# Subject

**Starting from the Tanagra's 1.4.11 version**, a new EXCEL add-in is available. It enables to define a data mining analysis directly from EXCEL spreadsheet without closing the EXCEL session.

The main asset of this functionality is that we can perform all data preparation (data transformation, feature construction, etc.) and basic descriptive statistics (mean, standard deviation, pivot table, etc.) in the spreadsheet. Then we call TANAGRA, from EXCEL, only for advanced machine learning technique.

In this tutorial, we show how to install and use this EXCEL add-in.

## Installing the new EXCEL add-in

### Checking presence of the Add-In

This new add-in is available from the 1.4.11 version. See in the TANAGRA directory if the **TANAGRA.XLA** really exists (with the standard installation, the TANAGRA's directory is usually *« c:\program files\tanagra »*). **Do not move this file**.

#### Installing the Add-In

In the next step, we must install the add-in in the spreadsheet. We click on the following menu.

N 12	Aicrosoft Exce	el - Classeur1							X
	<u>Fi</u> chier <u>E</u> dition	Affichage Ins	ertion Forma <u>t</u>	Outils Données Fer	jêtre <u>?</u>			_ 8	'×
	🖻 🖪 🍯	🗟 🚏 🐰 I	h 🛍 💅 🛛	Euro Conversion		ZI 🛍 4	3 🕄	°, G	» •
	E12 💌	=		<u>V</u> aleur cible					
	A	В	С	<u>S</u> olveur		F	G	Н	
1				Macro	•				_
2				Macros compléme	ntaires M	-			-
3 - A				×	W	R			-
5									- 1
6									
7									
8									_
9									_
10									
12									
13					i				
H A	Feuil	L / Feuil2 / Fe	uil3 /		•			•	Ī
Deg	ssin + 🗟 🌀	Formes automa	atiques 👻 🔨	ヽ □ ○ 🖾 🐗	🙍 🔌	<u>/</u> - <u>A</u> - :	= = ≓ □	<i>i</i> -	
Prêt	t						NUM		

A dialog box appears, we select the XLA file in TANAGRA's directory.

#### Didacticiel - Etudes de cas



The add-in is downloaded, we check if it is really activated.

Macro complémentaire	? 🗙
Macros complémentaires disponibles : Liaisons Access Mise à jour des liaisons de macros com MS Query Numérotation de formulaires ODBC V Solveur V Tanagra Utilitaire d'analyse Utilitaire d'analyse - VBA Utilitaires pour les modèles V	OK Annuler <u>P</u> arcourir
-Tanagra	

A new menu is now available in the EXCEL spreadsheet.

	Microsoft Exce	el - Classeur1	1				
	<u>Fi</u> chier <u>E</u> dition	Affichage Inse	ertion Forma <u>t</u> ;	<u>O</u> utils <u>D</u> onnées	Fe <u>n</u> être <u>?</u> T	anagra	<u>- 8 ×</u>
] 🗅	🖻 🖬 🎒	🗟 🚏 🐰 🛙	te 🛍 💅 🖂	) + Ci + 🍓	, Σ 🥂 🚟	Execute Tanagra	jG ₽
	E5 🔽	=				About	
	A	В	С	D	E –		
1							
2							
3							
4							
5						1	
6							
7							
8							-
	🕩 🕨 🕇 Feuild	L / Feuil2 / Fei	uil3 /		•		
De	<u>s</u> sin + 🔓 🌀	Formes automa	atiques 🗸 🔪		🐴 🙍 🔌 ·	• 🚣 • 📥 • 🔳	≡ 🕤 ᆠ
Prê	t					NUM	

From now, until we remove this add-in, this menu is always available when we start the spreadsheet.

### Working on a dataset

We use the Quinlan's WEATHER.XLS dataset (1993) in order to show the utilization of this add-in.

× N	licrosoft	Excel - weat	ther.xls					[	
	<u>Fi</u> chier <u>E</u> d	ition <u>A</u> ffichag	e <u>I</u> nsertion	Forma <u>t O</u> ul	tils <u>D</u> onnées	Fe <u>n</u> être	<u>?</u> Tanagra		_ 8 ×
	🛩 🔛	a 🕽 🖉	አ 🖻 🖻	S 🔊 🔊	- Ci - 🍓	Σf*		11 🕄	* G *
-	F9	-	=						
	A	В	С	D	E	F	G	Н	
1	Outlook	Temp	Humidity	Windy	Class				
2	sunny	75	70	yes	Play				
3	sunny	80	90	yes	DontPlay				
4	sunny	85	85	no	DontPlay				
5	sunny	72	95	no	DontPlay				
6	sunny	69	70	no	Play				
7	overcast	72	90	yes	Play				
8	overcast	83	78	no	Play				
9	overcast	64	65	yes	Play		_		
10	overcast	81	75	no	Play				
11	rain	71	80	yes	DontPlay				
12	rain	65	70	yes	DontPlay				
13	rain	75	80	no	Play				
14	rain	68	80	no	Play				
15	rain	70	96	no	Play				
16									
17									
	• • • •	eather /				•		1	<b>N</b> I
Deg	isin + <table-cell></table-cell>	🕝 Formes a	automatiques	- 🔨 🗙		4 🙎	<u>ð</u> - <u> -</u>	<mark>∧</mark> - ≡	= 🍘
Prêt	t						NU	M	

#### Selecting the dataset range

Before we activate the menu, we must **select the dataset range**.

The first row stands for the attribute name. The determination of the data type relies on the first row of the data: if it can be transformed in a numeric value, the variable is turned into a continuous attribute; it is defined as a discrete one otherwise.

N 12	Aicrosoft E	xcel - weat	ther.xls					
	<u>Fi</u> chier <u>E</u> diti	on <u>A</u> ffichage	e <u>I</u> nsertion	Forma <u>t</u> Out	ils <u>D</u> onnées	Fe <u>n</u> être <u>?</u>	Tanagra	_ 8 ×
0	🖻 🖬 👹	B 🛕 💖	አ 🖻 🛍	S 🔊 🗠	- 🤮 Σ ,	f≈ 🕃 ⋛↓	10. 🕐	°, G°,
-	E15	-	= Play					2
	Α	В	C	D	E	F	G	H.
1	Outlook	Temp	Humidity	Windy	Class			
2	sunny	75	70	yes	Play			
3	sunny	80	90	yes	DontPlay			
4	sunny	85	85	no	DontPlay			
5	sunny	72	95	no	DontPlay			
6	sunny	69	70	no	Play			
7	overcast	72	90	yes	Play			
8	overcast	83	78	no	Play			
9	overcast	64	65	yes	Play			
10	overcast	81	75	no	Play			
11	rain	71	80	yes	DontPlay			
12	rain	65	70	yes	DontPlay			
13	rain	75	80	no	Play			
14	rain	68	80	no	Play			
15	rain	70	96	no	Play ,			
16								
17								
18								<b>_</b>
	I ► ► \we	eather /			•			
Deg	ssin + 🗟 🙆	Formes a	automatiques	- 🔨 🛰 [		4 🙎 🗳	• 🚄 •	₽°,
Prêt	t		So	mme=2154	1	NU	M	

### The TANAGRA / EXECUTE TANAGRA menu

We click on the TANAGRA / EXECUTE TANAGRA menu in order to perform a data mining analysis.

× N	licrosoft E	xcel - weat	ther.xls					
	<u>Fi</u> chier <u>E</u> diti	ion <u>A</u> ffichage	e <u>I</u> nsertion	Forma <u>t</u> Out	ils <u>D</u> onnées	Fe <u>n</u> être <u>?</u>	Tanagra	_ 8 ×
	🛩 🖬 🤞	i 🕹 🗟 🕹	አ 🖻 🛍	s 💅 🗠 -	- 🍓 Σ	f≈ 🕃 🛃	Execut	e Tanagra
	E15	<b>T</b>	= Play				About.	
	Α	В	С	D	E	F		
1	Outlook	Temp	Humidity	Windy	Class			
2	sunny	75	70	yes	Play			
3	sunny	80	90	yes	DontPlay			
4	sunny	85	85	no	DontPlay			
5	sunny	72	95	no	DontPlay			
6	sunny	69	70	no	Play			
7	overcast	72	90	yes	Play			
8	overcast	83	78	no	Play			
9	overcast	64	65	yes	Play			
10	overcast	81	75	no	Play			
11	rain	71	80	yes	DontPlay			
12	rain	65	70	yes	DontPlay			
13	rain	75	80	no	Play			
14	rain	68	80	no	Play			
15	rain	70	96	no	Play			
16								
17								
18								-
4	► ► we	eather /	1		•			
Des	șin + 🍃 🤅	Formes a	automatiques	• \ 🔌 [		4 🙎 🎐	<mark>) - 🚄 - (</mark>	Ĵ ÷
Prêt			So	mme=2154	1	NU	M	

In the next dialog box, we can check and reset the range selection.

<b>N</b>	licrosoft E	xcel - weat	ther.xls					
	<u>Fi</u> chier <u>E</u> diti	on <u>A</u> ffichage	e <u>I</u> nsertion	Forma <u>t O</u> ut	ils <u>D</u> onnées	Fe <u>n</u> être <u>?</u>	Tanagra	_ 8 ×
	📽 🖬 (é	5 Q. V	አ 🖻 🖻	S 10.	- 🝓 Σ ;	f≈ 🔂 ĝ↓	1	° G °
	A1	-	= Play					
	A	В	С	D	E	F	G	- + <b>-</b>
1	Outlook	Temp	Humidity	Windy	Class			
2	sunny	75	70	yes	Play			
3	sunny	80	90	yes	DontPlay			
4	sunny	85	85	no	DontPlay			
5	SI Exect	ite Tanaor						
b 7	S	ne runugi	u					
8		Dataset rand	ie (includina th	he name of th	ne attributes -	- first row):		
9	OVE	\$0\$1.4F4	15					
10	OVE	1 100000	···				_	
11			-		OK	Ca	ncel	
12					010			
13								
14	rain	68	80	no	Play			
15	rain	70	96	no	Play			
16								
17								
18	N							<b>_</b> _
	► ► we	eather /			•			
Des	isin 🗸 🔓 🙆	Formes a	automatiques	• 🔨 🗡 [		4 🙎 🗳	<mark>)</mark> - <u> -</u> -	*
Poir	nter		So	mme=2154	ł	NU	M	

To **validate** the parameter settings, we click on the **OK button**. TANAGRA is automatically executed with the appropriate dataset.

🕎 TANAGRA 1.4.11 - [Dat	aset (tar	12.	txt)]					
Tile Diagram Component	Window	Help	)					- 8 ×
D 📽 🖬   🏪								
Analysis			Datase	et	desc	ription		~
Dataset (tan12.txt)			5 attribute	(5)				
			14 example	e(s)				
			Attribute	Cat	tegory	Informations		
			Outlook	Dis	crete	3 values		
			Temp	Cor	ntinue	-		_
			Humidity	Cor	ntinue	-		=
			Windy	Dis	crete	2 values		
			Class	Dis	crete	2 values		~
μ			0					
Data visualization	(	statis	tics:	nts	Nong	oarametric sta	tistics	
Instance selection	Featur	e cor	nstruction	Ì	F	eature select	ion	
Regression	Fact	orial	analysis			PLS		
Clustering	Sp	iv lea	irning	Ì	A	\eta-spv learn	ing	
Spv learning assessment		Scor	ing			Association		
Correlation scatterplot	Exp(	ort d	ataset		ţ,	Scatterplot		🔛 Viev
<								>

We see that the whole dataset (14 examples and 5 attributes) is really exported. The type of the variables is automatically defined according to the rules described above.

#### Working with TANAGRA

In this tutorial, we want to perform a basic descriptive statistic. First, we must **define the INPUT attributes**. We add the DEFINE STATUS component in the diagram using the short cut in the toolbar.

💯 TANAGRA 1.4.11 - [Da	taset (tar	n <b>12.</b> t	xt)]					
Tile Diagram Component	Window	Help						_ 8 ×
Analysis			Datase	t	desc	ription		^
🛄 Dataset (tan12.txt)			5 attribute 14 example	(s) e(s)				
			Attribute	Cat	tegory	Informations		
			Outlook	Dis	crete	3 values		
			Temp	Cor	ntinue	-		
			Humidity	Cor	ntinue	-		=
			Windy	Dis	crete	2 values		
			Class	Dis	crete	2 values		~
			Componer	nts				
Data visualization	:	Statist	tics		Nonp	oarametric sta	atistics	
Instance selection	Featur	e con	struction	Ì	F	eature select	ion	ĺ
Regression	Fact	orial a	analysis	Ì		PLS		
Clustering	Sp	ov lear	rning		N	leta-spv learn	ing	
Spv learning assessment		Scori	ng	Ì		Association		
Correlation scatterplot	📴 Exp	ort da	ataset		,	Scatterplot		🔛 Viev
<								>
								.:

We select the INPUT attributes in the following dialog box.

Define attribute statuses Parameters				
Attributes : Outlook C Temp C Humidity Windy Class	•	Target Outlook Windy Class	Input	Illustrative
ĘĘ Š			Clear select	ion
		ОК	Can	cel Help

We click on the contextual menu VIEW in order to display the results.

💇 TANAGRA 1.4.11 - [De	fine status 1]							
🕎 File Diagram Component	Window Help						-	₽ ×
D 📽 🖪   👪								
Analysis					Define			
🖃 🎹 Dataset (tan12.txt)					Baram	otors		
🕂 🔁 Define status 1		Target : 0			r ai aili	GLEIS		
Parameters.		Input:3						
Execute		Illustrative	0					
View								
			_/		Resi	utts		≡
		Attribute	Target	Input	Idustrative			
		Outlook	•	yes	-			
		Тетр		-	-			
		Humidity	1	-	-			
		Windy		yes	-			
		Class	. \	yes	1			
			i		-			~
		с	ompon	ients				
Data visualization	Statis	tics	No	onpara	ametric stati	stics	Instance selection	
Feature construction	Feature s	election		R	egression		Factorial analysis	
PLS	Cluste	ring		Sp	ov learning		Meta-spv learning	j.
Spv learning assessment	Scori	ing		A	ssociation			
Correlation scatterplot	🧖 Scatterp	lot		<mark>∛.</mark> ∶Vi	ew multiple	scatter	plot	
Export dataset	🔣 View dat	aset						
L								

The next component to insert into the diagram is UNIVARIATE DISCRETE STAT from the STATISTICS tab.

r 📾 🖪  🔛									
Analysis		r							
🖃 🎹 Dataset (tan12.txt)		Define status 1							
🖵 🏭 Define status 1		Target : 0 Input : 3 Illustrative	:0			Falan			
						Re			
		Attribute	Target	Input	Illustrative		Univariate discrete stat		
		Outlook		ves			Description		
		Тетр	2	-	2		Precondition		
		Humidity	-	-	-		One or more discrete attributes must be available in the dataset. The discrete attributes to be described must be set as INPLIT		
		Windy	-	yes	(#1)		Target attribute(s)		
		Class	-	yes			None.		
		<u></u>					Input attribute(s)		
					Componen	nts	Postcondition		
Data visualization	Stati	stics	No	npara	metric statistics	Instance s	None.		
Feature selection	Regre	ssion		Fact	orial analysis	PL:			
Spv learning	Meta-spv	learning	Sp	ov lear	ning assessment	Score	ng Hissocration		
Bartlett's test	🔠 Group	exploration		A	Normality Test	Luiu, T-T	est		
Brown - Forsythe's test	Levene	's test		lah,	One-way ANOVA	Lun T-T	est Unequal Variance		
Fisher's test	Linear Liter	correlation			One-way MANOVA	团Uni	ivariate continuous stat		
Group characterization	More L	Inivariate ci	ont stat	-	Paired 1-Test	iiii Uni	wariate discrete stat		

We select this component and drag it under the DEFINE STATUS component into the diagram. We click on the VIEW menu in order to display the results.



### **Recovering the results in EXCEL**

Because all the results are in the HTML format, we can copy and paste them in the spreadsheet. We click on the COMPONENT / COPY RESULTS menu.

💯 TANAGRA 1.4.11 - [Univ	ariate discrete st	at 1]												
💇 File Diagram Component V	Window Help									- 8 ×				
🗋 💕 🔜  🌇 Copy result	5 Eng													
Analysis						niusristo dic	orata stat 1			^				
🖃 🏢 Dataset (tan12.txt)	•					Darame				- 1				
😑 🚰 Define status 1		Attributes	:3											
🛄 Univariate disc	rete stat 1	Examples : 14												
						Dogu	H.c.							
		Attribute Gini Distribution												
				Value	:5	Count	Percent	Histogr	am					
		Outlook	0.6633	sunny		5	35.71%							
				overcast		4	28.57 %			=				
				rain		5	35.71 %							
				Value	s	Count	Percent	Histogr	am					
		Windy	0.4898	yes		6	42.86 %							
				no		8	57.14%							
		Class	0.4592	Value	15	Count	Percent	Histogr	am					
				Play		9	64.29 %							
				DontPlay		5	35.71 %							
										~				
				Compor	ients									
Data visualization	tion Statistics			etric statistics	Insta	nce selectio	on Feat	Feature construction						
Feature selection	Regression		Factor	rial analysis		PLS		Clustering						
Spv learning	Meta-spv learnin	g Si	ov learni	ng assessment		Scoring		Association						
🛱 Bartlett's test	erization	🛃 Li	near correlation		航 One-way	ANOVA	🗽 T-Test	l	📙 Univa					
Brown - Forsythe's test	tion	M.∐ ∧	ore Univariate c	ont stat	🐴 One-way	MANOVA	Later T-Test Unequ	al Variance						
aa rishers test	्र <sub>म्प</sub> Levenes test		<u>vy</u> N	ormality lest	me univariate cor	icindous stat								
<										>				

Then, in the spreadsheet, we insert a new sheet and paste the result. The appearance is kept more or less but the essential information is preserved.

	Aicrosoft I	Exce	el - wea	ther.	xls															X
1	Fichier Edit	tion	Affichag	je Ins	ertion	Format	Outils	Donn	ées	Fenêt	re ?	Таг	nagra						_ 18	
ĪD		4	 7	X	Ba 🙉	. 🛷 🛛	ю. •	Ci +	۵.	Σ	- f# 17	3 4	1 Z I	1	?	»	G	€	٩	- »
	H21	-		=								2 2	• • •		~	•	_	-		•
	A B		С			D		E			F			G			Н	-		
	Univariate																			
	discrete																			
1	stat 1												_							
2	Paramete	ers					_								_					_
3	åttrihutae -	. 2					_					_			-			-		
5	Examples :	14					_											-		-
6																				
7																				
8	Doculto											_			-			-		
10	Results	•					_													
11	Attribut	ute Gini		1	1			Distribution									-		-	
12					Va	lues	C	Count		Percent			Histogram							
13					sunny	/			6		35.71	%								
14				overc	ast		4			28.57%										
15	Outlook		0.6633 0.4898		8 rain				6		35.71	%								
16					Va	lues	C	ount		Pe	cent	:	Histo	igram						
47												~								
17					yes				0		42.80	70								
18	Windy				3 no				8		57.14	%								
19				Va	lues	C	ount		Pe	cent		Hista	gram							
20					Play				9		64.29	%								_
21	Class		0	.4592	2 DontF	Play			5		35.71	%								_
22					-		_					+						-		
23	24 Computation time : 0 ms.											+						-		-
25	25 Created at 22/11/2006 0:41:53																			
26	<b>P</b>	euil 1	View *	her /	/															Ē
De	ssin - La		Formes	autom	atiques	• \		10	≙1	<b>4</b> 1 🕼		× -	.1 -	Α.	=	+	<b>≓ □</b>			
Prê	t no		. 01.03			ì	<u> </u>			- 404 6.0	6   Y		_		_	NUM			•	

### **Conclusion -- Performance evaluation**

In this tutorial, we use a very small dataset in order to show the mechanism of this new functionality. The real question is: what is the quickness of this add-in when we treat a large dataset, knowing that the maximum size is anyway limited by the EXCEL capacities?

Another dataset (SHUTTLE.XLS) is distributed with this tutorial. It contains 58,000 examples and 10 attributes. By using the same method, we note that the computation time (data preparation and exportation towards TANAGRA) lasts only a few seconds.