



# Diskretisasi

# Diskretisasi

- Automatic oleh aplikasi Data Mining (kita tidak bisa mengatur)
- Manual

# Diskretisasi Manual

- a. Same Interval
- b. Same Frequency (berdasarkan data yang kita miliki)
- c. Dengan Clustering (misal k-means clustering)

# Same Interval

- Bila diketahui data umur pengkredit motor sebagai berikut:

24 56 32 21 45 36 24 36 60 45 40 20  
22 32 62 34

Kita bisa membuat kelas data

A untuk interval 20-29

B untuk interval 30-39

C untuk interval 40-49

D untuk interval 50-59

E untuk interval 60-69

# Same Interval

- Dengan demikian data bisa kita kodekan sebagai
- Terlihat data agak skew untuk training karena kelas D dan E jumlahnya relatif sedikit dibanding dengan kelas A, B dan C

24 A

56 D

32 B

21 A

45 C

36 B

24 A

36 C

60 E

45 C

40 C

20 A

22 A

32 B

62 R

34 B

# Same Frequency

- Kita bisa membuat kelas yang tidak skew dengan menggunakan konsep same frekuensi.

- Pertama, data kita urutkan, menjadi

20 21 22 24 24 32 32 34 36 36 40 45 45 56 60 62

- Kemudian, kita bagi menjadi  $n$  bagian yang sama. Misal kita akan membagi menjadi 5 bagian dengan kode A,B,C,D dan E, maka kelompok datanya adalah

A 20 21 22

B 24 24 32 32

C 34 36 36

D 40 45 45

E 56 60 62