

DATA MINING

Catatan Kuliah

Ikang FADHLI

ikanx101.com

30 March 2022

Contents

1	Week 1 Key Take Points	4
2	Week 2 Key Take Points	4
2.1	Standarisasi Data Numerik	4

List of Figures

1 Week 1 Key Take Points

Systems: physical process of interest.

Model: mathematical representation of the system.

Simulation: a type of model where computer is used to imitate the behavior of the system.

5 Tipe Data Mining:

1. Estimasi
2. Prediksi
3. Klasifikasi
4. Clustering
5. Asosiasi

Istilah Standar yang Perlu Dipahami:

1. Atribut; biasanya disebut dengan variabel atau satu kolom dalam dataframe.
2. Class atau label; biasanya disebut dengan isi (label) dari suatu atribut.
3. Data object; biasa disebut dengan respons atau baris data.

2 Week 2 Key Take Points

Tipe-tipe data:

1. Nominal.
2. Binary ada dua:
 - Symmetric binary (equally important) dan
 - Asymmetric binary (unequally important).

2.1 Standarisasi Data Numerik

Menggunakan formula $Z - score$, yakni $z = \frac{x-\mu}{\sigma}$.

Metode lainnya adalah dengan menghitung *mean absolute deviation*.

```
tes = runif(10,0,10)
tes
```

```
## [1] 9.6176276 3.1305043 3.8424873 9.4380110 7.4822160 9.2268074 0.1638225
## [8] 9.0644594 8.9462140 8.1329045
```

```
mean_abs_dev = function(tes){  
  m = sum(tes) / length(tes)  
  s = sum(abs(tes - m)) / length(tes)  
  z = (tes - m)/s  
  return(round(z,3))  
}  
  
mean_abs_dev(tes)
```

```
## [1] 0.999 -1.390 -1.128 0.933 0.213 0.855 -2.482 0.795 0.752 0.452
```