

Abstract of thesis presented to the Senate of Universiti Malaysia Terengganu in
fulfilment of the requirements for the degree of Master of Science

**OPTIMIZATION OF ARTIFICIAL NEURAL NETWORK (ANN) MODELS
USING DIFFERENT TRAINING ALGORITHMS FOR AIR QUALITY
PREDICTION**

NUR NATASHA BINTI ABDUL SAMAT

2022

Main Supervisor : Samsuri Abdullah, Ph.D

Co- Supervisor : Professor Marzuki Ismail, Ph.D

Associate Professor Nurul Adyani Ghazali, Ph.D

**Faculty : Faculty of Ocean Engineering Technology and
Informatics**

Forecasting particulate matter (PM₁₀) concentration is essential, which helps local authorities take precautionary measures and implement significant actions to improve air quality in specific locations. This study aims to establish the nonlinear model for PM₁₀ concentration forecasting in selected areas that apply three different training algorithms. Data for Pasir Gudang (S1), Kemaman (S2), Prai (S3) and Nilai (S4) from January 2007 until December 2014 was selected for forecasting PM₁₀ concentration levels. Parameters in this study are daily observations of PM₁₀, meteorological factors (wind speed, temperature, and relative humidity) and gaseous pollutants (SO₂, NO₂ and CO). Multilayer Perceptron (MLP) is a feed-forward artificial neural network model used in this study to establish a nonlinear model for PM₁₀ concentration by using three different training algorithms which are Levenberg-Marquardt (LM), Gradient descent with momentum and adaptive learning rate backpropagation (GDX) and Bayesian regularization (BR). The most critical part of MLP models is determining the number of neurons in the hidden layer. During the training of the models, it was found that the range of the number of neurons is 10 - 14 neurons. The best-fitted model was selected through the evaluation of performance indicators at which Coefficient of determination

(R^2) was used to check the accuracy of the model, Normalized Absolute Error (NAE) and Root Mean Square Error (RMSE) were used to find the error of the model. Higher accuracy is given by a value closer to 1. The model's error value closer to 0 indicated a better model. BR was selected as the best-fitted model for Station 1, Station 3 and Station 4. Meanwhile, for Station 2, LM was selected. R^2 value are range from 0.493 - 0.742 and for RMSE range from 6.764 - 12.466 $\mu\text{g}/\text{m}^3$. Overall, these models proved to increase the accuracy of forecasting with 30.78% (S1), 25.63% (S4), 14.40% (S3) and 9.51% (S2). It also denoted that these nonlinear models were able to reduce the error by 51.46% (S1), 25.32% (S4), 18.01% (S3) and 6.57% (S2).

Abstrak tesis yang dikemukakan kepada Senat Universiti Malaysia Terengganu
sebagai memenuhi keperluan untuk Ijazah Sarjana Sains

**PENGOPTIMUMAN MODEL RANGKAIAN NEURAL BUATAN (ANN)
MENGUNAKAN LATIHAN ALGORITMA YANG BERBEZA BAGI
PERAMALAN KUALITI UDARA**

NUR NATASHA BINTI ABDUL SAMAT

2022

Penyelia Utama : Samsuri Abdullah, Ph.D

Penyelia Bersama : Profesor Marzuki Ismail, Ph.D

Profesor Madya Nurul Adyani Ghazali, Ph.D

Fakulti : Fakulti Teknologi Kejuruteraan Kelautan dan Informatik

Ramalan kepekatan zarah terampai (PM_{10}) adalah sangat penting bagi membantu pihak berkuasa tempatan untuk mengambil langkah berjaga-jaga dan melaksanakan tindakan yang penting untuk meningkatkan kualiti udara di lokasi tertentu. Kajian ini bertujuan dalam penubuhan model tak linear untuk kepekatan ramalan PM_{10} di kawasan terpilih dengan menggunakan tiga algoritma latihan yang berbeza. Data Pasir Gudang (S1), Kemaman (S2), Prai (S3) dan Nilai (S4) dari Januari 2007 sehingga Disember 2014 telah dipilih untuk meramal tahap kepekatan PM_{10} . Parameter dalam kajian ini ialah pemerhatian harian PM_{10} , faktor-faktor meteorologi (kelajuan angin, suhu dan kelembapan relatif) dan pencemar gas (SO_2 , NO_2 dan CO). Perseptron Pelbagai Lapis (MLP) adalah suap depan tiruan model rangkaian neural yang digunakan untuk penubuhan model tak linear kepekatan PM_{10} menggunakan tiga algoritma latihan yang berbeza iaitu *Levenberg-Marquardt* (LM), *Gradient descent with momentum-adaptive learning rate backpropagation* (GDX) dan *Bayesian regularization* (BR). untuk melatih set data. Bahagian yang paling kritikal dalam model MLP adalah penentuan bilangan neuron dalam lapisan tersembunyi. Semasa latihan model, didapati bahawa julat bilangan neuron adalah 10 - 14 neuron. Model terbaik dipilih melalui penilaian petunjuk prestasi di mana pekali penentuan (R^2) digunakan untuk memeriksa ketepatan model, ralat mutlak normal multak (NAE) dan punca min ralat kuasa dua (RMSE)

digunakan untuk mencari ralat model. Ketepatan lebih tinggi diberikan oleh nilai lebih dekat dengan 1 dan ralat model di mana nilai lebih dekat dengan 0 menunjukkan model yang lebih baik. BR telah dipilih sebagai model yang terbaik untuk Stesen 1, Stesen 3 dan Stesen 4. Sementara itu, bagi Stesen 2, LM telah dipilih. Nilai R^2 adalah berjangka daripada 0.493 - 0.742 dan untuk RMSE 6.764 - 12.466 $\mu\text{g}/\text{m}^3$. Secara keseluruhan, model ini terbukti meningkatkan ketepatan ramalan dengan 30.78% (S1), 25.63% (S4), 14.40% (S3) dan 9.51% (S2). Ia juga ditandakan bahawa model tidak linear ini dapat mengurangkan ralat dengan 51.46% (S1), 25.32% (S4), 18.01% (S3) dan 6.57% (S2).