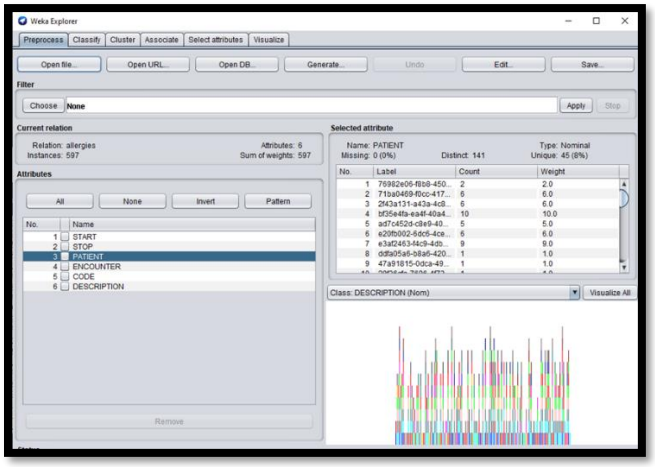
WEKA

Machine learning software to solve data mining problems



Links

- ✓ https://sourceforge.net/projects/weka/
- ✓ https://www.cs.waikato.ac.nz/ml/weka/
  https://lib.ugent.be/fulltxt/RUG01/000/842/101/RUG01-000842101_2010_0001_AC.pdf

License GNU General Public license (GPL 3.0 for Weka > 3.7.5)
Version 3.8.4
Last Update 12/20/2019
OS Linux, macOS, Windows

Description WEKA (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost any platform. The algorithms can either be applied directly to a dataset or called from your own Java code.
It incorporates representation and prescient investigation and displaying strategies, grouping, affiliation, relapse and order.
WEKA is an open source machine learning software that can be accessed through a graphical user interface, standard terminal applications, or a Java API. It is generally used for teaching, research, and industrial applications, contains a built-in tool for standard machine learning tasks, and additionally gives transparent access to well-known toolboxes such as scikit-learn, R, and Deeplearning4j.
WEKA contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to these functions.
WEKA supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. WEKA's techniques are predicated on the assumption that the data is available as one flat file or relation, where each data point is described by a fixed number of attributes (normally, numeric or nominal attributes, but some other attribute types are also supported). WEKA provides access to SQL databases using Java Database Connectivity and can process the result returned by a database query.

Features

- machine learning
- data mining
- preprocessing
- classification
- regression
- clustering
- association rules
- attribute selection
- experiments
- visualization
- workflow

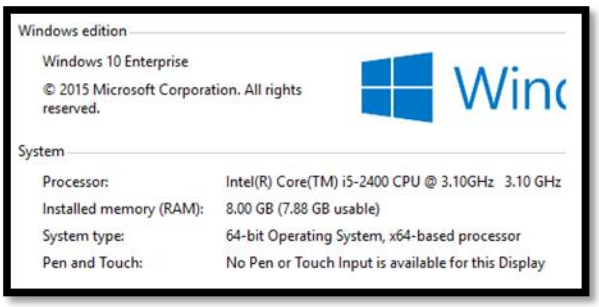
Connectivity / Supported Data Sources & Formats

- Arff, JSON, CSV, xrff, dat, data, names, and more
- Database using ODBC

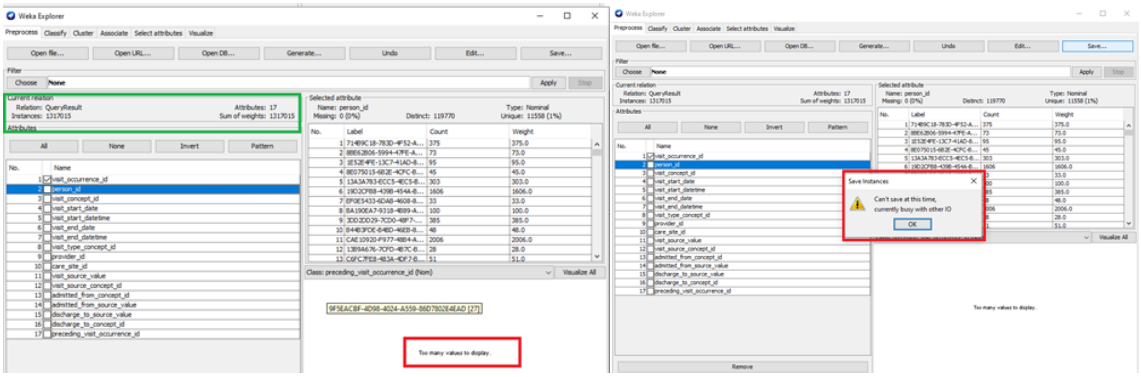
Limitations

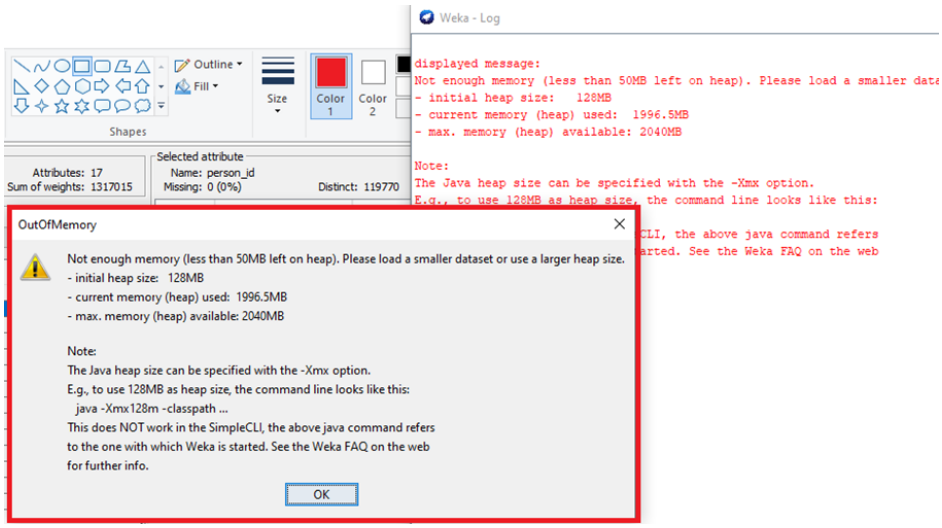
WEKA can only handle small data sets and is not capable of multi-relational data mining, but there is separate software for converting a collection of linked database tables into a single table suitable for processing using WEKA. Another important area that is currently not covered by the algorithms included in the WEKA distribution is sequence modeling.

Performance



Weka loads the 1.3 Million records but does not process the analysis and terminates with "Out-of-Memory" exception as shown in Pic. A machine with more memory and CPU capacity should be used



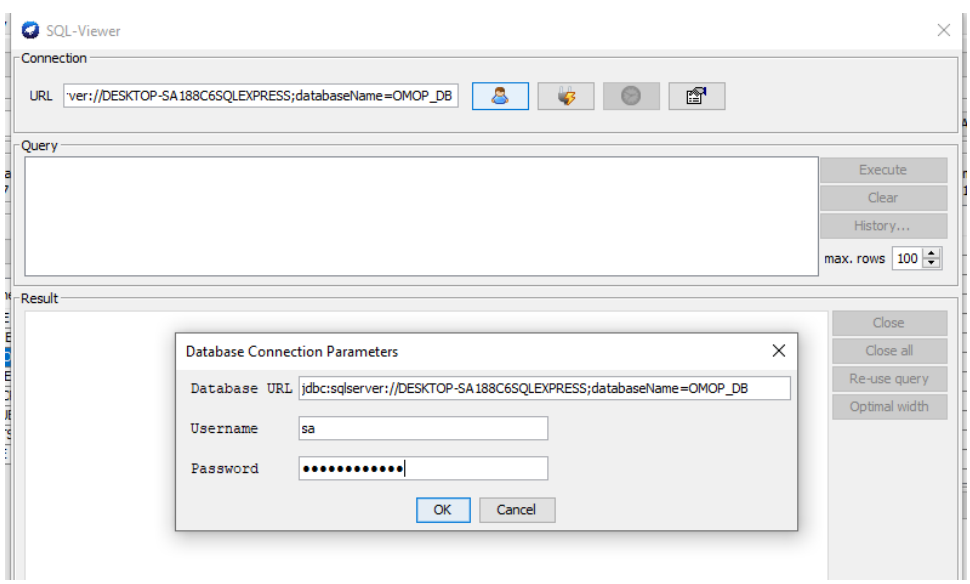


WEKA DATA ANALYSIS

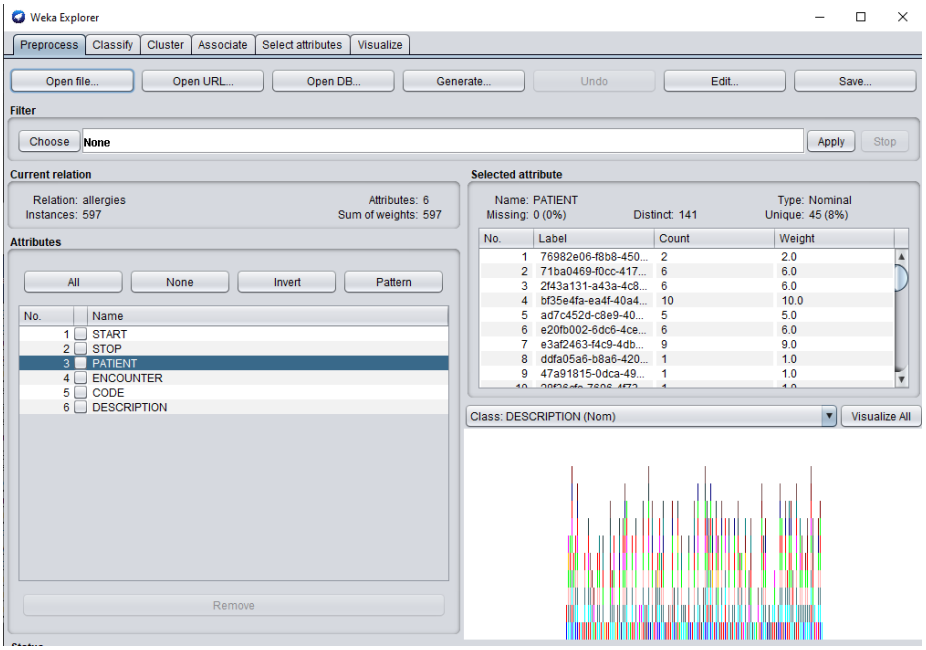
1. Download Weka from "https://waikato.github.io/weka-wiki/downloading_weka/"
2. Install Weka and click "Explorer", Select the "Open File" or "Open DB" options.
3. For CSV File select the file type as "CSV" and browse for the required CSV file.
4. For SQL DB give the connection string in the below format,

   jdbc:sqlserver://<ServerName>/<SQLInstanceName>;databaseName=<DBName>
   E.g:
   jdbc:sqlserver://DESKTOP-SA188C6SQLEXPRESS;databaseName=OMOP_DB

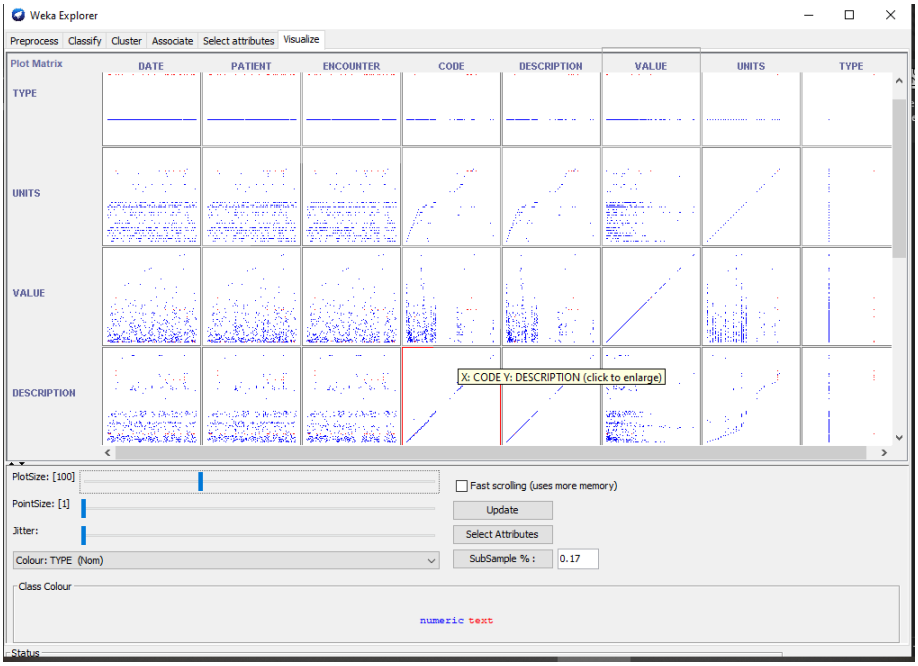
5. Give the required information like username and password.
6. Write a query to execute in the "Query" box and press "Execute".
7. View the data and the statistics as shown below,



SAMPLE WEKA STATISTICS:

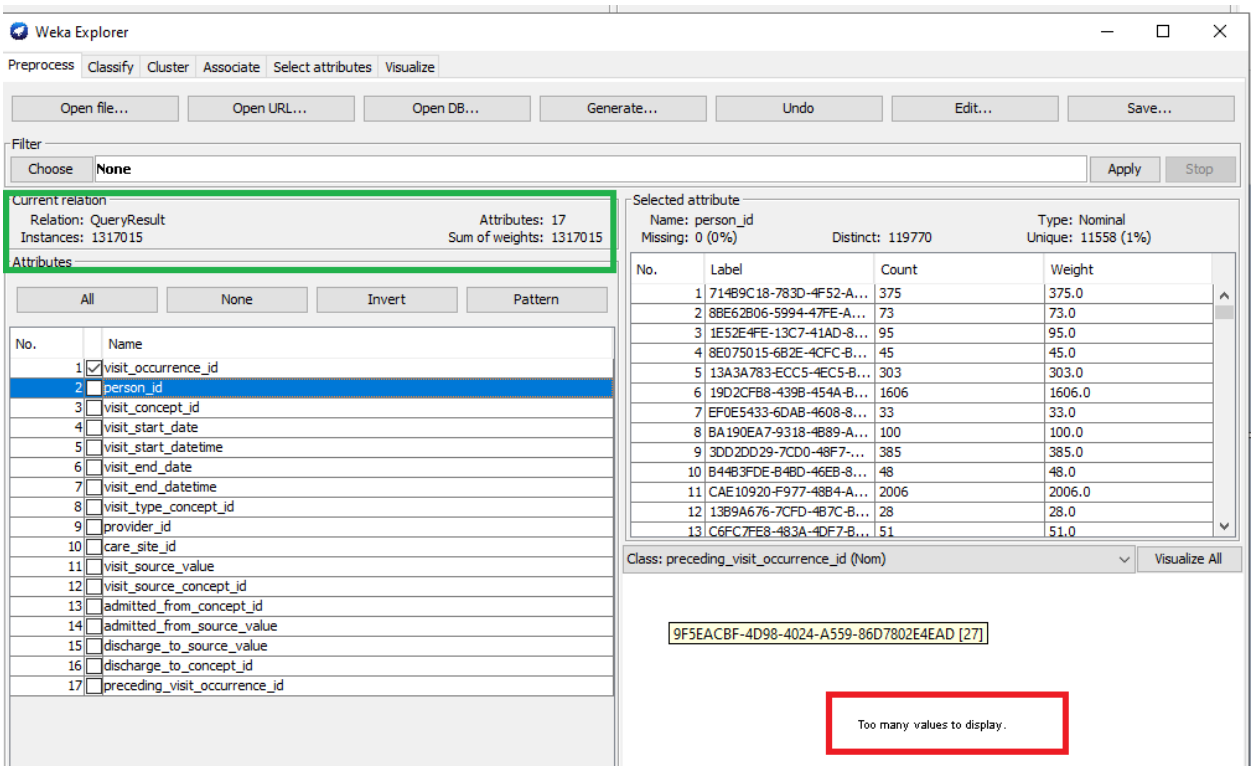


MIN AND MAX VALUES VISUALIZATION:



LIMITATIONS:

- Weka can load the 1,3M records but while trying to process "Too many values" warning is shown and expected results are not shown as shown in below,



- After freezing for some time, the "OutOfMemory" error is displayed.