

PERAMALAN KUALITI AIR DI TASIK CHINI MENGGUNAKAN MODEL RANGKAIAN NEURAL BUATAN

WAN ZUHAI DAH BINTI WAN IBRAHIM
AZURALIZA BINTI ABU BAKAR

Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia

ABSTRAK

Peramalan kualiti air adalah satu analisis ramalan yang berkaitan dengan tugas menganalisis data dalam aspek perlombongan data. Kaedah ini melibatkan analisis pemantauan data siri masa. Peranan peramalan kualiti air amat penting terutama sekali kepada hidupan air dan juga alam sekitar bagi mengelakkan berlakunya pencemaran air yang terlalu teruk. Kajian ini bertujuan untuk membangunkan sebuah sistem peramalan kualiti air di Tasik Chini. Dalam kajian ini, teknik rangkaian neural buatan digunakan bagi proses permodelan data dan fungsi multi aras perceptron (*multilayer perceptron*) digunakan untuk membangunkan model data yang terbaik dengan menggunakan Weka 3.9. Dalam kajian ini set data yang mengandungi 4 atribut input dan 1 atribut kelas dengan 511 data bagi setiap atribut telah digunakan. Hasil ujikaji menunjukkan model terbaik diperoleh pada bilangan aras tersembunyi iaitu 1 dan kadar pembelajaran dan momentum iaitu 0.3 dan 0.2 dengan peratusan ketetapan sebanyak 92.16%. Model terbaik yang dihasilkan digunakan dengan mengambil nilai-nilai pemberat bagi setiap aras untuk membangunkan sistem dengan menggunakan *Microsoft Visual Studio 2010*. Kajian ini mampu membangunkan sebuah model yang memperoleh peratusan ketetapan pengelasan yang tinggi.

PENGENALAN

Tasik Chini merupakan sebuah tasik paya yang terbina hasil daripada pelbagai proses semulajadi. Kandungan air di tasik chini ini juga berubah-ubah dan kewujudan tasik ini juga bergantung terus kepada persekitaran dan habitat semulajadi. Antara fungsi tasik ini ialah menjadi habitat bagi hidupan liar daratan dan air dan kebanyakannya spesies yang dilindungi. Selain itu, tasik ini juga merupakan kawasan tadahan air semulajadi dan merupakan sumber protein dan bekalan air untuk masyarakat orang asli yang tinggal berdekatan dengan tasik tersebut.

Kajian ini menggunakan kaedah Rangkaian Neural Buatan bagi meramal dan mengawal kualiti air di Tasik Chini. Rangkaian neural ini menyediakan cara-cara yang fleksibel dan cepat bagi mewujudkan model untuk anggaran kualiti air sungai. Sejak beberapa tahun kebelakangan ini, rangkaian neural buatan telah menunjukkan prestasi yang luar biasa sebagai alat regresi, terutamanya apabila digunakan untuk pengecaman corak dan fungsi anggaran. Rangkaian neural ini juga adalah satu kaedah komputasi yang diilhamkan oleh kajian otak dan sistem saraf dalam organisma biologi. Ia juga mewakili model matematik yang sangat unggul tentang pemahaman kita sekarang sebagai sistem yang kompleks (Sarkar, & Pandey 2015).

Oleh yang demikian, Kajian ini dijalankan untuk megelompokkan data kualiti air iaitu suhu, kekonduksian, pH, kekeruhan dan oksigen terlarut (DO) untuk memperoleh corak data yang boleh digunakan untuk membuat ramalan tersebut. Sehubungan dengan itu, proses perlombongan data dengan menggunakan model Rangkaian Neural Buatan akan dibangunkan.

PENYATAAN MASALAH

Kajian ini dijalankan di stesen Jerangking di Tasik Chini, Pahang. Kajian ini dilakukan bagi meramal tahap kualiti air supaya tidak berlakunya pencemaran air yang akan memberikan banyak kesan buruk kepada kehidupan di muka bumi terutamanya hidupan yang menetap di sana sama ada manusia, haiwan atau tumbuhan dan kualiti air di Tasik Chini ini juga dan ianya dapat dikawal pada peringkat awal.

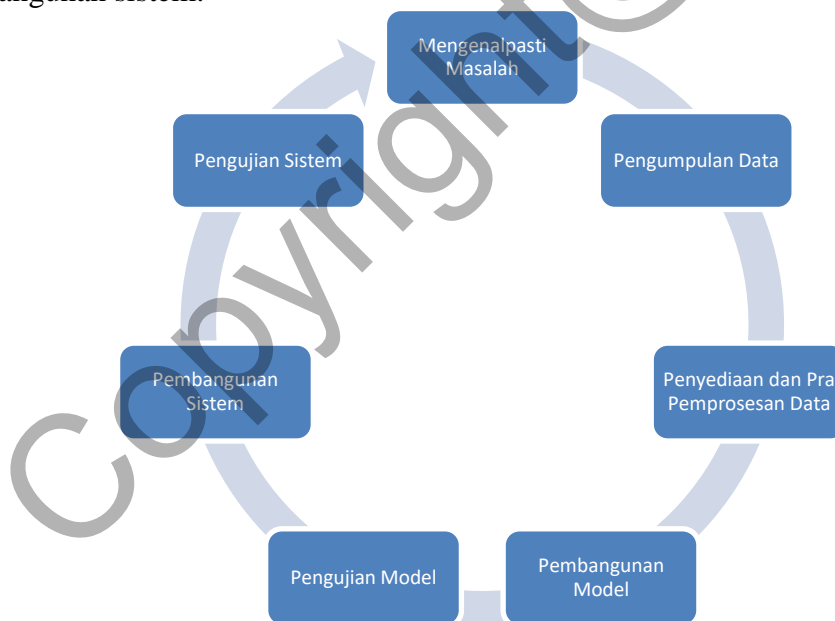
OBJEKTIF KAJIAN

Matlamat utama kajian ini ialah untuk membangunkan sistem cerdas bagi memantau tahap kualiti air di Tasik Chini dengan berdasarkan objektif berikut :

1. Mengenalpasti parameter-parameter yang bersesuaian dengan tahap kualiti air bagi stesen Jerangking di Tasik Chini.
2. Membangunkan satu model pengelasan tahap kualiti air berdasarkan tahap kekeruhan air menggunakan kaedah rangkaian neural buatan.
3. Membangunkan sebuah sistem cerdas bagi meramal tahap kualiti air di Tasik Chini berasaskan model Rangkaian Neural Buatan.

METOD KAJIAN

Perlombongan data merupakan salah satu kaedah di dalam bidang kecerdasan yang mempunyai kebolehan untuk mengenalpasti corak tertentu di dalam pangkalan data. Metodologi kajian ini mempunyai dua fasa penting iaitu fasa perlombongan data dan fasa pembangunan sistem.



Rajah 1 Metodologi Pembangunan Model

Fasa Mengenalpasti Masalah

Fasa ini melibatkan proses mengenalpasti masalah yang terdapat pada data tersebut. Data kualiti air di Tasik Chini yang diperolehi mempunyai banyak *missing value* dan data-data yang *redundant* kerana data yang diambil mempunyai tarikh dan masa yang berbeza. Oleh itu, beberapa parameter penting sahaja yang dipilih untuk dilombongkan.

Fasa Pengumpulan Data

Kajian ini menggunakan set data kualiti air iaitu yang terdiri daripada parameter suhu, kekonduksian, pH, kekeruhan dan oksigen terlarut (DO) yang telah diperolehi daripada Pusat Penyelidikan Tasik Chini UKM. Data diperolehi adalah dari bulan Februari 2011 sehingga Jun 2012, ia mengandungi 5 atribut dengan 512 data yang telah dibersihkan.

Fasa Penyediaan dan Pra Pemprosesan Data

Fasa ini melibatkan proses pembersihan data, mengenalpasti parameter dan perwakilan data siri masa. Fasa ini dijalankan bagi menyediakan set data yang bersih dan sedia untuk dilombong. Proses pembersihan data dilakukan dengan mengisi setiap ruang kosong data yang tidak lengkap dan membetulkan kesalahan yang terdapat pada data tersebut. Proses mengenalpasti parameter dilakukan dengan berdasarkan set data daripada kualiti air iaitu suhu, kekonduksian, pH, kekeruhan dan oksigen terlarut (DO). Selain itu, fasa penyediaan data melibatkan proses mendiskretkan data serta transformasi data untuk permodelan pengetahuan menggunakan rangkaian neural buatan.

a) Penggabungan Data

Langkah ini merupakan langkah pertama yang dilakukan dalam fasa pra-pemprosesan data. Seterusnya, ianya menepati objektif kedua kajian ini untuk memilih atribut yang sesuai untuk meramal tahap kualiti air. Proses ini menggabungkan lima atribut yang diperolehi daripada Pusat Penyelidikan Tasik Chini UKM antaranya suhu, kekonduksian, kekeruhan, pH dan oksigen terlarut. Atribut Tahun dan Masa tidak dipilih kerana tidak memberi kesan terhadap proses mendapatkan pengetahuan.

b) Pembersihan Data

Proses untuk pembersihan memerlukan masa yang lama berbanding proses-proses yang lain kerana ia merangkumi teknik untuk pemilihan atribut dan data yang berkaitan dengan kajian, mengisi data yang hilang dan membuang data yang tidak berkenaan. Oleh kerana permasalahan tersebut, ianya diatasi dengan beberapa cara iaitu dengan nilai tetap atau menggantikan nilai hilang tersebut dengan nilai purata. Nilai purata digunakan berdasarkan nilai purata pada bulan-bulan tersebut.

c) Pendiskretan Data

Proses pendiskretan data ialah proses untuk menukarkan data numerik kepada bentuk nominal iaitu dengan menggabungkan nilai-nilai atribut tersebut.

d) Normalisasi Data

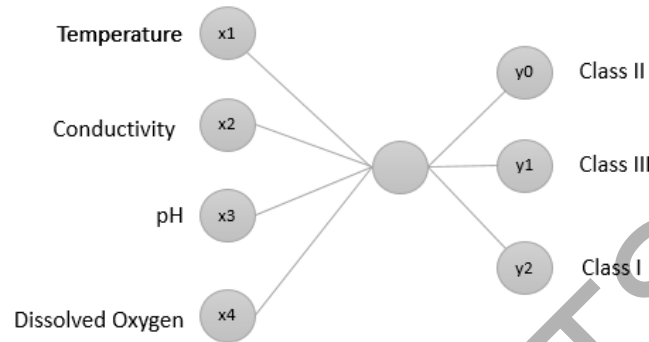
Normalisasi dan pengkategorian data sangat penting supaya set data yang digunakan mampu menghasilkan keputusan yang memuaskan serta dapat mengurangkan masalah kos yang tinggi dan masa yang lama. Data ditukarkan kepada bentuk normalisasi dalam julat 0 hingga 1. Hasil daripada transformasi data, perlombongan data sedia untuk dijalankan. Dalam proses perlombongan data, pengetahuan akan diekstrak untuk pembangunan model ramalan dan pemantauan yang paling efisien untuk digunakan bagi pembangunan sistem pemantauan kualiti air di Tasik Chini.

Fasa Pembangunan Model

Pembangunan model bagi kajian pemantauan kualiti air di Tasik Chini menggunakan kaedah Rangkaian Neural Buatan. Secara umumnya, model ini dinyatakan secara khusus oleh rangkaian topologi, ciri-ciri nod dan kaedah-kaedah pembelajaran atau yang dilatih. Ia adalah satu set pemberat yang berhubung dan mengandungi pengetahuan yang dijana oleh model tersebut. Model ini terdiri daripada sebilangan besar unit pemprosesan yang mudah dan ringkas yang berinteraksi dengan yang lain melalui *excitatory* atau *inhibitory connections* (Rajah 2). Tiga lapisan yang berbeza boleh dibezakan :

- i. Lapisan input – menghubungkan maklumat input kepada rangkaian. Dalam kajian ini

- dua sehingga enam nod input yang telah digunakan yang terdiri daripada oksigen terlarut, kekonduksian, suhu, kekeruhan dan pH.
- ii. Lapisan tersembunyi (satu atau lebih lapisan tersembunyi) – bertindak sebagai perantara kepada lapisan pengiraan.
 - iii. Lapisan output – menghasilkan output yang dikehendaki dalam kes ini iaitu kelas kualiti air.



Rajah 2 Contoh Konfigurasi ANN

Latihan ini dianggap sebagai pendekatan berhenti awal dan hanya set latihan yang digunakan untuk menentukan pemberat dan bias. Fungsi pengaktifan (activation function) yang digunakan dalam rangkaian ini ialah fungsi pengaktifan sigmoid. Fungsi sigmoid yang diujulkan adalah dari 0 hingga 1.

$$f(x) = \frac{1}{1 + e^{-x}}$$

Fasa Pengujian Model

Dalam proses pembangunan model, set data kualiti air dibahagikan kepada 2 iaitu data latihan dan data ujian. Langkah ini dilakukan bagi mendapatkan peratusan ketetapan yang paling tinggi. Nilai peratusan ketetapan yang paling tinggi ialah 100%, iaitu semua set data tersebut berjaya dikelaskan dengan betul dan tepat. Model data yang mempunyai peratusan yang tinggi dan terbaik akan digunakan, pemberat-pemberat bagi model tersebut akan diimplementasikan ke dalam untuk meramal tahap kualiti air di Tasik Chini.

Fasa Pembangunan Sistem

Pembangunan sistem cerdas ini menggunakan algoritma perambatan-balik dan bahasa pengaturcaraan yang digunakan ialah Microsoft Visual Studio 2010. Sistem peramalan kualiti air ini dibangunkan adalah untuk memberi keputusan yang cepat dan tepat dalam menentukan tahap kualiti air. Sistem ini membolehkan pengguna memasukkan nilai bagi parameter kualiti air dan sistem itu akan menghasilkan keputusan dengan menentukan tahap kualiti air tersebut dengan masa yang cepat. Sistem ini akan menghasilkan keputusan dengan membuat pengiraan nilai pemberat bagi setiap atribut bagi aras –aras rangkaian neural buatan.

Selepas proses permodelan data, model yang dibangunkan akan digunakan dalam fasa pembangunan sistem. Model tersebut akan menjadi asas kepada pembangunan bagi memantau kualiti air di Tasik Chini. Pemberat yang terdapat pada model rangkaian neural buatan akan dikodkan semula kepada bentuk Microsoft Visual Studio 2010.

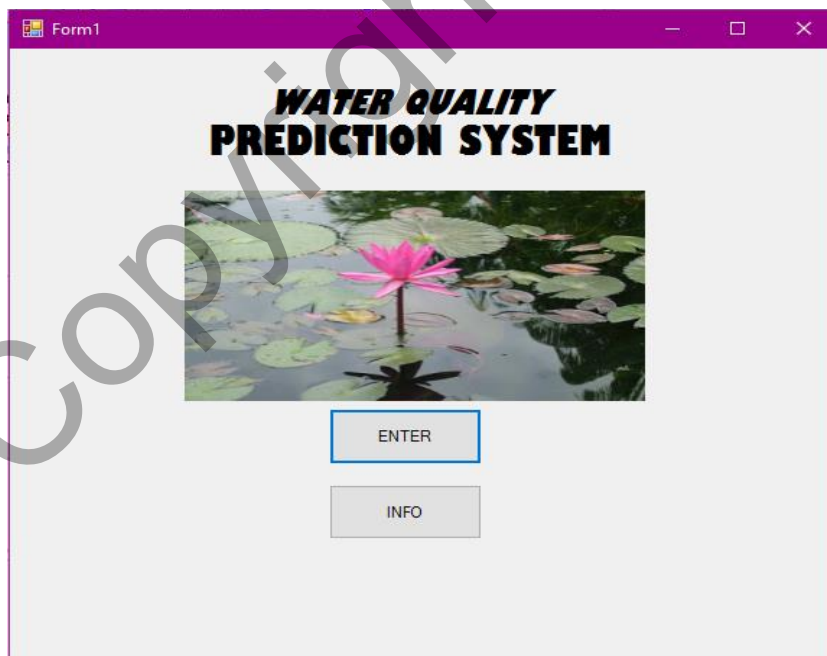
Fasa Pengujian Sistem

Fasa pengujian sistem ini bagi memastikan sistem melaksanakan apa yang pelanggan kehendaki. Dalam fasa ini, sistem akan diuji secara keseluruhan apabila Sistem Pemantauan Kualiti Air di Tasik Chini selesai dibina. Pengujian ini dilakukan oleh pengguna untuk memastikan sistem yang dibina dapat memenuhi keperluan pengguna.

HASIL KAJIAN

Pembangunan sistem ini membincangkan tentang mengimplementasikan model yang terbaik yang telah diperolehi melalui kaedah rangkaian neural buatan ke dalam sistem peramalan kualiti air. Pembangunan sistem ini dengan menggunakan bahasa pengaturcaraan Microsoft Visual Studio 2010. Rajah antaramuka sistem mempunyai 3 bahagian iaitu antaramuka pengenalan sistem, antaramuka peramalan kualiti air serta antaramuka info yang berkaitan dengan parameter-parameter kualiti air. Sistem peramalan kualiti air yang dibangunkan ini akan memaparkan hasil peramalan berdasarkan kepada data yang dimasukkan oleh pengguna. Rajah antaramuka sistem adalah seperti berikut.

Disini, pengguna akan menekan butang *ENTER* untuk ke antaramuka sistem peramalan kualiti air seperti pada Rajah 2. Rajah 3 menunjukkan contoh antaramuka bagi pengguna untuk memasukkkan nilai bagi setiap parameter-parameter kualiti air tersebut. Apabila pengguna menekan butang *RESULT*, hasil daripada pengiraan bagi model rangkaian neural buatan akan memperoleh hasil ramalan bagi tahap kualiti air tersebut seperti *clean*, *slightly polluted* atau *polluted*. Pengguna juga boleh menekan butang *RESET* untuk mengosongkan ruang input input dan output tersebut bagi memasukkkan nilai-nilai yang baru. Rajah 4 menunjukkan antaramuka info tentang parameter kualiti air bagi membantu menambah pemahaman pengguna.



Rajah 3 Antaramuka Utama Sistem Peramalan Kualiti Air

Rajah 4 Antaramuka Sistem Peramalan Kualiti Air

Temperature
Temperature affects aquatic organisms in a variety of ways. The body temperature of most aquatic organisms is the same as the surrounding water and fluctuates with the water temperature. Pollutants can become more toxic at higher temperatures.

pH
pH is the measure of hydrogen ions, or acidity, in the water. Water has hydrogen ions and hydroxyl ions. When there are equal numbers of both, the water is neutral. As the hydrogen ions increase, the water becomes more acidic; as the hydroxyl ions increase, the water becomes more basic.

Dissolved Oxygen
Most aquatic organisms need oxygen to survive. Dissolved oxygen is the oxygen present in water available to aquatic organisms. It is not the oxygen that is part of the water molecule but rather oxygen gas. Oxygen enters the water from the air through rain, turbulence and wind, and through the photosynthesis of aquatic plants. Decomposition of organic material is a major cause of low dissolved oxygen resulting in fewer species.

Turbidity
Turbidity is the measure of relative clarity of a liquid. It is an optical characteristic of water and is an expression of the amount of light that is scattered by material in the water when a light is shined through the water sample. The higher the intensity of scattered light, the higher the turbidity. Material that causes water to be turbid include clay, silt, finely divided inorganic and organic matter, algae, soluble colored organic compounds, and plankton and other microscopic organisms. Turbidity readings can be used as an indicator of potential pollution in a water body.

Conductivity
Conductivity is a measure of the ability of water to pass an electrical current. Conductivity in water is affected by the presence of inorganic dissolved solids such as chloride, nitrate, sulfate, and phosphate anions (ions that carry a negative charge) or sodium, magnesium, calcium, iron, and aluminum cations (ions that carry a positive charge). Organic compounds like oil, phenol, alcohol, and sugar do not conduct electrical current very well and therefore have a low conductivity when in water.

Rajah 5 Antaramuka Sistem Bagi Info

Proses bagi pengujian sistem sangat penting bagi menguji ketetapan sistem yang dihasilkan melalui hasil perbandingan hasil ramalan sistem daripada output dengan nilai yang sebenar. Data daripada pembangunan model rangkaian neural buatan akan diuji bagi membuktikan nilai pemberat yang dihasilkan bagi aras input, aras tersembunyi dan aras output adalah benar. Pengujian ini menggunakan nilai pemberat daripada model terbaik yang diperoleh iaitu model dengan peratus pembahagian bagi data latihan dan ujian ialah 90:10.

Data-data telah dimasukkan ke dalam sistem dan diuji satu persatu keatas sistem yang telah diimplementasikan model rangkaian neural buatan yang paling terbaik. Beberapa data telah digunakan dimana sebanyak 4 atribut input yang akan dimasukkan ke dalam sistem peramalan. Output yang paparkan oleh sistem akan dibandingkan dengan output yang sepatutnya diperoleh.

KESIMPULAN

Sistem peramalan kualiti air yang dibangunkan ini akan dibentuk sebelum dibenamkan model rangkaian neural buatan yang terbaik yang diperolehi hasil daripada pembangunan model bagi tujuan peramalan. Pemenaman model yang dilakukan dengan memasukkan nilai pemberat daripada model tersebut ke dalam sistem ini. Pengujian model ini dijalankan bagi menghasil sebuah model yang mencapai peratusan ketetapan yang paling tinggi. Bab ini juga menerangkan fungsi yang terdapat dalam sistem peramalan kualiti air di tasik chini yang dibangunkan. Hasil pengujian yang dikenalpasti dicatatkan pada laporan pengujian. Antara kualiti yang diambil kira ialah dari segi kelancaran operasi sistem, keberkesanan bagi mencapai keperluan pengguna.

Penggunaan Microsoft Visual Studio 2010 dalam menjalankan projrk ini dapat memudahkan kerja bagi menghasil antaramuka yang terbaik dan perlabagai tutorial dan latihan digunakan bagi meyempurnakan projek ini.

RUJUKAN

Sarkar, A. & Pandey, P. 2015. River Water Quality Modelling Using Artificial Neural Network Technique. *Aquatic Procedia*, 4(Icwrcoe), 1070–1077.
doi:10.1016/j.aqpro.2015.02.135