Tugas 4 Data Mining Pengukuran Jarak dalam Ruang Vector dan Metode Klasifikasi K-Nearest Neighbor (K-NN)

disusun untuk memenuhi
tugas mata kuliah Data Mining

oleh :

Khairul Umam Albi
2008107010072

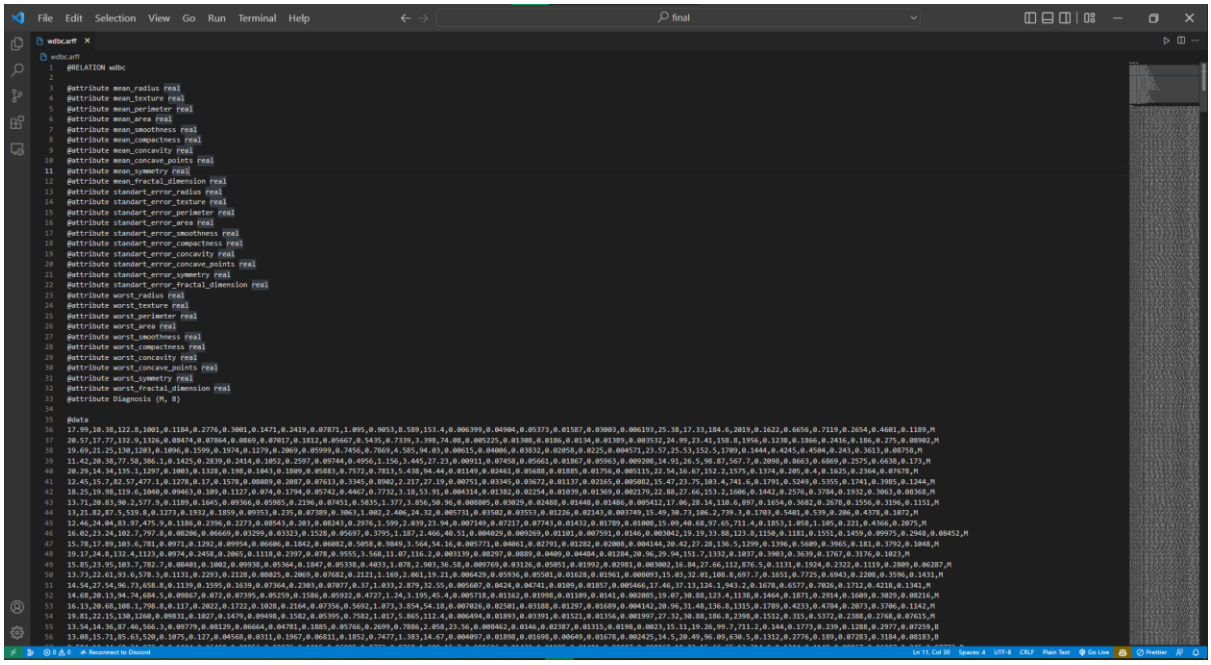
JURUSAN INFORMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS SYIAH KUALA
DARUSSALAM, BANDA ACEH
2022

1. Dataset WDBC

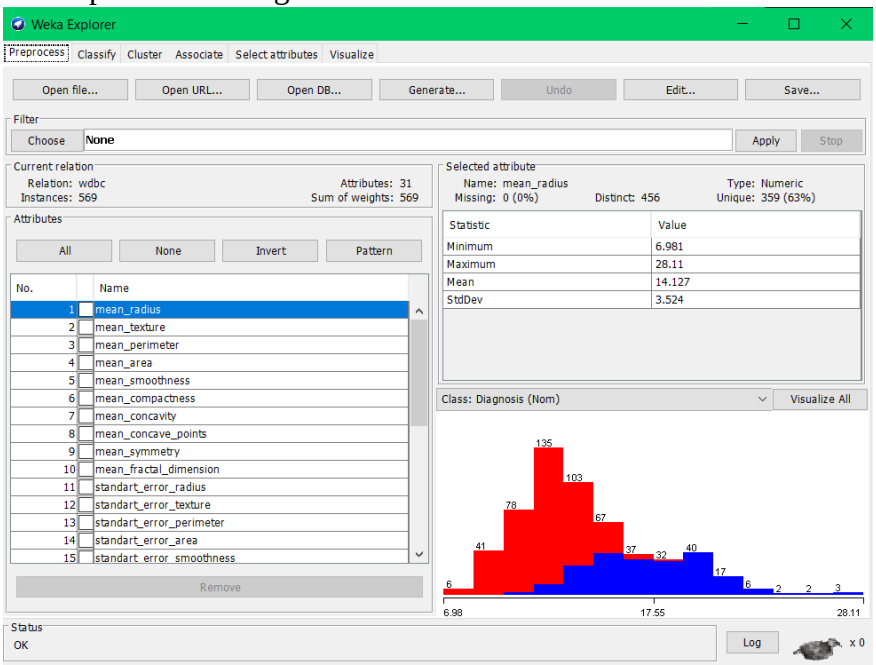
   Teknik mengubah data dari ektensi lain menjadi data yang bertipe arff

   - Hilangkan atribut ID sehingga tidak menjadi atribut dalam file ARFF (cek dalam dataset asli pada urutan berapa kolom ini berada).
   - Letakkan class label (M atau B) pada kolom terakhir.
   - Susun atribut lainnya dalam file ARFF sesuai urutan atribut pada data asli, hanya saja dalam file ARFF attribut ID dihilangkan dan atribut class label berada pada kolom terakhir.

   a. File arff



   b. data input oleh Perangkat Lunak Weka

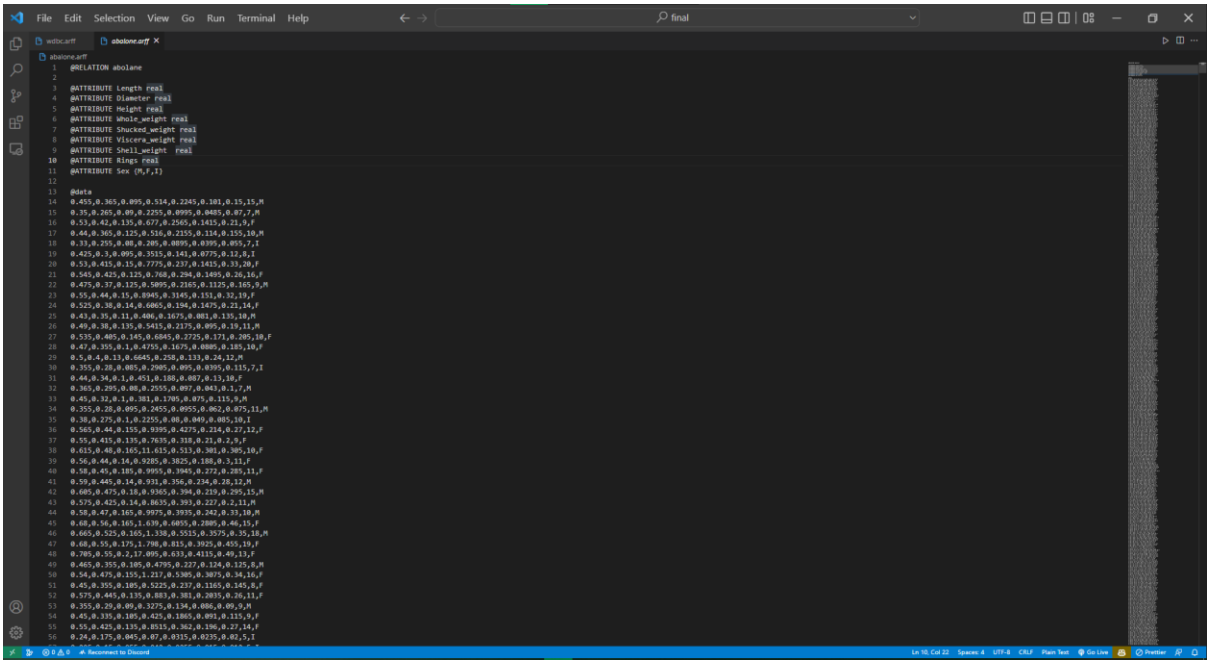


2. Dataset Abalone

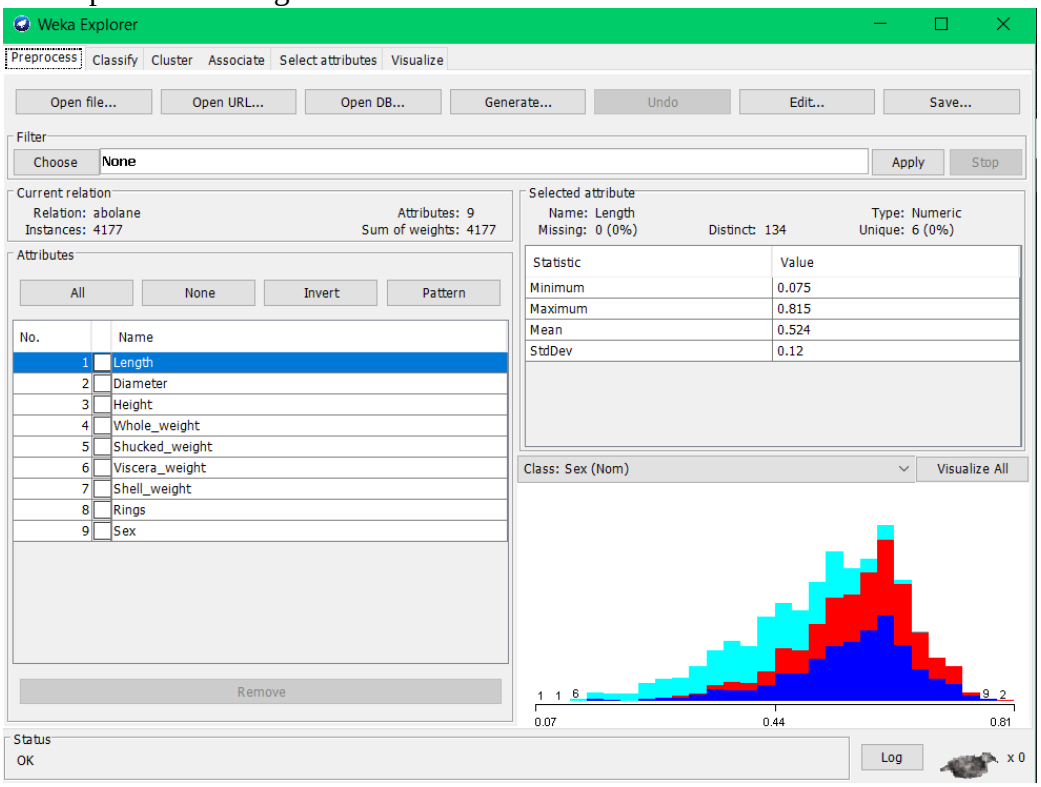
Teknik dalam mengubah data ke dalam bentuk arff

- Letakkan atribut nomimal (jenis kelamin) pada kolom terakhir.
- Selanjutnya, susun atribut dalam file ARFF dalam urutan sebagai berikut: Diameter, Rings, Height, Length, Shell weight, Shucked weight, Whole weight, Viscera weight.

a. File arff

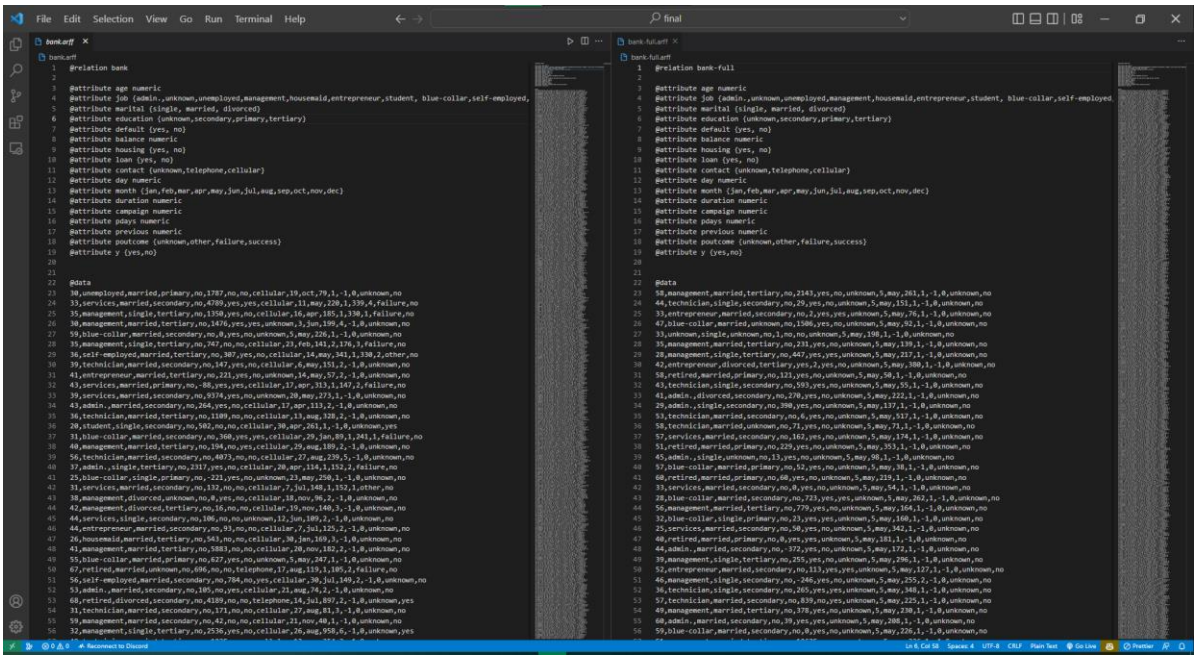


b. data input oleh Perangkat Lunak Weka



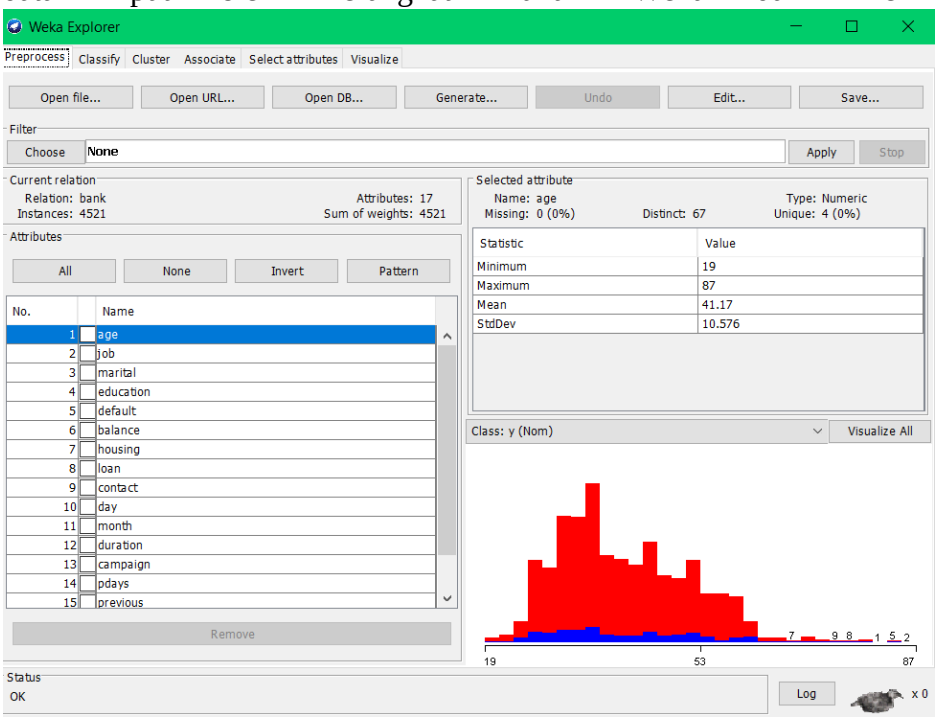
3. Dataset Bank Marketing
   Teknik mengubah data ke dalam bentuk arff
   - Susun atribut dalam file ARFF dalam urutan sebagaimana aslinya.
   - Ubah kedua file csv (bank.csv dan bank-full.csv) ke dalam format ARFF.

   a. File arff



   Note : kiri merupakan data arff dari nama file bank.arff dan sebelah kanan dengan nama bank-full.arff

   b. data input oleh Perangkat Lunak Weka dari file bank.arff



c. data input oleh Perangkat Lunak Weka dari file bank-full.arff