

Pendahuluan : Aplikasi*

Dr. rer. nat. Hendri Murfi

* Beberapa bagian dari slide ini adalah terjemahan dari slide Data Mining oleh I. H. Witten, E. Frank dan M. A. Hall

Intelligent Data Analysis (IDA) Group

Departemen Matematika, Universitas Indonesia – Depok 16424

Telp. +62-21-7862719/7863439, Fax. +62-21-7863439, Email. hendri@ui.ac.id

Rekomendasi Contact Lenses

Masalah Ideal

Age	Spectacle prescription	Astigmatism	Tear production rate	Recommended lenses
Young	Myope	No	Reduced	None
Young	Myope	No	Normal	Soft
Young	Myope	Yes	Reduced	None
Young	Myope	Yes	Normal	Hard
Young	Hypermetrope	No	Reduced	None
Young	Hypermetrope	No	Normal	Soft
Young	Hypermetrope	Yes	Reduced	None
Young	Hypermetrope	Yes	Normal	hard
Pre-presbyopic	Myope	No	Reduced	None
Pre-presbyopic	Myope	No	Normal	Soft
Pre-presbyopic	Myope	Yes	Reduced	None
Pre-presbyopic	Myope	Yes	Normal	Hard
Pre-presbyopic	Hypermetrope	No	Reduced	None
Pre-presbyopic	Hypermetrope	No	Normal	Soft
Pre-presbyopic	Hypermetrope	Yes	Reduced	None
Pre-presbyopic	Hypermetrope	Yes	Normal	None
Presbyopic	Myope	No	Reduced	None
Presbyopic	Myope	No	Normal	None
Presbyopic	Myope	Yes	Reduced	None
Presbyopic	Myope	Yes	Normal	Hard
Presbyopic	Hypermetrope	No	Reduced	None
Presbyopic	Hypermetrope	No	Normal	Soft
Presbyopic	Hypermetrope	Yes	Reduced	None
Presbyopic	Hypermetrope	Yes	Normal	None

Klasifikasi Bunga Iris

Data Bernilai Numerik Klasik

	Sepal length	Sepal width	Petal length	Petal width	Type
1	5.1	3.5	1.4	0.2	Iris setosa
2	4.9	3.0	1.4	0.2	Iris setosa
...					
51	7.0	3.2	4.7	1.4	Iris versicolor
52	6.4	3.2	4.5	1.5	Iris versicolor
...					
101	6.3	3.3	6.0	2.5	Iris virginica
102	5.8	2.7	5.1	1.9	Iris virginica
...					

3

Prediksi Kinerja CPU

Prediksi Bernilai Numerik

	Cycle time (ns)	Main memory (Kb)		Cache (Kb)	Channels		Performance
	MYCT	MMIN	MMAx	CACH	CHMIN	CHMAX	PRP
1	125	256	6000	256	16	128	198
2	29	8000	32000	32	8	32	269
...							
208	480	512	8000	32	0	0	67
209	480	1000	4000	0	0	0	45

4

Klasifikasi Negosiasi Kontrak

Contoh Yang Lebih Realistik

Attribute	Type	1	2	3	...	40
Duration	(Number of years)	1	2	3		2
Wage increase first year	Percentage	2%	4%	4.3%		4.5
Wage increase second year	Percentage	?	5%	4.4%		4.0
Wage increase third year	Percentage	?	?	?		?
Cost of living adjustment	{none,tcf,tc}	none	tcf	?		none
Working hours per week	(Number of hours)	28	35	38		40
Pension	{none,ret-allw,empl- pnt}	none	?	?		?
Standby pay	Percentage	?	13%	?		?
Shift-work supplement	Percentage	?	5%	4%		4
Education allowance	{yes,no}	yes	?	?		?
Statutory holidays	(Number of days)	11	15	12		12
Vacation	{below-avg,avg,gen}	avg	gen	gen		avg
Long-term disability assistance	{yes,no}	no	?	?		yes
Dental plan contribution	{none,half,full}	none	?	full		full
Bereavement assistance	{yes,no}	no	?	?		yes
Health plan contribution	{none,half,full}	none	?	full		half
Acceptability of contract	{good,bad}	bad	good	good		good

5

Aplikasi Pinjaman

Kasus Pada American Express

- Data: quisioner tentang informasi keuangan dan personal
- Pertanyaan: layakkah untuk diberi pinjaman ?
 - Metode statistik sederhana dapat menjawab hampir 90% kasus
 - Kasus-kasus pada garis batas (borderline) ditentukan oleh tenaga ahli yang berwenang
 - Akan tetapi, 50% kasus-kasus pada garis batas yang diberi pinjaman gagal dalam pengembalian
- Solusi:
 - Tolak semua kasus pada garis batas. Ini tidak mungkin karena kasus-kasus pada garis batas adalah nasabah aktif terbesar
 - Solusi lain → metode lain, misal machine learning

6

Aplikasi Pinjaman

Solusi Machine Learning

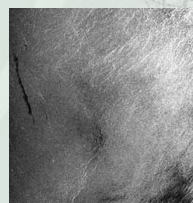
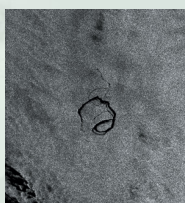
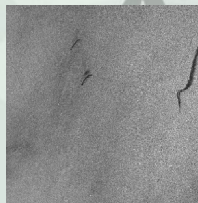
- Data training: 1000 sample untuk kasus-kasus pada garis batas
- 20 Atribut/Fitur:
 - Umur
 - Lamanya tinggal di alamat saat ini
 - Lamanya menjadi nasabah
 - Kepemilikan kartu kredit lain
 -
- Hasil:
 - Tenaga ahli memberikan akurasi 50%
 - If-then Rule memberikan akurasi 70%

7

Skrining Gambar

Pencemaran Perairan Pantai

- Data: gambar-gambar satelit dari perairan pantai
- Masalah: mendeteksi lapisan minyak pada gambar tersebut
 - Lapisan minyak muncul pada gambar sebagai area hitam dengan bentuk dan ukuran yang berubah-ubah
 - Persoalan ini tidak mudah, karena area hitam bisa juga disebabkan oleh kondisi cuaca, misal angin
 - Pendeteksian ini adalah proses yang mahal karena membutuhkan personel yang terlatih



8

Skrining Gambar

Solusi Machine Learning

- Data training: diekstrak dari area hitam dari gambar
- Atribut/Fitur:
 - Ukuran area
 - Bentuk area
 - Intensitas
 - Ketajaman dan lekukan dari batas
 - Kedekatan dengan wilayah lain
 - Info tentang latar belakang
- Kendala:
 - Sedikit data training
 - Data tidak seimbang : sebagian besar bukan lapisan minyak

9

Prediksi Beban Listrik

- Latar belakang: perusahaan pemasok listrik perlu memprediksi kebutuhan tenaga listrik pada masa yang akan datang
 - Peramalan beban min/max untuk setiap jam akan memberikan penghematan yang signifikan
- Data: berupa model beban yang dibangun secara manual dengan asumsi kondisi cuaca „normal“
 - Beban dasar dalam setahun
 - Periodesitas beban dalam setahun
 - Pengaruh hari libur
- Masalah: membuat model yang selaras dengan kondisi cuaca

10

Prediksi Beban Listrik

Solusi Machine Learning

- Atribut/Fitur:
 - Temperatur
 - Kelembaban
 - Kecepatan angin
 - Kondisi awan

11

Pemasaran dan Penjualan

- Data: perusahaan biasanya menyimpan data pemasaran dan penjualan
- Aplikasi:
 - **Loyalitas Pelanggan**, yaitu mendeteksi pelanggan yang akan menyeberang/pindah ke perusahaan lain berdasarkan perubahan tingkah laku
 - **Penawaran Khusus**, yaitu mengidentifikasi pelanggan yang potensial untuk keuntungan tertentu, misal pemilik kartu kredit yang membutuhkan uang selama liburan
 - **Analisa Market Basket**, yaitu mencari item-item yang biasanya terjadi secara bersamaan dalam transaksi

12