

# 1. Topic

## SIPINA add-on for Open Office et Libre Office Calc.

Combining a spreadsheet with the data mining tools is essential for the popularity of these last ones. Indeed, when we deal with a moderate sized dataset (thousands of rows and tens of variables), the spreadsheet is a practical tool for the data preparation. This is also a valuable tool for the preparation of the reports. It is thus not surprising that Excel, and generally speaking a spreadsheet, is one the most used tool by data miners<sup>1</sup>.

Both Tanagra and Sipina provide an add-on for Excel<sup>2</sup>. The add-on enables to insert a data mining tool menu into the spreadsheet. The user can select and send the dataset to Tanagra (or Sipina), which is automatically launched. But, only Tanagra provides an add-on for Open Office Calc and Libre Office Calc. It is not available for Sipina.

This omission has been corrected for this new version of Sipina (**Sipina 3.9**). In this tutorial, we show how to install and use the “**SipinaLibrary.oxt**” add-on for Open Office Calc (OOCalc) 3.3.0. The process is the same for Libre Office 3.5.1.

## 2. Installing Sipina

We can download the setup file here: <http://eric.univ-lyon2.fr/~ricco/sipina.html>

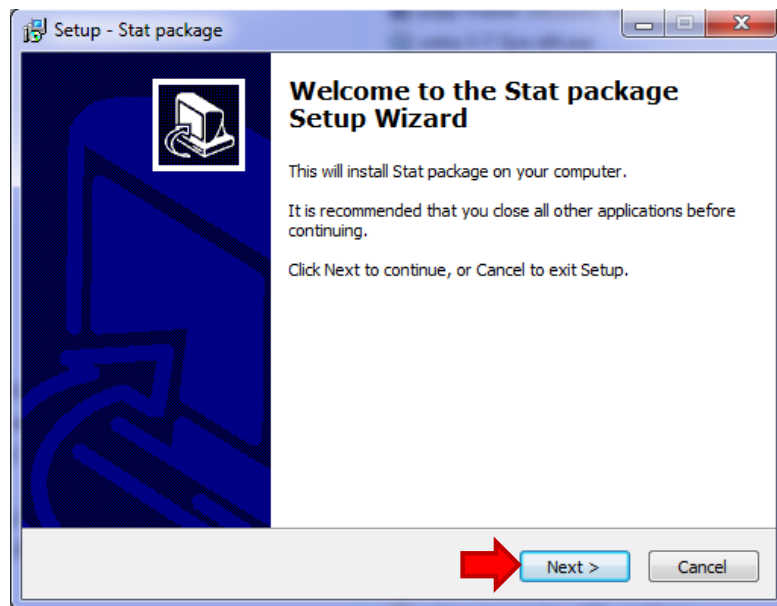
The screenshot shows a web browser window displaying the SIPINA website. The page has a dark green background and a sidebar on the left. The sidebar contains a logo with the word 'Research' and a navigation menu with items like 'Sipina Features', 'Sipina availability', and 'Download Sipina'. A red arrow points to the 'Download Sipina' link. The main content area is titled 'SIPINA DOWNLOAD' and includes a 'French website' link. Below this, there are two columns: 'Software' and 'Documentation'. The 'Software' section describes the 'Sipina Research version - 32 bits' and lists various supervised learning methods. The 'Documentation' section is divided into 'SIPINA Add-in for EXCEL spreadsheet' and 'Building decision tree interactively for the analysis of high blood pressure with SIPINA'. A table below the documentation lists files for download, including 'Sipina Research', '1-Add-In Installation', '2-How to use', and two 'Tutorial' files. A red arrow points to the 'Sipina Research' file link.

Software	File
Most recent <b>Sipina Research version - 32 bits</b> . Implements several supervised learning methods (decision tree, neural network, linear discriminant analysis,...), model assessments (cross-validation, bootstrap,...) and association rule algorithm.	<a href="#">Sipina Research</a>
Documentation	File
<b>SIPINA Add-in for EXCEL spreadsheet</b> (In english) An add-in for EXCEL(c) is incorporated in the SIPINA distribution. This add-in (SIPINA.XLA) enables to start a classification tree analysis, and more generally a data mining process, from your spreadsheet. This classification tree add-in appends a new menu in your spreadsheet. You select the cells range, activate the right menu: SIPINA is started and the selected dataset is automatically loaded.	<a href="#">1-Add-In Installation</a> <a href="#">2-How to use</a>
Building decision tree interactively for the analysis of high blood pressure with SIPINA.	<a href="#">Tutorial</a>
Using predefined learning (training) and test set for classifier performance evaluation.	<a href="#">Tutorial</a>
Definition of misclassification costs and the utilization of cost-sensitive	

<sup>1</sup> Kdnuggets polls, « [Data Mining/Analytic Tools Used](#) », 2<sup>nd</sup> place in 2010, 3<sup>rd</sup> place in 2011.

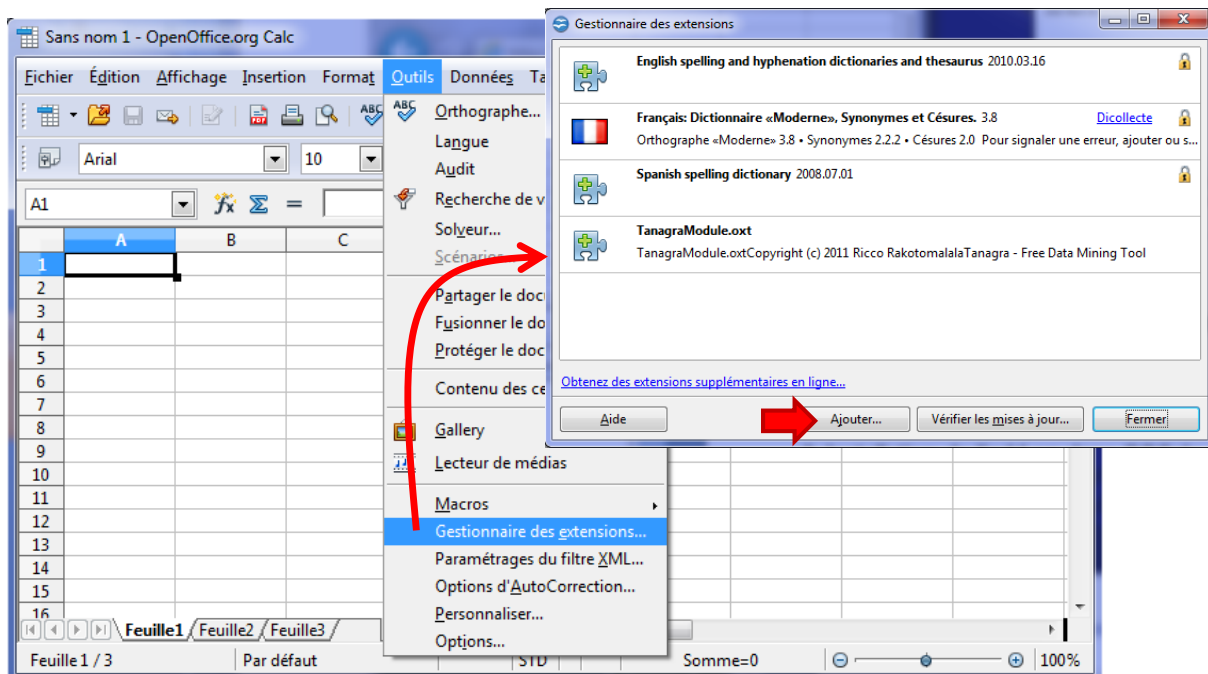
<sup>2</sup> <http://data-mining-tutorials.blogspot.fr/2010/08/tanagra-add-in-for-office-2007-and.html> (Tanagra, Excel 2007 and 2010); <http://data-mining-tutorials.blogspot.fr/2010/08/sipina-add-in-for-excel.html> (Sipina, Excel); <http://data-mining-tutorials.blogspot.fr/2011/07/tanagra-add-on-for-openoffice-calc-33.html> (Tanagra, OOCalc).

After we download the setup file, we start the installation process. We can simply click on the NEXT button throughout the process.



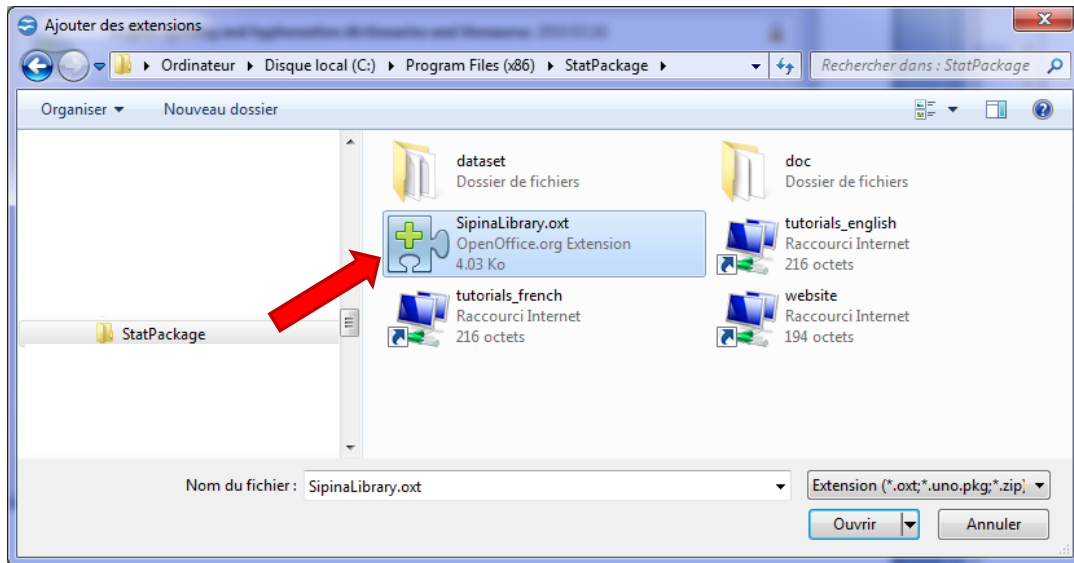
### 3. Installing the add-on for OOCalc

SIPINA being installed, we want to incorporate the add-on under OOCalc. We launch OOCalc. Into the Tools menu, we select the EXTENSIONS Manager menu.

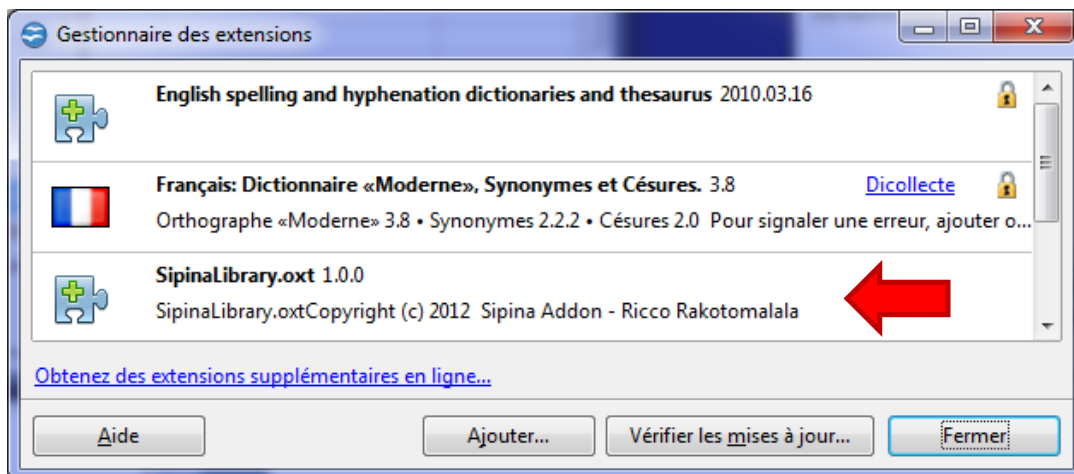


We pick the "**SipinaLibrary.oxt**" file. It is located into the Sipina installation directory i.e.

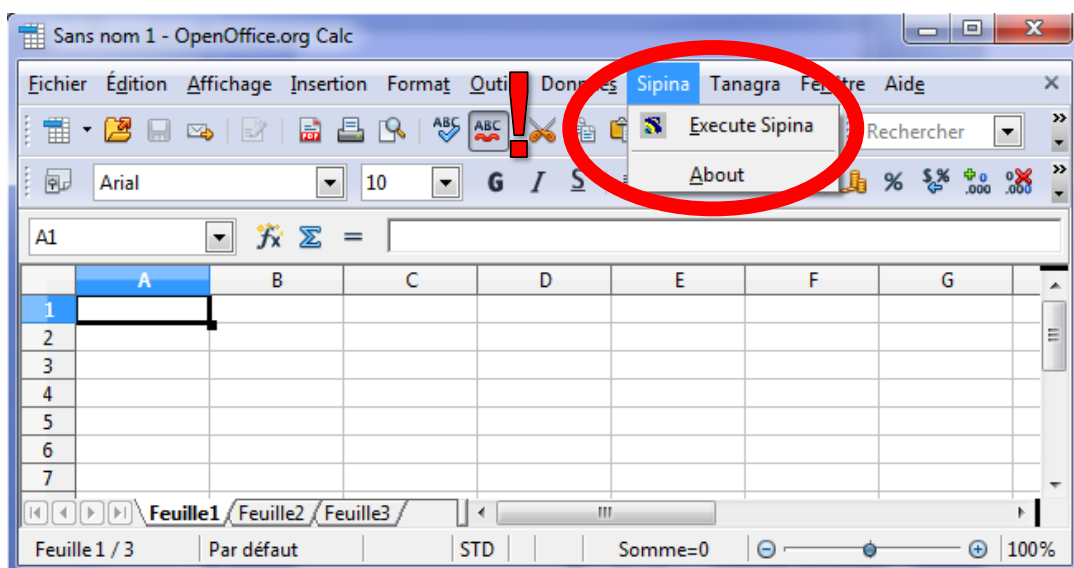
- « c:\Program Files\StatPackage » under 32 bit Windows ;
- « c:\Program Files (x86)\StatPackage » under 64 bit Windows.



Now, the add-on is incorporated into OOCalc.



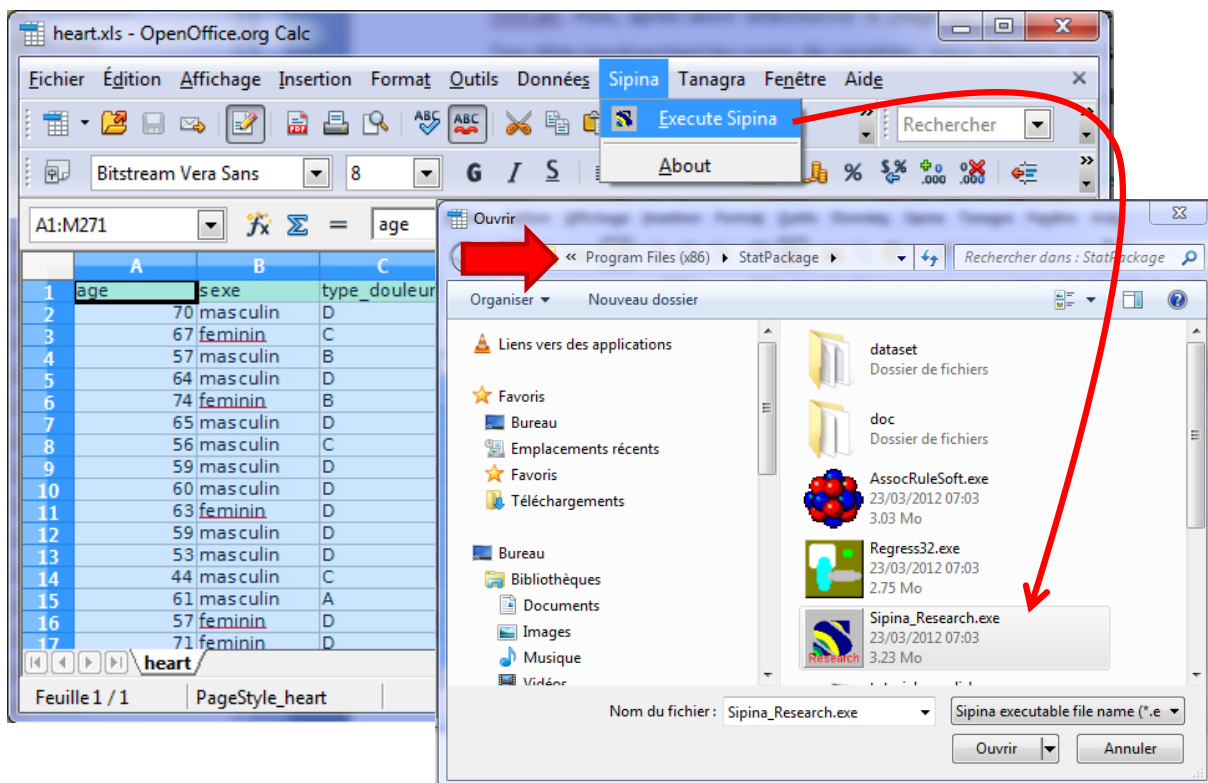
**To activate the add-on, we must close and restart OOCalc.** The SIPINA menu is now visible into the menu bar.



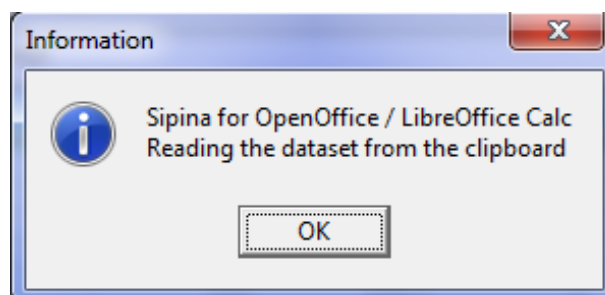
## 4. Using the add-on

To use the add-on, we must first load the data file (**heart.xls**). We select the data range, and we click on the SIPINA / EXECUTE SIPINA menu.

With the 32 bit version of Windows, SIPINA is started automatically. With the 64 bit version, a dialog box appears. We select the SIPINA executable file [**SIPINA\_RESEARCH.EXE**] (Note: if we select another EXE file, we can launch the Regression tool [**REGRESS32.EXE**] or the Association Rule learner tool [**ASSOCRULESOFT.EXE**]).



A dialog box confirms the data transferring.

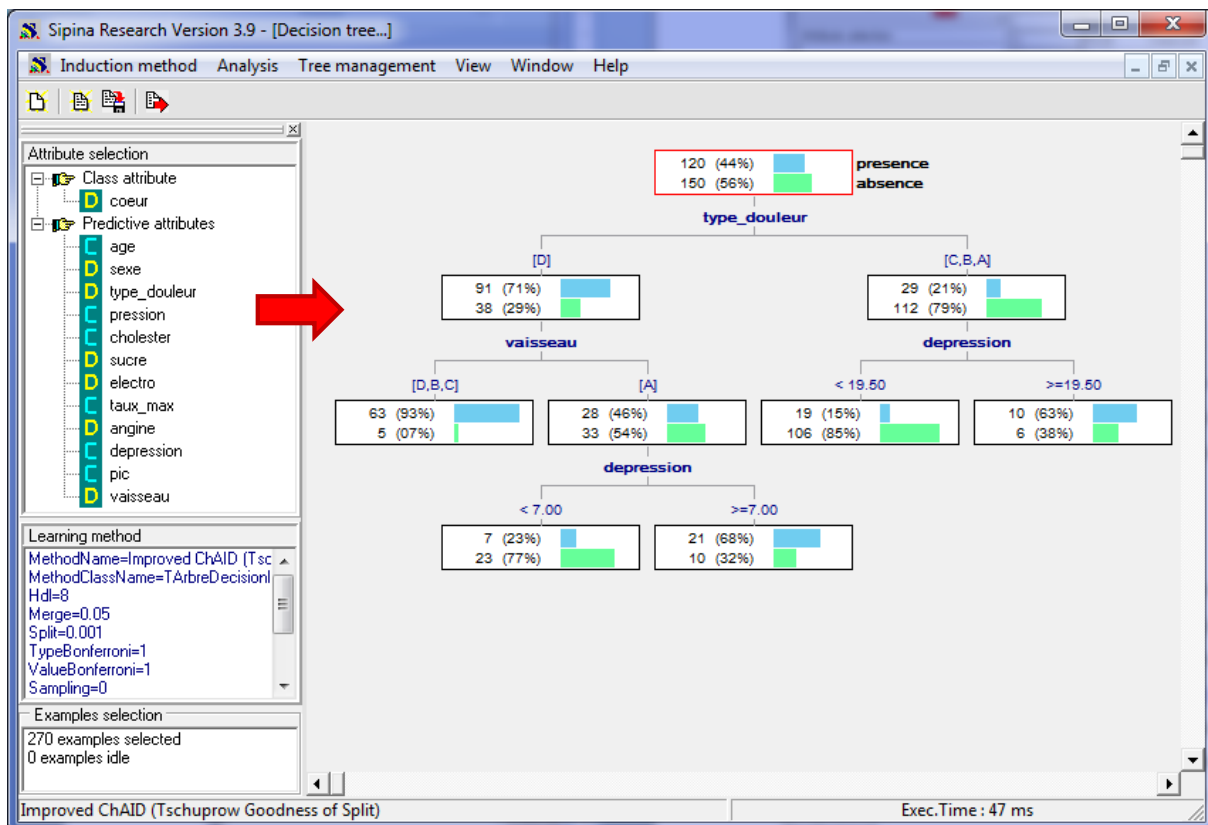


Last, SIPINA is visible. The dataset is loaded into the data grid.

The screenshot shows the 'Learning set editor' window. A red arrow points to the 'Data' menu. The main window displays a table with 16 rows and 8 columns: age, sexe, type\_douleur, pression, cholester, sucre, and electro. The 'Attribute selection' panel on the left is empty. The 'Learning method' panel shows 'Improved ChAID (Tschuprow Goodness of Split)'. The status bar at the bottom indicates 'Attributes : 13' and 'Examples : 27'.

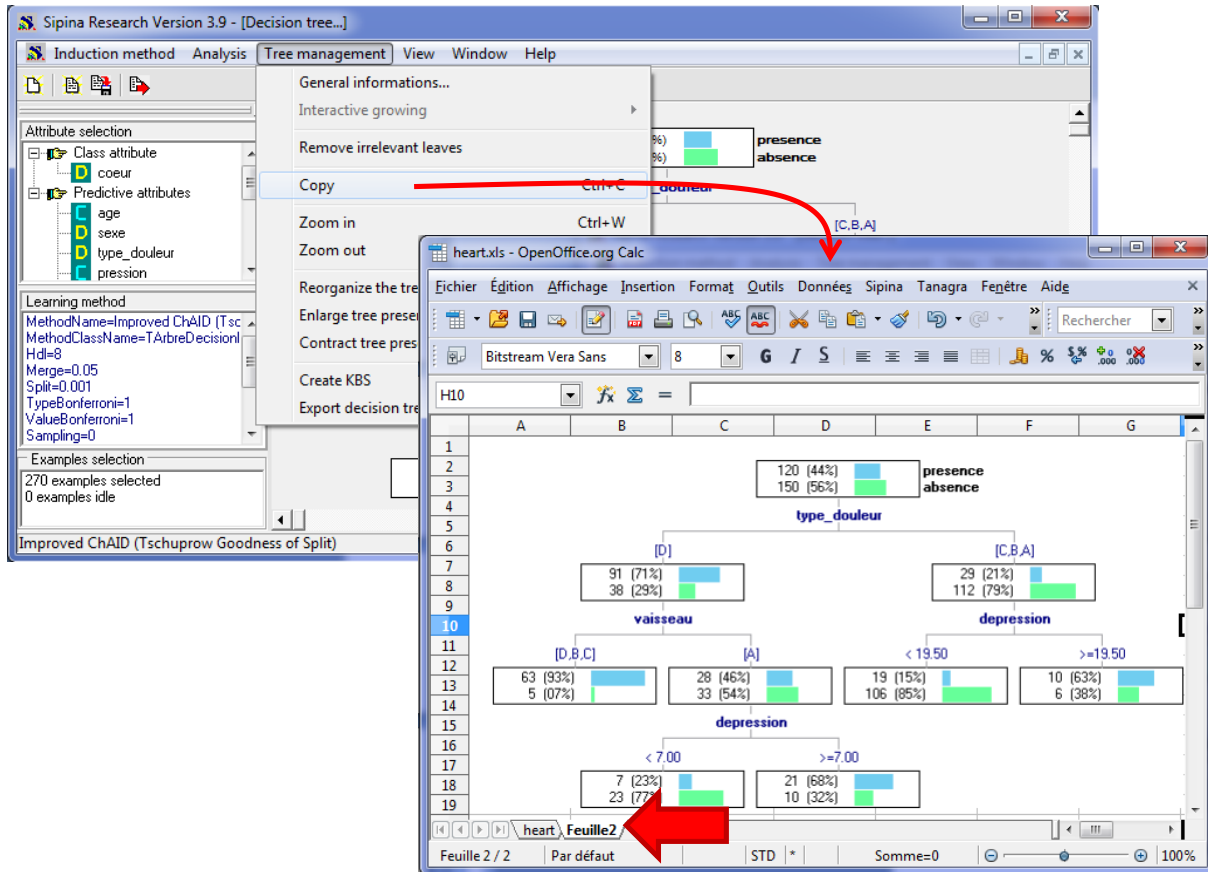
	age	sexe	type_douleur	pression	cholester	sucre	electro
1	70.00	masculin	D	130.00	322.00	A	C
2	67.00	feminin	C	115.00	564.00	A	C
3	57.00	masculin	B	124.00	261.00	A	A
4	64.00	masculin	D	128.00	263.00	A	A
5	74.00	feminin	B	120.00	269.00	A	C
6	65.00	masculin	D	120.00	177.00	A	A
7	56.00	masculin	C	130.00	256.00	B	C
8	59.00	masculin	D	110.00	239.00	A	C
9	60.00	masculin	D	140.00	293.00	A	C
10	63.00	feminin	D	150.00	407.00	A	C
11	59.00	masculin	D	135.00	234.00	A	A
12	53.00	masculin	D	142.00	226.00	A	C
13	44.00	masculin	C	140.00	235.00	A	C
14	61.00	masculin	A	134.00	234.00	A	A
15	57.00	feminin	D	128.00	303.00	A	C
16	71.00	feminin	D	112.00	149.00	A	A

The use of SIPINA is described in numerous tutorials<sup>3</sup>. We do not detail this in this tutorial. We note only that we obtain the following classification tree on our dataset.



<sup>3</sup> <http://data-mining-tutorials.blogspot.fr/search/label/Sipina>

We can copy the decision tree (TREE MANAGEMENT / COPY menu) and paste it into a new sheet of the workbook (or into a slide, into a word processor, etc.). This feature is very valuable when we want to write a report.



## 5. Conclusion

Adding advanced statistical features into a spreadsheet application is a good idea. We show how to install and use the Sipina add-on into OOCalc. A similar add-on is provided for Excel.

But Excel and OOCalc are not the only ones spreadsheets. Not so well known than OOCalc, [GNUMERIC](http://projects.gnome.org/gnumeric/doc/chapter-stat-analysis.shtml) spreadsheet is an attractive alternative. It differs on two points: it is a specific project, non-integrated into an office suite; the statistical features are directly implemented into the tool and not as extensions. The list of available functions is available online (<http://projects.gnome.org/gnumeric/doc/chapter-stat-analysis.shtml>). I think it will be made more complete with time.