

# Know4Business

Author: Roberto Bello, Engineer: Giulio Beltrami

## Introduction

The statistics so often are wrong or, at least, are wrong their users.

Are wrong in applying the tools of statistical science putting together fragments of informational objects or situations one from each other, completely different.

First they fragment, then mix and finally aggregate.

Finally they get a conclusion.

So the researchers on political trends break up the views of respondents, mix the individual responses, aggregate, and finally claim truth that are attributable only to the respondents that they have created in virtual way, non-existing in reality and certainly not attributable to individuals or homogeneous groups of respondents.

Similarly, the Business Intelligence makes available the tools of data analysis; it can cut the data and then reassembling it into multidimensional structures in which the peculiarities of information starting positions have been destroyed.

So the Business Intelligence brings together companies from different sectors, revenues, organic markets and payment habits, changing from time to time variables crossing data.

Which subject (or situations) may be applied then the decisions are taken, having destroyed the wealth of information overall subject (or situations) starts from?

For example, if you had an archive of animals including mammals, in which there were also men and primates, as a result I could get that mammals have on average about three feet.

Where can I find a mammal that has an average of three feet?

To make real statistics it's necessary to be maintained intact as much as possible the information assets of the initial data of the subject or the situation under review.

The techniques derived from Neural Networks use an approach to data analysis entirely respectful of the information assets of the initial data.

In fact do not require the user to define variables to cross, there by preventing him from making strange crosses.

Only require you to enter the maximum number of groups that the algorithm will create.

Do not destroy the wealth of information of the initial data, but work always subject's data (or situation) in relation to the data of other persons (or situations).

Retain all the information attributable to the subject or situation in question and create the categories of membership of the subjects (or situations) in which subjects (or situations) are similar to each other.

Induction is a very important learning mode of the living creatures.

A simple example can illustrate how one learns by induction.

Assume that you have in front of you a person who has never seen the containers in common use such as glasses, bottles, jars, cups, vases, boxes, bottles, jugs, cups, tetra, and so on.

Without comment in succession you show real examples of objects belonging to the categories described above.

The person may look, smell, touch and weigh the items shown.

After having examined a sufficient number of objects, the person will be easily able to group objects into categories that include objects similar to each other globally, focusing on some characteristics more than others deemed irrelevant because it does not discriminate.

For learning occurred, I could submit another object in the form of glass of different color, other materials and other weight still obtain the location of the object in the category of glasses.

Also for the person in the induction training could make two cups of categories: those without handle and with the handle (mug).

Learning has enabled the person to recognize the useful aspects of the object to move from the particular to the universal and neglecting the non-influential.

Know4Business realizes the acquisition of knowledge hidden in the data and is able to explain the characteristics of each identified group.

Know4Business is applicable in any field (medicine, market research, economics, fraud prevention, customer profiling, bank credit, market segmentation, political polls, pharmacological research, security systems, etc..).

Know4Business is able to determine an Index of Knowledge (*Knowledge Index*) measurement of the knowledge contained in the input data processed, allowing you to answer the question of *knowledge exists in the archive?*.

Know4Business makes available charts and statistics on the results.

Know4Business runs on Cloud Computing and will be available in a *pay per use*.

Know4Business is a Google App to find knowledge hidden in your data, text or numeric, using sophisticated algorithms derived from Neural Networks and Statistics.

The data are displayed as lists of entities homologous (the same class such as animals, disease, performance, faults), with their attributes, encoded in tables with:

- a first row of headers (alphanumeric), in which the first cell called the identity of the entity and the remaining attributes characteristic of the variables
- subsequent rows, in which the first cell identifies the entity and the subsequent contain the values of its attributes, indicated by the header of each column

Know4Business, after a series of checks, normalizations and adjustments of the input data, analyzes the entity with the aim to catalog the records into groups, for similarity; so as to maximize the homogeneity within each group trying to make different the groups between them, explaining also the grouping criteria.

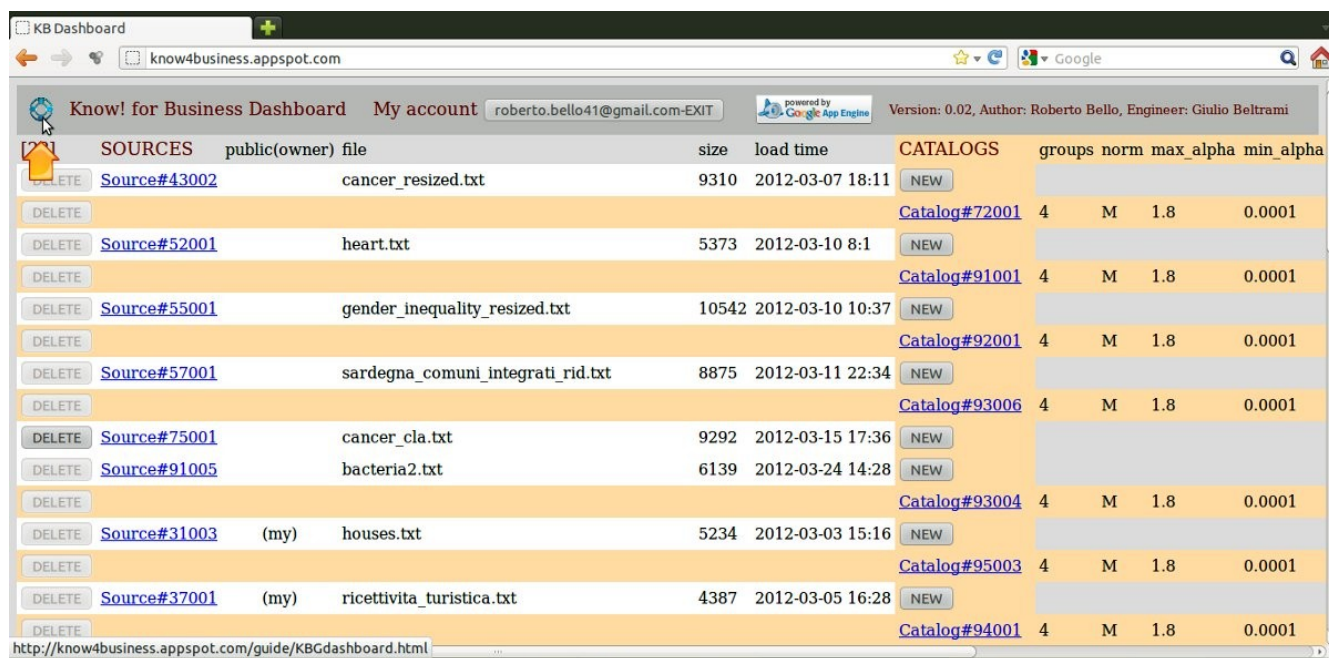
Know4Business provides six tools:

- SOURCE to acquire and maintain records to process available in tabular form
- CATALOGUE to catalog, with an algorithm based on the theory of neural networks (Kohonen), the sources
- STATISTICS to perform statistical analysis, to explore the goodness of classification, in addition to the results of the cataloging function (see SHOW)
- CLASSIFIER to classify new entities "on the fly", using the training matrix function obtained from the previous catalog
- SHOW to view the reports, analytical and descriptive summary of the groups created by cataloging and classification
- CHART that displays a vertical bar graph in the CV (coefficient of variation) of the variables / columns of the identified groups

## Dashboard

From the operational point of view, KB Dashboard is the main panel.

It opens when logged in via their Google Account, previously authorized.



The screenshot shows a web application titled "Know! for Business Dashboard". It features a navigation bar with a "My account" link and a "powered by Google App Engine" logo. The main content area displays two tables: "SOURCES" and "CATALOGS".

**SOURCES Table:**

	SOURCES	public(owner)	file	size	load time
DELETE	<a href="#">Source#43002</a>		cancer_resized.txt	9310	2012-03-07 18:11
DELETE	<a href="#">Source#52001</a>		heart.txt	5373	2012-03-10 8:1
DELETE	<a href="#">Source#55001</a>		gender_inequality_resized.txt	10542	2012-03-10 10:37
DELETE	<a href="#">Source#57001</a>		sardegna_comuni_integrati_rid.txt	8875	2012-03-11 22:34
DELETE	<a href="#">Source#75001</a>		cancer_cla.txt	9292	2012-03-15 17:36
DELETE	<a href="#">Source#91005</a>		bacteria2.txt	6139	2012-03-24 14:28
DELETE	<a href="#">Source#31003</a>	(my)	houses.txt	5234	2012-03-03 15:16
DELETE	<a href="#">Source#37001</a>	(my)	ricettivita_turistica.txt	4387	2012-03-05 16:28

**CATALOGS Table:**

	CATALOGS	groups	norm	max_alpha	min_alpha
NEW	<a href="#">Catalog#72001</a>	4	M	1.8	0.0001
NEW	<a href="#">Catalog#91001</a>	4	M	1.8	0.0001
NEW	<a href="#">Catalog#92001</a>	4	M	1.8	0.0001
NEW	<a href="#">Catalog#93006</a>	4	M	1.8	0.0001
NEW	<a href="#">Catalog#93004</a>	4	M	1.8	0.0001
NEW	<a href="#">Catalog#95003</a>	4	M	1.8	0.0001
NEW	<a href="#">Catalog#94001</a>	4	M	1.8	0.0001

To obtain context-sensitive help press the circle icon in the upper left corner.

The panel contains the tools in three sections of different colors from left to right:

- **SOURCES:** to manage the input of cataloging rather than classification, list the name, size and date / time of creation of data sources, the private rights of the user or public (available to all users if reading public), as described later.
- **CATALOGUES:** to manage one or more cataloging, for each SOURCE, suitable to act as sample for the training of the neural network, and allows to create lists (NEW button) or retrieve (link) cataloging, with their parameters and key indicators, such as described in the User Guide section of the same name.
- **CLASSIFIERS:** to manage the classification of SOURCE CATALOGUE consistent with the reference, and allows creating lists (NEW button) or call (link) classifications, with their main parameters and indicators, as described in the User Guide section of the same name.

The DELETE button allows the owner to remove his tools, with data referring and any other related tools.

Using the tools in the lower left you can upload files (SOURCES) for cataloging or classification

KB Dashboard

know4business.appspot.com

DELETE	Source#	(my)	File Name	Size	Date	NEW				
DELETE	Source#93009	(my)	anagrafica_esercizi_ricettivita_2010_rid.txt	16017	2012-03-27 15:7	NEW				
DELETE							Catalog#95004	4	M	1.8 0.0001
DELETE	Source#94002	(my)	furti_rapine_meno_gravi_x_1000_abitanti.txt	1863	2012-03-24 14:32	NEW				
DELETE							Catalog#97004	4	M	1.8 0.0001
DELETE	Source#94004	(my)	ricettivita_turistica.txt	4387	2012-03-24 14:46	NEW				
DELETE							Catalog#93007	4	M	1.8 0.0001
DELETE	Source#96003	(my)	mammalsize.txt	2971	2012-03-24 14:40	NEW				
DELETE							Catalog#94003	4	M	1.8 0.0001
DELETE	Source#97003	(my)	animals.txt	3866	2012-03-24 14:21	NEW				
DELETE							Catalog#98001	4	M	1.8 0.0001
DELETE	Source#98002	(my)	iscrizioni_cessazioni_imprese_rid.txt	9443	2012-03-27 15:16	NEW				
DELETE							Catalog#98003	4	M	1.8 0.0001
DELETE	Source#99001	(my)	rules_r.txt	4868	2012-03-29 14:53	NEW				
DELETE							Catalog#100001	4	M	1.8 0.0001

Fetch [public ☒] SOURCE

IMPORT Google SPREADSHEET will be available ...

UPLOAD csf FILE /home/roberto/kb\_gae/sport\_r.txt Sfoglia

To get the User Guide: click the header mini icons.

Press the browse button (Sfoglia) to locate the file to be imported from your computer, press the UPLOAD button csf FILE and wait until the upload is finished.  
The file is flagged for access to all (public).

## Constraints on sources

The files to be loaded must have the following characteristics:

- must be in tabular form without empty cells
- for cataloging, the product of the number of rows to the number of columns must not be higher than 2000 (limited to the current free version of Know4Business)
- the format must be of type text (txt, csv)
- fields are separated by tabs (tab)
- in cells containing the texts should not be spaces (San\_Marino instead of San Marino)
- the first line contains the descriptions of the columns separated by tabs (tab) and no spaces in descriptions
- the first column is intended to identify the identification code for each line of each record with no spaces (eg, customer number, product name, batch, etc..)
- to process the information contained in the cells at the intersection of record / case and variable / column from the second column, second row from the last column and last row all values, of all columns and all rows must be separated by tabs (tab)
- can not exist or empty fields containing only spaces
- a column to contain numeric data can not contain text data

To convert to text format of the tables available in xls (Excel) or OpenOfficeCalc (ODS) is sufficient to use programs that can read these formats and convert format (csv), choosing as the field delimiter (tab) and nothing (empty) as a text delimiter. There is no automatic transformation of dates and times. The dates should be processed by the user in pseudo continuous numeric variables by assigning the value 0 to the oldest date and increasing by one each subsequent date, or transforming the 365 year-round in parts per thousand, according to the formula:  $0.9999 \text{ day of the year} * / 365$ . Another variable could be given to the value of the year.

## Suggestions for sources

The pairs of variables in relation to each other are preferably to be expressed as the ratio, through a single value that provides more clear and immediate information; in this way are calculated of the variables obtained from the input variables. Suppose there are two variables: the weight and height of the person. Considered separately have little meaning, they would be best to get the coefficient of body mass that is definitely a good synthetic index of obesity (body weight in kilograms divided by height in meters squared).

Another important step in the preparation of preliminary data is to try to simplify the problem to be solved. To this end, it may be helpful to resize the space of input values, composed of all the possible links between the data input space grows exponentially with the growth data.

A technique often used to reduce the number of variables and improve the learning ability of neural networks is the principal component analysis, which seeks to identify a sub-space ( $m$  size) that is as meaningful as possible with respect to the input space ( $n$  size).

The  $m$  final variables are called principal components and are linear combinations of the  $n$  original variables. Other methods used to reduce the dimensionality of the problem to be solved are the elimination of variables highly correlated with each other and unimportant data for the goal to be achieved.

Often you must rearrange the archive of the input data to be cataloged by examining the results of initial processing, which often indicate that some variables / columns are completely irrelevant: their removal in subsequent processing often helps to improve the cataloging having mitigated the noise caused by variables / columns unnecessary.

In the processing of data related to clinical research, it was verified that the data on sex, nationality, residence, education, etc., not giving in those cases no contribution to the catalog, could be omitted to improve the quality of new learning.

A very important aspect to consider is related to the number of records contained in the archive input to cataloging.

Very often better results are obtained with input files of modest size that are better able to generalize and to produce arrays of training more predictive.

Unlike a file containing a considerable number of records could produce a flawed training resulting from overfitting (photo effect) only able to classify new records almost identical to those used in the identification phase.

In statistics it comes to overfitting (excessive adaptation) when a statistical model fits the observed data (the sample) using an excessive number of parameters.

You can not point out a priori the abundance of the best records in the files to be cataloged: too much depends on the number of variables and the information content of variables for all the recordings in the file.

The best suggestion is to make separate calculations with the original file and other files obtained from the random fewer recordings.

Other times it is convenient to remove, from the input file, the records contain information that clearly contradictory, absurd or missing: doing so reduces the size of the archive and improves quality by reducing noise.

## Start cataloging

The screenshot shows the KB Dashboard interface. At the top, there's a navigation bar with 'KB Dashboard' and a search bar. Below it, a table lists various sources. The last row, 'Source#101001 (my) sport\_r.txt', has a 'NEW' button circled in red. The table columns include source ID, name, type, size, date, and a 'NEW' button. Below the table, there are options to 'Fetch [public] SOURCE', 'IMPORT Google SPREADSHEET', and 'UPLOAD csf FILE'.

DELETE	Source	Type	Name	Size	Date	NEW	Catalog	Count	Model	Score
DELETE	Source#94002	(my)	furti_rapine_meno_gravi_x_1000_abitanti.txt	1863	2012-03-24 14:32	NEW	Catalog#95004	4	M	1.8 0.0001
DELETE	Source#94004	(my)	ricettivita_turistica.txt	4387	2012-03-24 14:46	NEW	Catalog#97004	4	M	1.8 0.0001
DELETE	Source#96003	(my)	mammalsize.txt	2971	2012-03-24 14:40	NEW	Catalog#93007	4	M	1.8 0.0001
DELETE	Source#97003	(my)	animals.txt	3866	2012-03-24 14:21	NEW	Catalog#94003	4	M	1.8 0.0001
DELETE	Source#98002	(my)	iscrizioni_cessazioni_imprese_rid.txt	9443	2012-03-27 15:16	NEW	Catalog#98001	4	M	1.8 0.0001
DELETE	Source#99001	(my)	rules_r.txt	4868	2012-03-29 14:53	NEW	Catalog#98003	4	M	1.8 0.0001
DELETE	Source#101001	(my)	sport_r.txt	13398	2012-03-31 15:14	NEW	Catalog#100001	4	M	1.8 0.0001

The new file (Source#101001 – sport\_r.txt) is shown in panel.

To request a catalog, press the NEW button (circled in red) and then confirm by pressing the OK button.

The screenshot shows the same KB Dashboard interface as before, but with a 'New Catalog?' dialog box open in the center. The dialog box has two buttons: 'Annulla' and 'OK'. The 'NEW' button for Source#101001 is still circled in red. The background table and interface elements are dimmed.

## Catalogue

The panel KB Catalogue is open and becomes active.

**Number of Groups:** Choose the number of classification groups (3 to 20).

The maximum potential number of groups is equal to the square of the number of selected groups, choosing the value 4, the groups catalog potential maximum will be 16.

It is useful to begin by choosing a low value of the Number of Groups (3 or 4), increasing later in the presence of heterogeneous groups.

Increasing the Number of Groups on the one hand causes an overall improvement of cataloging, but then a dilution of knowledge between groups, dilution which could not be desired.

It can happen that you want to investigate on their own minority groups to find out the exceptions and phenomena in progress, in this case a catalog using a high number of groups is suitable for purpose. The Number of Groups affects more than proportional the duration of the calculation to obtain the cataloging.

**Normalization** (Max, Std, None): choose the type of normalization (the maximum, the Standard Deviation, no normalization).

The data of SOURCES, before being cataloged, must be normalized.

The normalization which empirically gives the best results in the subsequent cataloging is the one on Max; with it the data of each column / variable are divided by the maximum value of the column / variable, thus obtaining new values that vary from zero to one, and what happening for all variables / columns, makes the variables / columns paritarie in the process of cataloging.

**Start Value of alpha** (1.8 - 0.9): choose the starting value of the parameter alpha (1.8 to 0.9). The parameter alpha is involved in the training phase of the neural network. It begins to work with the value defined by the Start Value of alpha. For each cycle decreases the value specified in parameter Decreasing step of alpha and ends to work when it reaches the value indicated in the End parameter value of alpha.

**End Value of alpha** (0.5 - 0.0001): choose the target value of the parameter alpha (1.8 to 0.9). The parameter alpha is involved in the training phase of the neural network. When the current value of the parameter alpha reaches the target value, the training phase of the neural network ends.

**Decreasing step of alpha** (0.1 - 0.001): choose the value of the parameter alpha decrease (from 0.1 to 0.001). The parameter is decreasing at each calculation cycle, the current value of alpha to the limit indicated by End Value of alpha. The lower the value of the parameter and the improvement of the current error becomes more sensitive, but with increasing computation time in a greater than proportional.

### **Effects of parameters on the results and training times**

Very accurate results, but with significant running time, are obtained:

- Number of Groups greater than 6
- Start Value of alpha equal to 1.8
- End Value of alpha equal to 0.0001
- Decreasing step of alpha equal to 0,001

Less accurate results, but with reduced running time, are obtained:

- Number of Groups less than 5
- Start Value of alpha less than 1.2
- Value End of alpha not less than 0,001
- Decreasing step of alpha step not less than 0.01

### **Start of training**

To start the process of cataloging click on the "CATALOGUE".



KB Dashboard KB Catalogue

know4business.appspot.com/KBCatalogue.py?sourcekey=ag9zfmtub3c0YnVzaW5lc3NyDgsSBINvdXJjZRIJlQYM

KB Catalogue Catalog #102001 Sample #101001 (sport\_r.txt)

Number of Groups (3 - 20) 4

Normalization(Max, Std, None) M

Start value of alpha (1.8 - 0.9) 1.8

End value of alpha (0.5 - 0.0001) 0.0001

Decreasing step of alpha (0.1 - 0.001) 0.01

CATALOGUE RESULTS? SHOW CHART STATISTICS

↑

You receive the message "*time computation in progress*".  
Press the button "RESULTS?" to see if the results are available.

KB Dashboard KB Catalogue

know4business.appspot.com/KBCatalogue.py?catalogkey=ag9zfmtub3c0YnVzaW5lc3NyDwsSB0NhdfGFsb2cY8ZwGDA&proi

KB Catalogue Catalog #102001 Sample #101001 (sport\_r.txt)

Number of Groups (3 - 20) 4

Normalization(Max, Std, None) M

Start value of alpha (1.8 - 0.9) 1.8

End value of alpha (0.5 - 0.0001) 0.0001

Decreasing step of alpha (0.1 - 0.001) 0.01

CATALOGUE RESULTS? SHOW CHART STATISTICS

time computation in progress ...

At the end of the process CATALOG you can view the results.

### Data in the panel

rows = number of records

cols = number of columns / variables

xstds = negligible

cat = type of end condition



epoch1 = negligible  
epoch2 = last epoch with minimum error  
CVbC = CV before cataloging  
CV = CVaC after cataloging  
KIndex = Knowledge Index (0 minimum, 1 maximum)  
Running time = minutes, and seconds

KB Dashboard KB Catalogue

know4business.appspot.com/KBcatalogue.py?catalogkey=ag9zftub3c0YnVzaW5lc3NyDws5B0NhdGFsb2cY8ZwGDA&proi Google

KB Catalogue Catalog #102001 Sample #101001 (sport\_r.txt)

Number of Groups (3 - 20) 4

Normalization(Max, Std, None) M

Start value of alpha (1.8 - 0.9) 1.8

End value of alpha (0.5 - 0.0001) 0.0001

Decreasing step of alpha (0.1 - 0.001) 0.01

CATALOGUE RESULTS? SHOW CHART STATISTICS

rows 142

cols 13

xsds [0.0, 1.3303958251750452, 1.1720407746596482, 0.99131333135830324, 0.95621905526551132, 1.1947429085587673, 0.83186742099359268, 0.77618983327340429, 0.9489633505030447, 0.8875825475455047, 0.68115339906047823, 0.9750541321635503, 0.9409404176331756, 0.89830617433604965]

cat Min alpha reached

epoch1 Epoch 70 min err 9.15454 min epoch 66 alpha 0.55800

epoch2 \*\*\*\* Epoch 91 WITH MIN ERROR 6.914 alpha 0.16200

CVbC 1.8958

CVaC 0.3899

KIndex 0.7943

time 01:16.929324

Pressing the button SHOW, you activate a new panel in the top row of the browser used with the name KB Show (see arrow). Other panels are activated by pressing the buttons CHART (KB Chart) and STATISTICS (KB Statistics).

KB Dashboard KB Catalogue

know4business.appspot.com/KBcatalogue.py?catalogkey=ag9zfmtub3c0YnVzaW5lc3NyDws5B0NhdfGfb2cY8ZwGDA&pr...

KB Catalogue Catalog #102001 Sample #101001 (sport\_r.txt)

Number of Groups (3 - 20) 4

Normalization(Max, Std, None) M

Start value of alpha (1.8 - 0.9) 1.8

End value of alpha (0.5 - 0.0001) 0.0001

Decreasing step of alpha (0.1 - 0.001) 0.01

CATALOGUE RESULTS? SHOW CHART STATISTICS

rows 142

cols 13

xstds [0.0, 1.3303958251750452, 1.1720407746596482, 0.99131333135830324, 0.95621905526551132, 1.1947429085587673, 0.83186742099359268, 0.77618983327340429, 0.9489633505030447, 0.8875825475455047, 0.68115339906047823, 0.9750541321635503, 0.9409404176331756, 0.89830617433604965]

cat Min alpha reached

epoch1 Epoch 70 min err 9.15454 min epoch 66 alpha 0.55800

epoch2 \*\*\*\* Epoch 91 WITH MIN ERROR 6.914 alpha 0.16200

CVbC 1.8958

CVaC 0.3899

KIndex 0.7943

time 01:16.929324

Pressing the button OUTSRT you view the table of input records cataloged in Groups (listed in the first column), the second column contains the identification of records, while the subsequent columns contain the original input values.

KB Dashboard KB Catalogue KB Show

know4business.appspot.com/KBshow.py?file=outsrt&key=ag9zfmtub3c0YnVzaW5lc3NyDws5B0NhdfGfb2cY8ZwGDA

KB Show Catalog#102001 MEDSD GRID CV OUTSRT SOURCE (sport\_r.txt)

*Group*	Rec_Id	Sport	Club	DB_Movement	DB_Max_Cinetic_Power	DB_Timing_MCP	DB_Max_Speed	DB_Pushing_Time	DB_Pulse	DB_Max_Dynamic_Strength	DB_Relative_Strength	DB_Max_Strength_Q	DB_Max_Strength_T	DB_Elasticity_Q
G_00_00	R_106	Football	Udinese	11.00	33.70	80.00	2.60	150.00	131.00	2690.00	3.30	2600.00	1180.00	7.00
G_00_00	R_113	Football	Udinese	16.00	26.90	80.00	2.49	180.00	138.00	2320.00	3.20	2000.00	1190.00	6.00
G_00_00	R_123	Football	Udinese	15.00	27.60	60.00	2.50	150.00	102.00	2480.00	3.70	2270.00	1270.00	6.00
G_00_00	R_137	Football	Zwolle	22.00	25.10	75.00	2.52	175.00	131.00	2250.00	3.40	2070.00	920.00	6.00
G_00_00	R_143	Football	Zwolle	21.00	31.40	50.00	2.69	145.00	103.00	2640.00	3.90	2460.00	1220.00	6.00
G_00_00	R_144	Football	Zwolle	21.00	31.10	65.00	2.71	160.00	113.00	2300.00	3.40	2170.00	1730.00	6.00
G_00_00	R_147	Football	Zwolle	11.00	22.60	75.00	2.37	170.00	103.00	2720.00	3.30	2640.00	1070.00	9.00
G_00_00	R_151	Football	Zwolle	12.00	28.80	90.00	2.44	155.00	190.00	2650.00	3.20	2380.00	1370.00	8.00
G_00_00	R_190	Football	Padova	14.00	29.20	95.00	2.49	145.00	116.00	2670.00	3.60	2510.00	1130.00	8.00
G_00_00	R_58	Football	Udinese	119.00	27.70	95.00	2.57	205.00	171.00	2420.00	3.10	2000.00	1340.00	5.00
G_00_00	R_72	Football	Udinese	22.00	30.20	65.00	2.77	185.00	125.00	2610.00	3.60	2300.00	1280.00	5.00
G_00_01	R_107	Football	Udinese	12.00	23.20	55.00	2.61	200.00	224.00	2560.00	3.40	2350.00	1120.00	8.00
G_00_01	R_135	Football	Udinese	2.00	22.20	75.00	2.35	190.00	318.00	2720.00	3.30	2180.00	1210.00	11.00
G_00_01	R_142	Football	Zwolle	9.00	23.10	60.00	2.32	180.00	185.00	2450.00	3.30	2270.00	1210.00	9.00
G_00_01	R_146	Football	Zwolle	7.00	24.00	55.00	2.25	140.00	150.00	2240.00	3.50	1970.00	1130.00	11.00
G_00_01	R_158	Football	Zwolle	7.00	18.70	70.00	2.08	140.00	153.00	2350.00	3.10	2120.00	1410.00	13.00
G_00_01	R_160	Football	Padova	23.00	29.60	80.00	2.75	215.00	239.00	2450.00	3.30	2030.00	1170.00	5.00
G_00_01	R_173	Football	Padova	28.00	25.60	85.00	2.69	195.00	185.00	2220.00	3.00	2130.00	1220.00	4.00
G_00_01	R_195	Football	Padova	18.00	24.20	75.00	2.52	215.00	166.00	2390.00	3.10	2140.00	1280.00	6.00
G_00_01	R_205	Football	Padova	7.00	25.80	55.00	2.18	150.00	129.00	2280.00	3.70	1790.00	1130.00	7.00
G_00_01	R_60	Football	Udinese	2.00	27.30	70.00	2.50	170.00	225.00	2650.00	3.40	2120.00	1390.00	8.00
G_00_01	R_74	Football	Udinese	8.00	28.70	85.00	2.60	165.00	180.00	2610.00	3.20	2400.00	1280.00	7.00
G_00_02	R_110	Football	Udinese	5.00	20.70	75.00	2.57	190.00	269.00	2060.00	3.00	1880.00	1440.00	6.00
G_00_02	R_69	Football	Udinese	28.00	16.40	80.00	2.45	250.00	243.00	2350.00	2.70	2060.00	1230.00	6.00
G_00_02	R_85	Football	Udinese	28.00	38.40	130.00	2.80	200.00	191.00	2410.00	3.40	2050.00	1310.00	4.00
G_00_03	R_101	Football	Udinese	18.00	30.90	115.00	2.73	185.00	199.00	1810.00	3.00	1540.00	1040.00	4.00
G_00_03	R_102	Football	Udinese	18.00	28.40	115.00	2.37	160.00	78.00	2270.00	3.10	2100.00	980.00	6.00
G_00_03	R_105	Football	Udinese	24.00	20.40	170.00	2.61	250.00	248.00	2000.00	2.50	1490.00	1540.00	3.00
G_00_03	R_105	Football	Udinese	24.00	20.40	170.00	2.61	250.00	248.00	2000.00	2.50	1490.00	1540.00	3.00

http://know4business.appspot.com/serve/A...U0M805\_0RJPDk-kL9bkdsbnnd50LX0dPdMd5393Q

Pressing the button OUTSRT a second time you can download the file on your computer.  
The same goes for the other output files (MEDSD, GRID, CV).

KB Show Catalog#102001

MEDSD GRID CV **OUTSRT** SOURCE (sport\_r.txt)

Apertura di catalog.102001.sport\_r.txt.outsrt

È stato scelto di aprire  
catalog.102001.sport\_r.txt.outsrt  
che è un: Documento in testo semplice (3,6 kB)  
da: http://know4business.appspot.com

Che cosa deve fare Firefox con questo file?

☐ Apriro con gedit (predefinita)

☒ Salva file

☐ Da ora in avanti esegui questa azione per tutti i file di questo tipo.

Annulla OK

*Group*	Rec Id	Sport	Club	DB Movement	DB Max Cinetic Power	DB Timing MCF	DB Max Speed	DB Pushing Time	DB Pulse	DB Max Dynamic Strength	DB Relativ Strength	DB Max Strength	DB Elasticity Q
G 00 00	R 106	Football	Udinese	11.00	33.70	80.00	2.60	150.00	131.00	2690.00	3.30	1180.00	7.00
G 00 00	R 113	Football	Udinese	16.00	26.90	80.00	2.49	180.00	138.00	2320.00	3.20	1190.00	6.00
G 00 00	R 123	Football	Udinese	15.00	27.60	60.00	2.50	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 137	Football	Zwolle	22.00	25.10	75.00	2.52	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 143	Football	Zwolle	21.00	31.40	50.00	2.69	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 144	Football	Zwolle	21.00	31.10	65.00	2.71	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 147	Football	Zwolle	11.00	22.60	75.00	2.37	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 151	Football	Zwolle	12.00	28.80	90.00	2.44	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 190	Football	Padova	14.00	29.20	95.00	2.49	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 58	Football	Udinese	119.00	27.70	95.00	2.57	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 00	R 72	Football	Udinese	22.00	30.20	65.00	2.77	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 107	Football	Udinese	12.00	23.20	55.00	2.61	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 135	Football	Udinese	2.00	22.20	75.00	2.35	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 142	Football	Zwolle	9.00	23.10	60.00	2.32	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 146	Football	Zwolle	7.00	24.00	55.00	2.25	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 158	Football	Zwolle	7.00	18.70	70.00	2.08	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 160	Football	Padova	23.00	29.60	80.00	2.75	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 173	Football	Padova	28.00	25.60	85.00	2.69	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 195	Football	Padova	18.00	24.20	75.00	2.52	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 205	Football	Padova	7.00	25.80	55.00	2.18	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 60	Football	Udinese	2.00	27.30	70.00	2.50	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 01	R 74	Football	Udinese	8.00	28.70	85.00	2.60	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 02	R 110	Football	Udinese	5.00	20.70	75.00	2.57	150.00	100.00	2100.00	3.20	1220.00	6.00
G 00 02	R 69	Football	Udinese	28.00	16.40	80.00	2.45	250.00	243.00	2350.00	2.70	2000.00	6.00
G 00 02	R 85	Football	Udinese	28.00	38.40	130.00	2.80	200.00	191.00	2410.00	3.40	2050.00	4.00
G 00 03	R 101	Football	Udinese	18.00	30.90	115.00	2.73	185.00	199.00	1810.00	3.00	1540.00	4.00
G 00 03	R 102	Football	Udinese	18.00	28.40	115.00	2.37	160.00	78.00	2270.00	3.10	2100.00	6.00
G 00 03	R 105	Football	Udinese	24.00	20.40	170.00	2.61	250.00	248.00	2000.00	2.50	1490.00	3.00
G 00 03	R 108	Football	Udinese	1.00	23.60	100.00	2.50	175.00	238.00	2140.00	3.10	1550.00	6.00
G 00 03	R 111	Football	Udinese	31.00	38.90	100.00	3.24	195.00	243.00	2380.00	3.20	2140.00	3.00
G 00 03	R 128	Football	Udinese	23.00	22.10	115.00	2.63	190.00	180.00	2460.00	2.70	2380.00	7.00
G 00 03	R 138	Football	Zwolle	11.00	25.50	110.00	2.45	175.00	143.00	2020.00	2.90	1940.00	5.00

Press the CV button to get the display of the table of coefficients of variation of the groups for each column / variable.  
Low values of CV signify high homogeneity in the group for the variable / column in question: value to zero means that the variable in the group / column has all the same values.

\*Mean\* The column contains the average CV of the group concerned considering all the variables / columns.

\*Means\* The last line contains the weighted average (with the number of records) of the CVs of all groups for the variable / column in question.

\*CVinp\* The last line contains the values of CV of the input data before cataloging.

The comparison of the second last row with the last line makes it possible to check on what variables / columns have been any combinations of the same or similar values.

Variables / columns with negligible improvement (decrease in the lower CV) could be removed in subsequent processing, because of their limited impact in the cataloging process.

KB Show Catalog#102001

MEDSD GRID **CV** OUTSRT SOURCE (sport\_r.txt)

*Groups*	Sport	Club	DB Movement	DB Max Cinetic Power	DB Timing MCF	DB Max Speed	DB Pushing Time	DB Pulse	DB Max Dynamic Strength	DB Relativ Strength	DB Max Strength	DB Elasticity Q	*Mean*	N recs	
G 00 00	0.0000	0.2635	0.6250	0.2576	0.1258	0.3370	0.5189	1.0536	0.3124	0.2170	0.3392	0.8091	0.1443	0.3845	000011
G 00 01	0.0000	0.4845	0.4736	0.3848	0.1069	0.6572	0.5442	0.4614	0.3644	0.1901	0.3282	0.4854	0.5074	0.3837	000011
G 00 02	0.0000	0.0000	0.1593	1.0498	0.4956	0.3681	0.2929	0.2126	0.4405	0.3441	0.1702	0.2974	0.1286	0.3045	000003
G 00 03	0.0000	0.2597	0.4844	0.7698	1.0207	0.5021	0.3773	0.3763	0.7839	0.3650	0.8707	0.5570	0.2009	0.5052	000022
G 01 00	0.0000	0.3077	0.7459	0.2561	0.1286	0.3374	0.5746	0.8599	0.2531	0.1619	0.2442	0.4968	0.3219	0.3606	000015
G 01 01	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	000001
G 01 02	0.0000	0.0000	0.4185	0.5918	0.5779	0.5631	0.1776	0.3474	0.2366	0.2646	0.6494	0.4819	0.4657	0.3673	000006
G 01 03	0.0000	0.5657	0.4761	0.5625	1.0039	0.3478	0.2796	0.2472	0.5756	0.3129	0.6994	0.8552	0.1811	0.4698	000009
G 02 00	0.0000	0.0000	0.7976	0.2544	0.3686	0.2101	0.4581	0.5109	0.1283	0.1781	0.1504	0.5011	0.1683	0.2866	000008
G 02 01	0.0000	0.5294	0.5867	0.3200	0.1131	0.3838	0.4674	0.2947	0.2304	0.1119	0.2600	0.3250	0.6231	0.3266	000010
G 02 02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	000001
G 02 03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	000001
G 03 00	0.0000	0.3191	0.5105	0.1310	0.1327	0.2767	0.5224	0.3833	0.1304	0.1891	0.2147	0.5562	0.4119	0.2906	000014
G 03 01	0.0000	0.5714	0.8556	0.2380	0.0837	0.1953	1.0069	0.6318	0.1117	0.1901	0.0868	0.5116	1.4708	0.4580	000005
G 03 02	0.0000	0.0000	0.4370	0.3711	0.1782	0.2564	0.2474	0.5328	0.0713	0.1675	0.0700	0.5804	0.7621	0.2826	000004
G 03 03	0.0000	0.0000	0.8435	0.9682	0.9210	0.7102	0.3464	0.4806	0.3076	0.5025	0.4379	0.5424	0.1320	0.4763	000021
*Means*	0.0000	0.2554	0.5939	0.4907	0.4742	0.4293	0.4368	0.4970	0.3441	0.2654	0.4055	0.5404	0.3362	0.3899	000142
*CVinp*	1.6286	2.2956	2.3207	1.9849	2.0105	1.7151	2.0499	1.9731	1.7760	1.5019	1.9878	2.0637	1.3376	1.8958	000142

Rows No = 19

Press the button KB Catalogue and then use the tool CHART.



KB Catalogue Catalog #102001 Sample #101001 (sport\_r.txt)

Number of Groups (3 - 20) 4  
 Normalization(Max, Std, None) M  
 Start value of alpha (1.8 - 0.9) 1.8  
 End value of alpha (0.5 - 0.0001) 0.0001  
 Decreasing step of alpha (0.1 - 0.001) 0.01

CATALOGUE RESULTS? SHOW CHART STATISTICS

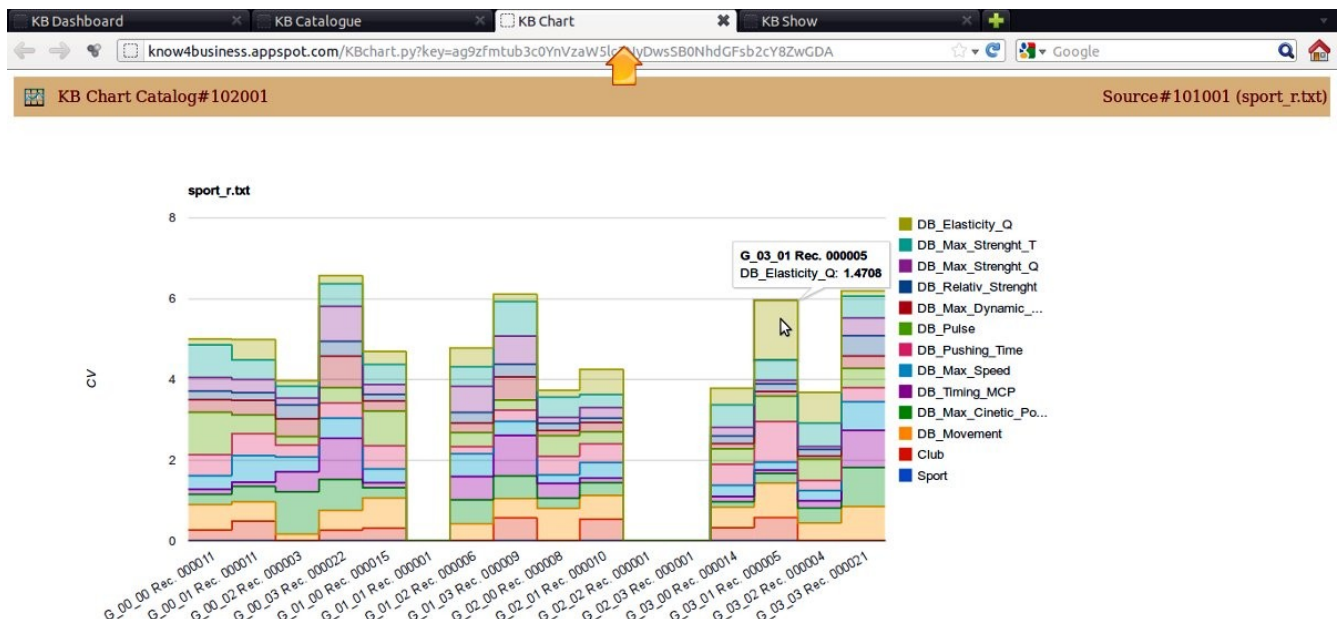
rows 142  
 cols 13  
 xsds [0.0, 1.3303958251750452, 1.1720407746596482, 0.99131333135830324, 0.95621905526551132, 1.1947429085587673, 0.83186742099359268, 0.77618983327340429, 0.9489633505030447, 0.8875825475455047, 0.68115339906047823, 0.9750541321635503, 0.9409404176331756, 0.89830617433604965]  
 cat Min alpha reached  
 epoch1 Epoch 70 min err 9.15454 min epoch 66 alpha 0.55800  
 epoch2 \*\*\*\* Epoch 91 WITH MIN ERROR 6.914 alpha 0.16200  
 CVbC 1.8958  
 CVaC 0.3899  
 KIndex 0.7943  
 time 01:16.929324

Press the button KB Chart.

## Chart

The CHART tool enables you to view the CV (coefficient of variation) of the variables / columns of the groups obtained from the catalog.

The CHART has the appearance of stacked vertical bars with different colors for each variable / column. In the abscissa are shown the groups with the frequencies of the records belonging to the group. The ordinate represents the scale of the cumulated CV: the lower the value the more the group is homogeneous inside. In the presence of CV values equal to zero, for constancy of the values in the group, the variable / column is not shown. Hovering the mouse over the colored bars are displayed the CV of the variable / column and the frequency of records in the group.



The graph shows that some variables / columns are not shown for all groups having a value equal to zero, as can be verified by viewing the output of the instrument CV SHOW.

Google Traduttore

KBGshow

KB Dashboard

KB Catalogue

KB Show

KBClassifier

KBGchart

know4business.appspot.com/KBshow.py?file=cv&key=ag9Zfmbut3c0YnVzaW5lc3NyDwsSB0NhdGFsb2Y8ZwGDA

google traduttore

KB Show Catalog#102001

MEDSD

GRID

CV

OUTSRT

SOURCE (sport\_r.txt)

Group	Sport	DB Max	DB Timing	DB Max Speed	DB Pushing Time	DB Pulse	DB Max Dynamic Strength	DB Relativ Strength	DB Max Strength Q	DB Max Strength T	DB Elasticity Q	*Mean*	N recs	
G 00 0	0.0000	0.2635	0.1258	0.3370	0.5189	1.0536	0.3124	0.2121	0.3392	0.8091	0.1443	0.3845	000011	
G 00 0	0.0000	0.4845	0.1069	0.6572	0.5442	0.4614	0.3644	0.1901	0.3282	0.4854	0.5974	0.3837	000011	
G 00 0	0.0000	0.0000	0.4956	0.3681	0.2929	0.2126	0.4405	0.3441	0.1702	0.2974	0.1286	0.3045	000003	
G 00 0	0.0000	0.2597	1.0207	0.5021	0.3773	0.3763	0.7839	0.3650	0.8707	0.5570	0.2009	0.5052	000022	
G 01 0	0.0000	0.3077	0.1286	0.3374	0.5746	0.8599	0.2531	0.1619	0.2442	0.4968	0.3219	0.3606	000015	
G 01 0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	000001	
G 01 0	0.0000	0.0000	0.5779	0.5631	0.1776	0.3474	0.2366	0.2646	0.6494	0.4819	0.4657	0.3673	000006	
G 01 0	0.0000	0.5657	1.0039	0.3478	0.2796	0.2472	0.5756	0.3129	0.6994	0.8552	0.1811	0.4698	000009	
G 02 0	0.0000	0.0000	0.2544	0.3686	0.2101	0.4581	0.5109	0.1283	0.1781	0.1504	0.5011	0.1683	0.2866	000008
G 02 0	0.0000	0.5294	0.1131	0.3838	0.4674	0.2947	0.2304	0.1119	0.2600	0.3250	0.6231	0.3266	000010	
G 02 0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	000001	
G 02 0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	000001	
G 03 0	0.0000	0.3191	0.1310	0.1327	0.2767	0.5224	0.3833	0.1304	0.1891	0.2147	0.5562	0.4119	0.2906	000014
G 03 0	0.0000	0.5714	0.2380	0.0837	0.1953	1.0069	0.6318	0.1117	0.1901	0.0868	0.5116	1.4708	0.4580	000005
G 03 0	0.0000	0.0000	0.3711	0.1782	0.2564	0.2474	0.5328	0.0713	0.1675	0.0700	0.5804	0.7621	0.2826	000004
G 03 0	0.0000	0.0000	0.9682	0.9210	0.7102	0.3464	0.4806	0.3076	0.5025	0.4379	0.5424	0.1320	0.4763	000021
*Mean*	0.0000	0.2554	0.4907	0.4742	0.4293	0.4368	0.4970	0.3441	0.2654	0.4055	0.5404	0.3362	0.3899	000142
*Cvmp*	1.6286	2.2956	1.9849	2.0105	1.7151	2.0499	1.9731	1.7760	1.5019	1.9878	2.0637	1.3376	1.8958	000142

Rows no = 19

## Classify

The tool allows the CLASSIFIER to classify an input file (SOURCE) by referring to a previous input file CATALOG similar in format and content: the same number of variables / columns and the same data format (number / text).

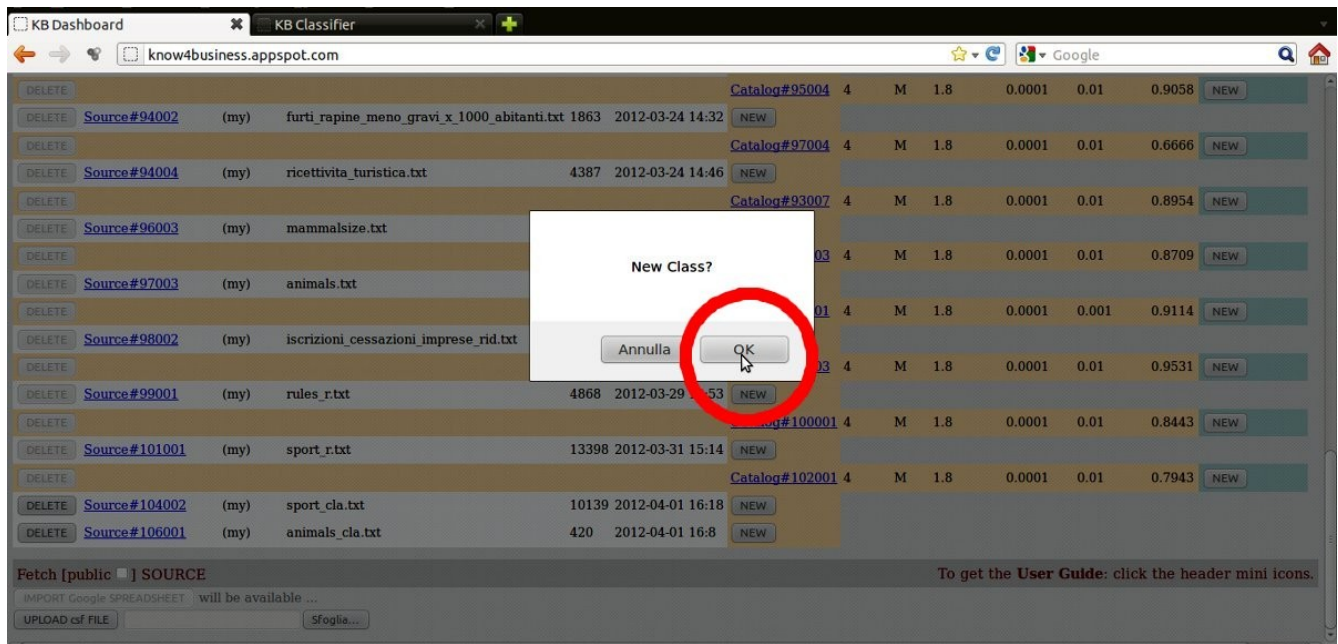
While the tool CATALOGUE requires non-trivial execution times, the Classifier tool works almost instantaneously simply by having to find the group most suitable placement of the recording in the matrix training GRID.

If the SOURCE that contains the data to be classified is not yet available, must choose (Sfoglia / Browse. ...) and then make the UPLOAD.

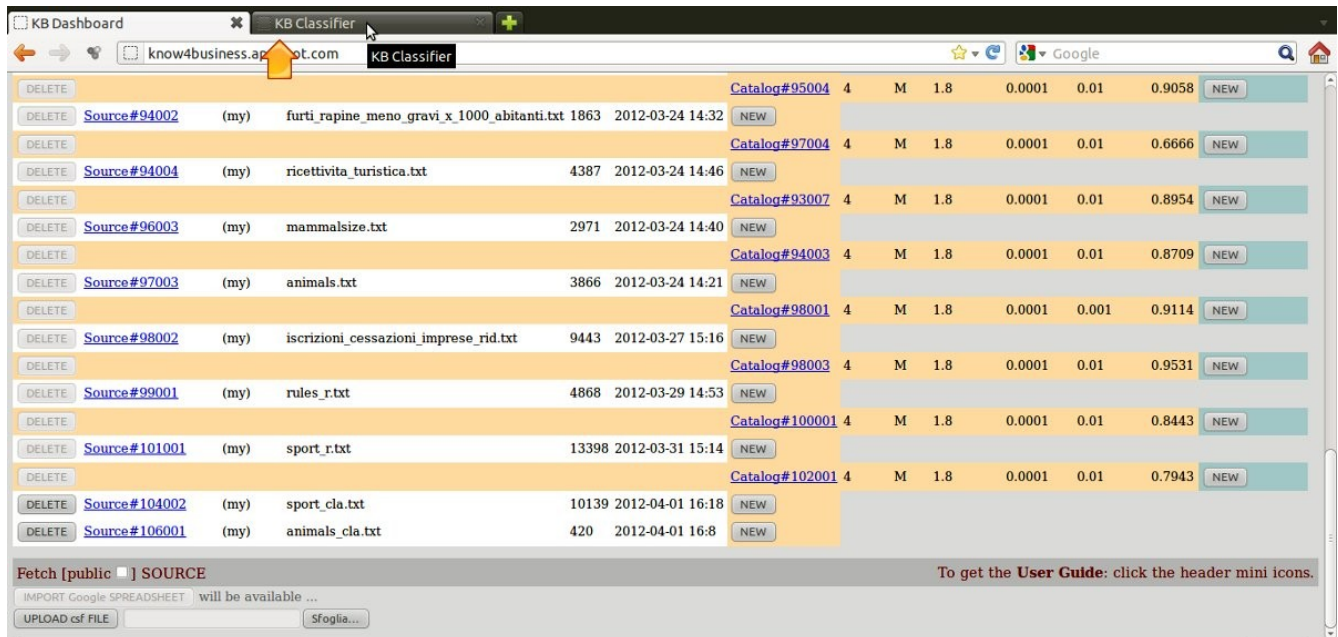
Referring to the File Catalog # 102001, press the NEW button to start a new classification.

Catalog#	Source	(my)	Name	Rows	Date	NEW
Catalog#95004				4	M 1.8	0.0001 0.01 0.9058
Catalog#94002	(my)	furti_rapine_meno_gravi_x_1000_abitanti.txt	1863	2012-03-24 14:32		
Catalog#97004			4387	2012-03-24 14:46		
Catalog#93007			2971	2012-03-24 14:40		
Catalog#94003	(my)	ricettivita_turistica.txt	3866	2012-03-24 14:21		
Catalog#98001	(my)	animals.txt	9443	2012-03-27 15:16		
Catalog#98003	(my)	iscrizioni_cessazioni_impres_rid.txt	4868	2012-03-29 14:53		
Catalog#100001	(my)	rules_r.txt	13398	2012-03-31 15:14		
Catalog#102001	(my)	sport_r.txt	10139	2012-04-01 16:18		
Catalog#104002	(my)	sport_cla.txt	420	2012-04-01 16:8		
Catalog#106001	(my)	animals_cla.txt				

Press the OK button to confirm.

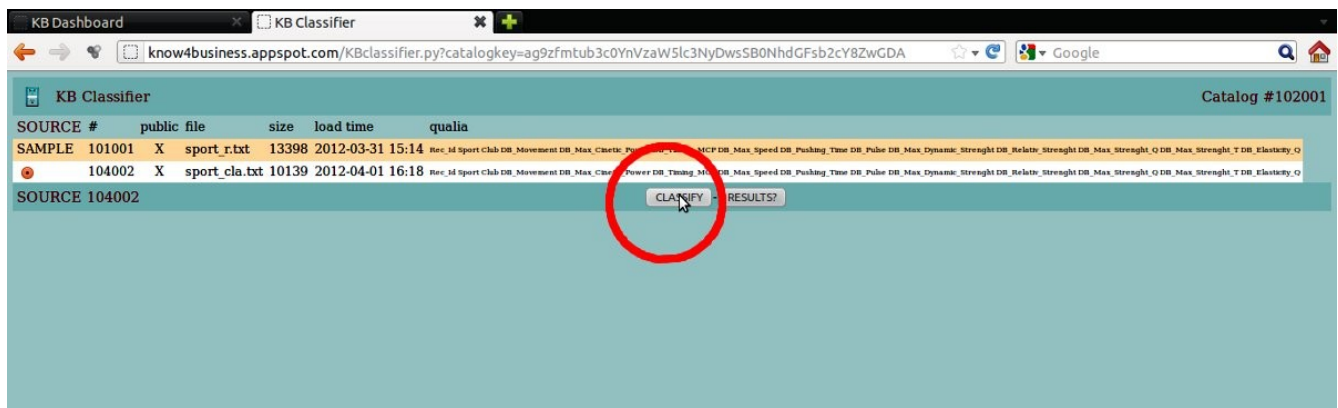


Press the KB Classifier panel.

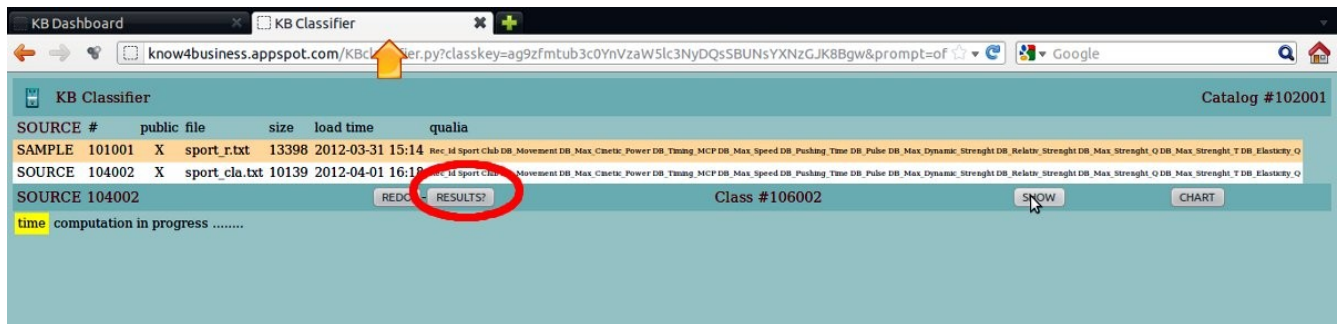


Choose File (SOURCE 104002) to be classified with respect to the reference file CATALOGUE.  
Press the button CLASSIFY to start classify.

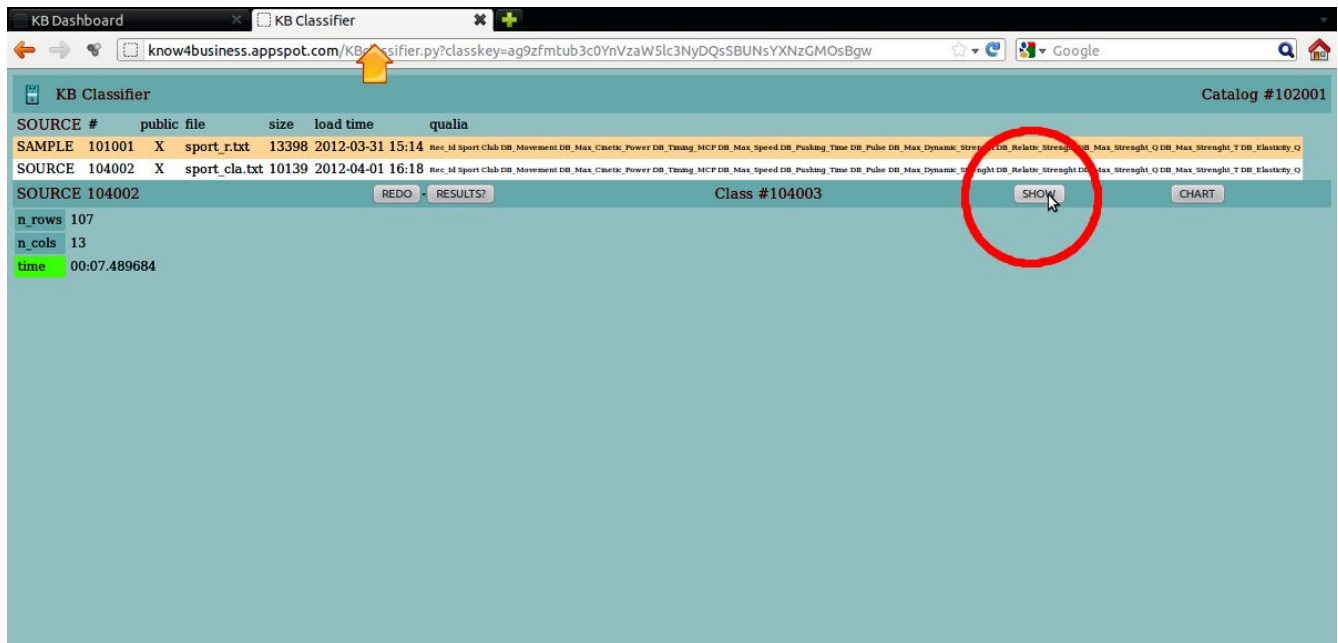




Press the button RESULTS? to check the results.



Press the SHOW button to access the results (OUTSRT, CV, MEDSD).



OUTSRT output in KB Show panel.

KB Dashboard KB Classifier KB Show

know4business.appspot.com/KBshow.py?file=outsr&key=ag9Zfntub3c0YnVzaW5lc3NyYXNzGMOsBgw

KB Show Class#104003

MEDSD CV **OUTSR** SOURCE (sport\_cla.txt)

Group	Rec Id	Sport	Club	DB Movement	DB Max Cyclic Power	DB Timing MCP	DB Max Speed	DB Pushing Time	DB Pulse	DB Max Dynamic Strength	DB Relate Strength	DB Max Weight G	DB Max Strength T	DB Elasticity G
G.00.00	R.106	Football	Udmere	11.00	33.70	60.00	2.60	150.00	131.00	2690.00	3.30	2600.00	1180.00	7.00
G.00.00	R.123	Football	Udmere	15.00	27.60	60.00	2.30	150.00	102.00	2480.00	3.70	2600.00	1270.00	6.00
G.00.00	R.137	Football	Zwalle	22.00	25.10	75.00	2.52	175.00	121.00	2230.00	3.40	2070.00	920.00	6.00
G.00.00	R.144	Football	Zwalle	21.00	31.10	65.00	2.71	160.00	113.00	2300.00	3.40	2170.00	1730.00	6.00
G.00.00	R.151	Football	Zwalle	12.00	28.80	90.00	2.44	155.00	190.00	2650.00	3.20	2380.00	1370.00	6.00
G.00.00	R.190	Football	Pudera	14.00	29.20	95.00	2.49	145.00	116.00	2670.00	3.60	2510.00	1130.00	6.00
G.00.00	R.36	Football	Udmere	119.00	27.70	95.00	2.57	205.00	171.00	2420.00	3.10	2000.00	1340.00	5.00
G.00.01	R.135	Football	Udmere	2.00	22.20	75.00	2.35	190.00	218.00	2720.00	3.30	2180.00	1210.00	11.00
G.00.01	R.142	Football	Zwalle	9.00	23.10	60.00	2.22	180.00	185.00	2450.00	3.30	2270.00	1210.00	9.00
G.00.01	R.146	Football	Zwalle	7.00	24.00	55.00	2.25	160.00	150.00	2240.00	3.30	1970.00	1130.00	11.00
G.00.01	R.158	Football	Zwalle	7.00	18.70	70.00	2.08	140.00	153.00	2150.00	3.10	2120.00	1410.00	13.00
G.00.01	R.160	Football	Pudera	23.00	29.60	80.00	2.75	215.00	239.00	2650.00	3.30	2030.00	1170.00	5.00
G.00.01	R.173	Football	Pudera	28.00	25.60	65.00	2.69	195.00	185.00	2220.00	3.00	2130.00	1220.00	4.00
G.00.01	R.193	Football	Pudera	18.00	24.20	75.00	2.52	215.00	166.00	2390.00	3.10	2160.00	1280.00	6.00
G.00.01	R.203	Football	Pudera	7.00	25.80	55.00	2.18	150.00	129.00	2280.00	3.70	1790.00	1130.00	7.00
G.00.02	R.110	Football	Udmere	5.00	20.70	75.00	2.37	190.00	209.00	2060.00	3.00	1880.00	1440.00	6.00
G.00.03	R.101	Football	Udmere	18.00	30.90	115.00	2.73	185.00	199.00	1810.00	3.00	1540.00	1040.00	4.00
G.00.03	R.102	Football	Udmere	16.00	28.40	115.00	2.37	160.00	78.00	2270.00	3.10	2100.00	980.00	6.00
G.00.03	R.105	Football	Udmere	24.00	20.40	170.00	2.61	250.00	248.00	2000.00	2.50	1490.00	1540.00	3.00
G.00.03	R.111	Football	Udmere	21.00	38.90	100.00	3.24	195.00	243.00	2380.00	3.20	2140.00	1410.00	3.00
G.00.03	R.138	Football	Zwalle	11.00	25.50	110.00	2.45	175.00	143.00	2020.00	2.90	1940.00	890.00	5.00
G.00.03	R.139	Football	Zwalle	27.00	22.80	135.00	2.69	265.00	207.00	2170.00	2.60	1790.00	1330.00	3.00
G.00.03	R.154	Football	Zwalle	30.00	31.40	135.00	2.69	195.00	156.00	2210.00	2.80	1930.00	1540.00	3.00
G.00.03	R.156	Football	Zwalle	7.00	19.20	105.00	2.47	205.00	274.00	2190.00	2.80	1600.00	1460.00	7.00
G.00.03	R.157	Football	Zwalle	19.00	27.40	100.00	2.88	195.00	222.00	2420.00	3.40	1920.00	1270.00	5.00
G.00.03	R.171	Football	Pudera	28.00	25.70	165.00	2.60	270.00	209.00	1820.00	2.70	1450.00	830.00	2.00
G.00.03	R.189	Football	Pudera	9.00	17.60	115.00	2.58	240.00	210.00	2040.00	2.90	1540.00	1440.00	4.00
G.00.03	R.49	Football	Udmere	16.00	18.30	100.00	2.26	220.00	173.00	1630.00	2.50	1450.00	950.00	4.00
G.00.03	R.51	Football	Udmere	23.00	19.90	100.00	2.60	230.00	160.00	1670.00	2.60	1710.00	1150.00	5.00
G.00.03	R.84	Football	Udmere	21.00	30.70	100.00	2.94	165.00	146.00	1860.00	3.10	1600.00	1020.00	4.00
G.00.03	R.85	Football	Udmere	28.00	30.40	130.00	2.80	200.00	191.00	2410.00	3.40	2050.00	1310.00	4.00
G.00.03	R.88	Football	Udmere	26.00	43.10	130.00	3.22	195.00	191.00	2430.00	3.00	2120.00	1240.00	4.00
G.00.03	R.96	Football	Udmere	3.00	18.70	150.00	2.62	255.00	328.00	2130.00	2.90	1170.00	1550.00	4.00
G.00.01	R.116	Football	Udmere	27.00	33.30	50.00	2.77	150.00	103.00	2750.00	3.90	2710.00	1170.00	7.00
											3.20	2300.00	1390.00	6.00

http://know4business.appspot.com/serve/AMI...E2R\_AqZv\_EtL2i9QtusNslu2xvFOhHfNv-HqjXZnjw

CV output in KB Show panel.

KB Dashboard KB Classifier KB Show

know4business.appspot.com/KBshow.py?file=cv&key=ag9Zfntub3c0YnVzaW5lc3NyYXNzGMOsBgw

KB Show Class#104003

MEDSD CV **OUTSR** SOURCE (sport\_cla.txt)

Group	Sport	Club	DB Movement	DB Max Cyclic Power	DB Timing MCP	DB Max Speed	DB Pushing Time	DB Pulse	DB Max Dynamic Strength	DB Relate Strength	DB Max Weight G	DB Max Strength T	DB Elasticity G	"Mean"	"N. row"
G.00.00	0.0000	0.3149	0.6632	0.2181	0.1139	0.2297	0.5914	1.0144	0.3142	0.3247	0.6241	0.1225	0.3628	0.00007	
G.00.01	0.0000	0.5931	0.3093	0.4451	0.1028	0.8113	0.6079	0.5754	0.3866	0.3071	0.4397	0.6071	0.4187	0.00008	
G.00.02	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.00001	
G.00.03	0.0000	0.2831	0.4315	0.7329	0.9781	0.4283	0.3941	0.4284	0.8061	0.4029	0.6015	0.1933	0.4993	0.00017	
G.01.00	0.0000	0.3464	0.8387	0.1436	0.1441	0.3011	0.5402	0.8906	0.2197	0.1614	0.2224	0.5790	0.1429	0.3501	0.00008
G.01.02	0.0000	0.4138	0.6091	0.0000	0.9863	0.5838	0.1065	0.4636	0.2401	0.2160	0.6362	0.8548	0.6316	0.4412	0.00004
G.01.03	0.0000	0.5756	0.5559	0.6011	0.9971	0.3504	0.2320	0.2537	0.6060	0.3431	0.6193	0.7210	0.1870	0.4801	0.00007
G.02.00	0.0000	0.0000	0.6950	0.2733	0.2714	0.1926	0.2530	0.5628	0.2430	0.2158	0.2296	0.4315	0.2306	0.2082	0.00006
G.02.01	0.0000	0.5099	0.4439	0.3652	0.1105	0.4183	0.5139	0.3554	0.2241	0.1332	0.2785	0.3286	0.2931	0.2133	0.00007
G.02.03	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.00001	
G.03.00	0.0000	0.2795	0.4914	0.1419	0.1213	0.2701	0.5564	0.3782	0.1367	0.1756	0.2210	0.5138	0.4478	0.2872	0.00012
G.03.01	0.0000	0.5712	0.8550	0.2452	0.0837	0.1790	1.0009	0.6710	0.1115	0.1961	0.0921	0.4863	1.4708	0.4572	0.00005
G.03.03	0.0000	0.0000	0.4370	0.8008	0.1782	0.2279	0.2474	0.5788	0.0716	0.1675	0.0756	0.5323	0.7621	0.2820	0.00004
G.03.03	0.0000	0.0000	0.8103	0.1055	0.0593	0.7154	0.3403	0.5217	0.2482	0.5230	0.4467	0.5119	0.1313	0.4766	0.00018
"Mean"	0.0000	0.2632	0.5828	0.4976	0.4869	0.4203	0.4378	0.5262	0.3422	0.2839	0.5317	0.2437	0.2943	0.2943	0.00107
"CVmp"	1.5623	2.1766	1.9244	1.7374	1.7961	1.4777	1.8391	1.7431	1.3787	1.3625	1.7466	1.7372	1.1921	1.6823	0.00107

Rows No = 17

Press the panel KB Classifier and then CHART button.

KB Classifier

Catalog #102001

SOURCE #	public	file	size	load time	qualia
SOURCE 101001	X	sport_r.txt	13398	2012-03-31 15:14	Rec_M Sport Club DB_Movement DB_Max_Cinetic_Power DB_Timing_MCP DB_Max_Speed DB_Pushing_Time DB_Pulse DB_Max_Dynamic_Streight DB_Relate_Streight DB_Max_Streight_Q DB_Max_Streight_T DB_Elasticity_Q
SOURCE 104002	X	sport_cla.txt	10139	2012-04-01 16:18	Rec_M Sport Club DB_Movement DB_Max_Cinetic_Power DB_Timing_MCP DB_Max_Speed DB_Pushing_Time DB_Pulse DB_Max_Dynamic_Streight DB_Relate_Streight DB_Max_Streight_Q DB_Max_Streight_T DB_Elasticity_Q

REDO RESULTS? SHOW CHART

Class #104003

n\_rows 107  
n\_cols 13  
time 00:07.489684

Press the panel KB Chart.

KB Classifier

Catalog #102001

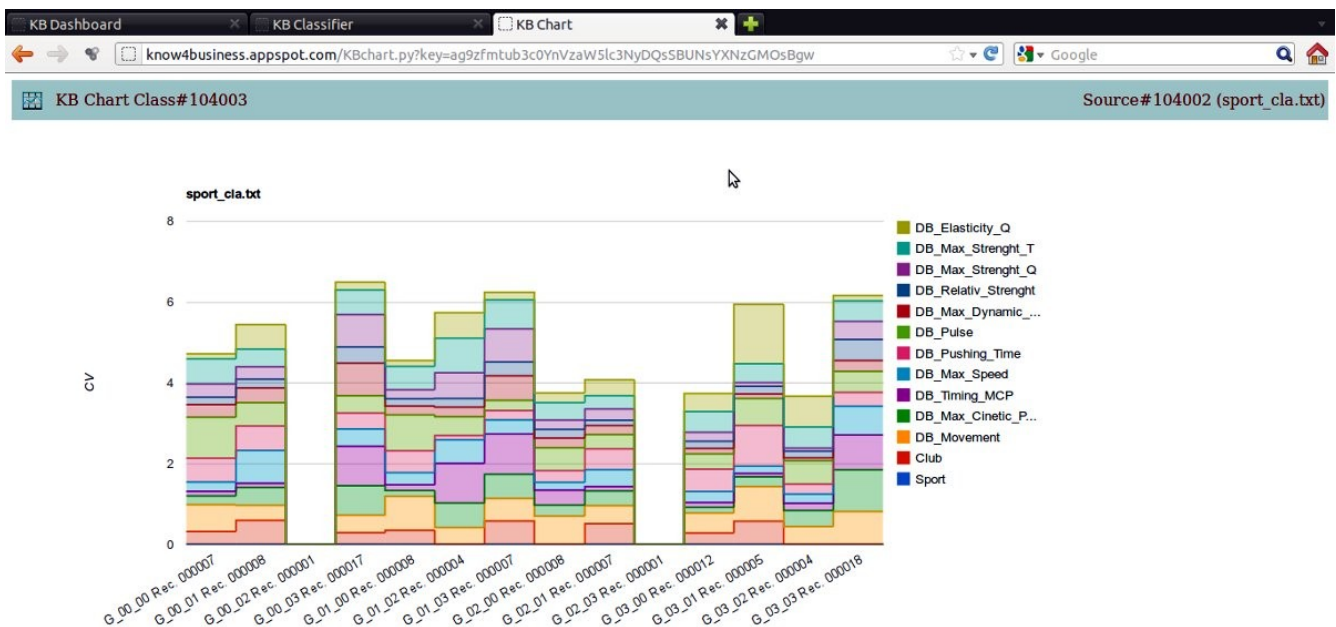
SOURCE #	public	file	size	load time	qualia
SOURCE 101001	X	sport_r.txt	13398	2012-03-31 15:14	Rec_M Sport Club DB_Movement DB_Max_Cinetic_Power DB_Timing_MCP DB_Max_Speed DB_Pushing_Time DB_Pulse DB_Max_Dynamic_Streight DB_Relate_Streight DB_Max_Streight_Q DB_Max_Streight_T DB_Elasticity_Q
SOURCE 104002	X	sport_cla.txt	10139	2012-04-01 16:18	Rec_M Sport Club DB_Movement DB_Max_Cinetic_Power DB_Timing_MCP DB_Max_Speed DB_Pushing_Time DB_Pulse DB_Max_Dynamic_Streight DB_Relate_Streight DB_Max_Streight_Q DB_Max_Streight_T DB_Elasticity_Q

REDO RESULTS? SHOW CHART

Class #104003

n\_rows 107  
n\_cols 13  
time 00:07.489684

Chart of sport\_cla.txt classified (#104003).



## Statistics

Press the KB Catalogue panel and then press STATISTICS button.

The screenshot shows the 'KB Catalogue' interface. The title bar indicates 'KB Dashboard' and 'KB Catalogue'. The address bar shows a URL from 'know4business.appspot.com'. The main content area displays a form for 'Catalog #102001' with parameters: Number of Groups (3 - 20) set to 4, Normalization(Max, Std, None) set to M, Start value of alpha (1.8 - 0.9) set to 1.8, End value of alpha (0.5 - 0.0001) set to 0.0001, and Decreasing step of alpha (0.1 - 0.001) set to 0.01. Below the form are buttons for 'CATALOGUE', 'RESULTS?', 'SHOW', 'CHART', and 'STATISTICS'. The 'STATISTICS' button is circled in red. Below the buttons is a table with columns: rows, cols, xsds, cat, epoch1, epoch2, CVbC, CVaC, KIndex, and time. The table contains data for 142 rows and 13 columns.

STATISTICS is the tool that calculates the statistics to display the characteristics of synthesis and breakdown of groups created by CATALOGUE.

You access the STATISTICS page where parameters are requested:

Group Consistency % (0 – 100) where 0 means no consistency

Variable Consistency % (0 – 100) where 0 means no consistency

Select groups containing records >= (to exclude groups with few records)

Select groups containing records <= (to exclude groups with many records)

Summary / Detail report (S/D) (if Detail, shows the data of variables / columns)

if Detail, display input records (Y / N) (if Y shows the detailed records in the SAMPLE)

Press the CALCULATE button.



KB Dashboard KB Catalogue KB Statistics

know4business.appspot.com/KBstatistics.py?catkey=ag9zfmtub3c0YnVzaW5lc3NyDwsSB0NhdGFsb2cy8ZwGDA

KB Statistics Catalog #102001 Sample #101001 (sport\_r.txt)

Stat #	omog_perc	omog_vari	rec_min	rec_max	est_rapp	inp_rapp
<input checked="" type="radio"/> 103001	10	10	1	100	S	
<input type="radio"/> 104001	10	10	1	100	D	Y

Stat #103001

Group Consistency (0 - 100)

Variable Consistency (0 - 100)

Select groups containing records >=

Select groups containing records <=

Summary / Detail report (S / D)

if Detail, display input records (Y / N)

time(micro) None

The output you get.

KB Dashboard KB Catalogue KB Statistics

know4business.appspot.com/KBstatistics.py?statkey=ag9zfmtub3c0YnVzaW5lc3NyDAs5BFN0YXQy2aQGD&catkey=ag9z

Group Consistency (0 - 100)

Variable Consistency (0 - 100)

Select groups containing records >=

Select groups containing records <=

Summary / Detail report (S / D)

if Detail, display input records (Y / N)

time(micro) 757397

rapport Min Perc. of group Consistency: 0 Min Perc. variable Consistency: 0 Min Number of records: 1 Max Number of records: 100

rapport by ROBERTO BELLO (COPYRIGHT MARCH 2011 ALL RIGHTS RESERVED)

rapport	G_00_00	Consistency	0.3845	%Consistency	80	Records	11	%Records	7.75
rapport	G_00_01	Consistency	0.3837	%Consistency	80	Records	11	%Records	7.75
rapport	G_00_02	Consistency	0.3045	%Consistency	84	Records	3	%Records	2.11
rapport	G_00_03	Consistency	0.5052	%Consistency	74	Records	22	%Records	15.49
rapport	G_01_00	Consistency	0.3606	%Consistency	81	Records	15	%Records	10.56
rapport	G_01_01	Consistency	0.0000	%Consistency	100	Records	1	%Records	0.70
rapport	G_01_02	Consistency	0.3673	%Consistency	81	Records	6	%Records	4.23
rapport	G_01_03	Consistency	0.4698	%Consistency	76	Records	9	%Records	6.34
rapport	G_02_00	Consistency	0.2866	%Consistency	85	Records	8	%Records	5.63
rapport	G_02_01	Consistency	0.3266	%Consistency	83	Records	10	%Records	7.04
rapport	G_02_02	Consistency	0.0000	%Consistency	100	Records	1	%Records	0.70
rapport	G_02_03	Consistency	0.0000	%Consistency	100	Records	1	%Records	0.70

Press the NEW button to get another statistic; fill in the form and press the CALCULATE button

KB Dashboard KB Catalogue KB Statistics

know4business.appspot.com/KBstatistics.py?catkey=ag9zfmtub3c0YnVzaW5lc3NyDwsSB0NhdGFsb2cy8ZwGDA&statkey=

KB Statistics Catalog #102001 Sample #101001 (sport\_r.txt)

Stat #	omog_perc	omog_vari	rec_min	rec_max	est_rapp	inp_rapp
<input type="radio"/> 103001	0	0	1	100	S	
<input type="radio"/> 104001	10	10	1	100	D	Y
<input checked="" type="radio"/> 109001	10	10	1	100	S	

Stat #109001

Group Consistency (0 - 100)

Variable Consistency (0 - 100)

Select groups containing records >=

Select groups containing records <=

Summary / Detail report (S / D)

if Detail, display input records (Y / N)

time(micro) None

The output you get.

KB Dashboard KB Catalogue KB Statistics

know4business.appspot.com/KBStatistics.py?statkey=ag9zfmtub3c0YnVzaW5lc3NyDA5S8FN0YXQyYdMGDA&catkey=ag9z

Stat #109001


Group Consistency (0 - 100)

Variable Consistency (0 - 100)

Select groups containing records >=

Select groups containing records <=

Summary / Detail report (S / D)

if Detail, display input records (Y / N)  

time(micro) 628549

rapport Min Perc. of group Consistency: 10 Min Perc. of variable Consistency: 10 Min Number of records: 1 Max Number of records: 100

rapport by ROBERTO BELLO (COPYRIGHT MARCH 2011 ALL RIGHTS RESERVED)

rapport G\_00\_00 Consistency 0.3845 %Consistency 80 Records 11 %Records 7.75

rapport \*\*\* Sport Consistency 0.0000 %Consistency 100.00

rapport Value Football Frequency 11 Percentage 100.00

rapport \*\*\* Club Consistency 0.2635 %Consistency 31.47

rapport Value Zwolle Frequency 5 Percentage 45.00

rapport Value Udinese Frequency 5 Percentage 45.00

rapport Value Padova Frequency 1 Percentage 9.00

rapport \*\*\* DB\_Max\_Cinetic\_Power Consistency 0.2576 %Consistency 33.00

rapport Min 22.60 Max 33.70 Step 2.78

rapport First Quartile (end) 25.38 Frequency % 18.18

rapport Second Quartile (end) 28.15 Frequency % 27.27

The output you get: frequencies for text data, quartiles for numeric data.

KB Dashboard KB Catalogue KB Statistics

know4business.appspot.com/KBStatistics.py?statkey=ag9zfmtub3c0YnVzaW5lc3NyDA5S8FN0YXQyYdMGDA&catkey=ag9z

Summary / Detail report (S / D)

if Detail, display input records (Y / N)

time(micro) 628549

rapport Min Perc. of group Consistency: 10 Min Perc. of variable Consistency: 10 Min Number of records: 1 Max Number of records: 100

rapport by ROBERTO BELLO (COPYRIGHT MARCH 2011 ALL RIGHTS RESERVED)

rapport G\_00\_00 Consistency 0.3845 %Consistency 80 Records 11 %Records 7.75

rapport \*\*\* Sport Consistency 0.0000 %Consistency 100.00

rapport Value Football Frequency 11 Percentage 100.00

rapport \*\*\* Club Consistency 0.2635 %Consistency 31.47

rapport Value Zwolle Frequency 5 Percentage 45.00

rapport Value Udinese Frequency 5 Percentage 45.00

rapport Value Padova Frequency 1 Percentage 9.00

rapport \*\*\* DB\_Max\_Cinetic\_Power Consistency 0.2576 %Consistency 33.00

rapport Min 22.60 Max 33.70 Step 2.78

rapport First Quartile (end) 25.38 Frequency % 18.18

rapport Second Quartile (end) 28.15 Frequency % 27.27

rapport Third Quartile (end) 30.92 Frequency % 27.27

rapport Fourth Quartile (end) 33.70 Frequency % 27.27

rapport \*\*\* DB\_Timing\_MCP Consistency 0.1258 %Consistency 67.28

rapport Min 50.00 Max 95.00 Step 11.25

rapport First Quartile (end) 61.25 Frequency % 18.18

rapport Second Quartile (end) 72.50 Frequency % 18.18

The output you get with the option: *Detail, display input records* = Y.



KB Dashboard KB Catalogue KB Statistics

know4business.appspot.com/KBstatistics.py?statkey=ag9zfmtub3c0YnVzaW5lc3NyDAs5BFN0YXQYsdsGDA&catkey=ag9zf Google

DELETE NEW CALCULATE

time(micro) 110229

rapport Min Perc. of group Consistency: 10 Min Perc. of variable Consistency: 10 Min Number of records: 1 Max Number of records: 100

rapport by ROBERTO BELLO (COPYRIGHT MARCH 2011 ALL RIGHTS RESERVED)

rapport G\_00\_00 Consistency 0.3845 %Consistency 80 Records 11 %Records 7.75

rapport \*\*\* report Consistency 0.0000 %Consistency 100.00

rapport G\_00\_00 ID record R\_106 Value Football

rapport G\_00\_00 ID record R\_113 Value Football

rapport G\_00\_00 ID record R\_123 Value Football

rapport G\_00\_00 ID record R\_137 Value Football

rapport G\_00\_00 ID record R\_143 Value Football

rapport G\_00\_00 ID record R\_144 Value Football

rapport G\_00\_00 ID record R\_147 Value Football

rapport G\_00\_00 ID record R\_151 Value Football

rapport G\_00\_00 ID record R\_190 Value Football

rapport G\_00\_00 ID record R\_58 Value Football

rapport G\_00\_00 ID record R\_72 Value Football

rapport Value Football Frequency 11 Percentage 100.00

rapport \*\*\* Club Consistency 0.2635 %Consistency 31.47

rapport G\_00\_00 ID record R\_106 Value Udinese

rapport G\_00\_00 ID record R\_113 Value Udinese

rapport G\_00\_00 ID record R\_123 Value Udinese

rapport G\_00\_00 ID record R\_137 Value Zwolle

## Indice

Know4Business.....	1
Introduction.....	1
Dashboard.....	2
Constraints on sources.....	4
Suggestions for sources.....	5
Start cataloging.....	6
Catalogue.....	7
Chart.....	12
Classify.....	13
Statistics.....	18