AN APPROACH TO THE USE OF MACHINE LEARNING TOOLS FOR THE PREDICTION OF ADVERSE EVENTS IN CANCER PATIENTS ON IMMUNOTHERAPY

eahp

A. Paradela Carreiro[1], L. Otero Millan[1], M. Alfonsín Lara[1], E.Y. Romero Ventosa[1], C. Veiga García[2], M.J. Villanueva Silva[3], A.M. Regueira Arcay[1], I. Agra Blanco[1], A. Romero Rendón[4], N. Martínez López De Castro[1].

1 Vigo University Hospital Group, Pharmacy Department, i-FARMA Vigo Research Group (IISGS).
2 Galicia Sur Health Research Institute, Senior Researcher, Vigo, Spain.
3 Vigo University Hospital Group, Oncology Department, Vigo, Spain.
4 Galicia Sur Health Research Institute, Research Fellow, Vigo, Spain.

INTRODUCTION

The FDA Adverse Reporting System (FAERS) is a tool to voluntary report adverse events (AE), These data can be downloaded and used to apply "Machine learning" (ML) techniques. The bibliography is very limited, although it has already been the subject of a systematic review (Kim et al, 2022). FAERS data set could be useful to elaborate potential predictive modeling.

PURPOSE

To test a tool of ML to develop a potential predictive model of AE caused by immune checkpoint Inhibitors (ICI), using FAERS data set.
To contrast and explain the ML results with a reference model (RM), obtained through conventional processing data (spreadsheet).

MATERIAL AND METHODS

All FAERS records from 2022 were downloaded, selecting those of the group ICIs notified as "main suspected drug" (inclusion criteria). Collected variables from FAERS data set were: AE, age, drug and sex. The ML decision tree classification algorithm J48 implemented in the Weka application (version 3.8.6) was used to elaborate the ML model. The RM was built using a spreadsheet to tabulate and analyze the data (pivot tables and descriptive statistics).

RESULTS

1,702,222 notifications were downloaded and 86,053 records were selected according to inclusion criteria. The J48 algorithm applied to a subset including "adverse effect", "sex" and "drug", allowed us to estimate, for each AE the most likely responsible ICI drug. The metrics of the ML model obtained were satisfactory and compatible with the RM analysis. The J48 algorithm produced a complex tree (to be expected given the large number of AE). The application of J48 on another subset that includes "adverse effect", "age" and "drug", had a lower predictive capacity, due to the lower consistency of the data (age is only recorded as younger or older than 65 years) and that there is a higher proportion of missing values. The RM allows the results obtained with ML to be easily explained and understood.

CONCLUSIONS

The results of the J48 algorithm were useful for the association between AE, sex and drug. Despite the inherent limitations of voluntary AE reporting, this study will serve as a starting point for applying ML techniques in any other group, using FAERS data.
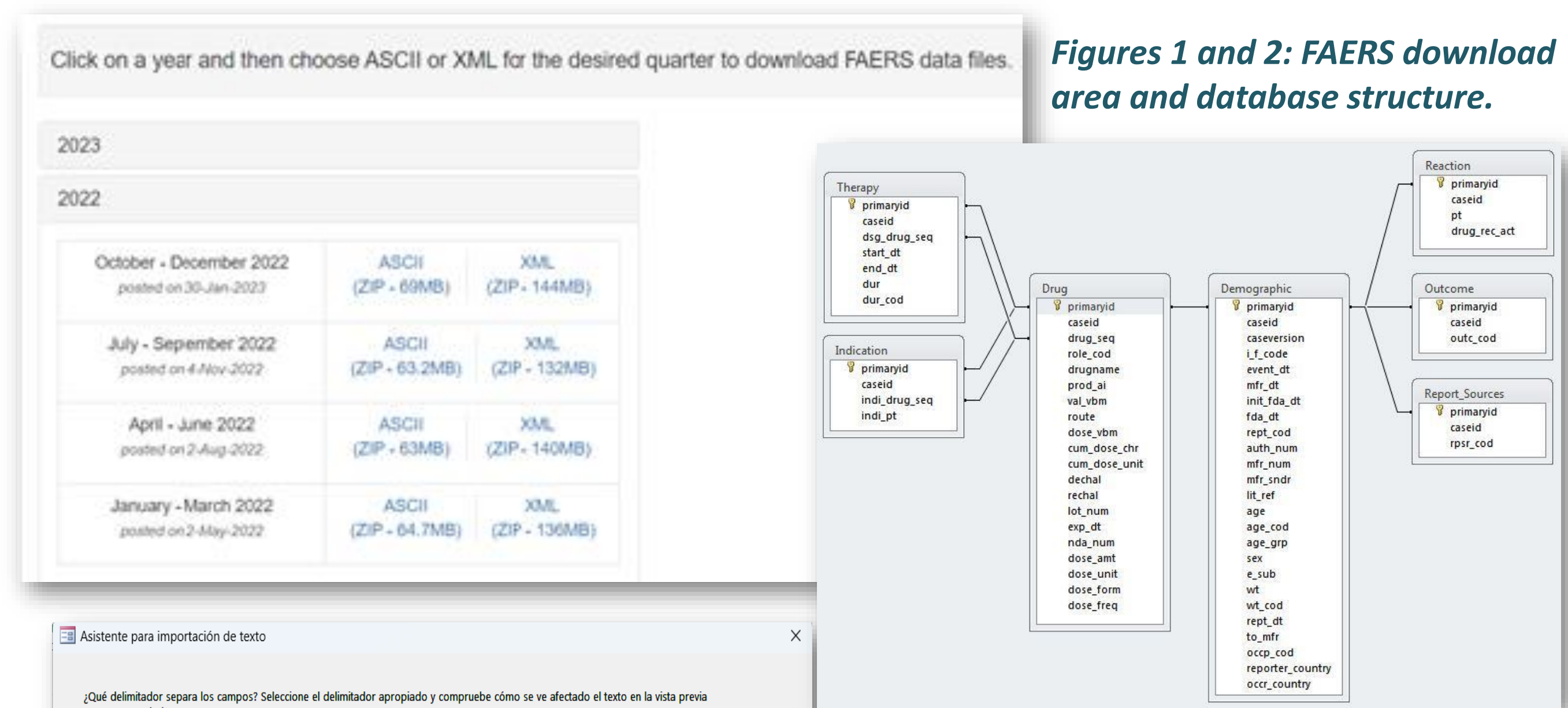
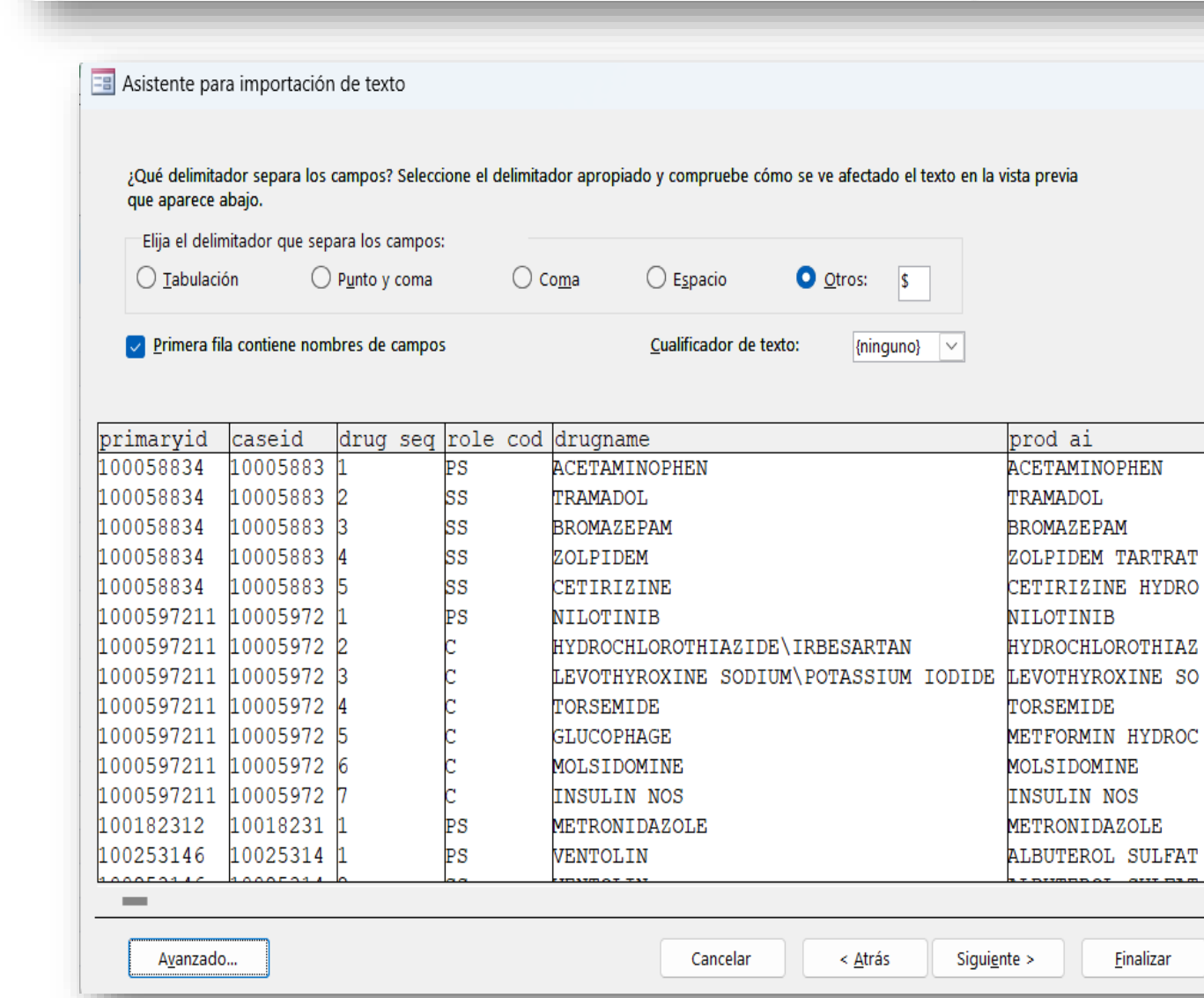Figures 1 and 2: FAERS download area and database structure.

Figure 3: FAERS data import tool to the local database in Access.
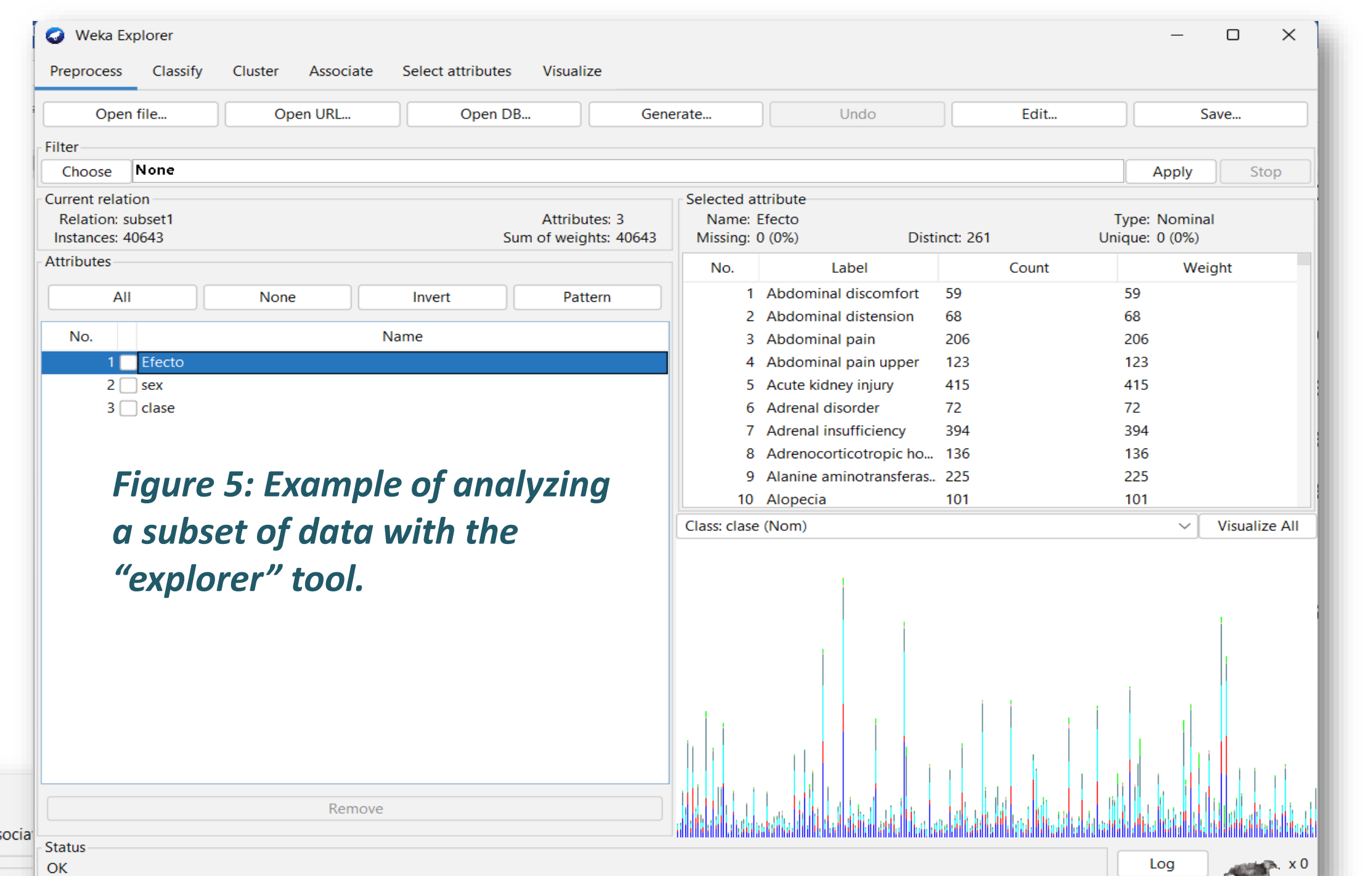

Figure 4: Weka "ML" software.

Figure 5: Example of analyzing a subset of data with the "explorer" tool.

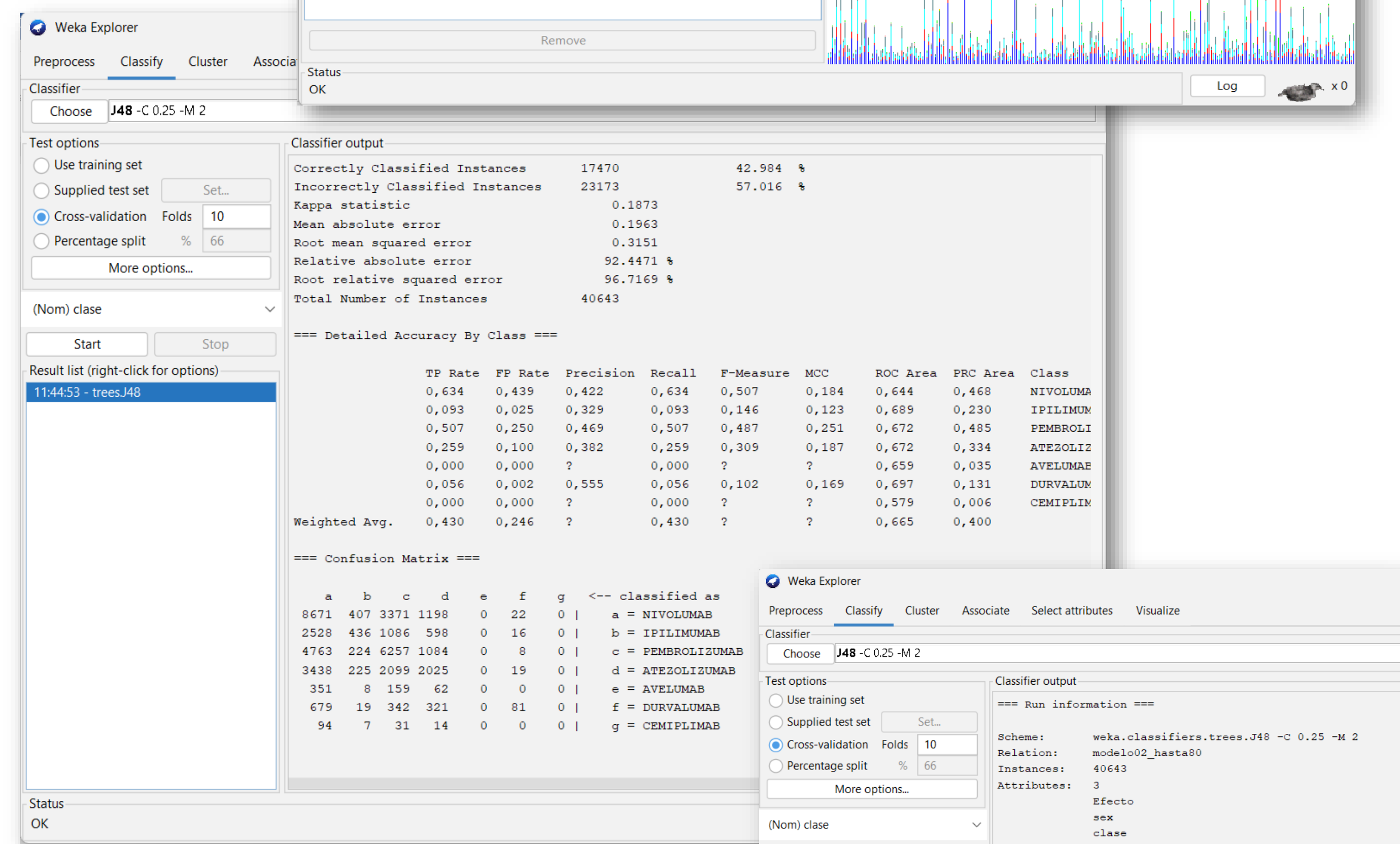Figure 6 and 7: Example of applying the J48 algorithm to a subset of data (metrics and tree).

Reference: Kim HR,Sung M,Park JA,Jeong K,Kim HH,Lee S,et al. Analyzing adverse drug reaction using statistical and machine learning methods:A systematic review. Medicine(Baltimore). June 24 2022;101(25).

5PSQ-097