

The Optimization of Neural Network Based PSO Feature Selection in the Classification of Graduates Working According to Their Field

Very Kurnia Bakti^{1,*} Dairoh²

¹Department of Computer Engineering, Polytechnic Harapan Bersama, Tegal, Indonesia

²Department of Informatics Engineering, Polytechnic Harapan Bersama, Tegal, Indonesia

*Corresponding author. Email: verykurniabakti@gmail.com

ABSTRACT

The many college graduates who work not by their field of knowledge. Obtained show that the horizontal alignment of < 80% in the last three years has not reached the ideal value. The question that often arises is why this can happen and what influences can determine the quality result of graduates why they don't work in their fields. we need a model that is used to see a pattern of graduates to work according to their scientific fields. In this research, the neural network algorithm method is used to see a pattern of graduates who work according to their scientific fields. The neural network is an algorithm method that can be used as a reliable classification algorithm but has shortcomings in its selection of features, wherewith the combination PSO has a good ability to solve problems that have non-linear and non-differentiable characteristics, multiple optima, large dimensions through good adaptations. derived from social psychological theory. We have obtained a higher accuracy of 71.51% using the combination of neural networks with PSO than that of one of the methods without PSO.

Keywords: *College, neural network, PSO Feature Selection, Graduate.*

1. INTRODUCTION

Computing technology is currently developing and facilitating human work in carrying out complex calculation due to large traffic and data, the presence of large data or commonly known as big data can benefit an institution to be taken into consideration in determining the direction of policies or decisions, of course, it must assist by data processing by utilizing algorithm methods as support [1].

Many university graduates who work are not by their field of knowledge, this is based on data obtained from the tracer study of one of the private universities, namely the joint hope Polytechnic. The figures obtained show that horizontal alignment < 80% in the last three years has not yet reached the ideal value. The question that often arises is why this can happen and what influences can determine the quality results of graduates why they don't work in their fields [2][3].

One algorithm method that can be used as a feature selection algorithm in processing data is PSO (Particle Swarm Optimization). The PSO algorithm can calculate

with its characteristic population-based optimization methods such as the genetic algorithm (GA). PSO is based on the behavior of a flock of birds, where social behavior consists of individual actions and the influence of other individuals in the group. PSO has a good ability to solve the problem that has nonlinear characteristics and non-diversity, multiple optima, large dimensions through adaptations derived from social psychology theory [4].

The combination method between the Neural Network and the PSO selection feature is expected that the output of the resulting model has high accuracy so that it can produce. The purpose of this study is to find a correlation between the influential attributes then increase the accuracy of the neural network algorithm at once finding the optimal model to solve problems for graduates so that they work according to their fields accurate decisions in making anticipatory decision decisions for the pre-graduate student so that they are suitable to work in the field [5][6].

2. METHOD

The data used is tracer study data, taken from the tracer study system of Harapan Bersama Polytechnic. Describe the method proposed in this study, namely the particle swarm optimization method on the neural network which shows a diagram of the activity.

In the initial data set processing, the data set is in the form of testing dan learning data, each particle represents a feature subset which is the solution of the candidate data set which is transformed into ranges 0 and 1, then the data set is divided by the 3-fold cross-validation method, after that the data set is tested with a neural network model and validates the resulting model to get the accuracy of the model if the accuracy criteria are met it will result in optimized features if the accuracy criteria are not met then look for attributes that are not optimal, next funding non-optimal and eliminating these suboptimal attribute so that the neural network optimization model is obtained so that it is retested with the neural network model. Data collection this study used secondary data from Harapan Bersama Polytechnic (graduates in 2017 to 2019). Data collection was collected from graduates. A total of 1281 data and label comparisons 657 inline or 52% and 624or 48% out of line.

The method proposed to find out the value of the accuracy of the data whether from this data the role of practical work is more effective in determining work according to the field through the application of the neural network method with the PSO algorithm using MATLAB Coding, form the obtained result that has been tested will be used as the basis for determining graduates to work accordingly field [7].

The data is tested using the neural network method with the PSO algorithm which will produce accuracy in the process of determining graduates to work according to their fields. Testing using validation techniques to get accuracy value. The data set was tested with the proposed method through rapid miner coding. After the testing phase is applied, the test result of the neural network method with the PSO algorithm can be seen as the level of accuracy matrix, so that the purpose of applying the neural network method with the PSO algorithm is to determine the graduates to work by their fields. The validation model we use is the 3-fold cross-validation method to study and test data. We divided the training data into three equal parts and then carried out the learning process three times, each time we selected another part of the data set for testing and used the remaining two parts for training. We use three-cross-validation, see Table 1.

Table 1. Illustration of 3-fold cross-validation

Validation to-n	Dataset
1	
2	
3	

The 3-fold cross-validation method has become the standard method for data learning and testing. The use of public data sets makes research that can be re-examined, undeniable, and verifiable [8].

The evaluation model that is applied is the determination of the accuracy obtained from the confusion matrix result. The confusion matrix is used to evaluate the classification model based on the calculation of the test object which is tabulated into a table where it will be predicted to be true and false. A confusion matrix is a data set having only two classes, one class as positive and the other as negative. Consist of four cells, namely True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN)[9] see Table 2.

Table 2. Confusion matrix for 2 class models

Classification	Predicted class		
	class = yes	class = no	
Observed class	class = yes	A (true positive - TP)	B (false negative- FN)
	class = no	C (false positive- FP)	D (true negative- TN)

So, to calculate accuracy the following formula is used:

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

3. RESULT AND DISCUSSION

The data obtained are private from the Academic section of Harapan Bersama Polytechnic and tracer study data. This study uses the neural network method with the PSO algorithm. The result of the study aims to determine the accuracy value of whether the role of practical work during college is more effective in predicting graduates to work according to their fields by applying the neural network method with the PSO algorithm. The attribute weights obtained from each attribute are as follow internship during college is more effective in predicting graduates to work according to their fields by applying the neural network method with the PSO algorithm. The attribute weights obtained from each attribute are as follows. see Table 3.

Table 3. Attribute weights obtained from each attribute

Attribute	Wight
graduate year	0.338
Study Program Code	0.192
IPK	0.511
pass on time	0.368
Scholarship	1
Gender	0
Age	0.432
Class	0.427
Practical Work (KP)	0.998

It can be concluded that the attributes that affect the graduates according to the field are year of graduates, study program code, IPK, graduation on time, scholarship, age, class, and practical work (KP). While the attribute that does not affect is Gender. The result of the experiments carried out on the data set formed the following neural network architecture. See Figure 1.

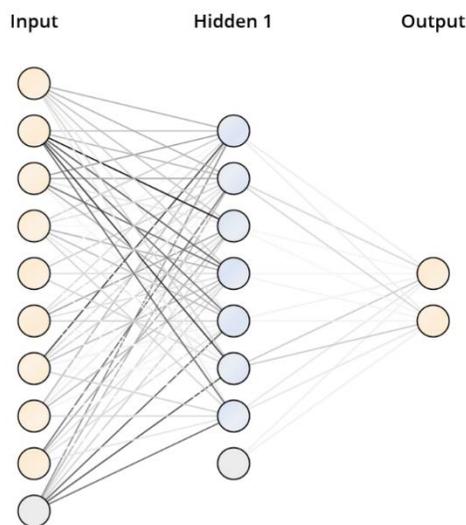


Figure 1 Neural network architecture that is formed

In the neural network algorithm model design, the training process is carried out with the following composition:

1. Hiden Layer: 7
2. Training Cycles:500
3. Learning rate:0.1
4. Momentum :0.9

Network-based on particle swarm optimization by comparing the accuracy value, in determining the level of accuracy of the comparison, the confusion matrix obtained is either the neural network model itself or the neural network based on particle swarm optimization based on 16384 processed data, with the result of the comparisons in Table 4 and Table 5.

Table 4. Accuracy value with NN

Accuracy 64.32%			
	true 1	true 0	class precision
predictions 1	784	320	71.01%
predictions 0	137	40	22.60%
class recall	85.12%	11.11%	

Based on table 4 it can be classified as follows:
 True Positive (TP) =784 records
 False Negative (FN) =137 records
 True Negative (TN) = 40 records
 False positive (FP) = 320

The classification results show that the level of accuracy using the neural network algorithm is 64.32% and the testing data with the neural network model optimized with PSO obtained the following results.

Table 5. Accuracy Value using NN with PSO

Accuracy 71.51%			
	true 1	true 0	class precision
predictions 1	915	359	71.82%
predictions 0	6	1	14.29%
class recall	99.35%	0.28%	

Based on table 5 it can be classified as follows:
 True positive (TP) = 915 records
 False Negative (FN) = 359 records
 True Negative (TN)= 1 records
 False Positive (FP) = 6 records

The PSO algorithm can calculate with its characteristic population-based optimization methods such as the genetic algorithm (GA). PSO is based on the behavior of a flock of birds, where social behavior consists of the actions of individuals and the influence of other individuals in a group, while the neural network is an adaptive system that can change its structure to solve problems based on external and internal information flowing through the network. The classification result shows that the level of accuracy using the neural network algorithm is 71.51%.

4. CONCLUSION

This study aims to determine the accuracy value of wart treatment through immunotherapy by applying the PSO algorithm-based neural network method. Based on the measurement of the level accuracy through the rapid miner application, it is known that the Neural Network (NN) model with the PSO algorithm has an accuracy value of 71.51% greater than the accuracy of the Neural Network method alone, which is 64.32%. The Neural Network method with PSO algorithm has better accuracy than the Neural Network method. Comparisons have also

been made to the decision tree algorithm. This research was conducted using the PSO algorithm, which has the advantage that this algorithm can increase the level of accuracy. The Neural Network method with the PSO algorithm can be used to classify graduates according to their fields.

ACKNOWLEDGMENTS

The resulting level of accuracy is not perfect because this study uses new objects, so it still requires further research. It is also necessary to explore other supporting attributes that can affect working according to fields, this research founded by Harapan Bersama Polytechnic.

REFERENCES

- [1] Y. Guo, "A Comprehensive Review of Deep Learning Methods and Applications," *Deep Learn. Vis. Underst.*, 2017, [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0925231215017634>.
- [2] A. Syafiq and S. Fikawati, "Tracer Study : Melacak Jejak Lulusan FKM UI (Hasil Study Kualitatif Tracer Sarjana Kesehatan Masyarakat FKM UI 2006)," *Kesmas Natl. Public Heal. J.*, vol. 1, no. 6, p. 252, 2007, doi: 10.21109/kesmas.v1i6.285.
- [3] V. Kurnia Bakti, M. Noval, and E. Purnomo Bayu Aji, "Sistem Pre Kompilasi Data Tracer Studi Online Ditjen Belmawa Ristekdikti(Studi Kasus: Politeknik Harapan Bersama)," *J. Inform. J. Pengemb. IT*, vol. 2, no. 1, pp. 50–53, 2017.
- [4] Ridwansyah and E. Purwaningsih, "Particle Swarm Optimization Untuk Meningkatkan Akurasi Prediksi Pemasaran Bank," *J. PILAR Nusa Mandiri*, vol. 14, no. 1, pp. 83–88, 2018.
- [5] P. Plant and O. Tinggi, "Implementasi Metode PSO-LDW untuk Optimasi Kontroler PID Pada Plant Orde Tinggi," pp. 2–5.
- [6] R. L. Rose, B. D. A. Selvi, and R. L. R. Singh, "Development of Hybrid Algorithm Based on PSO and NN to Solve Economic Emission Dispatch Problem," *Circuits Syst.*, vol. 07, no. 09, pp. 2323–2331, 2016, DOI: 10.4236/cs.2016.79202.
- [7] M. N. Alam, "Codes in MATLAB for Particle Swarm Optimization," no. March, pp. 8–11, 2016, DOI: 10.13140/RG.2.1.1078.7608.
- [8] H. Moayedi, A. Osouli, H. Nguyen, and A. S. A. Rashid, "A novel Harris hawks' optimization and k-fold cross-validation predicting slope stability," *Eng. Comput.*, no. 0123456789, 2019, DOI: 10.1007/s00366-019-00828-8.
- [9] S. Ruuska, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, and J. Mononen, "Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behavior of cattle," *Behav. Processes*, vol. 148, pp. 56–62, 2018, DOI: 10.1016/j.beproc.2018.01.004.