PERTEMUAN 3

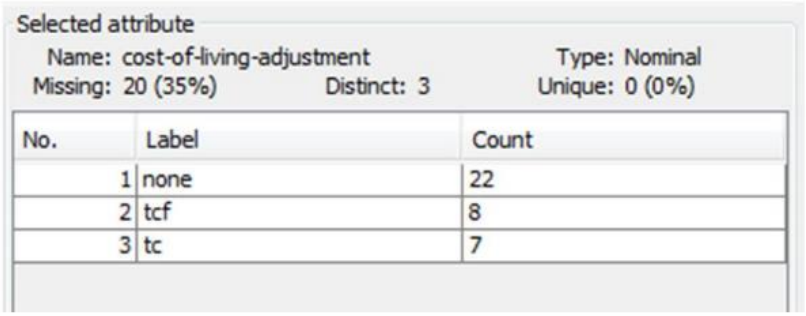
TEKNIK-TEKNIK PRAPROSES DATA MENGGUNAKAN WEKA

TUJUAN PRAKTIKUM

Mahasiswa akan dapat menggunakan teknik-teknik dalam praproses data meliputi pembersihan data, reduksi data, transformasi data dan diskretisasi data dengan tools Weka.

TEORI PENUNJANG

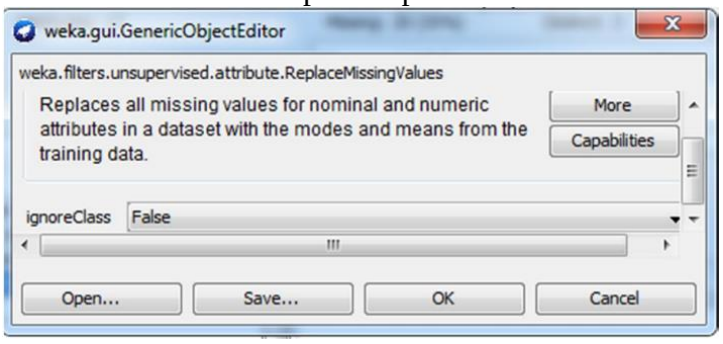
1. Pra-proses Data
   Pra-proses dilakukan karena dimungkinkan data set yang ada tidak lengkap, mengandung noise atau outlier, data tidak konsisten, atau ada data yang berulang. Tujuan penting dari pra-proses data adalah untuk meningkatkan kualitas data sehingga proses data mining juga menghasilkan pengetahuan baru yang lebih baik. Tugas utama dalam pra-proses data adalah pembersihan data, integrasi data, transformasi data, reduksi data, dan diskretisasi data.
2. Pembersihan Data
   a. Mengisi data nominal
      - Buka file labor.arff
      - Save as file menjadi labor2.arff
      - Buka kembali file labor2.arff
      - Perhatikan pada properties pada atribut cost-of-living-adjustment



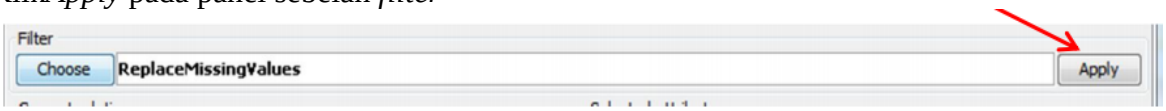
      - Pada filter, klik *Choose* > *filter* > *unsupervised* > *attribute* > *ReplaceMissingValue*
      - Klik pada panel sebelah *choose*

Klik,

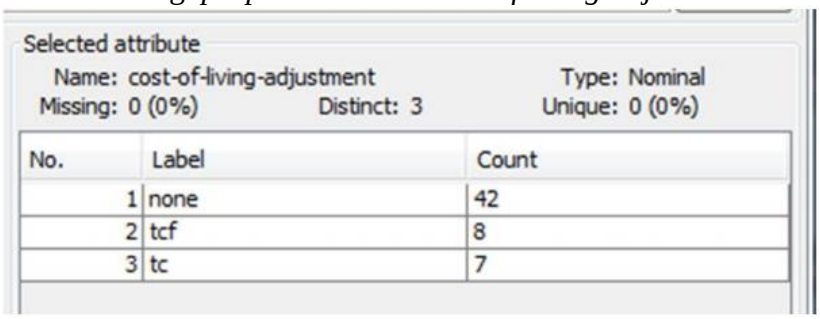
Maka akan muncul tampilan seperti berikut



- Klik OK
- Klik*Apply* pada panel sebelah *filter*

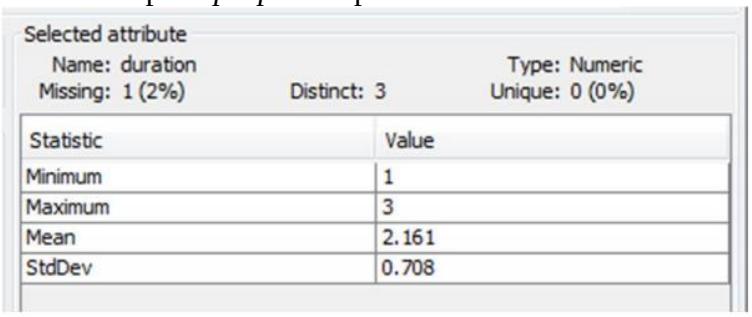


- Perhatikan lagi *properties* atribut *cost-of living-adjusment*

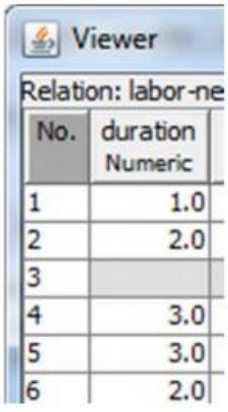


- *Missing value* pada atribut sudah hilang, dan jumlah data pada label none bertambah karena none merupakan modus dari data yang ada.

b. Mengisi data numerik
- Buka kembali file labor2.arff
- Perhatikan pada *properties* pada atribut *duration*

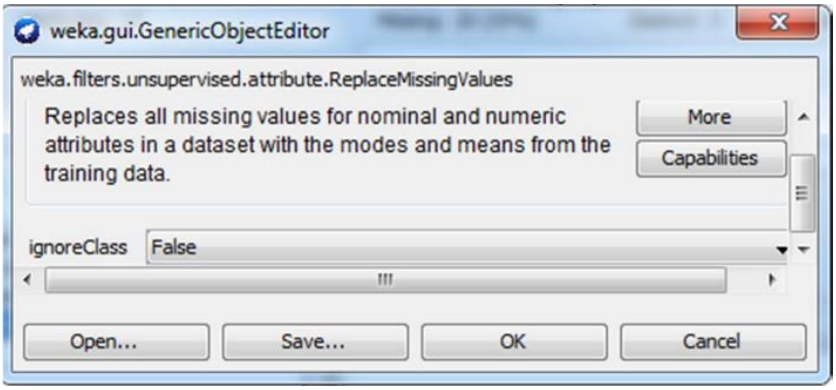


- Klik Edit untuk melihat posisi nilai yang hilang
- Perhatikan yang data *duration* yang nilainya kosong, kemudian tutup jendela viewer

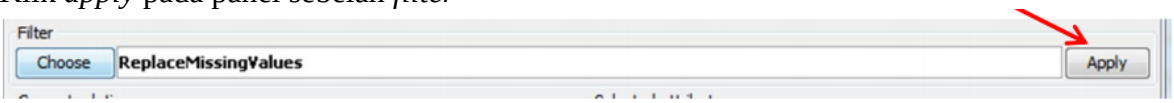


- Klik pada panel sebelah *choose*

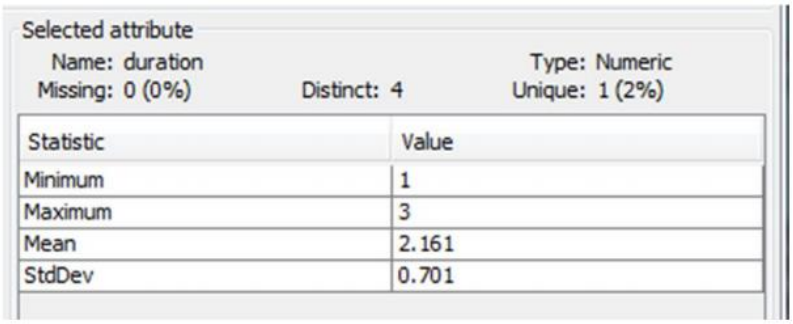
Klik, maka akan muncul tampilan seperti berikut



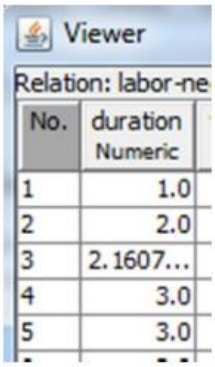
- Klik OK
- Klik *apply* pada panel sebelah *filter*



- Perhatikan lagi *properties* atribut *duration*



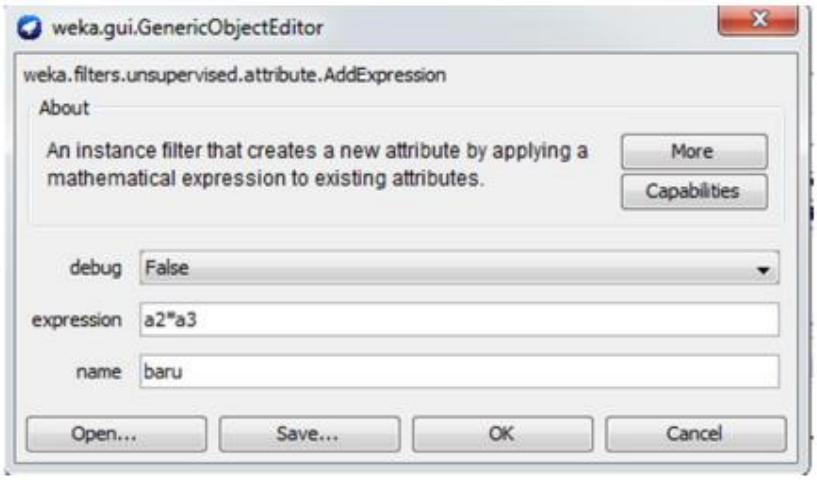
- Untuk melihat nilai yang mengisi data yang kosong, dapat dilakukan dengan mengeklik edit kembali
- Perhatikan kolom *duration* pada data ke-3



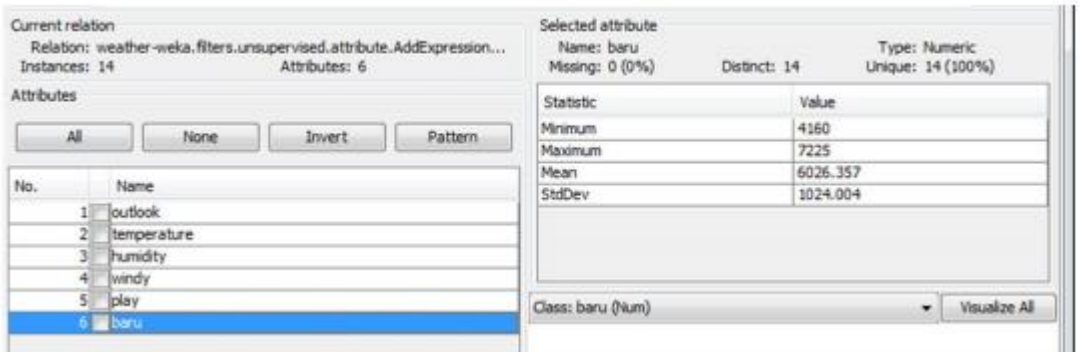
3. Integrasi Data
   Integrasi data biasanya tidak dilakukan pada Weka.

4. Transformasi Data
   a. Menambah atribut baru
      - Misal ingin menambah atribut baru yang merupakan hasil penjumlahan dari dua atribut yaitu temperature dan humidity pada data weather.numeric.arff
      - Buka file weather.numeric.arff
      - Pilih *Filter* > *Choose* > *filter* > *unsupervised* > *attribute* > *addExpression*
      - Klik panel disebelah *Choose*

dan akan muncul tampilan seperti berikut

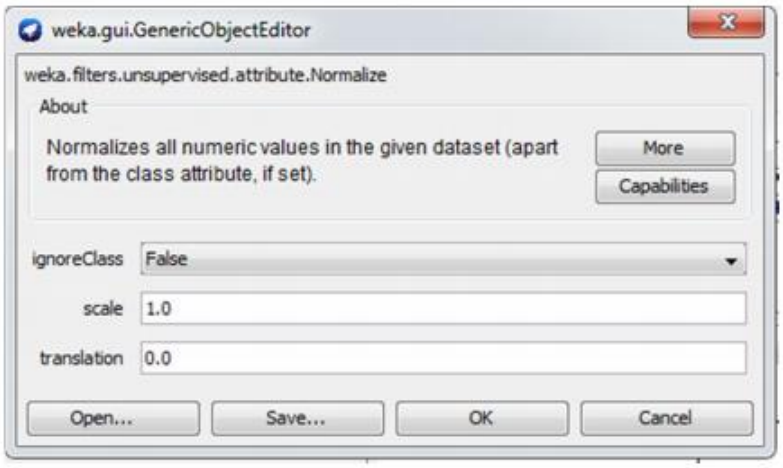


- Ganti expression menjadi sesuai kebutuhan dan cantumkan nama expression pada *name field*, klik OK
- Pilih *Apply*
- Perhatikan pada panel atribut, atribut pada weather.numeric.arff telah bertambah sesuai yang ditambahkan.

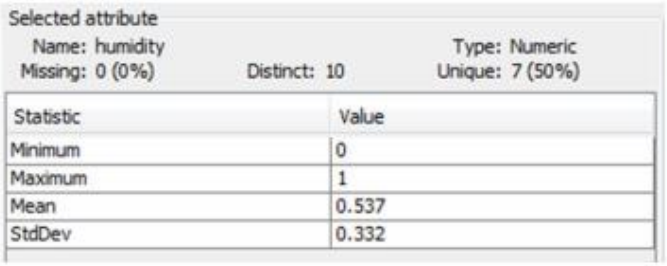


b. Melakukan normalisasi pada data set
   - Buka file weather.numeric.arff
   - Pilih *filter* > *Choose* > *filter* > *unsupervised* > *normalize*
   - Klik pada panel sebelah *Choose*

dan akan muncul tampilan seperti berikut

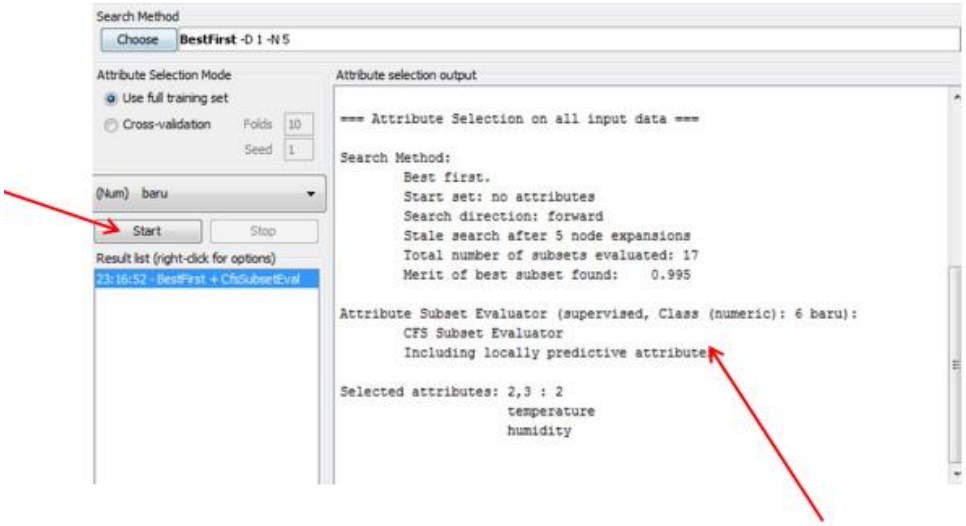


- Klik OK > Klik *Apply*
- Perhatikan pada atribut *temperature* dan *humidity*, selang nilai maksimumnya adalah satu karena sebelumnya skala di set 1.0

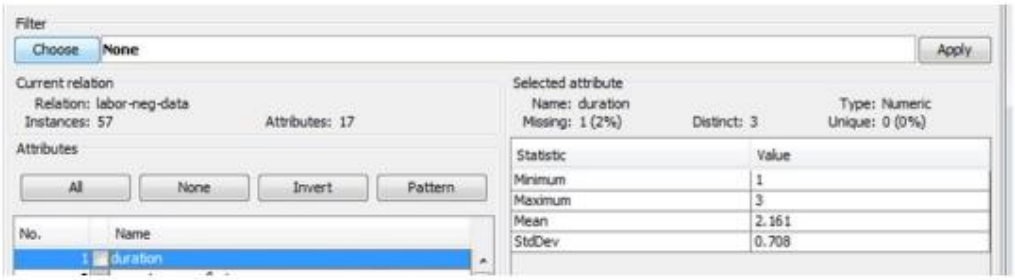


5. Reduksi Data
   a. Mereduksi dimensi data
      - Buka file labor2.arff
      - Pilih tab *Select Attributes*
      - Pada panel *Search Method*, Klik *Choose* > Pilih salah satu teknik yang akan dipakai (contoh : *Best First*)

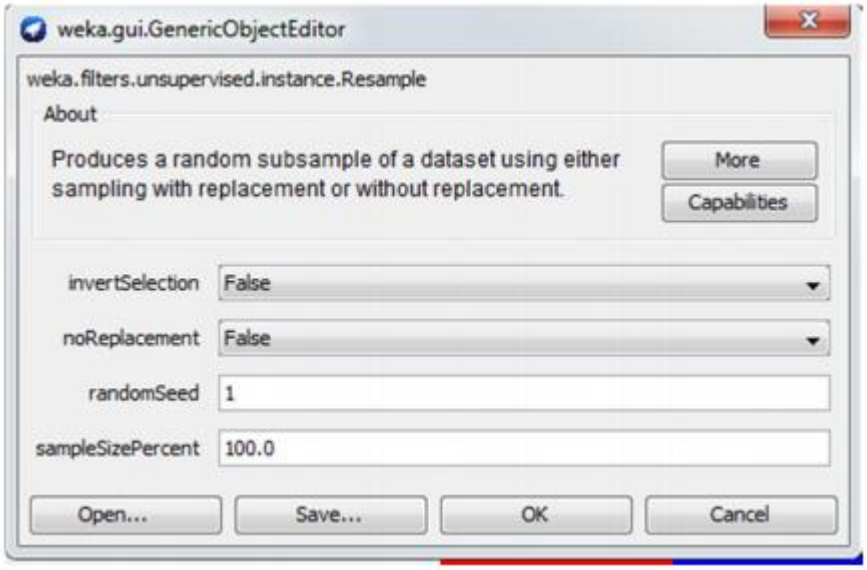
      - Setelah dipilih, Klik *Start*. Hasil seleksi atribut ada pada panel sebelah kanannya.



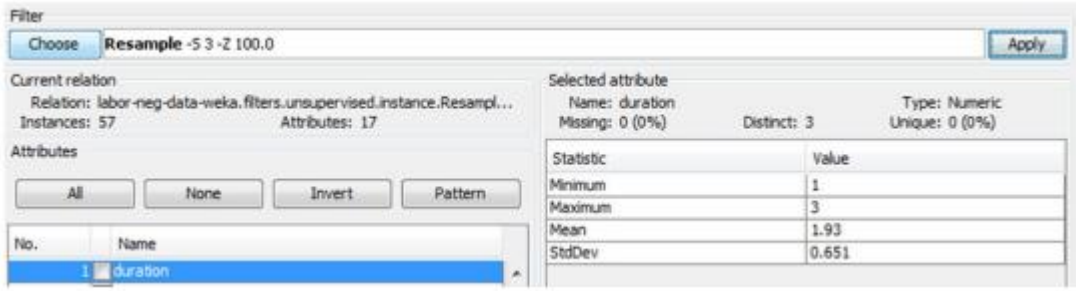
b. Mereduksi jumlah data
   - Buka file labor2.arff
   - Perhatikan *properties* dari data dan atribut *duration*



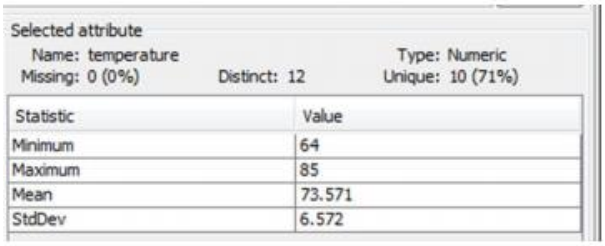
   - Pilih *filter* > *Choose* > *filter* > *unsupervised* > *instance* > *Resample*
   - Klik pada panel sebelah *Choose* dan akan muncul tampilan seperti berikut



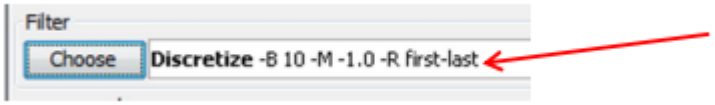
   - Klik OK, Klik *Apply*
   - Perhatikan *properties* dari data dan atribut *duration*



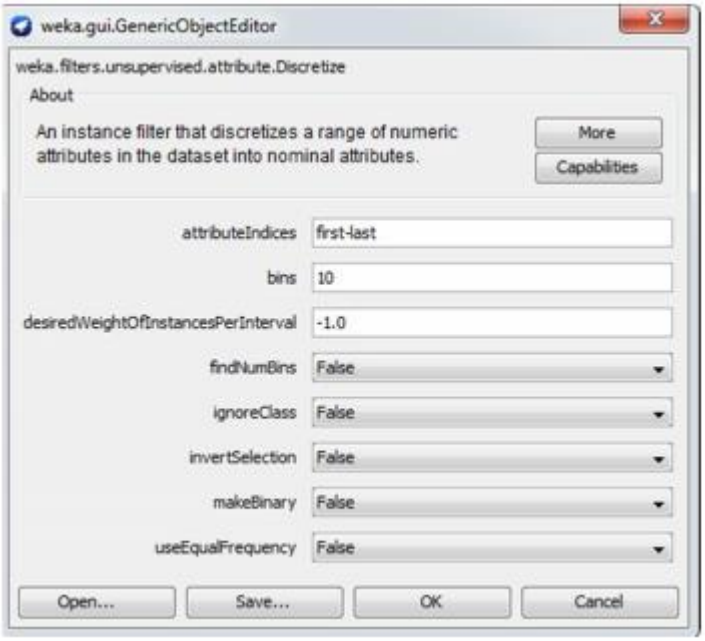
6. Diskretisasi Data
   - Buka file weather.numeric.arff
   - Perhatikan pada atribut *temperature*



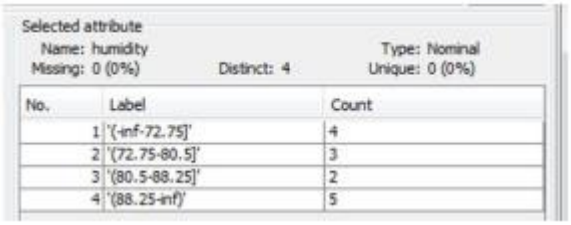
- Pilih *Filter* > *Choose* > *Unsupervised* > *attribute* > *Discretize*
- Klik panel di sebelah *Choose*



- Klik dan akan muncul tampilan seperti berikut



- Ubah *bins* untuk menyesuaikan jumlah selang yang diinginkan, klik OK
- Klik *Apply*, dan perhatikan kembali properties atribut *temperature*



- Tipe data atribut *temperature* telah berubah menjadi nominal dengan selang tertentu.

LAPORAN PENDAHULUAN

1. Apakah semua data dengan tipe.csv atau .arff dapat diproses untuk data mining
2. dengan menggunakan Weka?
3. Mengapa perlu dilakukan praproses data?
4. Apakah setiap akan melakukan tahapan data mining harus dilakukan praproses data?
5. Apakah semua teknik praproses data harus dilakukan ketika melakukan praproses data?

MATERI PRAKTIKUM

1. Praproses Data
2. Pembersihan Data
3. Integrasi Data
4. Transformasi Data

5. Reduksi Data
6. Diskretisasi Data

DAFTAR PUSTAKA

Bouckaert, R. R.. et al. 2013. *WEKA Manual for Version 3-6-9*. Edition of January 21, 2013. http://jaist.dl.sourceforge.net/project/weka/documentation/3.6.x/ WekaManual-3-6- 9.pdf . Accessed on 27 January 2013.