponding author



Student Graduation

In the study on student graduation, which has been conducted using several methods, including: (Riyanto et al., 2019) conducted research on student graduation using three methods: NN, SVM, and DT. The final results determined that the SVM method outperformed the other two methods with an accuracy rate of 85.18%. After that, researchers (Suhardjono et al., 2019) also tested student graduation data using the same method with the highest accuracy, which is SVM with PSO optimization, and obtained an accuracy rate of 86.57%. After successfully improving SVM accuracy with PSO, it was further tested by (Ridwansyah et al., 2020) with another optimization method, GA, and obtained an accuracy rate of 86.43%. Afterward, (Hendra et al., 2020) conducted further research by using optimization methods, but it did not improve the SVM method; instead, the DT method was improved. The final result obtained from the DT method and PSO with its accuracy rate was 87.56%. After that, PSO was used again by (Nurdin et al., 2023) with the NN method and obtained an accuracy rate of 86.94%. Therefore, in this study, the NN method will be tested with GA to achieve maximum accuracy.

METHOD

This section explains the neural network method, the optimization method used to improve student graduation accuracy, and the research flow. The research method employed consists of four steps: data acquisition, method application, method optimization, and result evaluation. Figure 1 clearly illustrates the research flow in this study.

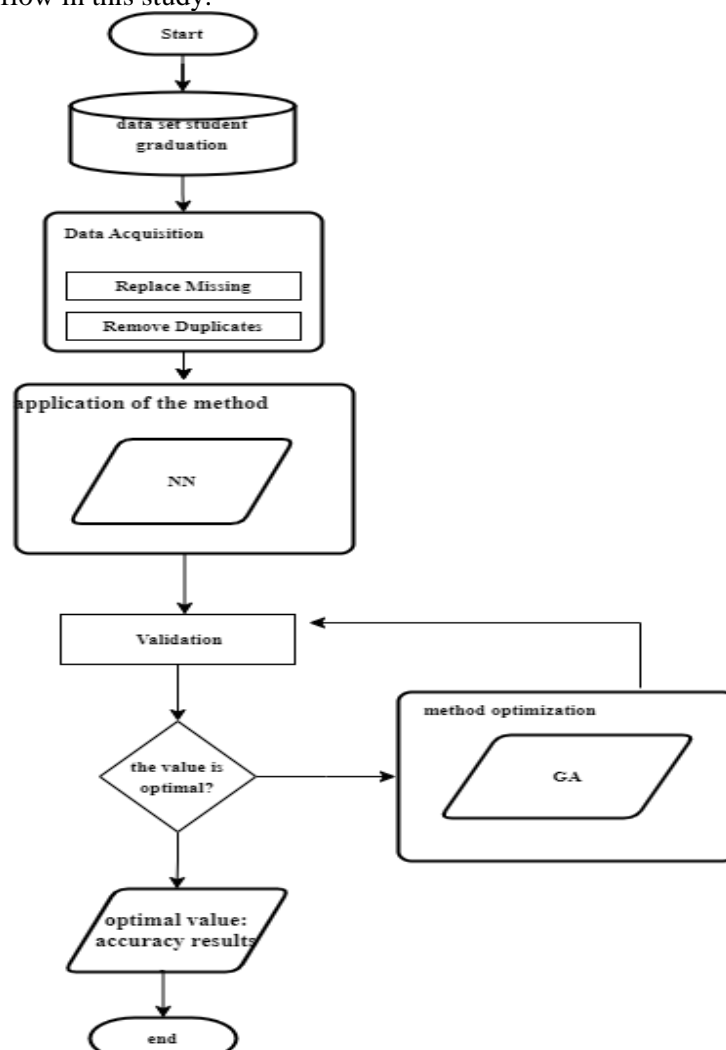


Fig. 1 Research Flow Diagram NN with GA

In Figure 1, it can be seen that the initial phase of the process begins with data collection based on primary data from private universities regarding 796 students who graduated from around 2013 to

*name of corresponding author

2018. The student graduation data has undergone prior data acquisition, where the data acquisition process includes replacing missing data and eliminating duplicate data, enabling the application of the neural network method in the test-ing. The graduation dataset has nine attributes as features for prediction and one attribute as the prediction la-bel, which indicates whether the student graduated on time or not. Subsequently, an optimization process is car-ried out using optimization methods such as genetic algorithms to discover and improve the best neural network model for predicting student graduations.

Data Acquisition

Graduation data was obtained from primary data at one of the private universities with a range of years from 2013 to 2018, where the collected data pertains to student graduations. From this dataset, there are 796 records of student graduations, including those who graduated on time and those who did not. The dataset contains 10 attributes, including predictor attributes. Subsequently, missing data was replaced, and duplicate data was removed from the graduation dataset.

Application of the Method

Student graduation data is tested using the neural network method, following the data acquisition phase. The data is tested using a neural network to obtain accuracy and Area Under Curve (AUC) results. With the following formula:

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN} \quad (1)$$

TN = True Negative

TP = True Positive

FN = False Negative

FP = False Positive

Method Optimization

The data that has been tested with the neural network method will be optimized using the genetic algorithm method to achieve better accuracy and Area Under Curve (AUC) results compared to the previous method's implementation.

Evaluation of Results

After optimizing the neural network method with the genetic algorithm, an evaluation will be conducted on both methods to determine whether there has been an improvement in accuracy and AUC or if there has been no improvement in these metrics.

RESULT

Data Acquisition

Based on the research flowchart shown in Figure 1, the data used in this study was obtained from primary data sources originating from one of the private universities. This data will be used in a series of experiments using the NN method and an accuracy enhancement method optimized using GA. Prior to this, the data underwent a process of replacing missing values with the mean value and removing irrelevant duplicate data to ensure data validity. The results of this process can be found in Table 1, which displays sample student graduation data used in this research.

Table 1. Sample Student Graduation Data

GENDE R	HISCH DEP	FROM HISCH	CAI 1	CAI 2	CAI 3	CAI 4	CAI 5	CAI 6	ON TIME
1	7	3	2,95	2,93	3,1	3,25	3,25	3,22	YES
1	9	3	2,27	2,79	2,9	2,95	3,21	3,19	YES
2	3	3	2,82	3,12	3,31	3,4	3,45	3,46	YES
1	5	3	2,59	2,79	2,76	2,9	2,93	2,96	NO
1	10	3	3,23	3,05	3,1	3,14	3,23	3,28	NO
1	4	3	3,23	3,43	3,56	3,53	3,55	3,51	YES
1	4	3	2,55	2,81	2,77	2,83	3,07	3,15	YES

*name of corresponding author



1	4	3	2,77	2,67	2,82	3	3,12	3,12	YES
1	5	3	1,86	2,62	2,87	2,96	3,12	3,11	YES
1	4	3	2,77	2,57	2,58	2,92	3,19	3,19	YES
2	4	2	2,95	3,17	3,18	3,34	3,46	3,52	YES
2	4	1	2,18	2,21	2,76	3,01	3,12	3,2	YES
2	4	3	1,82	1,62	1,65	1,75	2,28	2,6	NO
2	5	2	2,18	2,6	2,63	2,9	3,11	3,16	YES
1	4	1	1,82	2,07	2,11	2,33	2,68	2,65	YES
2	10	3	3,09	3,19	3,35	3,28	3,32	3,35	YES
2	3	3	2,68	2,76	2,9	3,23	3,17	3,21	YES
1	5	3	2,91	3,05	3,03	3,24	3,37	3,35	YES
2	10	3	3,27	3,19	3,06	3,34	3,31	3,34	YES
1	5	4	2,55	2,33	2,29	2,8	3,04	3,07	YES
2	7	3	2,41	3,1	3,18	3,19	3,23	3,35	YES

Application of the Method

The data in Table 1 will undergo testing using the NN method. The testing and validation process of student graduation will be conducted using neural network techniques through RapidMiner software with a 10-fold cross-validation method. After the testing process of the student graduation dataset using the NN method is completed, the research results will be reflected in Table 2.

Table 2. Confussion Matrix

	TP	TN	class precision
FP	594	83	87.74%
FN	40	79	66.39%
class recall	93.69%	48.77%	

From the presented confusion matrix, it can be concluded that there are 594 students predicted to graduate on time, and this prediction matches reality. However, there are 83 students predicted to graduate on time, but they actually did not. Conversely, 79 students were predicted not to graduate on time, and this prediction matches reality, but 40 students predicted not to graduate on time actually graduated on time. As a result, the accuracy rate of this confusion matrix reaches 84.55%. In addition, Figure 2 displays the AUC curve that illustrates the student graduation data graph.

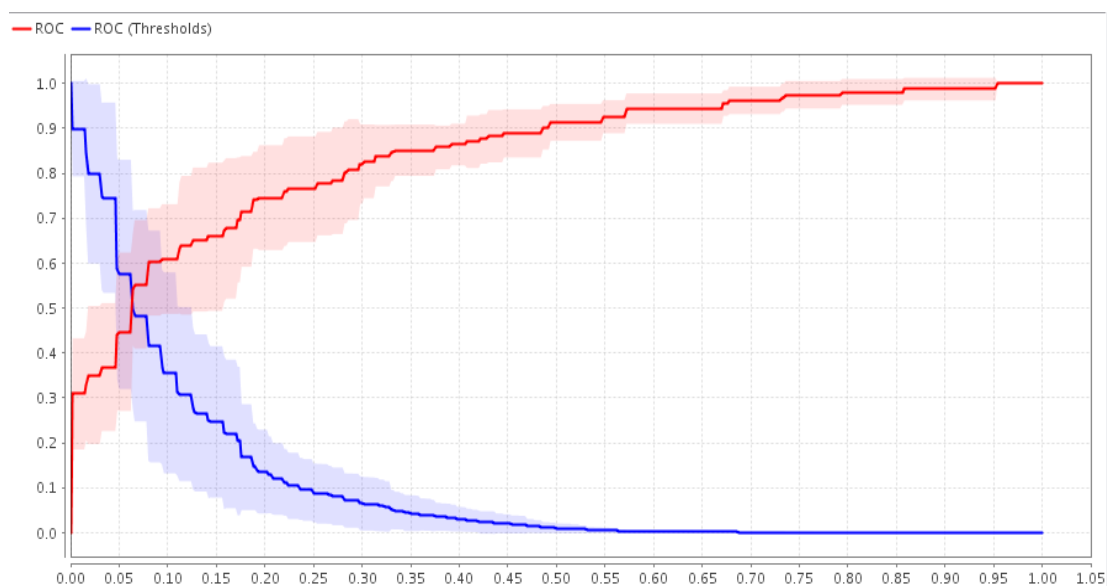


Fig. 2. AUC Graduation of students with neural networks

In Figure 2, we can observe that the results of using the Neural Network (NN) method yield an AUC value of 0.845%. This AUC value reflects the quality of the curve generated by the method, and

*name of corresponding author



this number indicates that the NN method performs very well. In this context, the very good quality of the curve indicates that the NN method effectively and accurately models and predicts student graduation data.

Furthermore, by experimenting with student graduation data using the NN method, the result is the formation of a neural network model architecture. Figure 3 presents a visualization of this architecture, which is a representation of how the NN method organizes and processes data to make predictions about student graduation.

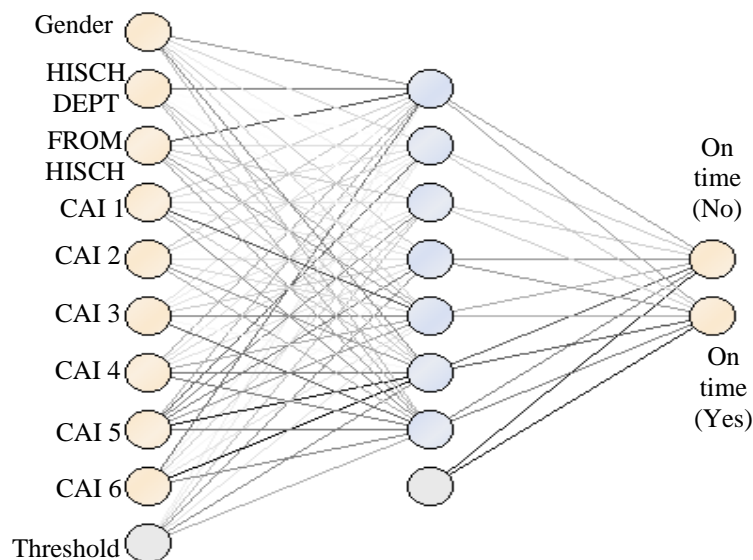


Fig. 3. Student Graduation Architecture with neural networks

In Figure 3, we can clearly see the structure of the Neural Network architecture used in data processing. The first layer, which is the input layer, consists of ten elements with distinct roles. Each element in this layer represents relevant attributes such as gender, high school major, high school origin, and GPA values from semester 1 to semester 6. There is also one element that functions as a threshold value used in the calculation process.

Next, in the hidden layer, a single hidden layer is applied. Within this hidden layer, there are weights assigned with predetermined values for each attribute originating from the input layer. The calculation process in this hidden layer is a crucial part of data processing in the neural network, and these weights play a role in generating the final predictions.

Furthermore, this network architecture produces two outputs that can be interpreted as predictions related to student graduation. These two outputs can indicate whether a student is estimated to graduate on time or not based on the data processed by this neural network. Thus, this depiction provides a more detailed understanding of how this artificial neural network works in data processing and prediction.

Method Optimization

After applying the neural network method to student graduation data, the next step is to improve the accuracy of this data using a genetic algorithm method through RapidMiner software with a 10-fold cross-validation method. The genetic algorithm method is specifically designed to enhance the weights applied in the neural network method. Therefore, by combining these two methods, we can obtain more accurate results in processing and analyzing student graduation data. After the testing process of the student graduation dataset using the NN method is completed, the research results will be reflected in Table 3.

Table 3. Confussion Matrix

*name of corresponding author



	TP	TN	class precision
FP	602	69	89.72%
FN	32	93	74.40%
class recall	94.95%	57.41%	

TN = 93
TP = 602
FN = 69
FP = 32

$$Accuracy = \frac{602 + 93}{602 + 32 + 93 + 69} = \frac{695}{796} = 87.33$$

From the presented confusion matrix, it can be concluded that there are 604 students predicted to graduate on time, and this prediction matches reality. However, there are 69 students predicted to graduate on time, but they actually did not. Conversely, 93 students were predicted not to graduate on time, and this prediction matches reality, but 32 students predicted not to graduate on time actually graduated on time. As a result, the accuracy rate of this confusion matrix reaches 87.33%. In addition, Figure 4 displays the AUC curve that illustrates the student graduation data graph.

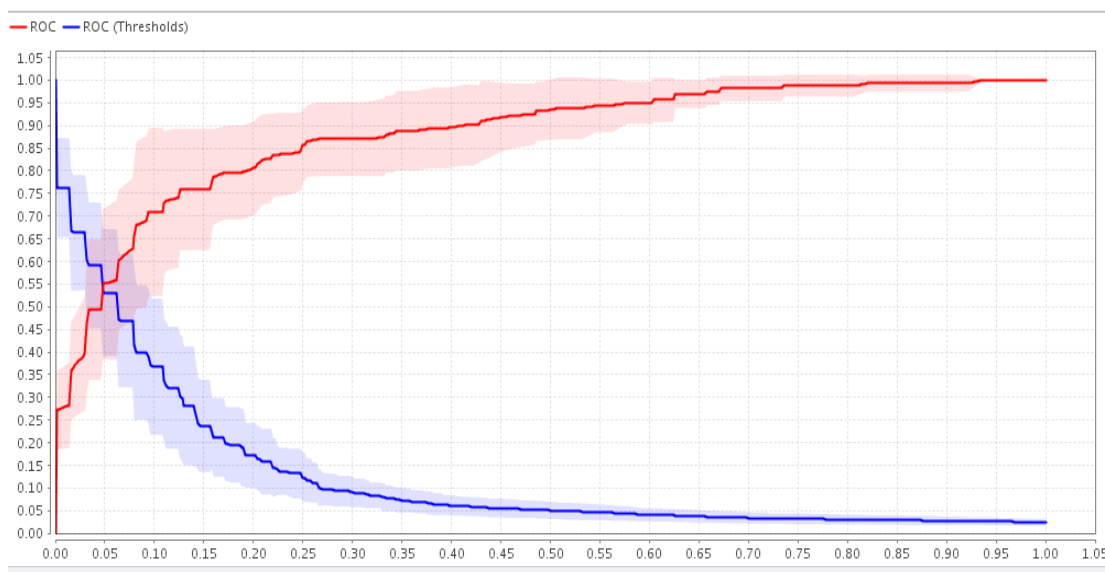


Fig 4. AUC Graduation of students Improved NN with GA

In Figure 4, we can observe that the results of using the Neural Network (NN) method with Genetic Algorithm (GA) yielded an AUC value of 0.882%. This AUC value reflects the quality of the curve generated by the method, and this number indicates that the performance of the GA method applied to NN is very good. In this context, the very good quality of the curve indicates that the GA method applied to NN effectively and accurately models and predicts student graduation data.

Evaluation of Results

The student graduation data tested using the NN method with GA obtained a significant increase in accuracy and AUC, with the performance of each method assessed through metrics in each method to evaluate the quality of the model, as seen in the comparison of accuracy and AUC in Table 4.

Table 4. Hasil Perbandingan Algoritma

*name of corresponding author



Algorithm	Neural Network	NN Optimasi GA
TP	594	602
TN	83	69
FP	79	93
FN	40	32
Accuracy	84.55%	87.33%
AUC	0.845%	0.882%

Table 4 shows that the accuracy of the two methods increased by 02.78% from 84.55% to 87.33%. And the AUC value increased by 0.037% from 0.845% to 0.882%. so it can be seen in figure 5, the diagram formed by the results of table 4.

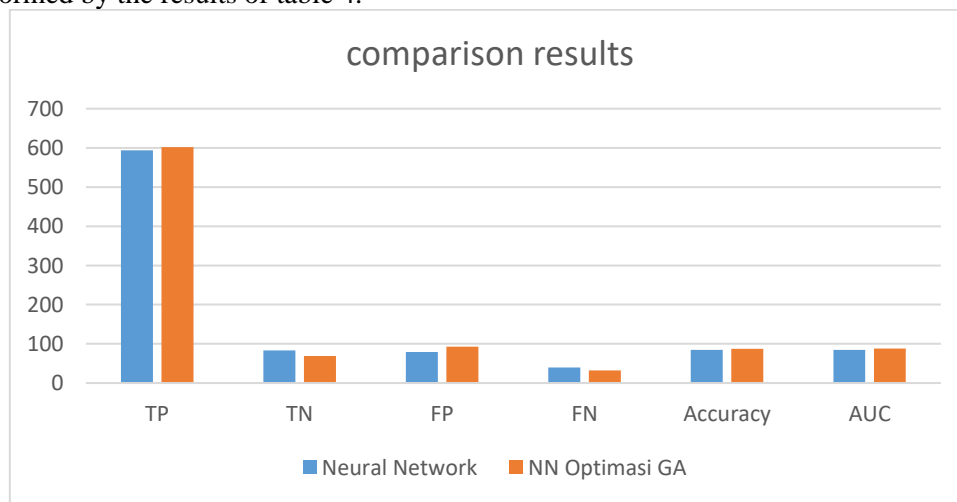


Fig. 5. Comparison results

DISCUSSIONS

The graduation data of the students used has undergone a data preprocessing stage, including steps such as removing duplicate data, filling in missing data with the average values of each attribute, and conducting tests using the NN method. The initial test results showed an accuracy rate of 84.55%. Furthermore, the data was retested by optimizing the NN method using genetic algorithms (GA), resulting in an accuracy rate of 87.33% and an AUC value of 0.882%, an increase from the previous 0.845%.

This experiment demonstrates a significant improvement in the accuracy of classifying student graduations through the application of GA to enhance NN performance. These results can be valuable in providing more accurate diagnoses and enabling students to graduate on time before the semester ends. Therefore, this research has the potential to provide significant benefits to students pursuing higher education.

One of the main contributions of this research is the customized use of GA to fit the characteristics of student graduation data. The GA method allows NN to describe more complex relationships within the data, ultimately enhancing accuracy in predicting graduations, which is a crucial step in the context of higher education.

Future research can focus on efforts to further improve the performance of neural network models themselves. This can be done by exploring more complex neural network architectures, using larger datasets, or employing other optimization techniques.

Therefore, this research makes a significant contribution to improving student graduation prediction using the enhanced NN method with GA. The improvement in accuracy has the potential to greatly assist universities in managing and monitoring student graduation more effectively.

*name of corresponding author



CONCLUSION

In conclusion, this journal has successfully demonstrated that the use of genetic algorithms in optimizing the parameters of artificial neural network models can significantly improve student graduation prediction. The prediction accuracy for passing increased by 02.78%, and the Area Under the Curve (AUC) value also increased by 0.037%. These results provide strong evidence that genetic algorithms are an effective approach to enhancing the performance of student graduation prediction models.

REFERENCES

- Ariyati, I., Ridwansyah, & Suhardjono. (2018). Implementasi Particle Swarm Optimization untuk Optimalisasi Data Mining Dalam Evaluasi Kinerja Asisten Dosen. *Jurnal Informatika Dan Komputer (JIKO)*, 3(2), 70–75.
- Chen, H., Zhang, J., Xu, Y., Chen, B., & Zhang, K. (2012). Performance comparison of artificial neural network and logistic regression model for differentiating lung nodules on CT scans. *Expert Systems with Applications*, 39(13), 11503–11509. <https://doi.org/10.1016/j.eswa.2012.04.001>
- Hendra, Azis, M. A., & Suhardjono. (2020). ANALISIS PREDIKSI KELULUSAN MAHASISWA MENGGUNAKAN DECISSION TREE BERBASIS PARTICLE SWARM OPTIMIZATION. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(1), 102–107. <https://doi.org/https://doi.org/10.32736/sisfokom.v9i1.756>
- K, J., & M, P. (2012). Optimizing Weights of Artificial Neural Networks using Genetic Algorithms. *Int. J. Adv. Res. Comput. Sci. Electron. Eng*, 1(10), 2277–9043.
- Moro, S., & Laureano, R. M. S. (2011). Using Data Mining for Bank Direct Marketing: An application of the CRISP-DM methodology. *European Simulation and Modelling Conference*, 117–121.
- Nazari, A., & Pacheco Torgal, F. (2013). Predicting compressive strength of different geopolymers by artificial neural networks. *Ceramics International*, 39(3), 2247–2257. <https://doi.org/10.1016/j.ceramint.2012.08.070>
- Nurdin, H., Sartini, Sumarna, Maulana, Y. I., & Riyanto, V. (2023). Prediction of Student Graduation with the Neural Network Method Based on Particle Swarm Optimization. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 8(4), 2353–2362. <https://doi.org/10.33395/sinkron.v8i4.12973>
- Ridwansyah, & Purwaningsih, E. (2018). Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank. *Jurnal PILAR Nusa Mandiri*, 14(1), 83–88.
- Ridwansyah, R., Wijaya, G., & Purnama, J. J. (2020). Hybrid Optimization Method Based on Genetic Algorithm for Graduates Students. *Jurnal Pilar Nusa Mandiri*, 16(1), 53–58. <https://doi.org/10.33480/pilar.v16i1.1180>
- Riyanto, V., Hamid, A., & Ridwansyah. (2019). Prediction of Student Graduation Time Using the Best Algorithm. *Indonesian Journal of Artificial Intelligence and Data Mining*, 2(2), 1–9. <https://doi.org/http://dx.doi.org/10.24014/ijaidm.v2i1.6424>
- S, R., & N, R. (2016). Diagnosis of chronic kidney disease using machine learning algorithms. *International Journal of Innovative Research in Computer and Communication Engineering*, 4(1), 812–820.
- Salmu, S., & Solichin, A. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naïve Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta. *Seminar Nasional Multidisiplin Ilmu (SENMI) 2017*, (April), 701–709.
- Satapathy, S. C., Chittineni, S., Krishna, S. M., Murthy, J. V. R., & Reddy, d P. V. G. D. P. (2012). Kalman particle swarm optimized polynomials for data classification. *Appl. Math. Model*, 36(1), 115–126.
- Suwardjono, Wijaya, G., & Hamid, A. (2019). PREDIKSI WAKTU KELULUSAN MAHASISWA MENGGUNAKAN SVM BERBASIS PSO. *Bianglala Informatika*, 7(2), 97–101.
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>

*name of corresponding author



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.