

Meramal Kualiti Air Sungai Menggunakan Kaedah Analisis
Diskriminan Dua Kumpulan Berdasarkan Komposisi Alga

Sharifah Nasriah Binti Wan Obeng

Universiti Sains Malaysia

2006

Abstrak

Alga merupakan kumpulan terbesar organisma akuatik Ia mempunyai klorofil dan dapat menghasilkan makanannya sendiri melalui proses fotosintesis. Ia boleh dijumpai di serata dunia seperti di dalam laut, sungai dan di tempat yang lembap di darat. Kewujudan komposisi alga ini mempengaruhi kualiti air sungai di Malaysia. Beberapa pendekatan tradisional dalam menentukan kualiti air sungai telah dilakukan sejak dulu tetapi ia mengambil masa yang lama untuk mengklasifikasikan air tersebut sama ada bersih atau tercemar. Oleh itu, satu kaedah pintar dilakukan bagi meramal kualiti air sungai dengan tepat dan mudah berdasarkan komposisi alga. Di dalam projek ini, kaedah analisis diskriminan dua kumpulan digunakan untuk meramal kualiti air sungai. Analisis diskriminan ini boleh mengklasifikasikan kualiti air sungai tersebut sebanyak 94.9%. Satu fungsi diskriminan diperolehi dan ketepatan keberkesanannya dalam mengklasifikasikan air telah dibuktikan di akhir projek dan ia boleh digunakan untuk tujuan peramalan di masa akan datang. Secara kesimpulannya, analisis diskriminan dua kumpulan ini merupakan satu kaedah pintar yang boleh meramal kualiti air sungai dengan tepat kepada bersih atau tercemar berdasarkan komposisi alga.

Abstract

Algae are large and diverse group of aquatic plantlike organisms. The algae contain chlorophyll and can manufacture their own food through the process of photosynthesis. They are distributed worldwide throughout the sea, freshwaters, rivers and wetland. The existences of the alga composition have a greater influence to the river's water in Malaysia. There are several traditional approaches used a long time ago in predicting the river's water quality but it is time consuming to classify whether the water is clean or polluted. So, we need a smarter way to predict the quality of river's water accurately and easily based on the alga composition. In this project, the two group discriminant analysis is used to predict the river's water quality. The results show that the proposed system classifies the river's water with 94.9% of accuracy. The results obtained prove that the discriminant analysis is suitable and has high capability to be used as intelligent classifier to classify the river's water quality based on algae composition. As a conclusion, the two group discriminant analysis is one of the smarter ways to predict the quality of the river's water whether it is clean or polluted accurately based on the alga composition.

PENGHARGAAN

Bismillahirrahmanirahim.

Segala puji-pujian bagi Allah, Tuhan seru sekalian salam. Selawat dan salam buat Nabi Muhammad SAW, ahli keluarganya dan para sahabatnya. Alhamdulillah, bersyukur ke hadrat Illahi kerana dengan limpah kurniaNya, projek ini telah dilaksanakan dengan jayanya.

Pertama kali adalah penghargaan khas saya tujukan buat insan yang sentiasa saya sanjungi dan sayangi iaitu kedua ibu bapa saya, Puan Sharifah Aminah Wan Zainal Abidin dan Encik Wan Obeng Wan Hashim di atas segala doa, pengorbanan dan ikhtiar kalian dikala saya lemah dan memerlukan sokongan. Terima kasih juga kepada nenda tercinta, Puan Sharifah Fatimah Wan Sadi dan keluarga yang juga sentiasa memberi kata-kata semangat dan sokongan. Terima kasih semua.

Ucapan terima kasih juga saya tujukan buat penyelia saya, Dr. Nor Ashidi Mat Isa dan Puan Norizah Mohamad yang telah banyak membantu saya dari awal sehinggalah tamat projek ini. Berkat bimbingan, dorongan, pandangan, teguran dan nasihat daripada mereka, segala kekeliruan dan masalah yang dihadapi telah dapat ditangani dengan baik. Sekali lagi ucapan terima kasih saya berikan buat penyelia-penyelia saya kerana sudi menyediakan pengumpulan data yang telah diambil dari Jabatan Sains Biologi, Universiti Sains Malaysia.

Selain itu, terima kasih juga saya tujukan kepada Puan Dzati Athiar Ramli, Encik Aeizal Azman Abdul Wahab dan Prof Madya Mohd Shukri Yahya kerana telah banyak membantu saya di dalam permasalahan matematik yang pada awalnya agak rumit buat saya. Pengajaran dan pembelajaran yang diberikan telah banyak membantu projek ini.

Penghargaan khas diberikan kepada En. Fakhroul Ridhwan dan rakan-rakan lain yang bersama-sama membantu dan mengharungi projek ini. Terima kasih kepada USM dan semua yang menjayakan projek ini. Semoga projek ini akan dirujuk dan boleh dimanfaatkan oleh semua.

-Sharifah Nasriah Binti Wan Obeng-

KANDUNGAN

Muka Surat

ABSTRAK	ii
ABSTRACT	iii
PENGHARGAAN	iv
JADUAL ISI KANDUNGAN	v
SENARAI RAJAH	vi
SENARAI JADUAL	vii

BAB 1 PENGENALAN

1.1	Pendahuluan.....	1
1.2	Objektif Projek.....	2
1.3	Panduan Laporan Projek.....	3

BAB 2 KAJIAN ILMIAH

2.1	Pengenalan.....	5
2.2	Alga Sungai.....	5
2.3	Kualiti Air Sungai di Malaysia.....	6
2.4	Pendekatan Tradisional dalam Penentuan Air Sungai.....	7
2.5	Analisis Diskriminan Dua Kumpulan.....	10
2.5.1	Contoh Kehidupan Sebenar.....	11
2.5.2	Asas kepada Analisis Diskriminan.....	13
2.5.3	Ringkasan.....	16

BAB 3 METODOLOGI

3.1	Pengenalan.....	17
3.2	Analisis Diskriminan Dua Kumpulan.....	17

3.2.1	Persampelan Data.....	17
3.2.2	Pembahagian Kualiti Air Sungai.....	20
3.2.3	Statistik Penting yang Berkaitan dengan Analisis Diskriminan.....	20
3.2.4	Mengkaji Tahap Keertian Pembolehubah Diskriminan.....	23
3.2.5	Kaedah Titik Pemisahan	23
3.2.6	Kesimpulan.....	24

BAB 4 KEPUTUSAN

4.1	Pengenalan.....	25
4.2	Ujian Univariat.....	25
4.3	Ujian Multivariat.....	27
4.4	Fungsi Diskriminan.....	28
4.5	Ujian Matriks Kovarians.....	30
4.6	Penentuan Titik Pemisahan.....	31
4.7	Penggunaan Analisis Diskriminan.....	31
4.8	Ulasan.....	36

BAB 5 KESIMPULAN.....	37
------------------------------	-----------

RUJUKAN.....	39
---------------------	-----------

LAMPIRAN

SENARAI RAJAH

2.1 Taburan normal yang baik dan tidak baik bagi analisa diskriminan.....	14
2.2 Serakan data.....	15
2.3 Jumlah penindanan bagi paksi x dan y.....	15
2.4 Jumlah penindanan bagi paksi baru.....	16

SENARAI JADUAL

4.1 Keputusan ujian univariat untuk pembolehubah diskriminan.....	26
4.2 Keputusan ujian multivariat untuk pembolehubah diskriminan.....	28
4.3 Pekali pembolehubah diskriminan.....	28
4.4 Nilai eigen dan kolerasi kanonik.....	29
4.5 Keputusan pengkelasan.....	30
4.6 Keputusan pengujian menggunakan model tersebut.....	35

BAB 1

PENGENALAN

1.1 Pendahuluan

Secara umumnya, alga merupakan kumpulan terbesar organisma akuatik. Sebanyak 34 spesis alga didapati wujud di dalam sungai-sungai di Malaysia. Komposisi alga ini sangat penting di dalam menentukan kualiti air sungai ini. Di dalam projek ini, hanya 21 spesis alga sahaja yang akan digunakan. Bilangan alga ini telah dikumpul di beberapa stesen (sungai) dan telah diambil dari tahun 1998 hingga ke tahun 2000 [Pusat Sains Kajahayat, USM, 2005]. Spesis-spesis alga ini boleh wujud sama ada di dalam air yang bersih atau tercemar atau kedua-duanya sekali. Sesetengah spesis alga hanya boleh wujud hanya di dalam air bersih atau tercemar sahaja. Tetapi sesetengahnya boleh wujud di dalam kedua-duanya.

Beberapa parameter telah digunakan sejak dulu lagi bagi menentukan kualiti air sungai ini seperti oksigen terlarut (Dissolved Oxygen, DO), permintaan oksigen biokimia (Biological Oxygen Demand, BOD), Permintaan oksigen kimia (Chemical Oxygen Demand, COD), dan pepejal terampai (Suspended Solids, SS). Di dalam projek ini, analisis diskriminan dua kumpulan digunakan untuk meramal atau mengklasifikasikan sama ada air sungai tersebut bersih atau tercemar dengan mengenalpasti komposisi alga (pembolehubah-pembolehubah) yang terbaik yang memberikan diskriminan terbaik di antara dua kumpulan iaitu kepada kumpulan air bersih atau tercemar.

1.2 Objektif Projek

Objektif projek ini dijalankan adalah untuk meramal kualiti air sungai berdasarkan komposisi alga yang wujud di dalam air sungai. Dengan menggunakan kaedah analisis diskriminan dua kumpulan (two-group diskriminant analysis), kualiti air sungai dapat diklasifikasikan dan dikesan sama ada dalam keadaan bersih atau tercemar. Perisian SPSS iaitu (Statistical Package for the Social Sciences) digunakan untuk menjalankan kaedah analisis ini.

Projek ini bertujuan untuk meramal kualiti air sungai sama ada bersih atau tercemar berdasarkan komposisi alga yang telah wujud di dalamnya. Analisis diskriminan adalah salah satu teknik yang terbaik untuk menjayakan ramalan tersebut dengan mencapai beberapa objektif di bawah:

1. Menentukan pembolehubah (komposisi/ spesis) yang terbaik untuk didiskriminasi di antara dua kumpulan.
2. Menggunakan pembolehubah/ faktor yang telah dikenalpasti untuk menerbitkan persamaan atau fungsi untuk pengkomputeran satu pembolehubah atau indeks baru yang mana akan mewakili perbezaan di antara dua kumpulan.
3. Menggunakan pembolehubah yang telah dikenalpasti atau indeks yang telah dikomputerkan untuk menerbitkan satu peraturan untuk mengklasifikasikan pemerhatian masa depan ke dalam salah satu dari dua kumpulan tersebut.

Seterusnya, kewujudan projek ini adalah salah satu cara alternatif untuk mengklasifikasikan kualiti air sungai dengan tepat untuk kegunaan jangka masa yang panjang dan memudahkan pihak tertentu dalam menguji kualiti sesuatu air di masa akan datang.

1.3 Panduan Laporan Projek

Projek ini mengandungi 5 bab utama yang mana akan menjelaskan secara terperinci keseluruhan perjalanan projek ini.

Bab 1 merupakan pengenalan dan menjelaskan tentang latar belakang projek ini. Selain itu, bab ini akan menerangkan serba sedikit tentang apa yang diperlukan di dalam projek ini, objektif dan apakah pendekatan yang berkaitan dengan projek ini secara umum.

Bab 2 pula menerangkan dengan jelas berkaitan dengan kajian ilmiah yang diperolehi berdasarkan maklumat-maklumat yang telah dikumpul dari buku-buku. Kajian ilmiah ini akan menerangkan apakah itu alga, bagaimanakah komposisi alga ini dapat mempengaruhi kualiti air sungai dan sebagainya. Seterusnya, penerangan dengan jelas berkaitan dengan analisis diskriminan dua kumpulan dan kegunaanya di dalam meramal kualiti air sungai ini.

Seterusnya bab 3 menunjukkan metodologi dan teknik yang digunakan di dalam projek ini. Bab ini memfokus kepada kaedah analisis diskriminan dua kumpulan. Di dalam analisis diskriminan, bab ini menunjukkan bagaimana data disampelkan,

membentuk fungsi diskriminan dan seterusnya menentukan titik pemisah untuk analisis ini.

Bab 4 menerangkan tentang keputusan projek ini berdasarkan metodologi yang telah dinyatakan di dalam bab 3 dan di akhir bab ini, keberkesanan kaedah analisis diskriminan ini dalam meramal air sungai akan dibuktikan.

Akhir sekali, bab 5 akan menyimpulkan keseluruhan projek ini dan beberapa cadangan akan dikemukakan bagi menambahbaikkan lagi projek ini.

BAB 2

KAJIAN ILMIAH

2.1 Pengenalan

Bab ini akan menerangkan dengan jelas berkaitan dengan kajian ilmiah untuk empat sub topik utama iaitu alga, kualiti air sungai, pendekatan tradisional dalam penentuan kualiti air sungai dan analisis diskriminan dua kumpulan. Bab ini dimulakan dengan definisi alga dan spesis-spesis yang digunakan di dalam projek ini. Seterusnya, penjelasan tentang kualiti air sungai di Malaysia akan diterangkan pada sub topik ini. Sub topik 2.3 pula, membincangkan tentang pendekatan tradisional dalam penentuan kualiti air sungai. Sub topik yang terakhir akan mendefinisikan serta menerangkan dengan jelas berkaitan dengan analisis diskriminan dua kumpulan.

2.2 Alga Sungai

Alga merupakan kumpulan terbesar organisma akuatik dan kebanyakannya adalah multisel dan sebahagiannya pula sel tunggal. Alga mempunyai klorofil dan ia dapat menghasilkan makanannya sendiri melalui proses fotosintesis. Ia juga bertindak sebagai pengeluar bahan organik dalam persekitaran akuatik. Sebatian tak organik seperti karbon dioksida, ammonia, nitrat dan fosfat merupakan punca makanan untuk mensintesis sel alga baru bagi menghasilkan oksigen.

Alga ini boleh dijumpai di serata dunia seperti di dalam laut, sungai dan di tempat yang lembap di darat. Alga yang terdapat di dalam air yang tercemar atau adakalanya bertoksid boleh membahagi dan seterusnya membiak dan menyebabkan ekosistemnya tidak seimbang secara seriusnya. Projek ini memfokuskan alga yang terdapat di dalam sungai-sungai di Malaysia sahaja. Berdasarkan data yang telah dikumpul dari sungai-sungai (stesen–stesen), di dapati bahawa terdapat 21 spesies alga wujud di dalamnya. Diantaranya adalah *Acnantes minutissima*, , *Amphora libica*, *Anomoeoneis brachysira*, *Cocconeis pediculus*, *C. placentula*, *C. thumensis*, *Cymbella turgida*, *Diatoma sp.*, *Eunotia grunowi*, *Fragilaria sp. 1*, *Frustulia rhomboids*, *gomphonema acuminatum*, *G. gracile*, *Navicula sp.*, *N. cryptocephala*, , *Pinnularia sp.1*, *P. braunii var Psammothidium bioreii*, *Surirella sp.1*, *S. linearis* and *S. tenuissima*. Populasi alga ini bergantung kepada kualiti air sungai tersebut iaitu bersih atau tercemar. Dalam pada itu, alga ini membiak bergantung kepada persekitaran dan musim. Selain itu, pembiakan alga ini bergantung dengan kehadiran cahaya matahari, nitrat, fosforus dan juga suhu.

2.3 Kualiti Air Sungai di Malaysia

Kualiti air sungai di Malaysia boleh ditentukan dengan merujuk kepada Indeks Kualiti Air Negara dan ianya dibahagikan kepada 5 kelas iaitu sangat baik, baik, sederhana, tercemar dan sangat tercemar.

Sungai yang diklasifikasikan sebagai tercemar dan sangat tercemar kebanyakannya terletak di bahagian hilir sungai. Kawasan ini merupakan kawasan yang

pesat membangun dengan perindustrian, perumahan dan pertanian. Sungai ini menjadi tercemar kerana ia menjadi pengumpul dan menerima sisa-sisa buangan dari bahagian hulu sungai. Antara sungai-sungai yang sangat tercemar di Malaysia ialah Sungai Juru (Pulau Pinang), Sungai Kelang (Selangor), Sungai Lingi (Negeri Sembilan), Sungai Kempas (Johor) dan Sungai Perai (Pulau Pinang) [Program pemantauan kualiti air daratan (sungai), 2006]. Selain itu, sungai-sungai yang dikelaskan sebagai sederhana biasanya lokasinya terletak sebelum memasuki kawasan bandar tetapi sungai-sungai tersebut masih menerima sisa buangan dari aktiviti di persekitarannya.

Sungai-sungai yang bersih pula terletak di kawasan hulu sungai di mana kegiatan yang dilakukan di kawasan ini tidak mengakibatkan pencemaran yang ketara. Walaubagaimanapun, di bahagian hulu ini berlaku pula hakisan diakibatkan aliran yang deras. Ini seterusnya menyebabkan kelodak dan sendimen dibawa ke bahagian hilir sungai.

Di dalam projek ini, kualiti sungai tertumpu kepada dua kelas sahaja iaitu bersih (baik) atau tercemar sahaja.

2.4 Pendekatan Tradisional dalam Penentuan Kualiti Air Sungai

Pendekatan tradisional terhadap penentuan kualiti air dan sistem kawalan air sungai tercemar telah wujud sejak 2 dekad lalu seperti kaedah simulasi, penentuan dan statistik. Majoriti dari kewujudan model kualiti air ini adalah secara mekanikal dan bergantung kepada persamaan kebezaan yang telah diterbitkan dengan mengaplikasi lebihan jisim (O'connor et. al, 1973) dan juga kaedah-kaedah kimia seperti:

1. Oksigen terlarut (Dissolved Oxygen, DO)

Ia merupakan satu parameter yang penting bagi menentukan tahap pencemaran air. Kandungan oksigen terlarut dipengaruhi oleh suhu. Pertambahan suhu menyebabkan penurunan kandungan oksigen terlarut. Oksigen terlarut ini diperlukan oleh semua organisma organik sepertimana pentingnya oksigen kepada hidupan di darat. Kandungan oksigen terlarut yang tidak mencukupi mengakibatkan organisma akuatik akan mati.

2. Permintaan oksigen biokimia (Biological Oxygen Demand, BOD)

BOD ialah pengukuran kuantiti oksigen yang diperlukan oleh mikroorganisma dan bakteria untuk menstabilkan bahan organik biorosot terutamanya yang mempunyai kandungan karbon dan nitrogen. Proses pengoksidaan berlaku apabila bahan organik biorosot diurai oleh mikorganisma dan tenaga dihasilkan untuk kegunaan mikroorganisma itu sendiri.

Terdapat beberapa faktor yang mempengaruhi nilai BOD iaitu suhu, masa dan cahaya. Suhu memberi kesan ke atas pengambilan oksigen kerana aktiviti metabolik bertambah dengan pertambahan suhu. Masa yang ditetapkan untuk membuat ujian juga penting kerana jumlah oksigen yang digunakan akan bertambah dengan bertambahnya masa. Selain itu, cahaya juga memberikan kesan ke atas nilai BOD kerana kebanyakan air asli mengandungi alga dan oksigen boleh digantikan semula jika terdapatnya cahaya.

3. Permintaan oksigen kimia (Chemical Oxygen Demand, COD)

Ia adalah satu kaedah menentukan kandungan sesuatu bahan pencemar kimia di dalam air. Ianya digunakan bagi memendekkan masa ujian kerana bahan organik dioksidakan secara kimia dan bukannya secara biologi sepertimana BOD yang memerlukan masa sehingga 5 hari. Oleh kerana hampir kesemua bahan organik dioksidakan dalam ujian COD, nilainya adalah lebih tinggi jika dibandingkan dengan nilai BOD kerana hanya sebahagian sahaja bahan organik diuraikan.

4. Pepejal terampai (Suspended Solids, SS)

Ia biasanya terdiri daripada zarah-zarah organik dan zarah-zarah tidak organik yang bersaiz lebih besar daripada 0.001mm dan tidak larut dalam air. Zarah organik ini terdiri daripada alga, bakteria, protozoa dan lain-lain manakala pepejal terampai tak organik pula terdiri daripada tanah liat, kelodak dan sebagainya. Kehadiran pepejal terampai boleh menyebabkan air menjadi keruh, berbau dan beracun. Mikroorganisma aerob mengurai bahan organik di bahagian atas air lalu menghasilkan karbon dioksida, air dan sel yang baru. Di bahagian anaerobik pula, mikroorganisma menguraikan bahan organik dan sel yang terenap kepada asid organik dan seterusnya kepada karbon dioksida, air, hidrogen sulfid dan metana.

Penentuan kuantiti pepejal terampai boleh dilakukan melalui penurasan seterusnya penimbangan berat zarah-zarah yang tertinggal di atas kertas turas.

Kandungan pepejal terampai yang tinggi menunjukkan tahap pencemarannya adalah tinggi.

2.5 Analisis Diskriminan Dua Kumpulan

Analisis diskriminan menitikberatkan masalah mengenal beza di antara dua populasi yang bergantung kepada pemerhatian terhadap beberapa pembolehubah. Kita akan menentukan kedua-dua populasi tersebut sebelum mengumpul data dan mendapatkan sampel bagi setiap individu untuk setiap populasi. Masalah statistik ini digunakan untuk membangunkan peraturan atau fungsi diskriminan bergantung kepada ukuran yang diperolehi bagi setiap individu yang membolehkan kita untuk memasukkan sesuatu individu yang baru apabila ianya tidak diketahui sama ada dari populasi manakah letaknya individu baru itu.

Data yang diperolehi bagi setiap individu akan mengandungi pelbagai nilai bagi pembolehubah rawak kolerasi. Ia akan menunjukkan kolerasi-antara yang akan mengambil kira pembolehubah bersama (pembolehubah yang banyak) dan bukannya satu pembolehubah sahaja pada satu masa. Melalui pendekatan yang umum seperti analisis regresi, satu cara optima hendaklah dibina dalam menghasilkan kombinasi linear bagi pembolehubah-pembolehubah dan kombinasi linear ini kemudiannya dapat digunakan untuk mengklasifikasikannya. Penukaran masalah asas ini dari pembolehubah berbilang yang kompleks kepada satu pembolehubah yang senang untuk digunakan akan menghasilkan dan mengelaskan seseorang individu kepada salah satu daripada dua populasi adalah bergantung kepada nilai kombinasi linear bagi individu tertentu. Manipulasi statistik bergantung kepada kebolehan membina kombinasi linear

yang baik dan dengan membangunkan metodologi berkaitan yang berlainan daripada analisis regresi ini, akan diklasifikasikan di bawah analisis diskriminan.

Pada peringkat awal penggunaan, analisis diskriminan ini telah melibatkan aplikasi biologi dan perubatan. Walaubagaimanapun, ia telah digunakan dengan meluas dalam bidang bisnes, pembelajaran, kejuruteraan dan psikologi telah juga diambil perhatian dan telah berkembang sehingga ke hari ini.

2.5.1 Contoh Kehidupan Sebenar

Telah terbukti beberapa penyiasatan penyelidikan kehidupan sebenar yang melibatkan masalah analisis statistik boleh digunakan secara semulajadi menggunakan analisis diskriminan dua kumpulan ini iaitu:

1. Hulka et al, 1972

Di dalam pembelajaran penjagaan kesihatan peringkat rendah, ia berminat untuk membolehkan diskriminan di antara orang yang mempunyai simptom penyakit yang mendapatkan rawatan (populasi 1) dan orang yang mempunyai simptom yang tidak mendapatkan rawatan (populasi 2). Beberapa pembolehubah (iaitu X 's) boleh digunakan dalam menjelaskan perbezaan di antara kedua-dua populasi yang diukur iaitu berdasarkan masa bagi simptom itu, keseriusan simptom, jumlah kebimbangan terhadap simptom, perasaan seseorang terhadap kebolehan doktor dalam memberi komen, bilangan hari kehilangan-katil atau *bed-loss* (iaitu hari yang digunakan untuk berehat di katil dan ini memberi keputusan kepada kehilangan masa hari bekerja) dan kepelbagaian sosioekonomi

dan lokasi tempat. Teknik analisis diskriminan boleh menjadi sangat berguna dalam menentukan apakah kombinasi linear pembolehubah-pembolehubah ini dapat sebaiknya membantu dalam meramal adakah seseorang itu mempunyai simptom penyakit terus mendapatkan rawatan atau tidak.

2. Kleinbium, 1971

Di dalam pembelajaran epidemiologik, ia memberi perhatian terhadap insiden wabak koronari hati (Coronary Heart Disease, CHD) yang direkabentuk melibatkan sekumpulan orang yang bebas dari CHD berdasarkan bilangan tahun. Pada permulaan masa bagi mengikuti perkembangan, pembolehubah (iaitu X's) seperti tahap kolestrol, tekanan darah, pekerjaan, tinggi dan pelbagai sosioekonomi dan lokasi tempat telah direkodkan bagi setiap individu. Di penghujung masa mengikuti perkembangan ini, ia telah menentukan sama ada subjek telah membangunkan CHD (populasi 1) atau tidak (populasi 2). Dengan menggunakan teknik analisis diskriminan ini, kombinasi linear bagi pembolehubah-pembolehubah boleh membantu dalam mengenal beza sama ada populasi CHD atau bukan CHD dapat ditentukan.

3. McMichael et al., 1975

Di dalam pembelajaran berkaitan kesihatan pekerja di dalam industri tertentu, matlamat penyelidikan adalah untuk menentukan sama ada sejarah kerja untuk majikan (sebelum dan sekarang) yang mempunyai leukemia (populasi 1) mempunyai cara bekerja yang berbeza secara signifikan atau pun tidak dengan majikan yang tidak mempunyai leukemia (populasi 2). Selain itu, ia mengambil kira pelbagai pembolehubah yang mungkin seperti umur, pendidikan dan sejarah

merokok. Teknik diskriminan analisis sangat berguna dalam menentukan kepentingan sejarah kerja sesuatu majikan di dalam kombinasi dengan beberapa faktor yang berkemungkinan sebagai agen yang berhubung-leukimia

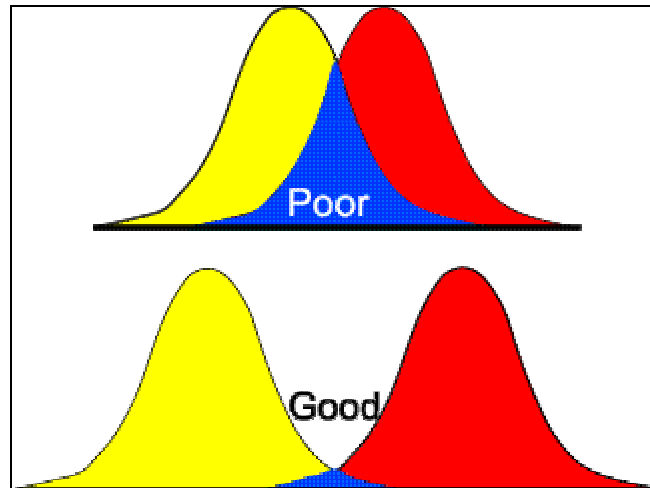
Berdasarkan contoh di atas, dapat dilihat bahawa pelbagai persoalan penyelidikan statistik secara semulajadi boleh dilihat dalam konteks permasalahan analisis diskriminan. Secara jelasnya, kekangan utama di dalam ketiga-tiga masalah ini diperlukan untuk menjadi peraturan statistik dalam menentukan kombinasi yang manakah bilangan pembolehubah yang terpenting yang menyediakan pendiskriminasi yang terbaik di antara kedua-dua populasi yang ditentukan.

Selain itu, analisis diskriminan boleh digunakan hanya apabila pembolehubah- X adalah berterusan (continuous) dengan taburan yang tidak terlampau.

2.5.2 Asas kepada Analisis Diskriminan

Analisis diskriminan boleh digunakan jika pembolehubah bersandar boleh dikategorikan (nominal) dan pembolehubah tak bersandar adalah selanjar. Analisis diskriminan ini berupaya untuk mengelola 2 atau lebih banyak bilangan pembolehubah bersandar.

Untuk mengukur kejayaan di dalam kaedah ini, skor diskriminan hendaklah di dalam taburan normal dan darjah tindanan di antara skor diskriminan boleh digunakan seperti yang ditunjukkan dalam rajah 2.1:



Rajah 2.1: Taburan normal yang baik dan tidak baik bagi analisa diskriminan

Seterusnya, skor diskriminan ini adalah hasil tambah linear pemberat untuk pembolehubah diskriminasi. Kesemua pembolehubah boleh diapiawakan dengan memastikan perbezaan skala di antara pembolehubah dihapuskan.

Skor diskriminan dikira dari fungsi diskriminan yang diperolehi dan ditulis dalam bentuk berikut:

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n \quad (2.1)$$

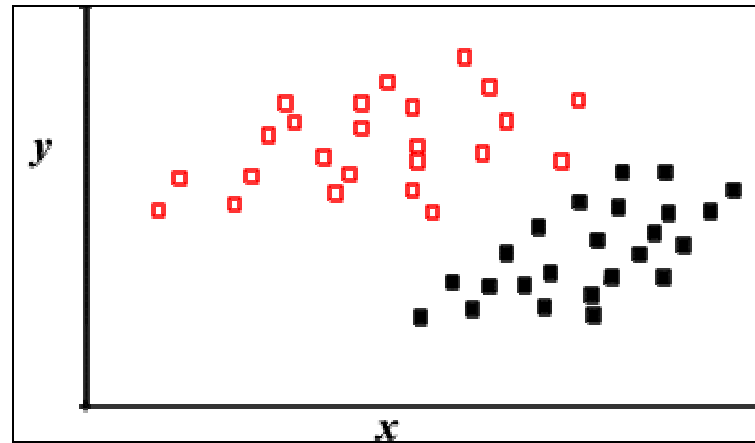
Di mana

Z adalah skor diskriminan;

α_n adalah koefisien (pemberat) untuk pembolehubah i;

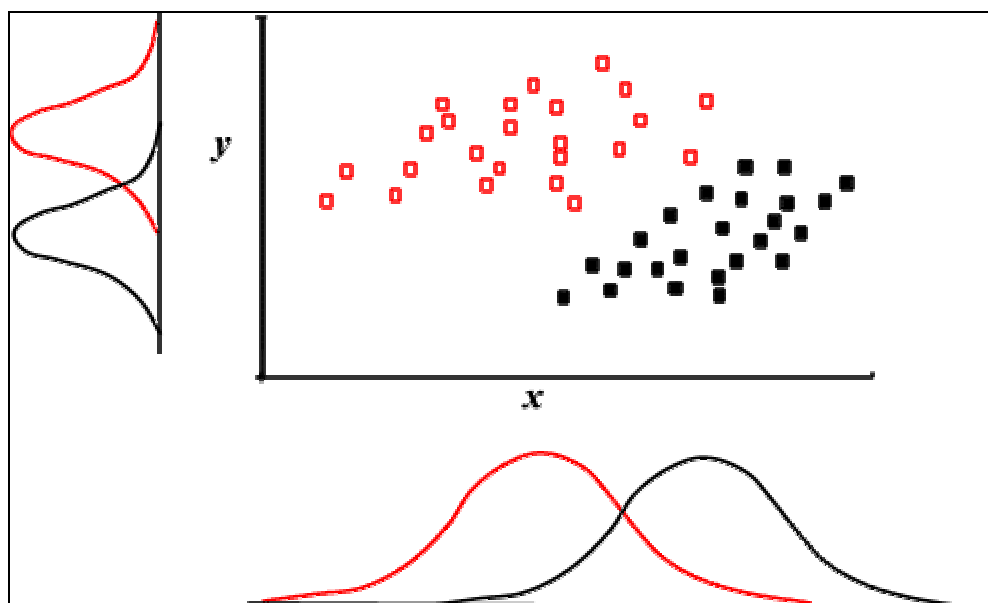
X_n adalah skor terpiawai untuk pembolehubah i (pembolehubah piawai mempunyai min= 0 dan sisihan piawai = 1).

Rajah 2.2 menunjukkan contoh serakan set data yang mudah iaitu 2 kumpulan dengan dua pembolehubah.



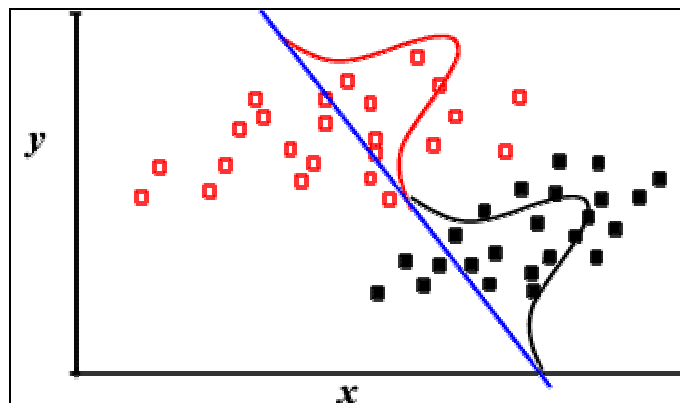
Rajah 2.2: Serakan data

Daripada rajah 2.3, dapat dilihat dengan jelas bahawa 2 kumpulan tersebut boleh dipisahkan oleh dua pembolehubah. Tetapi terdapat jumlah penindanan yang besar pada setiap paksi. Oleh itu, satu paksi baru hendaklah dibina supaya kedua-dua kumpulan tidak bertindan di antara satu sama lain.



Rajah 2.3: Jumlah penindanan bagi paksi x dan y

Rajah 2.4 menunjukkan paksi baru yang dihasilkan mewakili pembolehubah yang baru dan merupakan kombinasi linear bagi x dan y yang merupakan skor diskriminan.



Rajah 2.4: Jumlah penindanan bagi paksi baru

2.5.3 Ringkasan

Secara ringkasnya, bab ini telah menerangkan tentang beberapa alga yang wujud di dalam air sungai dan kualiti air sungai ini boleh ditentukan dengan merujuk kepada Indeks Kualiti Air Negara. Beberapa pendekatan tradisional seperti penggunaan beberapa kaedah kimia telah dibincangkan di dalam bab ini. Akhir sekali, penerangan yang jelas tentang analisis diskriminan dua kumpulan serta beberapa contoh kehidupan sebenar yang berkaitan dengannya dibincangkan. .

BAB 3

METODOLOGI

3.1 Pengenalan

Bab ini akan menerangkan bagaimana persampelan data dilakukan, statistik yang penting, mengkaji tahap keertian serta kaedah untuk mencari titik pemisahan bagi analisis diskriminan dengan menggunakan SPSS.

3.2 Analisis Diskriminan Dua Kumpulan

Analisis diskriminan dua kumpulan merupakan salah satu kaedah yang sesuai digunakan untuk mengklasifikasikan komposisi alga ini kepada dua kumpulan air iaitu air tercemar dan air bersih. Peratusan klasifikasi bagi kaedah ini bergantung kepada kebolehannya mendiskriminasi beberapa pembolehubah iaitu alga kepada dua kumpulan tersebut dan akan diketahui setelah analisis ini dilakukan.

3.2.1 Persampelan Data

Sebelum projek ini dijalankan, pihak Pusat Pengajian Sains Kajahayat Universiti Sains Malaysia telah mengumpul sebanyak 43 spesis alga dari sungai-sungai (stesen-stesen) yang berlainan yang terdapat di Pulau Pinang. Data ini telah dikumpul selama tiga tahun iaitu dari tahun 1998 sehingga 2000.

Untuk projek ini, sebanyak 21 spesis alga digunakan. Berdasarkan komposisi alga yang telah diambil sepanjang tempoh tiga tahun tersebut, kualiti air telah dibahagikan kepada dua kumpulan iaitu bersih dan tercemar. Dengan adanya komposisi alga ini, satu persamaan telahan Z seperti pada persamaan (2.1) yang terbaik dapat dihasilkan bagi mewakili kedua-dua kumpulan air tersebut.

Spesis-spesis alga ini telah dikumpul secara individu (satu persatu) dalam jangka masa tiga tahun iaitu dari tahun 1998 sehingga 2000 dan telah diuji dan dikelaskan kepada bersih atau tercemar. Sebanyak 86 sampel air tercemar dan 50 sampel air bersih digunakan untuk tujuan data latihan manakala 5 sampel masing-masing dari kedua-dua kumpulan digunakan sebagai data ujian.

Kesemua 21 spesis alga telah wujud di dalam kedua-dua kumpulan sama ada bersih atau tercemar. Seperti yang diketahui, projek ini bertujuan meramal kualiti air berdasarkan komposisi alga. Oleh itu, adalah penting mencari komposisi alga yang boleh mewakili kedua-dua kumpulan dan seterusnya menghasilkan satu persamaan telahan Z yang terbaik.

Untuk menentukan komposisi alga yang terbaik yang akan dipilih di dalam projek ini, kita perlu melihat saling hubungan atau kolerasi antara komposisi alga tersebut dengan menggunakan matriks kolerasi. Oleh itu, matriks kolerasi yang diperolehi untuk setiap tahun iaitu dari tahun 1998 hingga 2000 akan dikaji berdasarkan saling hubungan yang kuat antara komposisi alga yang berkait rapat dengan air tercemar.

Didapati kesemua komposisi alga telah dikenalpasti mempunyai saling hubungan yang kuat dengan air tercemar. Komposisi alga tersebut adalah

X ₁	Achanthes
X ₂	Amphora Libica
X ₃	Anomoeoneis Brashyshira
X ₄	Cocconeis
X ₅	Coscinodiscus
X ₆	Cyclotella
X ₇	Cymbella
X ₈	Diatoma
X ₉	Diplonris Decipiens
X ₁₀	Eunotia
X ₁₁	Fragilaria
X ₁₂	Frustulia
X ₁₃	Gomphonema Acuminatum
X ₁₄	Hantzschia Amphioxys
X ₁₅	Navicula
X ₁₆	Neidium Affine
X ₁₇	Nitzschia
X ₁₈	Pinnularia
X ₁₉	Psammothidium Bioretti
X ₂₀	Stauroneis Obtusa
X ₂₁	Surirella

3.2.2 Pembahagian Kualiti Air Sungai

Kualiti air sungai boleh dibahagikan kepada dua kumpulan iaitu kumpulan air tercemar (polluted) dan kumpulan air bersih (clean). Kumpulan pertama terdiri daripada kumpulan air yang tercemar yang mempunyai beberapa komposisi alga yang dapat dikesan dalam tempoh tiga tahun tersebut manakala kumpulan kedua iaitu kumpulan air bersih yang mengandungi beberapa komposisi alga yang lain.

3.2.3 Statistik yang Penting Berkaitan dengan Analisis Diskriminan

1) Matriks Klasifikasi

Dikenali sebagai matrik penganggar. Mengandungi bilangan cerapan yang diklasifikasikan dengan betul dan yang tidak diklasifikasikan dalam kumpulan yang tepat.

2) Kolerasi Kanonik

Kolerasi kanonik menyukat sejauh mana hubungan yang wujud antara nilai skor diskriminan dengan setiap kumpulan.

3) Koefisien Fungsi Diskriminan

Merupakan pekali bagi pembolehubah yang tidak dipiawaikan apabila kesemua pembolehubah masih lagi dalam unit ukuran yang asal.

4) Skor diskriminan

Nilai koefisien yang belum dipiawaikan didarab dengan nilai setiap pembolehubah. Hasil darab tersebut akan dijumlahkan dan ditambah kepada satu nilai malar untuk mengira skor diskriminan.

5) Koefisien Fungsi Diskriminan yang dipiawaikan

Ia merupakan koefisien fungsi diskriminan apabila kesemua pembolehubah telah dipiawaikan dan mempunyai nilai min sifar dan nilai varians satu.

6) Sentroid

Sentroid merupakan nilai min bagi skor diskriminan bagi setiap kumpulan priori. Bilangan sentroid bergantung kepada bilangan kumpulan priori yang terlibat di dalam projek.

Kelebihan analisis ini adalah ia boleh mengurangkan ruang dimensi dalam analisis (ruang dimensi analisis adalah dengan bilangan pembolehubah tidak bersandar) menjadi $G-1$ dimensi, yang mana G ialah bilangan kumpulan priori.

Analisis ini melakukan proses transformasi untuk menjadikan model dalam bentuk yang paling mudah iaitu satu dimensi.

Fungsi diskriminan yang terbentuk seperti yang telah dinyatakan di dalam bab 2
 yaitu

$$Z = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n \quad (2.1)$$

akan mentransformasikan nilai pembolehubah individu sehingga meninggalkan satu
 nilai diskriminan tunggal Z yang boleh digunakan untuk mengelaskan objek
 dengan

$X_1, X_2, \dots, X_n$ adalah pembolehubah tak bersandar

$\alpha_1, \alpha_2, \dots, \alpha_n$ adalah pekali diskriminan

Diskriminan analisis dua kumpulan akan menghitung dan menghasilkan nilai
 diskriminan, α_n , manakala pembolehubah tak bersandar, X_n adalah nilai sebenar
 dari sampel dengan $n = 0$ atau 1 .

Setelah nilai pekali diskriminan diperolehi, nilai diskriminan untuk sebarang
 serapan dalam sampel, ataupun sebarang jenis air dapat dihitung berdasarkan
 persamaan Z tersebut. Seterusnya, jenis air akan diletak ke dalam kumpulan yang
 bersesuaian berdasarkan kepada nilai diskriminan.

3.2.4 Mengkaji Tahap Keertian Pembolehubah Diskriminan

Untuk mengkaji tahap keertian setiap pembolehubah diskriminan, kita hendaklah mengetahui adakah setiap pembolehubah diskriminan ini berbeza secara beerti bagi kedua-dua kumpulan tersebut. Oleh itu, satu ujian hipotesis perlu dilakukan untuk melihat perbezaan nilai min bagi kedua-dua kumpulan tersebut. Ujian hipotesis adalah seperti berikut

$$\begin{aligned}H_0 &: \mu_1 = \mu_2 \\H_a &: \mu_1 \neq \mu_2\end{aligned}\tag{3.2}$$

di mana μ_1 dan μ_2 masing-masing merupakan min bagi populasi kumpulan satu dan kumpulan dua masing-masing. Ujian hipotesis ini boleh dijalankan dengan menggunakan ujian statistik Wilks' λ . Jika nilai Wilks' λ yang semakin kecil maka semakin besar kebarangkalian untuk menolak hipotesis nol.

3.2.5 Kaedah Titik Pemisahan

Nilai skor diskriminan yang diperolehi boleh digunakan untuk mencari nilai titik pemisahan antara kedua-dua kumpulan tersebut. Nilai ini sangat berguna untuk pemerhatian di masa akan datang. Secara amnya, kualiti air mesti mempunyai nilai skor diskriminan yang melebihi titik pemisahan untuk berada di dalam kumpulan satu manakala mempunyai nilai skor diskriminan yang kurang untuk berada di dalam kumpulan yang kedua.

Kaedah yang biasa digunakan untuk mencari nilai titik pemisahan ialah

$$\text{Titik pemisahan, } Z_p = \left(\frac{\overline{Z}_1 + \overline{Z}_2}{2} \right) \quad (3.3)$$

$\overline{Z}_1$ dan $\overline{Z}_2$ merupakan purata nilai skor diskriminan bagi kedua-dua kumpulan satu dan dua.

3.2.6 Kesimpulan

Secara kesimpulannya, satu fungsi diskriminan terbentuk dan satu titik pemisahan diperolehi di dalam menentukan kualiti air sungai dengan menggunakan kaedah analisis diskriminan dua kumpulan ini.