



## UJIAN TENGAH SEMESTER (UTS) GENAP 2020/2021

MATA KULIAH	: Pengantar Data Mining	DOSEN	: Syaiful Anam, S.Si., MT, Ph.D Zuraidah Fitriah, S.Si., M.Si
KELAS	: A	TANGGAL	: 6 April 2021
PROGRAM STUDI	: Matematika	WAKTU	: 45 Menit
SIFAT UJIAN	: Open book		

Tuliskan pernyataan berikut dan berilah tandatangan

*Saya menyatakan bahwa jawaban UTS ini adalah hasil dari pemikiran saya sendiri, jika terbukti tidak benar maka saya sanggup menerima resiko dengan mendapatkan nilai nol.*

Yang menyatakan,  
Tanda tangan  
Nama

Kerjakan soal-soal dibawah ini. Nama File Gunakan Nim\_Nama

### Soal 1

Jelaskan definisi dan pentingnya *data cleansing* pada *preprocessing data*.

### Soal 2

Transaksi	Item yg dibeli
1	Susu, Teh, Gula
2	Teh, Gula, Roti
3	Teh, Gula
4	Susu, Roti
5	Susu, Gula, Roti
6	Teh, Gula
7	Gula, Kopi, Susu
8	Gula, Kopi, Susu
9	Susu, Roti, Kopi
10	Gula, Teh, Kopi

Tentukan *frequent itemset* dari data penjualan di samping dengan menggunakan algoritma apriori dan dapatkan pula *association rules*nya. Diketahui min.support 50% dan min.confidence 50%.

Selamat Mengerjakan



KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN  
UNIVERSITAS BRAWIJAYA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
JURUSAN MATEMATIKA

Jl. Veteran, Malang 65145, Jawa Timur, Indonesia, Telp-fax : +62-341-571142  
<http://matematika.ub.ac.id>, e-mail: [jurmatub@ub.ac.id](mailto:jurmatub@ub.ac.id)

---

## UJIAN TENGAH SEMESTER (UTS) GENAP 2020/2021

MATA KULIAH : Pengantar Data Mining

DOSEN : Syaiful Anam, S.Si., MT, Ph.D  
Zuraidah Fitriah, S.Si., M.Si

KELAS : A

TANGGAL : 6 April 2021

PROGRAM STUDI : Matematika

WAKTU : 45 Menit

SIFAT UJIAN : Open book

---

Tuliskan pernyataan berikut dan berilah tandatangan

*Saya menyatakan bahwa jawaban UTS ini adalah hasil dari pemikiran saya sendiri, jika terbukti tidak benar maka saya sanggup menerima resiko dengan mendapatkan nilai nol.*

Yang menyatakan,  
Tanda tangan  
Nama

Kerjakan soal-soal dibawah ini. Nama File Gunakan Nim\_Nama

### Soal No 3.

Diketahui data dengan fitur sebagai berikut.

Data ke	x	y
1	1	1
2	4	1
3	3	4
4	5	4

- Tuliskan satu iterasi dari metode K-means jika jarak yang digunakan adalah **Jarak Manhattan** dan diketahui jumlah cluster adalah 2 serta *centroid* /pusat cluster adalah (1,1) dan (5,4) (Hati-hati: **bukan jarak Euclid**)
- Gunakan Metode *Complete Linkage* untuk mengcluster data diatas menjadi dua *cluster* dimana jarak yang digunakan **Jarak Manhattan**. Gambarkan dendogramnya.

Selamat Mengerjakan.