**MAKALAH**

**DATASET PROFIL PELANGGAN RESTORAN**

**Dosen pengampu: FHIRA NHITA, S.T., M.T.**

**Disusun Oleh :**

**-**

**-**

**-**

**-**

**-**

****

**FAKULTAS TEKNIK INFORMATIKA**

**TELKOM UNIVERSITY**

**2018**

Daftar Isi

[KATA PENGANTAR 3](#_Toc524422472)

[BAB I PENDAHULUAN 4](#_Toc524422473)

[Latar Belakang 4](#_Toc524422474)

[Rumusan Masalah 4](#_Toc524422475)

[Tujuan 4](#_Toc524422476)

[BAB II PEMBAHASAN 5](#_Toc524422477)

[Dataset 5](#_Toc524422478)

[Preprocessing 5](#_Toc524422479)

[Kasus: Dataset Profil Pelanggan Sebuah Restoran 5](#_Toc524422480)

[Sumber Data 5](#_Toc524422481)

[Deskripsi Data 5](#_Toc524422482)

[Masalah Data 6](#_Toc524422483)

[Solusi Masalah Data 6](#_Toc524422484)

[PENUTUP 7](#_Toc524422485)

# KATA PENGANTAR

# BAB I PENDAHULUAN

## Latar Belakang

## Rumusan Masalah

1. Apa itu dataset?
2. Apa itu *missing values* dalam dataset?
3. Apa itu *preprocessing*?
4. Bagaimana cara mengidentifikasi sebuah dataset?
5. Bagaimana cara memperbaiki *missing values*?
6. Bagaimana menggunakan *tools data mining*?

## Tujuan

1. Mengetahui dan memahami dataset.
2. Memahami masalah yang dapat timbul dalam sebuah dataset.
3. Dapat memberikan solusi terhadap masalah yang ada pada dataset.
4. Memahami dan mengaplikasikan tahapan *preprocessing* dalam mengolah sebuah dataset.
5. Eksplorasi *software* dan *tools* yang digunakan untuk *data mining*.

# BAB II PEMBAHASAN

## Dataset

Data dalam data mining merupakan kumpulan objek data dan atributnya. Dataset merupakan sekumpulan banyak data dalam table database. Data-data ini memiliki property atau karakteristik yang dinamakan atribut. Pada dataset mungkin timbul sebuah anomaly data, contohnya adalah *missing values*. *Missing values* adalah sebuah keadaan dataset dimana nilai data kosong atau tidak ada. Beberapa sebab terjadinya *missing values* antara lain, informasi tidak terkumpul secara lengkap atau atribut tidak bisa diterapkan untuk semua kasus. Salah satu solusi untuk penanganan *missing values* adalah dengan estimasi nilai *missing values* menggunakan mean dari data tersebut.

## Preprocessing

Strategi dan teknik yang saling berkaitan untuk membuat data lebih mudah/cocok untuk digunakan pada *data mining*. Tujuannya adalah untuk meningkatkan hasil analisis data mining terkait masalah waktu, cost dan kualitas.

## Kasus: Dataset Profil Pelanggan Sebuah Restoran

### Sumber Data

<https://archive.ics.uci.edu/ml/datasets/Restaurant+%26+consumer+data#>

### Deskripsi Data

* Instances: 138
* Attributes: 19
* userID: Nominal
* latitude: Numeric
* longitude: Numeric
* the\_geom\_meter: Nominal (Geospatial)
* smoker: Nominal, Missing: 3, 2 [false,true]
* drink\_level: Nominal, 3 [abstemious,social drinker,casual drinker]
* dress\_preference:Nominal, Missing: 5, 4 [informal,formal,no preference,elegant]
* ambience: Nominal, Missing: 6, 3 [family,friends,solitary]
* transport: Nominal, Missing: 7, 3 [on foot,public,car owner]
* marital\_status: Nominal, Missing: 4, 3 [single,married,widow]
* hijos: Nominal, Missing: 11, 3 [independent,kids,dependent]
* birth\_year: Nominal
* interest: Nominal, 5 [variety,technology,none,retro,eco-friendly]
* personality: Nominal, 4 [thrifty-protector,hunter-ostentatious,hard-worker,conformist]
* religion: Nominal, 5 [none,Catholic,Christian,Mormon,Jewish]
* activity: Nominal, Missing: 7, 4 [student,professional,unemployed,working-class]
* color: Nominal, 8 [black,red,blue,green,purple,orange,yellow,white]
* weight: Numeric
* budget: Nominal, Missing: 7, 3 [medium,low,high]
* height: Numeric

### Masalah Data

Beberapa atribut diatas memiliki missing values, seperti di atribut smoker, dress\_preference, ambiance, transport, marital\_status, hijos, activity, dan budget.

### Solusi Masalah Data

Solusi dalam penanganan *missing values* adalah dengan estimasi nilai menggunakan mean dari dataset yang ada sehingga didapat sebuah nilai yang akan digunakan pada atribut yang memiliki *missing values.*

# PENUTUP