



RESPONSI DIKUMPULKAN JAM 16.00 DI MEJA, Pengerjaan dengan kertas folio

Kerjakan dengan jelas, lengkap dan teliti, alat bantu yang diperkenankan hanyalah Kalkulator (Handphone/Tablet tidak diijinkan) !

1. Jelaskan istilah-istilah dalam data mining berikut ini :[POIN 15]
 - a. Data Mining
 - b. Supervised Learning
 - c. Association Rules
2. *Cross Industry Standard Process for Data Mining* (CRISP-DM) terdiri dari enam fase yang saling terkait. Jelaskan proses yang dilakukan pada fase keempat *Data Modelling*, seberapa penting fase keempat pada keberhasilan proses data mining, dan dampaknya jika terjadi kesalahan pada fase tersebut! [POIN 15]
3. Berikut adalah dataset Pizza Delivery Order, Buatlah Decision Tree (Pohon Keputusan) dari sample contoh data set berikut ini, dimana atribut measurement yang digunakan adalah *information gain*, dan pembentukan tree dilakukan dengan tanpa *pre pruning* terlebih dahulu. [POIN 30]

Tabel 1 Pizza Delivery Order

No.	Jumlah Pesanan	Cuaca	Kondisi Jalanan	Keterangan
1.	Banyak	Cerah	Macet	Terlambat
2.	Banyak	Hujan	Macet	Tepat
3.	Sedang	Cerah	Lancar	Tepat
4.	Sedikit	Cerah	Macet	Terlambat

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n p_i * \log_2 p_i$$

$$\text{Gain(S,A)} = E(S) - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * E(S_i)$$

S= himpunan kasus

Pi=proporsi dari Si terhadap S

A=Atribut

|Si|=jumlah kasus pada partisi ke-i

N = jumlah partisi atribut A

4. Tentukan kelas label dari sample dataset (X) dimana X= (Outlook=Sunny, Temperature=Cool, Humidity=High,Wind=Weak) pada dataset dibawah ini, dengan menggunakan algoritma Naive Bayes Classifier! [POIN 25]

Rumus :

1. $P(x|C_i)$ untuk Class i

2. $P(X|C_i)$ untuk setiap Class : $P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$

3. $P(X|C_i) * P(C_i)$

Tabel 2 Dataset Pertandingan Tennis

Outlook	Temperature	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Overcast	Mild	High	Strong	Yes

5. Studi Kasus.

Dibawah ini ada sample contoh dataset pada **tabel 3 Database Obat Poliklinik UDINUS** menurut anda termasuk ke dalam kategori learning manakah contoh kasus dibawah dalam machine learning, kemudian tentukan metode dan algoritma data mining yang tepat untuk memproses dataset di bawah, Jelaskan jawaban anda dengan detail dan lengkap! **[POIN 15]**

Tabel 3. Database Jenis Obat Keluar di Poliklinik UDINUS

Tanggal	NIM	Obat1	Obat2	Obat3	Obat4	Obat5
22/10/2014	A11.001	Ibuprofen	Amoxicicilin	Vitamin C	-	-
22/10/2014	A12.003	Cefixime	Paracetamol	Multivitamin	-	-
23/10/2014	A14.023	Multivitamin	Paracetamol	-	-	-
24/10/2014	A21.034	Antiseptik	Fentanil	Paracetamol	Vitamin A	Aspirin
24/10/2014	A11.901	Amoxicicilin	Paracetamol	-	-	-
24/10/2014	A14.098	Metadon	Paracetamol	Antibiotik	-	-
25/10/2014	A11.111	Antiseptik	Ciprofloaxin	Multivitamin	-	-

----- Do the best and Get the best-----

Diperiksa Oleh :	Disahkan Oleh :
Koordinator Mata Kuliah	Ka. Prodi
<u>Ardytha Luthfiarta, M.Kom, M.CS.</u>	<u>Heru Agus Santoso, P.hD</u>