LAPORAN PRAKTIKUM DATA MINING

WEKA

(Waikato Environtment for Knowledge Analysis)

Disusun oleh:

Nama : Sevty Ayu Saputry

NPM :1808107010093

JURUSAN INFORMATIKA

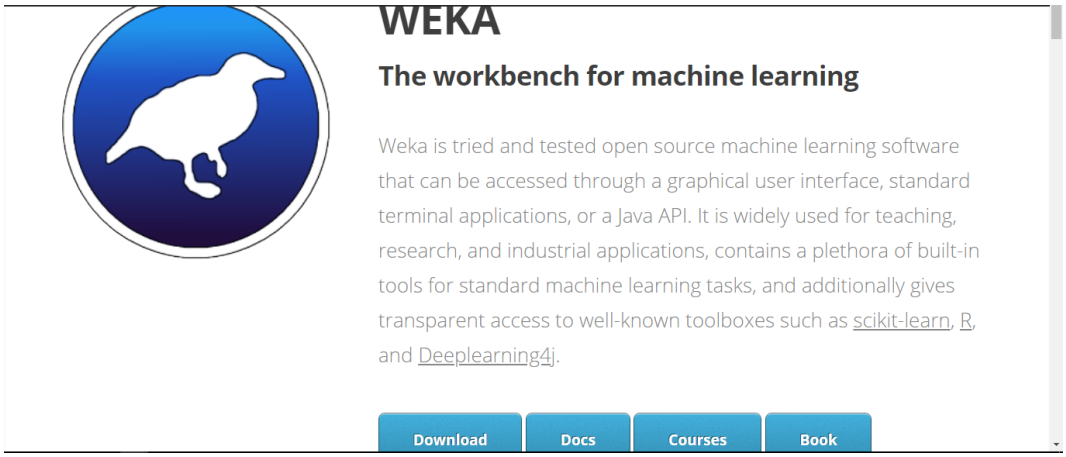
FAKULTAS MIPA

UNIVERSITAS SYIAH KUALA

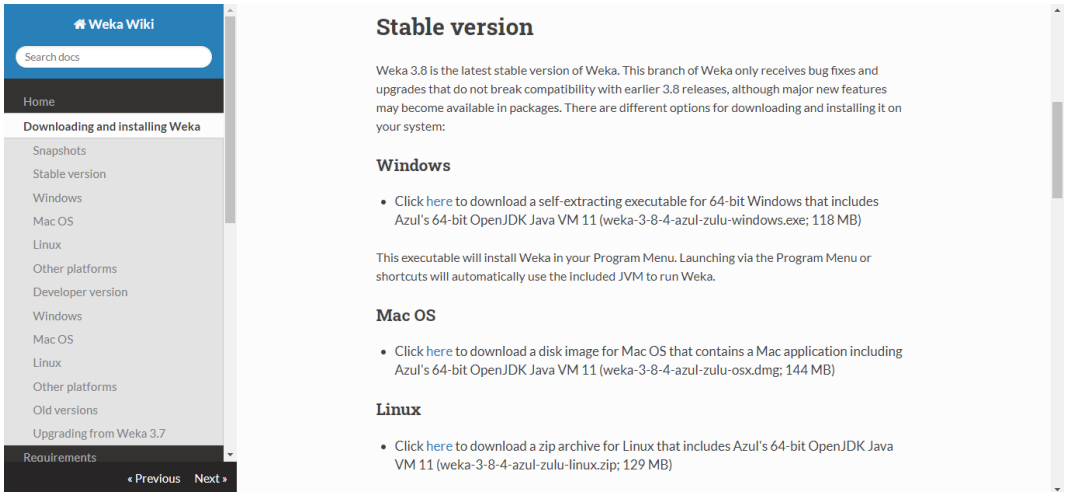
2020

Mengubah abalone dan wdbc ke format ARFF

1. Mengunduh aplikasi WEKA 3.8
   https://www.cs.waikato.ac.nz/ml/weka/

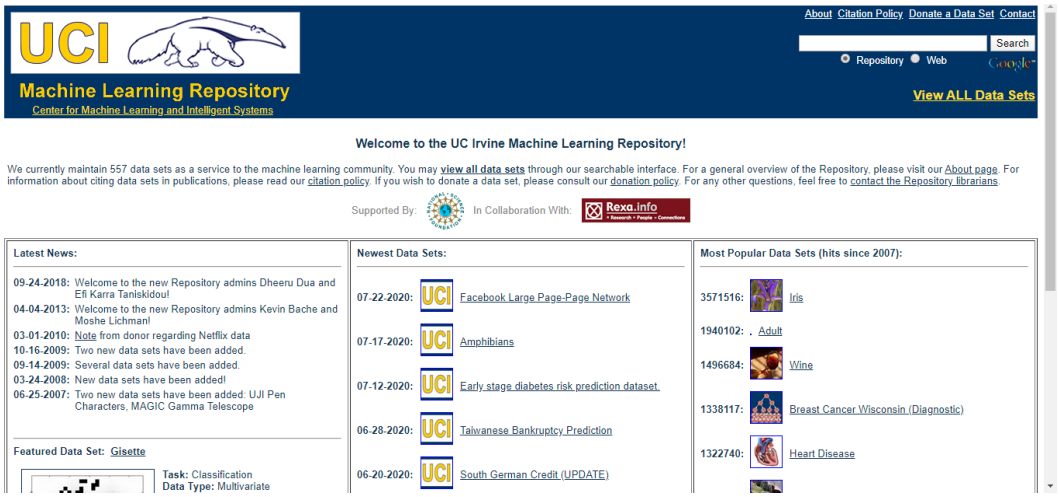


   - Klik “Download”

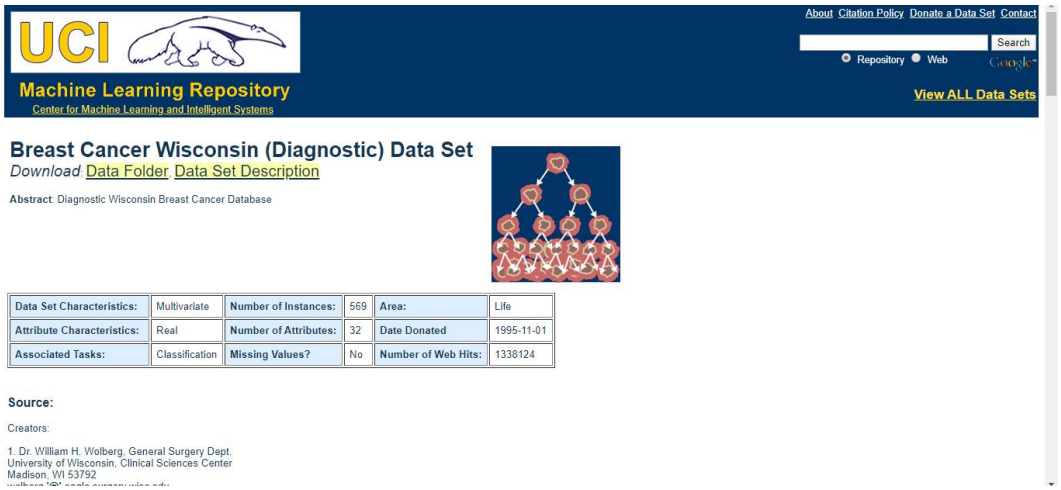


   - Pilih stable version disesuaikan dengan sistem operasi yang ada di laptop
   - WEKA 3.8 berhasil diunduh

2. Mengunduh datasets
   http://archive.ics.uci.edu/ml/index.php



- Pilih *Breast Cance Wisconsin (Diagnostic)* dan *Abalone*



- Pilih “Data Folder”

Index of /ml/machine-learning-databases/breast-cancer-wisconsin

- Parent Directory
- Index
- breast-cancer-wisconsin.data
- breast-cancer-wisconsin.names
- unformatted-data
- wdbc.data
- wdbc.names
- wpbc.data
- wpbc.names

Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips SVN/1.7.14 Phusion_Passenger/4.0.53 mod_perl/2.0.11 Perl/v5.16.3 Server at archive.ics.uci.edu Port 80

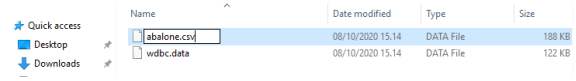
Index of /ml/machine-learning-databases/abalone

- Parent Directory
- Index
- abalone.data
- abalone.names

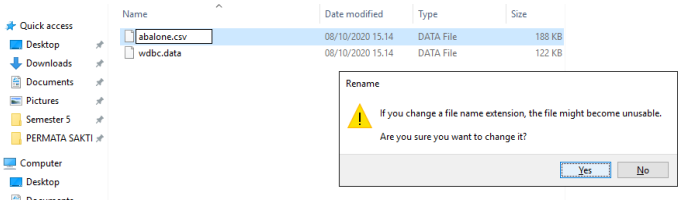
Apache/2.4.6 (CentOS) OpenSSL/1.0.2k-fips SVN/1.7.14 Phusion_Passenger/4.0.53 mod_perl/2.0.11 Perl/v5.16.3 Server at archive.ics.uci.edu Port 80

- Pilih ”wdbc.data” dan “abalone.data” untuk mengunduh

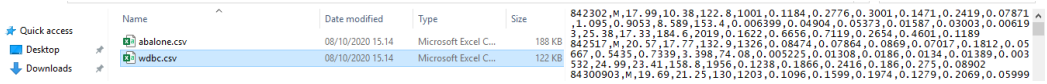
3. Mengubah ke format .csv
   - *rename* file dengan format .csv



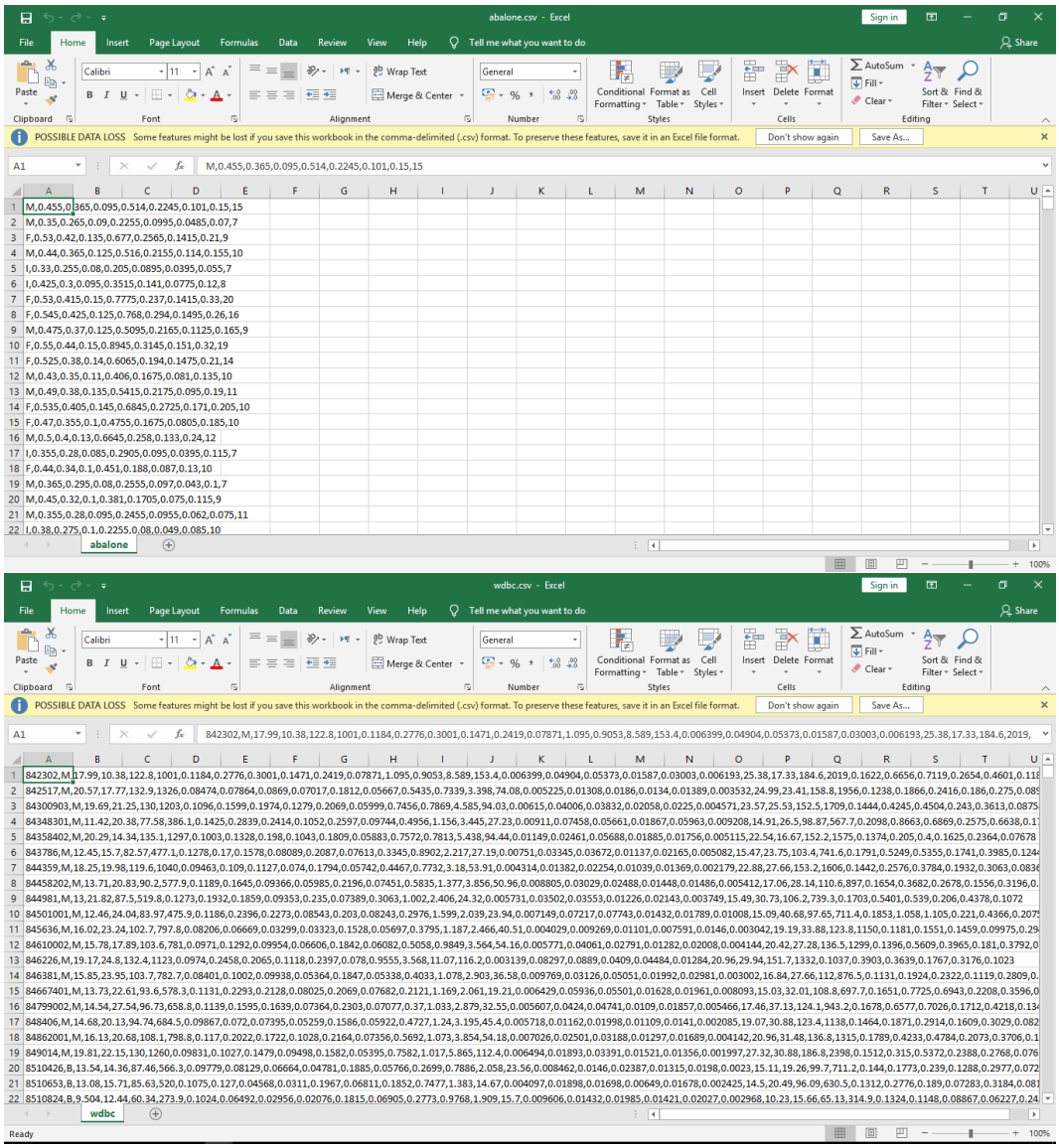
   - klik “yes”



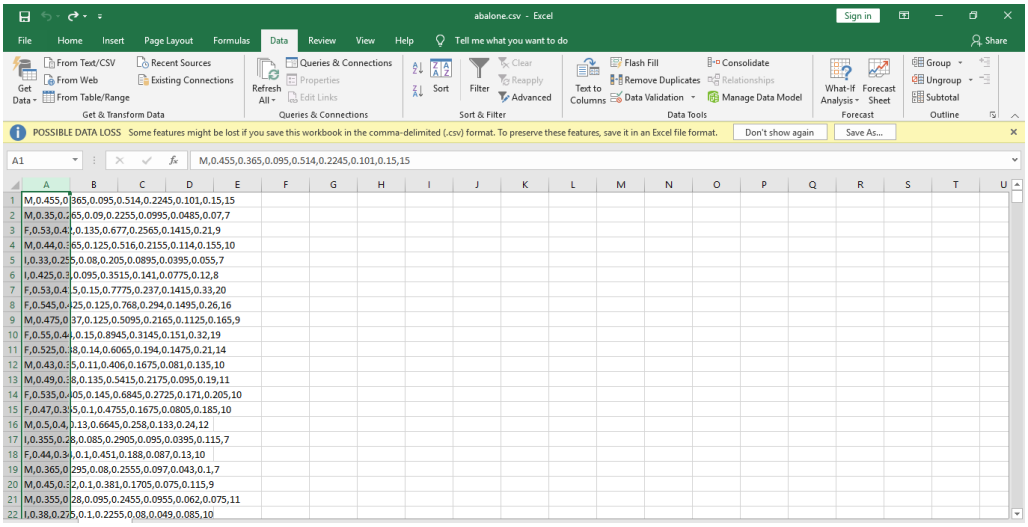
   - file akan berubah menjadi seperti tampilan di bawah



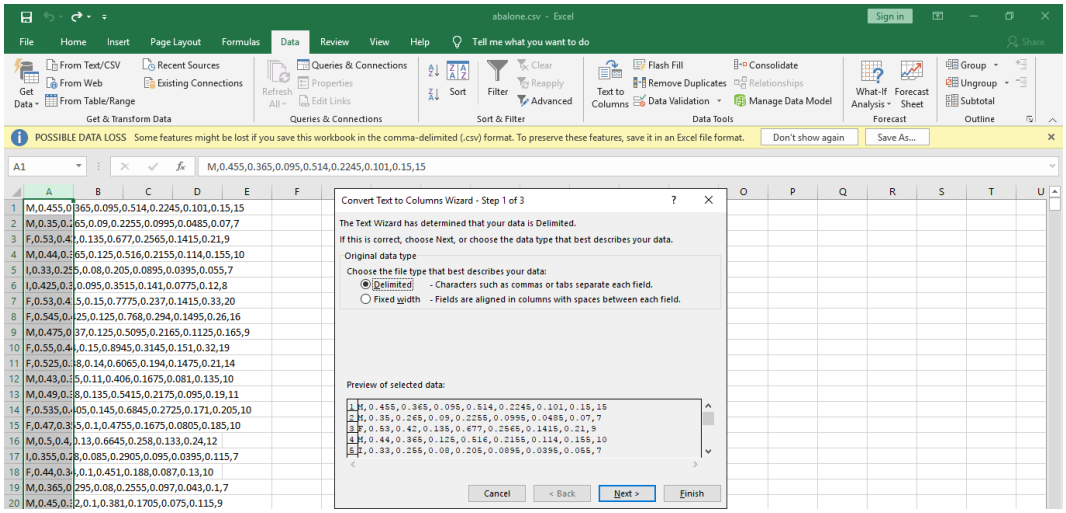
4. Edit file format .csv
   - buka file abalone.csv dan wdbc.csv



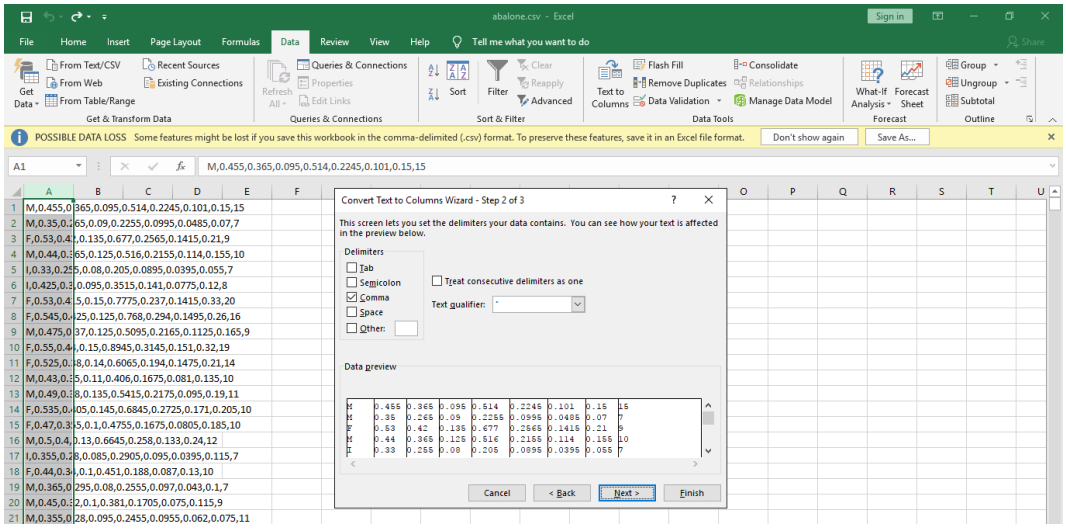
   - Blok seluruh kolom



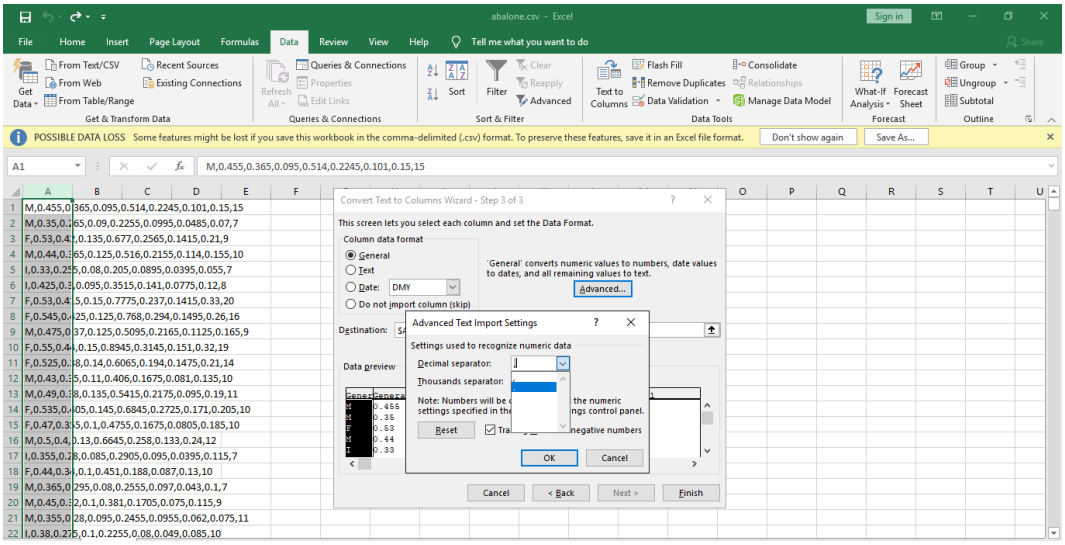
- Klik Data>Text to Column



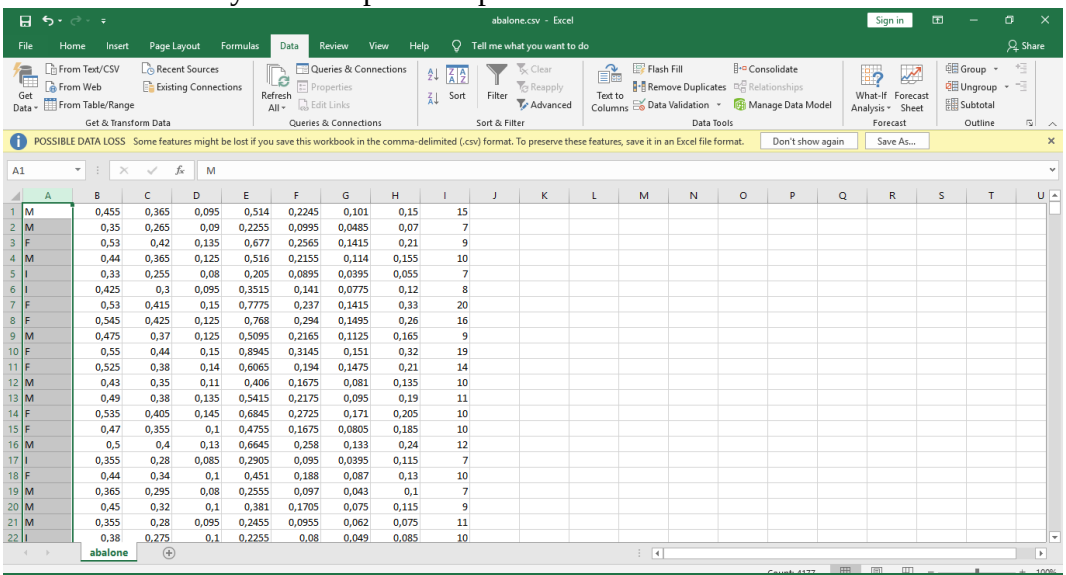
- Klik next, pilih "*comma*", klik next

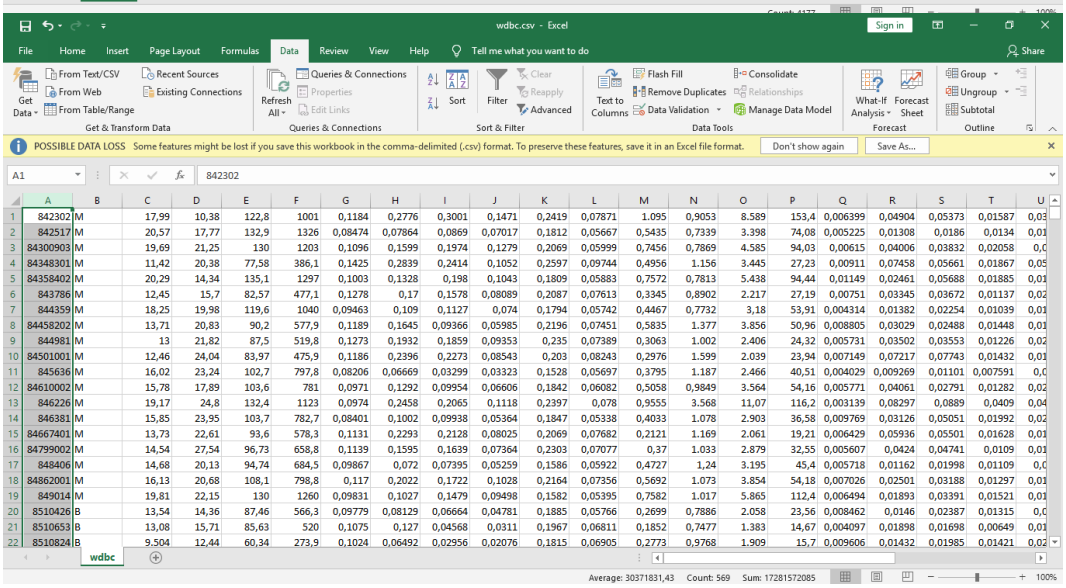


- Tampilan akan seperti di bawah ini, kemudian klik *advance*, pilih *decimal separate* tanda ".", klik oke, dan *finish*



- Hasilnya akan seperti tampilan di bawah





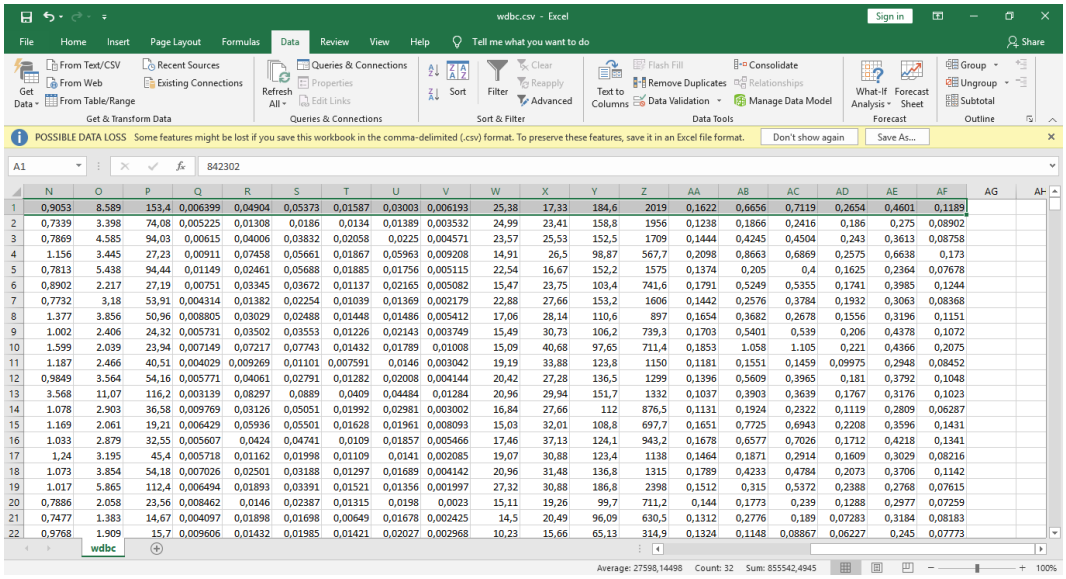
5. Mengedit kolom berdasarkan soal
   a. WDBC

      Secara garis besar, dataset WSDL terdiri atas 569 sampel. Setiap sampel memiliki 32 atribut (ID, diagnosis, 30 fitur nilai real). Atribut **Diagnosis** adalah class label dari dataset WDBC dengan nilai (M = malignant, B = benign). Sepuluh atribut pertama (kolom 3 sd 12) merupakan nilai "mean" dari setiap pengamatan pada sel nucleus, sepuluh atribut yang kedua (kolom 13 sd 22) merupakan nilai "standard error" (SE) dari setiap pengamatan pada sel nucleus, dan sepuluh atribut yang terakhir (kolom 23 sd 32) merupakan nilai "worth" dari setiap pengamatan pada sel nucleus.

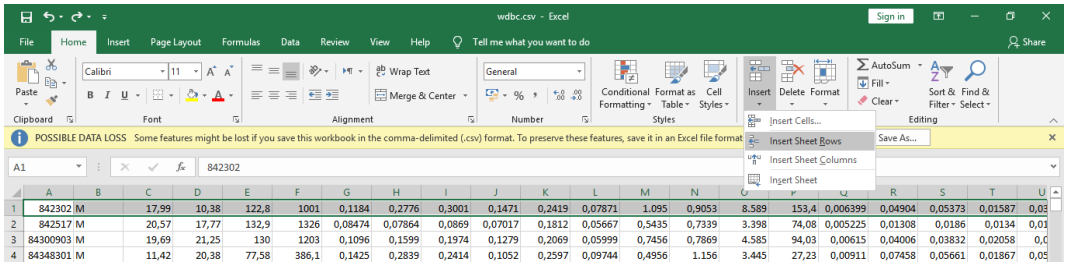
      Dataset WDBC

      - Hilangkan atribut ID sehingga tidak menjadi atribut dalam file ARFF (dalam dataset asli berada pada kolom 1).
      - Letakkan class label (M atau B) pada kolom terakhir.
      - Susun atribut 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.

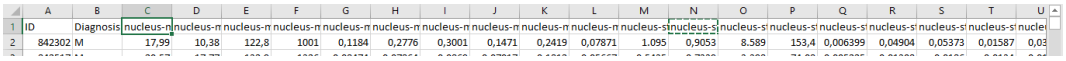
      - Blok baris pertama



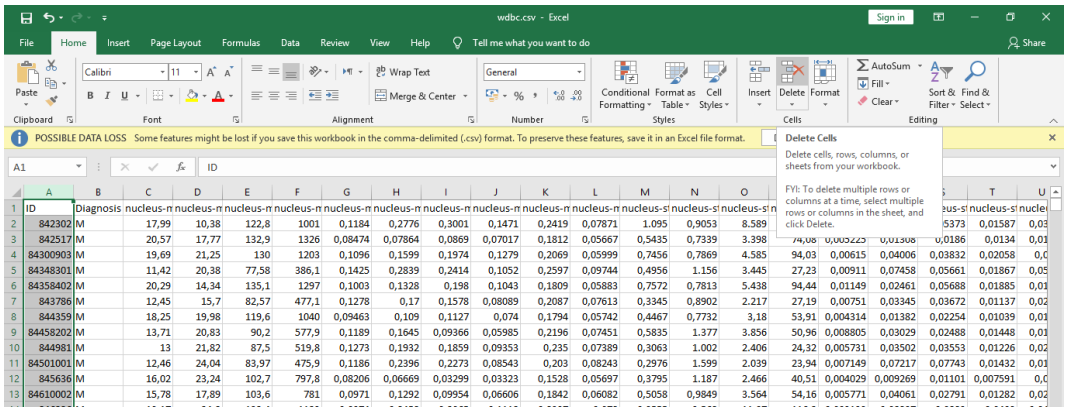
- Klik *Insert Sheet Rows*



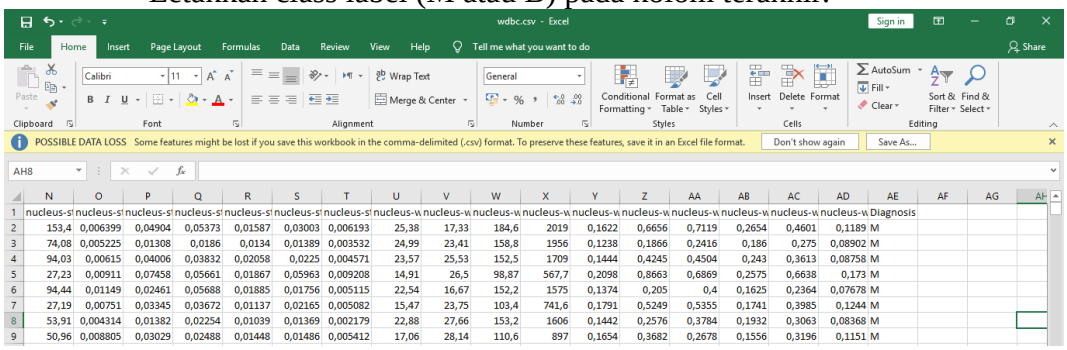
- Isi baris pertama sesuai nama setiap kolom



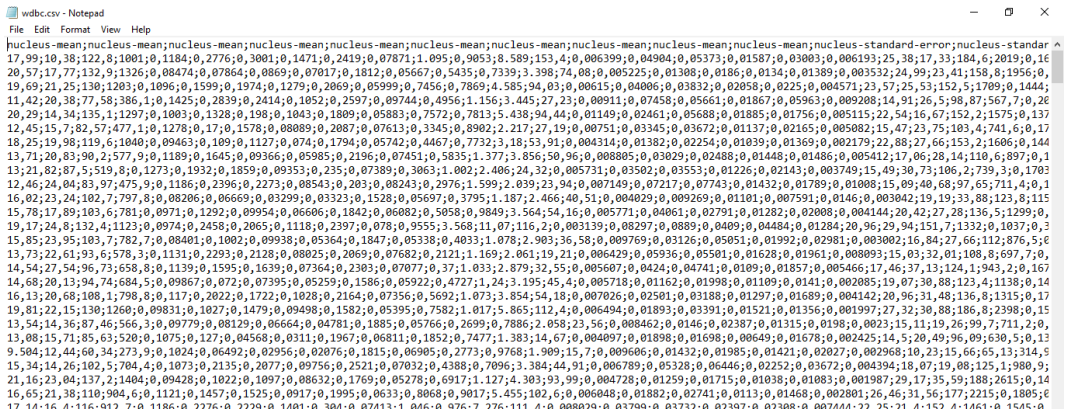
- Hilangkan atribut ID sehingga tidak menjadi atribut dalam file ARFF (dalam dataset asli berada pada kolom 1). Dengan cara blok kolom ID klik *delete*



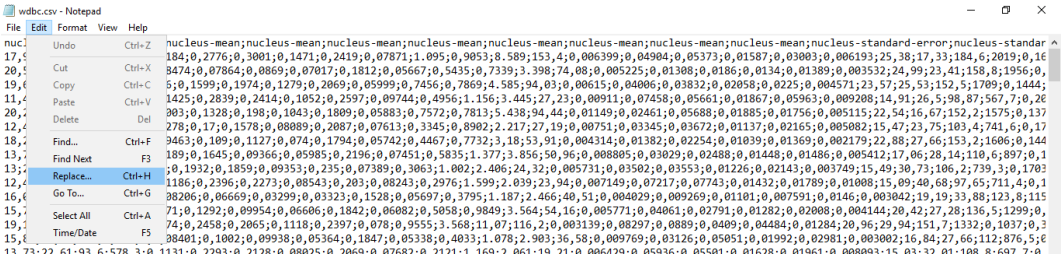
- Letakkan class label (M atau B) pada kolom terakhir.



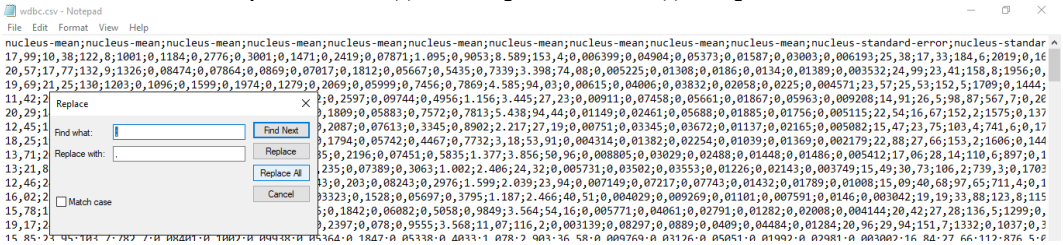
- Simpan file yang sudah diubah
- Kemudian buka dengan aplikasi *notepad*



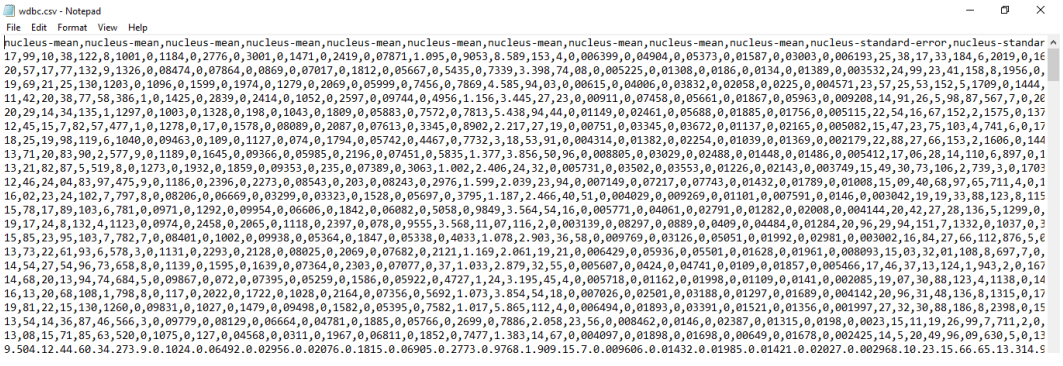
- klik edit>replace

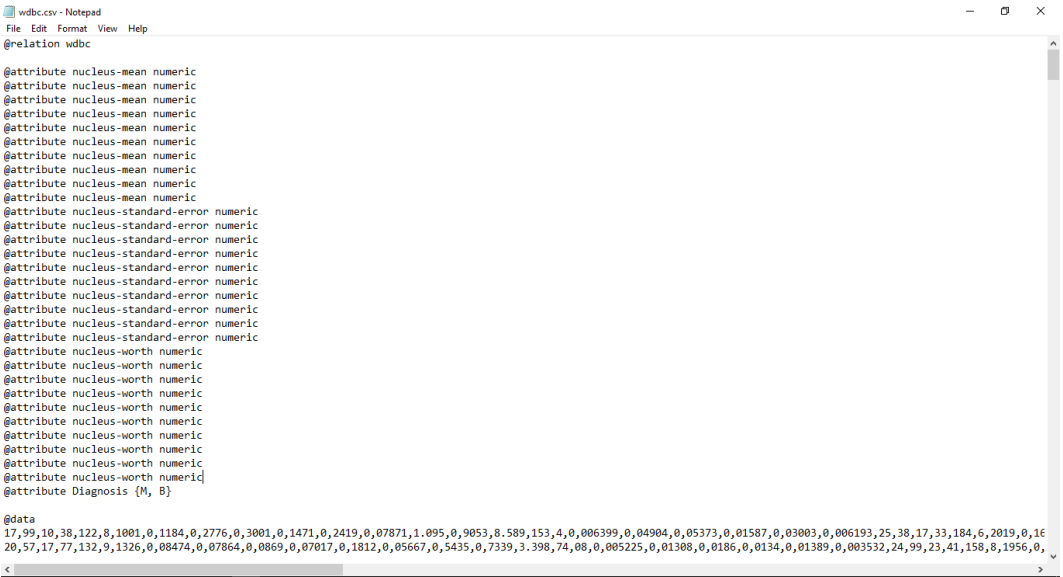


- ketik di *find what*: (;) dan *replace with*: (,)>*Replace All*



- hasil akhirnya akan terlihat seperti tampilan di bawah ini





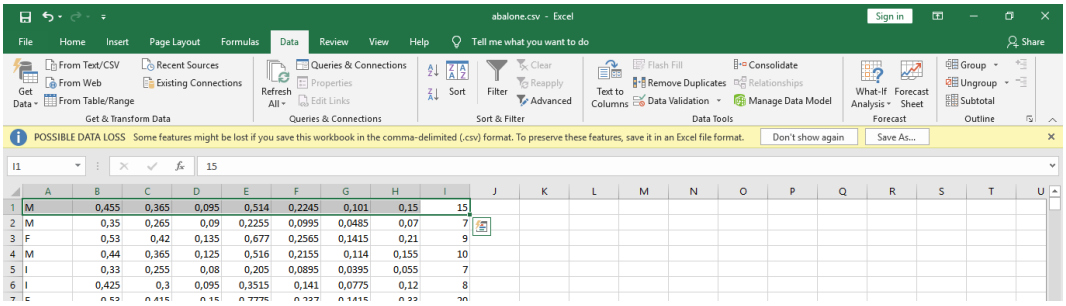
b. Abalone

Dataset Abalone memiliki 4177 sampel dan terdiri dari 8 atribut. Tidak ditemukan "missing value" pada dataset ini. Detail tentang dataset ini juga dapat dilihat pada tautan

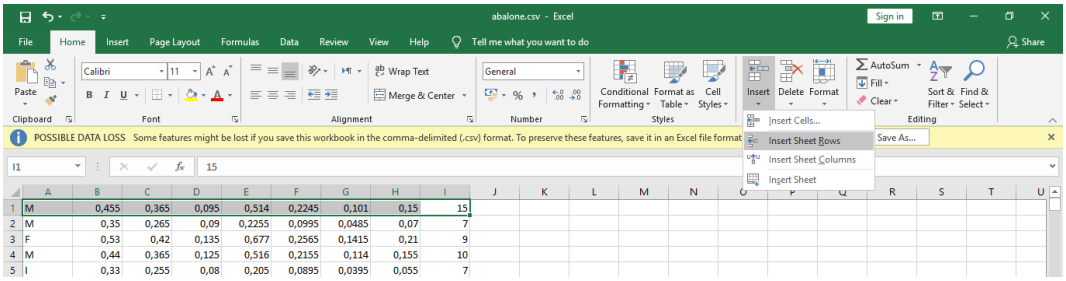
Data Set Description.

- 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.

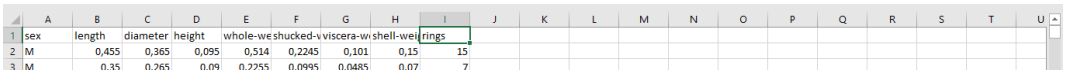
- Blok baris pertama



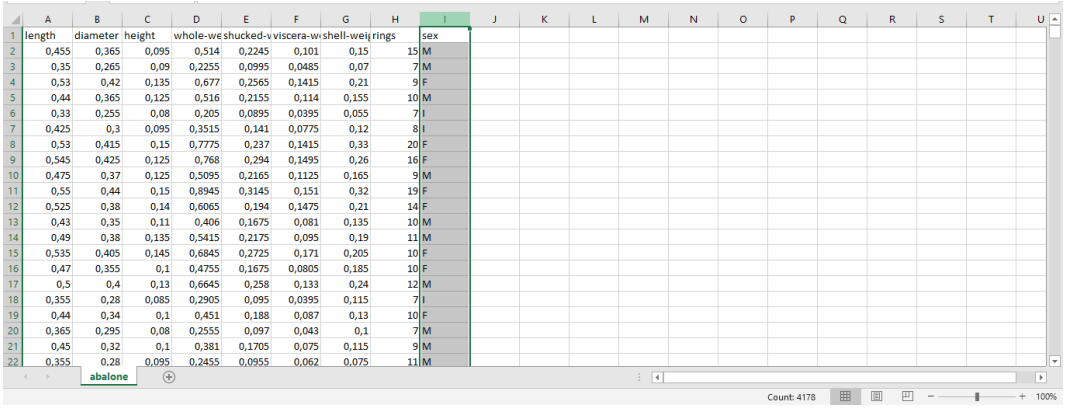
- Klik *Insert Sheet Rows*



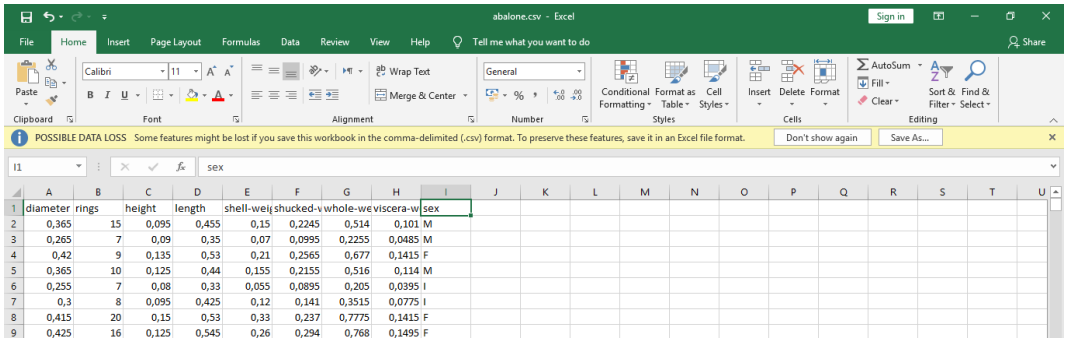
- Isi baris pertama sesuai nama setiap kolom



- Letakkan atribut nomimal (jenis kelamin) pada kolom terakhir.

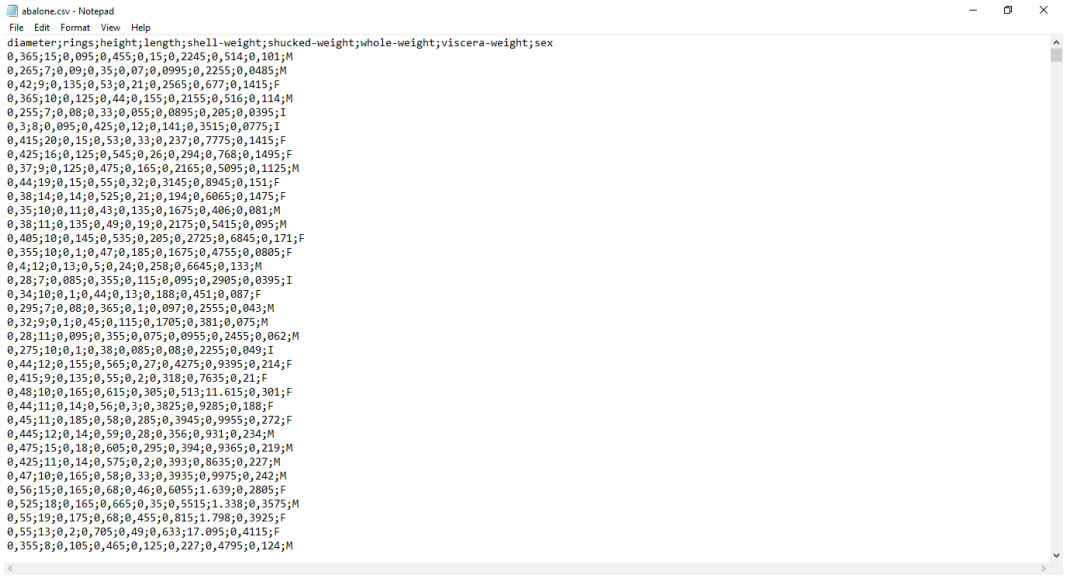


- Susun atribut dalam file ARFF dalam urutan sebagai berikut: Diameter, Rings, Height, Length, Shell weight, Shucked weight, Whole weight, Viscera weight.

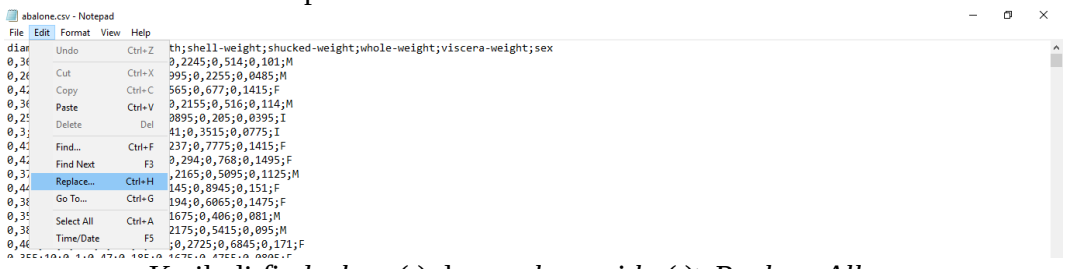


diameter	rings	height	length	shell-wei	shucked-v	whole-we	viscera-w	sex
0,365	15	0,095	0,455	0,15	0,2245	0,514	0,101	M
0,265	7	0,09	0,35	0,07	0,0995	0,2255	0,0485	M
0,42	9	0,135	0,53	0,21	0,2565	0,677	0,1415	F
0,365	10	0,125	0,44	0,155	0,2155	0,516	0,114	M
0,255	7	0,08	0,33	0,055	0,0895	0,205	0,0395	I
0,3	8	0,095	0,425	0,12	0,141	0,3515	0,0775	I
0,415	20	0,15	0,53	0,33	0,237	0,7775	0,1415	F
0,425	16	0,125	0,545	0,26	0,294	0,768	0,1495	F

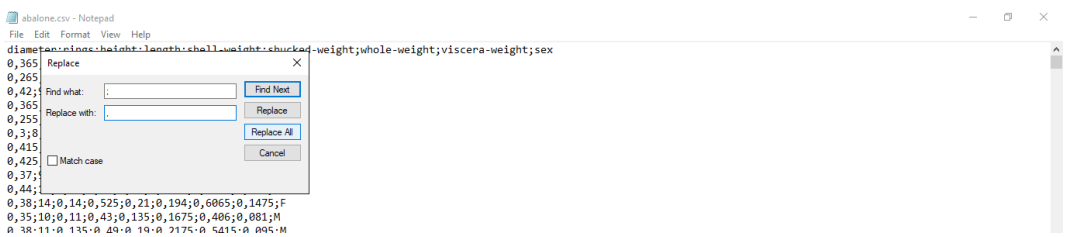
- Simpan file yang sudah diubah
- Kemudian buka dengan aplikasi *notepad*



- Klik edit>replace



- Ketik di *find what*: (;) dan *replace with*: (,)>*Replace All*



- Hasil akhirnya akan terlihat seperti tampilan di bawah ini

abalone.csv - Notepad

File Edit Format View Help

```
diameter,rings,height,length,shell-weight,shucked-weight,whole-weight,viscera-weight,sex
0,365,15,0,095,0,455,0,15,0,2245,0,514,0,101,M
0,265,7,0,09,0,35,0,07,0,0995,0,2255,0,0485,M
0,42,9,0,135,0,53,0,21,0,2565,0,677,0,1415,F
0,365,10,0,125,0,44,0,155,0,2155,0,516,0,114,M
0,255,7,0,08,0,33,0,055,0,0895,0,205,0,0395,I
0,3,8,0,095,0,425,0,12,0,141,0,3515,0,0775,I
0,415,20,0,15,0,53,0,33,0,237,0,7775,0,1415,F
0,425,16,0,125,0,545,0,26,0,294,0,768,0,1495,F
0,37,9,0,125,0,475,0,165,0,2165,0,5095,0,1125,M
0,44,19,0,15,0,55,0,32,0,3145,0,8945,0,151,F
0,38,14,0,14,0,525,0,21,0,194,0,6065,0,1475,F
0,35,10,0,11,0,43,0,135,0,1675,0,406,0,081,M
0,38,11,0,135,0,49,0,19,0,2175,0,5415,0,095,M
0,405,10,0,145,0,535,0,205,0,2725,0,6845,0,171,F
0,355,10,0,1,0,47,0,185,0,1675,0,4755,0,0805,F
0,4,12,0,13,0,5,0,24,0,258,0,6645,0,133,M
0,28,7,0,085,0,355,0,115,0,095,0,2905,0,0395,I
0,34,10,0,1,0,44,0,13,0,188,0,451,0,087,F
0,295,7,0,08,0,365,0,1,0,097,0,2555,0,043,M
0,32,9,0,1,0,45,0,115,0,1705,0,381,0,075,M
0,28,11,0,095,0,355,0,075,0,0955,0,2455,0,062,M
0,275,10,0,1,0,38,0,085,0,08,0,2255,0,049,I
0,44,12,0,155,0,565,0,27,0,4275,0,9395,0,214,F
0,415,9,0,135,0,55,0,2,0,318,0,7635,0,21,F
0,48,10,0,165,0,615,0,305,0,513,11.615,0,301,F
0,44,11,0,14,0,56,0,3,0,3825,0,9285,0,188,F
0,45,11,0,185,0,58,0,285,0,3945,0,9955,0,272,F
0,445,12,0,14,0,59,0,28,0,356,0,931,0,234,M
0,475,15,0,18,0,605,0,295,0,394,0,9365,0,219,M
0,425,11,0,14,0,575,0,2,0,393,0,8635,0,227,M
0,47,10,0,165,0,58,0,33,0,3935,0,9975,0,242,M
0,56,15,0,165,0,68,0,46,0,6055,1.639,0,2805,F
0,525,18,0,165,0,665,0,35,0,5515,1.338,0,3575,M
0,55,19,0,175,0,68,0,455,0,815,1.798,0,3925,F
0,55,13,0,2,0,705,0,49,0,633,17.095,0,4115,F
0,355,8,0,105,0,465,0,125,0,227,0,4795,0,124,M
```

abalone.csv - Notepad

File Edit Format View Help

```
@relation abalone

@attribute diameter numeric
@attribute rings numeric
@attribute height numeric
@attribute length numeric
@attribute shell-weight numeric
@attribute shucked-weight numeric
@attribute whole-weight numeric
@attribute viscera-weight numeric
@attribute sex {M, F}

@data
0,365,15,0,095,0,455,0,15,0,2245,0,514,0,101,M
0,265,7,0,09,0,35,0,07,0,0995,0,2255,0,0485,M
0,42,9,0,135,0,53,0,21,0,2565,0,677,0,1415,F
0,365,10,0,125,0,44,0,155,0,2155,0,516,0,114,M
0,255,7,0,08,0,33,0,055,0,0895,0,205,0,0395,I
0,3,8,0,095,0,425,0,12,0,141,0,3515,0,0775,I
0,415,20,0,15,0,53,0,33,0,237,0,7775,0,1415,F
0,425,16,0,125,0,545,0,26,0,294,0,768,0,1495,F
0,37,9,0,125,0,475,0,165,0,2165,0,5095,0,1125,M
0,44,19,0,15,0,55,0,32,0,3145,0,8945,0,151,F
0,38,14,0,14,0,525,0,21,0,194,0,6065,0,1475,F
0,35,10,0,11,0,43,0,135,0,1675,0,406,0,081,M
0,38,11,0,135,0,49,0,19,0,2175,0,5415,0,095,M
0,405,10,0,145,0,535,0,205,0,2725,0,6845,0,171,F
0,355,10,0,1,0,47,0,185,0,1675,0,4755,0,0805,F
0,4,12,0,13,0,5,0,24,0,258,0,6645,0,133,M
0,28,7,0,085,0,355,0,115,0,095,0,2905,0,0395,I
0,34,10,0,1,0,44,0,13,0,188,0,451,0,087,F
0,295,7,0,08,0,365,0,1,0,097,0,2555,0,043,M
0,32,9,0,1,0,45,0,115,0,1705,0,381,0,075,M
0,28,11,0,095,0,355,0,075,0,0955,0,2455,0,062,M
0,275,10,0,1,0,38,0,085,0,08,0,2255,0,049,I
0,44,12,0,155,0,565,0,27,0,4275,0,9395,0,214,F
0,415,9,0,135,0,55,0,2,0,318,0,7635,0,21,F
```