

# PERAMALAN INDEKS KUALITI AIR DI TASIK CHINI DENGAN MENGGUNAKAN KAEDAH POHON KEPUTUSAN

NUR FATIN HAZWANI SAIMI  
AZURALIZA ABU BAKAR

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia*

## ABSTRAK

Masalah pencemaran tahap kualiti air di Tasik Chini berada di satu tahap yang membimbangkan. Pencemaran yang berlaku menjadi semakin teruk dari hari ke hari. Projek yang dibangunkan ini adalah salah satu cadangan solusi yang digunakan untuk menyelesaikan masalah pencemaran kualiti air dengan memfokuskan pembangunan sebuah sistem peramalan bagi mengenal pasti tahap kualiti air di Tasik Chini. Ini dapat membantu pihak yang berkaitan memantau tahap kualiti air dari masa ke semasa. Sistem peramalan ini dibangunkan berasaskan petua melalui kaedah perlombongan data. Data-data yang dipilih digunakan untuk menghasilkan petua dan model ramalan. Hasil petua yang diperolehi daripada permodelan data akan dimuatkan ke dalam sistem untuk tujuan peramalan. Data-data Tasik Chini ini belum pernah digunakan dalam penyelidikan perlombongan data. Dalam kajian ini, sistem dibangunkan bagi membantu Pusat Penyelidikan Tasik Chini bagi meramal tahap kualiti air dan melaksanakan pelan perawatan. Hasil pengujian sistem menunjukkan bahawa sistem ini boleh diterima oleh responden yang menguji sistem ini dalam proses pengujian.

## PENGENALAN

Air merupakan salah satu sumber semulajadi yang sangat penting bagi kehidupan di bumi terutama kepada manusia, haiwan dan tumbuhan. Perkembangan negara dari segi ekonomi, perdagangan, industri dan perkhidmatan telah menyebabkan perkembangan bandar dan penduduknya telah meningkatkan pencemaran alam sekitar sehingga ke tahap kritikal terutama bagi indeks kualiti air. Sumber air di bumi ini adalah terdiri daripada sungai, tasik, pantai dan laut yang mana mempunyai kepentingan-kepentingan tertentu seperti kawasan tadahan hujan, tempat habitat haiwan dan sebagainya. Ekoran daripada kesedaran mengenai kepentingan kualiti air ini, pelbagai pihak cakna dengan mengambil tanggungjawab untuk menjaga dan mengawal kualiti air antaranya adalah seperti Jabatan Alam Sekitar iaitu sebuah badan NGO mahupun orang awam yang menjadi sukarelawan.

Kajian ini dijalankan di Tasik Chini iaitu sebuah tasik asli di Semenanjung Malaysia yang terletak di Negeri Pahang kira-kira 100 km dari Kuantan. Tasik ini merupakan tasik air tawar kedua terbesar di Malaysia dan terdiri daripada 12 siri tasik. Tasik ini juga adalah antara sumber utama untuk Sungai Pahang dan diempang untuk mengekalkan kedalaman tasik pada musim kemarau. Ia bertindak sebagai tanah lembap yang memainkan peranan penting dalam ekosistem semulajadi

Ramalan kualiti air sangat penting bagi memastikan sumber air yang digunakan adalah bersih dan perlu di pantau dari semasa ke semasa. Secara umumnya, Jabatan Alam Sekitar Malaysia menggunakan Indeks Kualiti Air sebagai standard untuk mengukur kebersihan dan kualiti bekalan air yang boleh digunakan untuk kegunaan domestik, akuakultur dan pengairan (JAS, 2001). Antara parameter yang akan digunakan dalam kajian ini adalah seperti suhu, kekonduksian, pH, kekeruhan, jumlah bubaran pepejal dan oksigen terlarut.

## PENYATAAN MASALAH

Setelah pemerhatian dilakukan, indeks kualiti air di Tasik Chini telah menurun dan semakin tercemar. Hal ini disebabkan oleh berlakunya perkembangan teknologi, penduduk dan industri di sekitar kawasan Tasik Chini yang telah mempengaruhi kualiti air tasik tersebut. Penambahan penduduk orang asli di Tasik Chini yang dikenali sebagai orang Jakun juga telah menyumbang sedikit sebanyak kepada pencemaran air di tasik tersebut. Setiap data yang diambil dan dikumpul oleh Pusat Penyelidikan Tasik Chini (TCPP) Universiti Kebangsaan Malaysia sebelum ini masih belum pernah dikaji oleh mana-mana pihak.

## OBJEKTIF KAJIAN

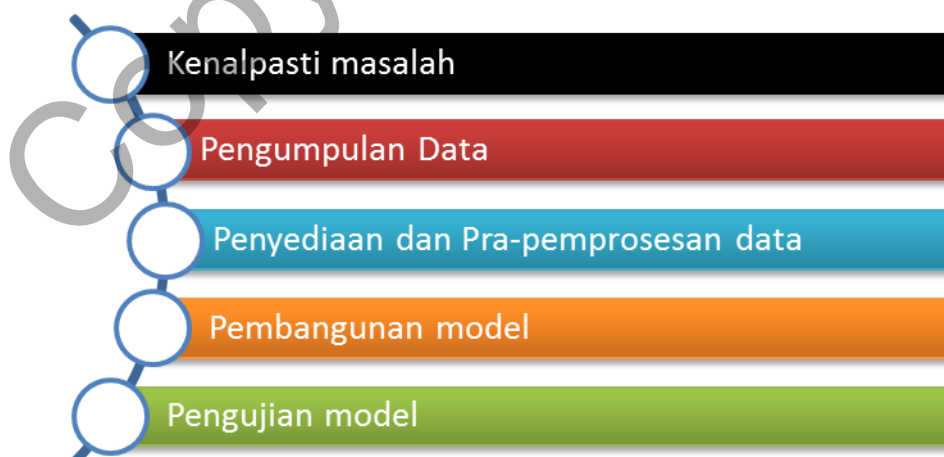
Secara keseluruhannya, objektif utama projek ini dijalankan adalah untuk membangunkan sebuah perisian yang akan membantu pihak penyelidik di Pusat Penyelidikan Tasik Chini (TCPP) Universiti Kebangsaan Malaysia untuk meramal dan mengawal indeks atau tahap kualiti air di Tasik Chini berdasarkan parameter *Dissolved Oxygen*(ODO).

Antara objektif lain bagi kajian ini adalah :

- i. Bagi membangunkan sebuah sistem cerdas yang menggunakan kaedah pohon keputusan bagi peramalan tahap kualiti air.
- ii. Mengetahui parameter-parameter yang penting yang memberi kesan terhadap kualiti air.

## METOD KAJIAN

Metodologi kajian memainkan peranan yang sangat penting dalam memastikan kajian yang dilakukan berjalan dengan lancar dan teratur serta memenuhi setiap spesifikasi yang diperlukan. Model ramalan sistem cerdas ini dibangunkan melalui dua fasa yang penting iaitu fasa perlombongan data dan fasa pembangunan sistem. Rajah 1 merujuk kepada fasa yang pertama iaitu fasa perlombongan data yang merangkumi fasa kenal pasti masalah, fasa pengumpulan data, fasa penyediaan dan pra pemprosesan data, fasa pembangunan model serta fasa pengujian model manakala Rajah 2 pula merupakan fasa pembangunan sistem yang terdiri daripada fasa pembangunan sistem dan fasa pengujian.



Rajah 1 Metodologi Pembangunan Model

## FASA PEMBANGUNAN MODEL

### **Fasa Mengenalpasti Masalah**

Antara tunjang utama bagi setiap kajian yang dilakukan adalah dengan mengenalpasti masalah yang ingin diselesaikan. Masalah yang ingin dikaji dalam kajian ini adalah kaedah peramalan kualiti air yang dapat diaplikasikan di Tasik Chini dan algoritma yang sesuai untuk digunakan dalam pembangunan model. Fasa ini juga menentukan objektif dan pernyataan masalah dalam kajian ini.

### **Fasa Pengumpulan Data**

Fasa ini adalah merupakan fasa dimana semua data dikumpul untuk digunakan dalam membangunkan model ramalan. Data untuk kajian ini diperoleh daripada Pusat Penyelidikan Tasik Chini (TCPP) Universiti Kebangsaan Malaysia. Data kajian adalah data yang bercirikan data siri masa yang dikumpulkan sepanjang dua (2) tahun iaitu dari tahun 2011 sehingga tahun 2012.

### **Fasa Penyediaan Dan Pra Pemrosesan Data**

Dalam fasa ini, data asal yang diperoleh dari fasa pengumpulan data akan melalui beberapa proses sebelum ia digunakan untuk membangunkan model ramalan. Antara proses yang dilakukan adalah:

#### **i. Pembersihan Data**

Proses ini kebiasaannya adalah untuk mengenal pasti data yang tidak lengkap dan memenuhi ruang tersebut dengan data purata. Selain itu, data-data yang rosak akan dibuang seterusnya memperbaiki data yang tidak konsisten. Semua attribut akan dinilai dan dipilih mengikut kesesuaian kajian.

#### **ii. Pendiskritan Data**

Proses ini akan memudahkan proses permodelan dimana data yang ada mudah untuk difahami dan digunakan untuk mengkaji corak data bagi pembangunan model ramalan.

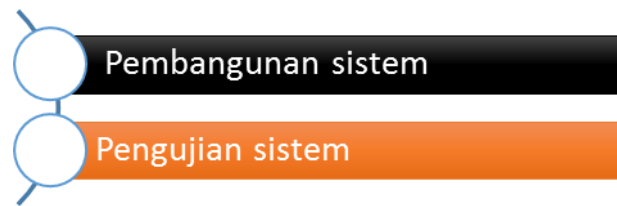
### **Fasa Pembangunan Model**

Dalam proses ini, model ramalan berasaskan petua dibangunkan. Data-data yang terhasil dari fasa sebelum ini akan digunakan untuk menghasilkan petua. Kaedah perlombongan petua pohon keputusan adalah salah satu kaedah perlombongan data yang digunakan untuk menghasilkan model ramalan.

### **Fasa Pengujian Model**

Menerusi proses ini, model yang terhasil daripada fasa pembangunan model akan diuji terlebih dahulu sebelum digunakan. Hal ini penting adalah kerana setiap model yang dipilih akan dijadikan asas dalam pembangunan sebuah sistem yang akan membentuk peraturan. Model yang dipilih seharusnya adalah model yang dapat menghasilkan keputusan yang baik dengan jumlah ketepatan yang tinggi.

## FASA PEMBANGUNAN SISTEM



Rajah 2 Metodologi Pembangunan Sistem

### Fasa Pembangunan Sistem

Dalam fasa pembangunan sistem ini, lakaran cadangan sistem akan dilakukan yang akan merangkumi senibina sistem dan perwakilan antara muka yang akan digunakan antara pihak pengguna dan sistem. Reka bentuk model akan dihasilkan secara terperinci termasuk dari segi logikal dan fizikal sistem tersebut. Seterusnya, reka bentuk tersebut akan dibangunkan dengan mengkodkan aturcara untuk sistem ini.

### Fasa Pengujian Sistem

Pada fasa ini, aturcara yang dibina akan diuji bagi memastikan tiada kesalahan output dan memenuhi justifikasi yang ditetapkan. Selain itu, fasa ini juga akan menguji komunikasi dua hala antara sistem dan pengguna. Sistem ini akan diselenggara dari semasa ke semasa bagi memastikan pihak pengguna mencapai tahap kepuasan terhadap sistem ini dan selesa menggunakannya. Jika objektif kajian gagal untuk dicapai, sebuah imbasan semula akan dilakukan bermula dari fasa pertama bagi membuat penambahbaikan kajian yang mendalam.

Bagi menghasilkan sebuah sistem yang efektif, perisian yang digunakan dalam membangunkan sistem tersebut hendaklah dititikberatkan. Pemilihan perisian yang sesuai akan memudahkan lagi proses pembangunan serta melancarkan sistem. Antara perisian yang digunakan dalam membangunkan Sistem Peramalan Tasik Chini adalah seperti berikut:

- i. Perisian Weka versi 3.9.1
- ii. *Microsoft Excel 2016*
- iii. *Microsoft Visual Basic Studio 2010*

## 5 HASIL KAJIAN

Bahagian ini membincangkan hasil daripada proses pembangunan Sistem Peramalan Tasik Chini. Selain itu, ia juga memberikan penerangan yang mendalam mengenai reka bentuk dan model sistem peramalan. Seperti yang telah dijelaskan dalam bahagian di atas, pembangunan sistem ini terbahagi kepada dua (2) fasa iaitu fasa penghasilan model dan fasa pembangunan sistem. Perisian Weka telah digunakan dalam menghasilkan sebuah model melalui kaedah perlombongan data yang terpilih iaitu model pohon keputusan. Model pohon keputusan digunakan untuk memperolehi petua-petua yang berguna berdasarkan corak yang terdapat dalam data kajian. Model akan diuji berdasarkan beberapa kriteria iaitu pembahagian mod ujian atau *percentage split* dan *cross validation* sesebuah data. Sebelum data dimasukkan ke dalam perisian Weka, data yang akan dimasukkan akan melalui proses pra-pemprosesan data yang mana akan dilakukan menerusi perisian *Microsoft Excel 2016*. Proses pra-pemprosesan data ini merangkumi proses pembersihan data, proses pemilihan parameter serta jumlah data

yang akan digunakan dan seterusnya di diskritisakan. Fail data yang di proses oleh *Microsoft Excel 2016* akan disimpan di dalam format yang ditetapkan iaitu format *Comma delimited* atau lebih dikenali sebagai CSV. Rajah 3 menunjukkan data Tasik Chini yang telah dibersihkan manakala Rajah 4 menunjukkan data yang telah didiskritisakan.

Date	Temperature(°C)	Cond(uS/cm)	TDS(mg/l)	pH(unit)	Turbidity(NTU)	ODO(mg/l)
1/2/2011	26.07	29.7	0.019	5.78	51.7	1.72
2/2/2011	24.49	34.2	0.019	6.52	177.7	6.48
3/2/2011	24.7	37.3	0.019	6.42	151.4	6.03
4/2/2011	24.59	39.3	0.019	6.25	115.7	5.02
5/2/2011	24.76	37.7	0.019	5.97	95.2	2.82
6/2/2011	25.21	37.3	0.019	5.93	85.9	1.24
7/2/2011	25.61	32	0.019	5.78	61.9	0.72
8/2/2011	25.93	29.7	0.019	5.71	48.8	1.12
9/2/2011	26.81	29	0.019	5.72	48.2	1.29
10/2/2011	27.46	28.7	0.019	5.72	42.8	2.21
11/2/2011	27.57	28.3	0.019	5.72	199.3	2.46
12/2/2011	27.41	27.7	0.019	5.7	26.5	2.19
13/2/2011	27.21	27.7	0.019	5.63	24.9	1.94
14/2/2011	27.21	28.3	0.019	5.62	24.2	1.37
15/2/2011	27	27.7	0.019	5.62	21.8	1.13
16/2/2011	27.24	28	0.019	5.63	20.2	0.74
17/2/2011	27.31	27	0.019	5.63	20.3	0.69
18/2/2011	27.06	27.3	0.019	5.63	18	0.6
19/2/2011	27.3	28.7	0.019	5.69	21.8	0.5
20/2/2011	27.26	28.3	0.019	5.69	18	0.53
21/2/2011	27.36	29.3	0.019	5.69	38.5	0.43

Rajah 3 Data yang telah dibersihkan

Data yang telah didiskritisakan ini dimasukkan ke dalam perisian Weka untuk proses seterusnya yang mana adalah untuk mengklasifikasikan data yang terdapat dalam kajian melalui format CSV. Antara langkah yang penting dalam mengklasifikasikan data adalah sesi pemilihan mod ujian. Dalam kajian ini, mod ujian yang digunakan adalah sebanyak 80:20. Petua yang terhasil dalam proses klasifikasi ini akan digunakan dalam membangunkan sistem peramalan kualiti air. Petua-petua ini akan menjadi asas kepada keputusan yang akan diambil dalam menentukan tahap kualiti air. Berikut merupakan petua-petua yang terhasil setelah klasifikasi berlaku.

Date	Temperature(°C)	Cond(uS/cm)	TDS(mg/l)	pH(unit)	Turbidity(NTU)	ODO(mg/l)
1/2/2011	low	low	low	acidic	low	low
2/2/2011	low	low	low	acidic	low	high
3/2/2011	low	low	low	acidic	low	high
4/2/2011	low	low	low	acidic	low	high
5/2/2011	low	low	low	acidic	low	medium
6/2/2011	low	low	low	acidic	low	low
7/2/2011	low	low	low	acidic	low	low
8/2/2011	low	low	low	acidic	low	low
9/2/2011	low	low	low	acidic	low	low
10/2/2011	medium	low	low	acidic	low	low
11/2/2011	medium	low	low	acidic	low	low
12/2/2011	medium	low	low	acidic	low	low
13/2/2011	medium	low	low	acidic	low	low
14/2/2011	medium	low	low	acidic	low	low
15/2/2011	low	low	low	acidic	low	low
16/2/2011	medium	low	low	acidic	low	low
17/2/2011	medium	low	low	acidic	low	low
18/2/2011	medium	low	low	acidic	low	low
19/2/2011	medium	low	low	acidic	low	low
20/2/2011	medium	low	low	acidic	low	low
21/2/2011	medium	low	low	acidic	low	low

Rajah 4: Data Diskrit

### J48 pruned tree

-----

Temperature(°C) = low

| Turbidity(NTU) = low

| | pH(unit) = acidic: low (22.0/8.0)

| | pH(unit) = strong acidic

| | | TDS(mg/l) = low: medium (39.0/23.0)

| | | TDS(mg/l) = medium: low (7.0/3.0)

| | | TDS(mg/l) = high: high (6.0/1.0)

| | pH(unit) = neutral: low (0.0)

| | pH(unit) = alkaline: low (0.0)

| | pH(unit) = strong alkaline: high (3.0)

| Turbidity(NTU) = medium: low (33.0/1.0)

| Turbidity(NTU) = high: low (0.0)

Temperature(°C) = medium

| pH(unit) = acidic: low (79.0/8.0)

| pH(unit) = strong acidic

| | TDS(mg/l) = low

| | | Turbidity(NTU) = low: medium (50.0/11.0)

| | | Turbidity(NTU) = medium: low (2.0)

| | | Turbidity(NTU) = high: medium (0.0)

| | TDS(mg/l) = medium: low (73.0/22.0)

| | TDS(mg/l) = high: medium (10.0/3.0)

| pH(unit) = neutral: medium (1.0)

| pH(unit) = alkaline: low (8.0/3.0)

| pH(unit) = strong alkaline: medium (36.0/17.0)

Temperature(°C) = high: medium (125.0/24.0)

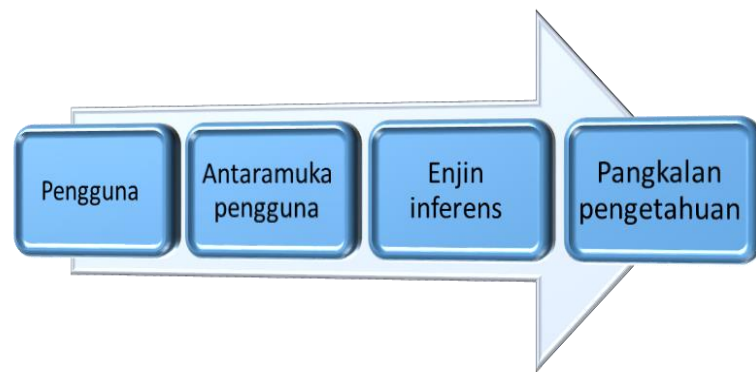
Petua-petua diatas akan diaplikasikan dalam pembangunan sistem peramalan agar sistem yang dibangunkan mampu meramal tahap kualiti air Tasik Chini dengan mudah dan tepat.

Sistem adalah penting dalam menjadi perwakilan antara kita dan pihak pengguna yang akan menggunakan sistem yang telah dibangunkan. Ia membantu untuk berinteraksi dengan pengguna dengan menggunakan bahasa aras tinggi yang lebih mudah difahami dan dekat dengan pengguna. Metodologi bagi pembangunan sistem ini terdiri daripada dua (2) peringkat iaitu fasa pembangunan sistem dan fasa pengujian sistem.

Setelah pembangunan model dibangunkan, model yang terhasil akan digunakan pula untuk pembangunan sistem. Dalam kajian ini, model tersebut akan menjadi asas bagi membina sistem peramalan indeks kualiti air di Tasik Chini. Setiap petua yang dipilih dalam membentuk model akan di kodkan semula kepada bentuk *Visual Basic.NET* dengan menggunakan sebuah perisian yang popular iaitu *Microsoft Visual Studio 2012*. *Microsoft Visual Studio 2012* adalah sebuah perisian yang akan menjadi platform dalam pembangunan sistem bagi kajian ini.

Bagi membangunkan sebuah sistem, terdapat empat (4) peranan utama yang perlu dipenuhi iaitu peranan pengguna, antaramuka pengguna, enjin inferens dan juga pangkalan

pengetahuan. Rajah 5 menunjukkan cara sistem ini berfungsi.

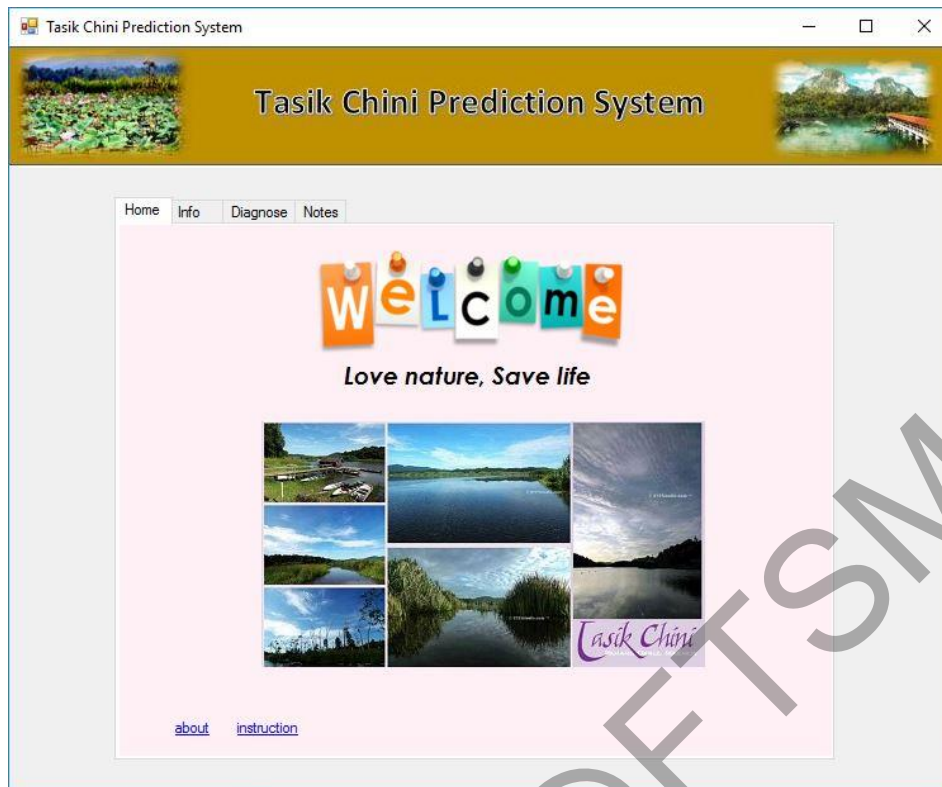


Rajah 5 Rekabentuk sistem peramalan

Pihak pengguna akan memberi input melalui interaksi dengan antaramuka pengguna. Input seterusnya akan disalurkan kepada enjin inferens. Enjin inferens berperanan untuk melakukan analisis terhadap input tersebut dan juga sebagai pemproses maklumat yang akan menghubungkan pengguna dan pangkalan pengetahuan bagi memberikan output kepada pengguna.

Seperti yang telah dinyatakan, pembangunan antaramuka sistem ini akan dilakukan dengan menggunakan *Microsoft Visual Studio 2012*. Fungsi *drag and drop* yang terdapat dalam perisian ini dapat memudahkan proses pembangunan sistem ini sekaligus dapat menjimatkan masa selain daripada kos pembinaan. Setelah itu, set petua atau maklumat akan diimplikasikan dalam sistem. Maklumat-maklumat yang terdapat dalam sistem ini akan digunakan bagi membuat kesimpulan kepada input yang dimasukkan oleh pengguna dan seterusnya memberi keputusan. Natiujahnya, jika sistem ini berjaya untuk memberi keputusan yang tepat, maka sistem ini mampu untuk membuat peramalan.

Setelah sistem dibangunkan, sistem ini akan diuji beberapa kali sebelum ia digunapakai kepada pengguna. Ujian yang paling penting yang perlu dilakukan adalah dengan menguji kebolehan sistem untuk berinteraksi dengan pengguna. Sistem hendaklah boleh menerima sebarang input daripada pengguna sebelum input tersebut dapat diproses. Selain itu, sistem ini juga akan diuji sama ada output yang dikeluarkan adalah tepat ataupun tidak. Sekiranya tiada kesalahan output yang diberikan, sistem ini sudah boleh digunakan secara keseluruhan. Sesetengah sistem akan menitikberatkan juga tahap mesra sistem terhadap pengguna. Hal ini disebabkan oleh, sistem yang baik bukan sahaja mampu memberikan output yang tepat namun adalah mesra pengguna. Rajah 6, Rajah 7, Rajah 8, Rajah 9 dan Rajah 10 merupakan antaramuka sistem peramalan Tasik Chini.



Rajah 6 Antaramuka depan sistem peramalan

Rajah 7 Antaramuka tab *Info* sistem peramalan



Tasik Chini Prediction System

Home Info Diagnose Notes

Temperature : Please Select

Conductivity : Please Select

TDS : Please Select

pH : Please Select

Turbidity : Please Select

Result :

Generate Reset

Rajah 8 Antaramuka tab *Generate* sistem peramalan

Tasik Chini Prediction System

Home Info Diagnose Notes

Temperature

The water temperature of a river is very important, as many of the physical, biological, and chemical characteristics of a river are directly affected by temperature.

Conductivity

Conductivity is a measurement of the ability of an aqueous of an element that has gained or lost an electron which will create a negative or positive state.

Total Dissolved Solids

This is a measure of the solid materials dissolved in the river water. Concentrations of total dissolved solids that are too high or too low may limit growth and lead to the death of many aquatic life forms.

pH

The pH level is a measure of the acid content of the water. Most forms of aquatic life tend to be very sensitive to pH. Water containing a great deal of organic pollution will normally tend to be somewhat acidic.

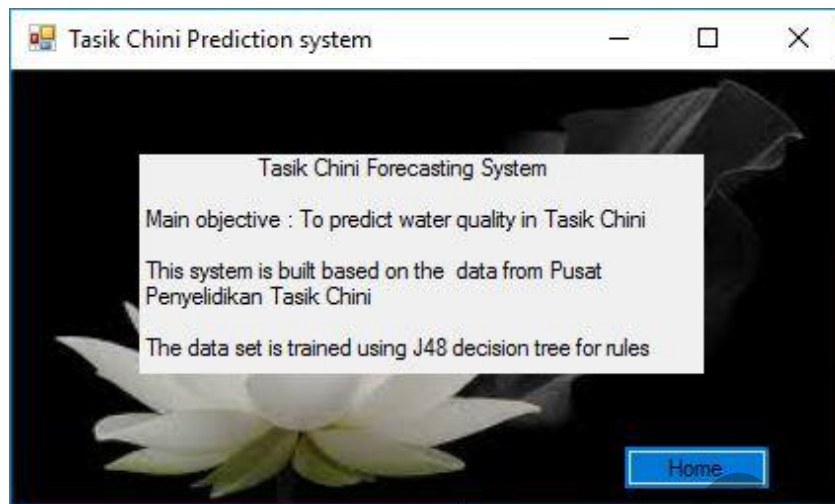
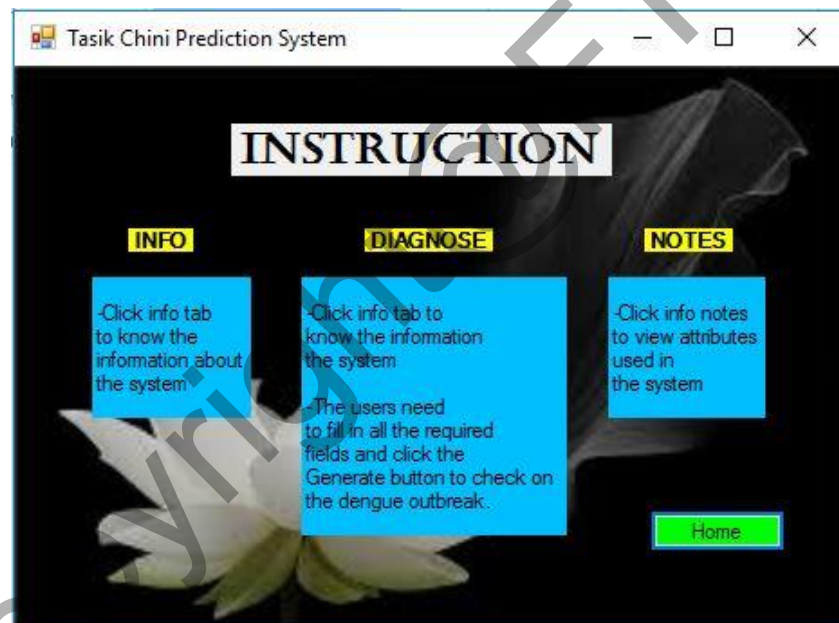
Turbidity

Turbidity is a measure of the dispersion of light in a column of water due to suspended matter. The higher the turbidity, the cloudier the water appears. If water becomes too turbid, it loses the ability to support a wide variety of plants and other aquatic organisms.

Dissolved Oxygen

The dissolved oxygen test measures the amount of life-sustaining oxygen dissolved in the water. Low levels of dissolved oxygen in water are a sign of possible pollution.

Rajah 9 Antaramuka tab *Notes* sistem peramalan

Rajah 10 Antaramuka link label *About* sistem peramalanRajah 11 Antaramuka link label *Instruction* sistem peramalan

## KESIMPULAN

Sistem peramalan kualiti air Tasik Chini ini dijangka dapat membantu pihak-pihak yang berkaitan bagi meramal tahap kualiti air di Tasik Chini. Hal ini amat penting bagi rakyat Malaysia dalam usaha menjaga dan memantau tahap kualiti air tasik semulajadi itu di samping dapat memulihara kawasan tersebut dari pencemaran apatah lagi sejak kebelakangan ini, tasik tersebut sering dikaitkan dengan air tercemar. Penjagaan tasik ini bukan sahaja untuk keperluan rakyat Malaysia sebagai tempat perlindungan malah ia juga menyumbang kepada sektor pelancongan. Penggunaan teknik perlombongan data memudahkan data-data tersebut diproses dan mendapatkan pengetahuan berguna. Teknik ini merupakan sebuah teknik yang utama dan digunakan secara meluas dalam era sekarang.

## RUJUKAN

- Asas Perlombongan Data. (n.d.). Retrieved from [http://hariz98.tripod.com/lan/Asas\\_Perlombongan\\_Data.htm](http://hariz98.tripod.com/lan/Asas_Perlombongan_Data.htm)
- Han, J. Kamber, M. 2006. Data Mining Concept and Techniques. San Francisco: Morgan Kaufmann Publisher
- Jaloree, S., Rajput, A., & Gour, S. (2014). Decision tree approach to build a model for water quality, 4, 25–28.
- Othman, M. S., & Lim, E. C. (2006). Keadaan eutrofikasi di Tasik Chini, Pahang. Sains Malaysiana, 35(2), 29–34.

Copyright@FTSM