

The Excel Data Mining Add-in. Applications in Audit and Financial Reports

Daniel HOMOCIANU,

"Alexandru Ioan Cuza" University of Iaşi, E-mail: daniel.homocianu@feaa.uaic.ro

Dinu AIRINEI, "Alexandru Ioan Cuza" University of Iaşi, E-mail: adinu@uaic.ro

Abstract

Performance reasons in decision making based on business data usually requires a good management of multiple data formats and also processing speed, flexibility, portability, automation, power of suggestion and ease of use. The paper comes with theoretical ideas and practical examples in favor of using the Excel Data Mining Add-in's for the aforementioned reasons. Most of the examples include figures linked to video scenarios constructed by the authors and part of an interactive online list with eighteen pieces. Together they contribute to understanding most of the requirements to fulfill in order to have valid examples and useful results.

Keywords: business and financial data, spreadsheets, Data Mining (DM), examples

JEL Classification: C61, D81, D83, M42

To cite this article:

Homocianu, D. and Airinei, D. (2017), The Excel Data Mining Add-in. Applications in audit and financial reports, *Audit Financiar*, vol. XV, no. 3(147)/2017, pp. 451-468, DOI: 10.20869/AUDITF/2017/147/451

To link to this article:

http://dx.doi.org/10.20869/AUDITF/2017/147/451 Received: 17.03.2017 Revised: 13.04.2017 Accepted: 20.04.2017

and Romania

Introduction

This paper starts from some techniques used by most Data Mining tools when dealing with large data from databases and presents the advantages of using spreadsheets as client applications (msdn.microsoft. com/...dn282385.aspx). The last ones are so familiar to the end users and have an interface that integrates programming or scripting languages for office applications such as VBA meaning Visual Basic for Applications for Microsoft Excel (msdn.microsoft.com/.../ee814737.aspx), and Google Apps Script for Google Sheets (developers.google.com/.../sheets), many functions and advanced facilities for processing, analysis, representation and simulation all based on the interactivity and dynamics principles with great impact on users' ability to perceive, interpret, understand and manage complex information in different cases.

The concept of Data Mining essentially means the supervised identification of undiscovered patterns and hidden relationships in huge data sets

(searchsqlserver.techtarget.com). Inmon which is a wellknown guru in data warehousing (computerweekly.com) gave one of the most concise definitions of a Data Mining (Inmon and Linstedt, 2014) namely analysis of large quantities of data to find patterns such as groups of records, unusual records and dependencies. The Data Mining initiatives usually come from marketing and retail sales departments and are suited for organizations having very large databases (Airinei, 2002).

This concept is closely related to that of data oriented Decision Support Systems (DSS) and Business Intelligence (BI) – especially the one for strategic purposes (dssresources.com/...id=174) that requires huge amounts of data (bi-insider.com). Although the BI term is known as a set of concepts and methods for improving decisionmaking emerging in the 90's (Howard Dresner from the Gartner Group - dssresources.com/.../dsshistory.html), the evidences from the specialty literature indicate approaches from 15 years earlier (Cleland and King, 1975; Pearce, 1976) containing clear references to BI, business planners and managers and decision making.

As concluded by Dan Power

(dssresources.com/...id=199), Data Mining tools include: case-based reasoning, data visualization (mostly graphs, trees, and clusters), fuzzy queries and analyzes, genetic algorithms, and neural networks.

Starting a few years ago we are witnessing implementations of this concept and related models not only in applications dedicated to database and data

warehouse management systems, but in modules of spreadsheet applications that are working with these above as suggested even from this paper's title. This seems obvious when thinking that such dedicated products allowed the construction of DM structures and models starting simply from one table (usually as aggregation of many others from a database).

The applicability of the theoretical and practical elements of this article in auditing, especially the one of performance (Fraser, 1998) and financial reports is justified starting from a specific need to valorize the existing data structures (often data records in tables and tables in databases) and get rapidly and at minimum cost reports able to present clear information on causality related to effectiveness (actual / estimated results compared to those proposed) and efficiency (consumed resources compared to achieved / estimated results).

The concrete examples in this article support certain conclusions drawn from the literature review, namely: the utility of approaching the audit engagements by using data mining techniques (Vintilescu Belciug et al., 2010) as a complement to traditional methods of risk analysis and intervention on site, the consecration of existence of possible areas of integration between data mining and audit processes grouped by stages (Sirikulvadhana, 2002) such as: planning, execution, documentation and completion) or by specific examples (Wang and Yang, 2009) as neural networks for: risk assessment, finding errors and fraud, determining the going concern of a company, evaluating financial distress, and making bankruptcy predictions and decision trees for: analysis of bankruptcy, bank failure, and credit risk), the advance of the latest software tools that implement data mining algorithms and the fact that many users considered them until recently being not very friendly (Chersan et al., 2013) and requiring technical skills advanced enough.

The paper also aims to eliminate some confusions on using time stamp values (e.g. calendar dates, parts of it or replacement values) when operating on time series containing business data (e.g. sales amount recognized as a factor of direct influence for the level of certain financial indicators such as the operating income).

1. The research methodology

The source data for the examples presented in this paper come from two Microsoft samples databases. The first set of examples was created starting from an Access database file called foodmart (sites.google.com/.../ supp4excel2datamining) originally available on the



installation CD of a previous version of Microsoft (MS) SQL Server. The second one is from a MS SQL Server sample database called "AdventureWorksDW2012_Data.mdf" already installed and prepared for use inside a Windows 8.1 32 bits virtual machine (y2u.be/Xs2SWtBqdzI) that we have used for this article. This machine benefited from the Microsoft Imagine / formerly Dream Spark educational software license for all applications installed inside and it was optimized for Oracle Virtual Box. In fact SQL Server 2012 (or 2008) is a prerequisite for the installation of the Excel Data Mining add-in which is detailed in the second video tutorial (playlist mentioned below). Although they serve for building the examples and related video support materials (tutorials – the playlist created by the authors and available at goo.gl/JDDtFp), such data only have a guide purpose in this research with high applicative nature, the similarities to reality being merely coincidental.

2. From intuitive patterns to deep analysis starting from simple sources of data as tables

The first example we have chosen was meant to classify by generating a decision tree where the estimated variable was a categorical one with two possible values (house_owner: Yes or No – Y or N) depending on some other fields (see **Figure no. 1**) containing information about customers (an export to Excel from the *customer* table in the *foodmart* database).

Figure no. 1. Example of export followed by the use of classification option of the Excel Data Mining add-in and the configuration of the input fields

RLE HOME	CREATE EX	een Ternal I	DATA DATABASE TOOLS	TABLE TOOLS S FIELDS TAB	LE 18-19 El Access					? - 0	3 X Sign in		
	同か日日か日	er d	EXML File	98 98 96									
Saved Linked Table Ex moorts Manager	icel Access O	DBC B	More* Exports	File File or XP	s More -								
inparte interrigio	eport & Link	acters -	Labora	Export	-						~		
Tables ® «			Expo	rt to Excel spreadshee	et					customer	~		
Tubics	custome	id - la	account num - Expo	t selected object to an	Excel address1	- add	- at - 1	addrer - c	ity	- state prov - postal e	ode		
- second		1	87452024688 h	iheet in a Microsoft Ex	el file. Bailey Road	10,011	and the set	Tlaxia	100	Oaxaca 15057	1000		
category		2	87470586299 Whelp!	y Derrick I.	2219 Dewing Aver	NJ.		Sook	8	BC 17172			
currency	1.	3	87475757600 Derry	Jeanne	7640 First Ave.			Issaq	uah.	WA 73980			
i customer		4	87500482201 Spence	Michael J.	337 Tosca Way		1.1	Burna	aby	BC 74574			
FILE HOME	INS Classific	ation			+	PERS	PINOT	DATAMINING	TE	AM			Sign in
	2					12	1	To I	2		<		
Data Data - Data	Column 1	o analyce	houseowe	er	•		Model	8.4	odels	W81-328/TS) *			
Data Preparation	Input col	unec					Model Usi	ige Mari	gene	ent Connection Hel	Þ		
at * :		•	Column Name			-							
			customer_id										
A B			account_num			K	L	M	N	0 P	Q R	5	T L
1 customer_account	t_r		inane			*_pr	o postal_	co-country o	uston	mer_phone1 phone2 bir	thdate marital_	st yearly_incge	ender total
2 1 8.75E+	10		Inane			9C9	15057	Mexico		30 271-555-9 119-555-1	8/26/1961 M	\$30K - \$50 F	
3 2 8.75E+	10		m			11111	17172	Canada	-	101 211-555-7/807-555-9	7/3/1915 S	\$70K - \$90 M	
4 3 8.75E+	10		address1			1000	73980	USA	~		Classify wrizeru		
5 4 8.75E*	10		addess2			12210	14514	LIFA	Spli	it data into training and testing se	ts .		*
0 5 8.73E*	10	<u> </u>	address3			10000	91793	USA					
7 9.755+	10		addept4			100	13343	Mexico				_	-
9 8 8.75€+	10	2	chu			11012	12942	Mexico	Pe	scentage of data for testing		3	30 0
10 9 8.75E+	10	2	state receives			11124	17256	Canada	Ма	evinum number of rows:			•
11 10 8.76E+	10	<u> </u>	milial ande			1993	82017	USA	0	escription			
12 11 8.76E+	10	2	porta_score			1157	35890	USA		The weard will split your data in the default, the data mining engine will u	mining structure into a train use the training set to train t	ing set and a testing he mining models, a	g set. By and the testing
13 12 8.76E+	10		country id	9	Algorithm P	arame	eters			set to test the accuracy of the mode	th.	the Andrew Market	i and here
14 13 8.76E+	10	-	customer_region_to						11.7	testing. If you set both options, the v	sizard will use both limits. Fo	or example, if the M	lavinum
15 14 8.76E+	10	<u>–</u>	phone i	Algorithm:	Microsoft Decision Tr	ees			11.3	number of rows' is less than the 'Peri rows' will be used for testing. If you t	centage of data for testing' set the Maximum number of	then the Maximum sows' to 0, this limit	t number of t will not be
16 15 8.76E+	10	<u>ц</u>	phonec	Earameters:					11.3	used			6
17 16 8.76E+	10	1990 1990	DETIGATE AND A DETIGATE	Parameter	Val.		Delarit	Bance		and testing sets. Because the partiti	ons are created in your min	ing structure, your s	iource data is
18 17 8.762+	10	120 C	States and a state of the state	COMPLEXITY P	ENALTY			0.0.1.0	1.1	not affected.			
18 8.762+	10		yeany_income	FORCE_REGRE	SSOR			1					
19 8.7/2+	10		gendel	MADOMUM_INPU	T_ATTRIBUTES		255	[0.85535]			< Back	Next>	Cancel
20 6.7/2*	10		tota_chidren	MAXIMUM_DUT	PUT_ATTRIBUTES		255	0.65535	L	1476.000.71127.005.5	11///1518 C	C359 . CC514	
22 8 774	10		num_children_at_hone	SCORE METHO	D		4	1.3.4		610-555-6 921-555-6	10/7/1947 M	\$30K - \$50 E	
23 8 775+	10	2	education	SPLIT_METHOD			3	[1.3]		444-555-1/625-555-9	10/24/1916 S	\$30K - \$50 F	
24 8.77E+	10	2	date_accrit_opened	Parameter descri	them					102-555-7-721-555-7-	12/4/1921 S	\$30K - \$50 F	
25 8.77 +	10		member_card	Passee descri						506-555-4 333-555-71	9/26/1952 \$	\$70K - \$90 F	
27 26 8.77E+	10	2	occupation						<u> </u>	425-555-3 452-555-3	8/25/1954 S	\$10K - \$30 M	
28 27 8.77E+	10		houseowner						4	852-555-8'986-555-1	3/17/1934 M	\$30K - \$50 F	
29 28 8.77E+	10	2	num_cart_owned	Lucia and and a second						794-555-6-467-555-9	3/12/1970 S	\$10K - \$30 F	
30 29 8.77E+	10			bbA	Benove		OK.	Cancel	E.S.	583-555-4 125-555-6	10/4/1937 S	\$10K - \$30 F	
30 8.77E+	10								_	548-555-9:529-555-3	7/8/1917 S	\$50K - \$70 F	
31 8.77E+	10				1	FIRE	63996	USA	1000024	46 598-555-1 448-555-2	7/2/1978 M	\$50K - \$70 F	
32 8.77E+	10					100	93949	Canada		96 937-555-2 326-555-4	1/8/1930 5	\$10K - \$30 F	
· · cust	om Param	eles.		<back n<="" td=""><td>ext> Cancel</td><td></td><td></td><td>1</td><td></td><td></td><td></td><td></td><td></td></back>	ext> Cancel			1					

Source: The video tutorial created by the authors: y2u.be/Nx9xqCX1DjY



The results (**Figure no. 2**) of this classification above by using default settings (Microsoft Decision Trees algorithm and 30 percent of data for testing) consist in: (1) a decision tree and (2) a dependency network that indicate the most important variables that influence the

houseowner value, namely marital_status (married or single – M or S) and yearly_income (eight thresholds: '\$10K - \$30K', '\$30K - \$50K', '\$50K - \$70K', '\$70K -\$90K', '\$90K - \$110K', '\$110K - \$130K', '\$1300K -\$150K', '\$150K +'), in this order of importance.

Figure no. 2. Example of result of classification starting from data in a table with customers and made by using the Microsoft Decision Trees algorithm



Source: The video tutorial created by the authors: y2u.be/Nx9xqCX1DjY

As seen above (left side of Figure no. 2) the branches that indicate a higher probability for Yes (Y – houseowner) are darker, the rest of them being colored with a lighter shade. We can also observe that the houseowner as a variable depends essentially on the marital_status (right side of Figure no. 2 – the slide bar on Strongest Links) and then on the yearly_income (the slide bar on All Links). And that can also be deduced directly from the decision tree in which the node closest to the root expresses a test (inf.ucv.ro) corresponding to the marital_status attribute. When clicking on marital_status='M' (terminal node) we have got a probability more than 74% in all ten tests we have done in the same configuration (input columns, column to analyze, algorithm, percentage of data for testing).



Figure no. 3. Examples of discriminative analysis after applying the logistic regression (profs.info.uaic.ro) for the same conditions above and specifying those two already identified major impact input variables and some of their values

Input:					Input		- Duár	w.e	
Attribute	Value	Output			Attribute	Value	Out		
marital_status	s	Output Attribute:	houseowner	¥	marital_status	S	Out	put Attribute: houseowner	¥
yearly_income	\$10K - \$30K	Value 1:	Y	v	yearly_income	\$150K +	Val.	e 1: Y	~
		Value 2:	N	v			Valu	ve 2: N	v
<		>			<		>		
Variables:					Variables:				
Attribute	Value	Favors Y 🔍	Favors N	٨	Attribute	Value	Favors Y 🔍	Favors N	^
state_province	Oaxaca				state_province	Oaxaca			
state_province	Mexico				city	Tlaxiaco			
state_province	Veracruz				city	Mexico City			
city	Walla Walla				state_province	Mexico			
city	Camacho				state_province	Veracruz			
city	Hidalgo				city	Orizaba			
city	Oakland				city	Acapulco			
city	Tixapan				city	Walla Walla			
city	Coronado				city	Camacho			
state_province	Guerrero				state_province	Zacatecas			
city	Mil Valley				city	Hidalgo			
state_province	Jalisco				city	Oregon City			
city	Seattle				city	National City			
city	Langley				city	Oakland			
city	Vancouver				city	Port Orchard			
city	Kirkland				city	Tixapan			
state_province	Yucatan			~	city	Corvallis			~
<				>	<				>
					0.15.1				a
Copy to Excel				Llose	Lopy to Excel				Llose

Source: The video tutorial created by the authors: y2u.be/-6jzQuyTjlo

In the previous images (Figure no. 3) we tried to show how we have predicted probabilities that the customers fall into those two categories of the binary response (onlinecourses.science.psu.edu): house owner or not, depending on some explanatory variables and their values. We have done the discriminative analysis partially captured above (Figure no. 3) starting from another algorithm, namely Logistic Regression implemented by Microsoft using a variation of the Neural Network algorithm (msdn.microsoft.com/.../ms174828.aspx) which is easier to train.

The "Copy to Excel" functionality helped us to send the results back to Excel as new sheets containing screen shots (left side of Figure no. 4 for decision trees) or most importantly data sets with visual effects usually involving conditional formatting done automatically (e.g. discriminative analysis based on logistic regression - right side of Figure no. 4).

Based on the results described above (Figures no. 1-4) on can develop similar examples to address also the problem of customer classification (corresponding to the acceptance / maintaining phase of the audit approach) into one of two categories: acceptable / unacceptable, starting from a validated log of such decisions, in tabular format and including many other descriptive attributes (geography, industry, average number of employees, turnover, evolutions of certain indicators, amount of fee, etc.).



Figure no. 4. Example of results of the "Copy to Excel" functionality

Source: The video tutorials created by the authors: y2u.be/Nx9xqCX1DjY and y2u.be/-6jzQuyTjlo

3. Cumulating historical data and using descriptive fields from many tables of a database

The dynamic and interactive reports responding to many information needs that we are so familiar with as well as the older static ones as snapshots of information at precise moments and generating more questions than answers (Rasmussen et al., 2002) may use both current and historical data. The 1st category is represented by data from Transaction Processing Systems (TPS) and commonly referring to the current year while the second essentially means data involving a larger period as time reference. The proportion of using those two categories essentially depends on decisional needs (at operational, tactical or strategic level). For minimizing the redundancy and dependency of data or because of storage space and write speed needs (deshpande.mit.edu) the schema of a traditional



relational data source is usually thought as many tables obtained by applying the principles of normalization (w3schools.in). Moreover, because of further performance reasons (read, respectively write speed needs) historical data must be separated from current data. Both categories essentially include records from transaction tables (e.g. expenses, sales, exams, etc.) the difference being made by the value of the time stamp. That explains why those tables loaded only with historical data are being renamed with a time indication, archived and separated from the rest of the transactional system in order to improve its operational (current) performance. When needing large amounts of historical data for analyzes based on ad-hoc queries the systems must do vice versa by aggregating into a single table (source for a fact table in a data warehouse) all the records from the historical archives of the transaction tables (of the same type as the resulting one). In most cases, that generates the advantage of an increased potential to identify patterns. But it also comes with difficulties related to putting data together in a common and consistent format especially when the applications and the structure of the data source have also changed in time.





Source: The video tutorial created by the authors: y2u.be/kTuYLuav3Eo



The **Figure no. 5** is presenting an example of inventory data gathering (applications including freight audit) in two major steps corresponding to two SQL queries in Microsoft Access: 1st - based on cumulating (UNION clause) the records from two transaction tables of the same type and corresponding to just two years (1997 and 1998) and adding an necessary id column (inventory_id with values generated automatically – AutoNumber type) in the resulting persistent table (INTO clause); 2nd - based on temporarily retrieving values of

descriptive fields from all the tables related or suitable for a relation (**Figure no. 5** – INNER JOIN clause) with the one resulting from the 1st query above, namely *inventory_fact*. In this case the resulting tabular data consisting in the second set of just 11352 records won't get into a persistent table of the database (a kind of denormalization - searchoracle.techtarget.com) otherwise needed to save time at the expense of storage space and it will serve for external export (Excel) just after executing / running the query itself.

Figure no. 6. Results of consecutively using 2 Data Mining models - derived target field with only 2 values



		Browse			
Input		-	Output		
Attribute	Value		Uutu		
warehouse_sales	0.000 - 22.667		Output Attribute:	ordshp	*
		~	Value 1:	1	~
				0	
			Value 2:	0	v
、		,			
Variables:					
Attribute	Value	Favors 1 V		Favors 0	^
store_fax	801-555-4321				
wa_address1	9116 Tice Valley Blv.				
store_street_address	1250 Coggins Drive				
product_name	High Quality Toilet Paper				
store_manager	Green				
store_name	Store 5				
store_phone	801-555-4324				
wa_address1	1893 Northridge Drive				
product_category	Canned Oysters				
product_name	Blue Medal Small Brown Eggs	1			
store_street_address	1120 Westchester PI			-	
brand_name	Best				
warehouse_fax	804-555-6674				

Source: The video tutorials created by the authors: y2u.be/4nOMMRoC2BU and y2u.be/wce_aoTTsbw

Moreover, for speed of design reasons we have chosen all source fields without selecting them explicitly but indicating that by using the most flexible wildcard character, asterisk / "*", after the table name (Figure no. 5), both SQL and design mode (techrepublic.com). For the same reasons above the new derived column needed for analysis (Figure no. 6 - output attribute for both models: classification-top and logistic regression-bottom) was then defined directly in Excel by using the IF function (top of Figure no. 6, ord<>shp as 1 or 0 meaning that units ordered and units shipped are different compared to each other or equal).

4. Association rules for identifying behavioral patterns

In the theory and practice of data warehouses and multidimensional modeling the examples below reminds of the "snowflake" schema meaning that the source for a dimension (perspective of analysis based on descriptive columns organized in hierarchies) is not represented by just a single table but many related ones (in one-tomany relations: e.g. product category, product subcategory, and product – **Figure no.** 7) able to support the analysis with more than just one descriptive field per dimension. In order to be able to apply the association rules algorithm in this case below we have also needed repetitive values for the *SalesOrderNumber* field to be associated to different product categories / subcategories / names.

The main reason for gathering those descriptive data residing in multiple tables from the database in the example above (**Figure no. 7**) is to determine association rules type "If I buy the product X, I will buy the product Y too." in the purchasing behavior (FactInternetSales source table) and the most important dependencies (**Figure no. 8**).

From the results in Figures no. 7 and 8 we can understand why the applications of the algorithms able to identify association rules can contribute to audit and fraud detection and prevention. As example, if the set of inputs would have attributes such as: Claim identificator, Insurance type, Name of the insurance product, Name of the insured person, Insurer, Name of the examiner agent and Solution (total or partial loss and reject) and the algorithm would identify "IF Casco insurance, Insured person X and Examiner agent Y THEN total loss" as association with high probability and importance, it would not necessarily mean a fraud alarm but it would worth at least the effort to investigate further.



Figure no. 7. Gathering both types of data: transactional about sales and descriptive about products by using a single MS SQL Server query

<u>Eile Edit Format View Hel</u> SELECT [SalesOnderNumb [EnglishProductSubcate EROM [dbo] [FactIntern	p er], [EnglishProductCateg	zacu Alomo]			
SELECT [SalesOrderNumb [EnglishProductSubcate EROM [dbo] [EactIntern	er], [EnglishProductCate	zapułlo ma l			
[DimProductSubcategory WHERE [dbo].[FactInter [dbo].[DimProduct].[Pr [ProductSubcategoryKey [dbo].[DimProductCateg	<pre>goryName], [EnglishProduc etSales], [dbo].[DimProduct] netSales].[ProductKey]=[d oductSubcategoryKey]=[dbo] AND [dbo].[DimProductSu ory].[ProductCategoryKey]</pre>	goryName], ctName], [UnitPr uctCategory], [d dbo].[DimProduct o].[DimProductSu ubcategory].[Pro]	ice] bo].].[Produc bcategor] ductCateg	tKey] AND /]. goryKey]=	^ ~
ql - MV32bits\admin (56))* 🗙					•
SELECT [SalesOrderNumber], [E FROM [dbo].[FactInternetSales WHERE [dbo].[FactInternetSale] Messages	nglishProductCategoryHame], [En], [dbo].[DimProductCategory], s].[ProductKey]=[dbo].[DimProdu	glishProductSubcateg [dbo].[DimProductSub ct].[ProductKey] AND	oryName], [category], [dbo].[Dim	EnglishProduc [dbo].[DimPro Product].[Pro	tName], [UnitP duct] ^ ductSubcategor
rderNumber EnglishProductCategoryN	ame EnglishProductSubcategoryName	EnglishProductName	UnitPrice		^
97 Bikes	Road Bikes	Road-150 Red, 62	3578.27		
98 Bikes	Mountain Bikes	Mountain-100 Silver, 44	3399.99		
99 Bikes	Mountain Bikes	Mountain-100 Silver, 44	3399.99		
00 Bikes	Road Bikes	Road 650 Black, 62	699.0982		
'01 Bikes	Mountain Bikes	Mountain-100 Silver, 44	3399.99		
02 Bikes	Road Bikes	Road-150 Red, 44	3578.27		
03 Bikes	Road Bikes	Road-150 Red, 62	3578.27		
'04 Bikes	Mountain Bikes	Mountain-100 Black, 48	3374.99		
105 Bikes	Mountain Bikes	Mountain-100 Silver, 38	3399.99		
06 Bikes	Road Bikes	Road-150 Red, 48	3578.27		
07 Bikes	Road Bikes	Road-150 Red, 48	3578.27		
08 Bikes	Road Bikes	Road-650 Red, 52	699.0982		
09 Bikes	Road Bikes	Road-150 Red, 52	3578.27		
10 Bikes	Road Bikes	Road-150 Red, 56	3578.27		
State Contraction	and the second	Service and the service of the servi	Route Steam of the		
	[dbo].[DimProduct].[Pr [ProductSubcategoryKey [dbo].[DimProductCategoryKey [dbo].[DimProductCategory] SELECT [SalesOrderNumber], [E FROM [dbo].[FactInternetSales [] Messages IderNumber EngishProductCategoryN 97 Bikes 99 Bikes 99 Bikes 00 Bikes 00 Bikes 01 Bikes 02 Bikes 03 Bikes 04 Bikes 05 Bikes 05 Bikes 06 Bikes 06 Bikes 07 Bikes	[dbo].[DimProduct].[ProductSubcategoryKey]=[dbo] [ProductSubcategoryKey] AND [dbo].[DimProductSubcategoryKey] [dbo].[DimProductCategory].[ProductCategoryKey] [l+MV32bits\admin (50)* × SELECT [SalesOrderNumber], [EnglishProductCategoryName], [En FROM [dbo].[FactInternetSales], [dbo].[DimProductCategory], MHERE [dbo].[FactInternetSales].[ProductKey]=[dbo].[DimProdu Messages rderNumber EnglishProductCategoryName 97 Bikes Road Bikes 98 Bikes Mountain Bikes 99 Bikes Road Bikes 00 Bikes Road Bikes 01 Bikes Road Bikes 02 Bikes Road Bikes 03 Bikes Road Bikes 04 Bikes Road Bikes 05 Bikes Road Bikes 06 Bikes Road Bikes 07 Bikes Road Bikes 08 Bikes Road Bikes 09 Bikes Road Bikes 09 Bikes Road Bikes 09 Bikes Road Bikes 09 Bikes	[db0].[DimProduct].[ProductSubcategoryKey]=[db0].[DimProductSubcategory].[Pro [ProductSubcategoryKey] AND [db0].[DimProductSubcategory].[Pro [db0].[DimProductCategory].[ProductCategoryKey] [I-MV32bit\admin(56)* × SELECT [SalesOrderNumber], [EnglishProductCategoryName], [EnglishProductSubcateg FROM [db0].[FactInternetSales], [db0].[DimProductCategory], [db0].[DimProductSubcategory], [db0].[DimProductSubcategory], [db0].[DimProductSubcategory], [db0].[DimProductSubcategory], [db0].[DimProductSubcategory], [db0].[ProductKey].MDD Messages FactInternetSales].[ProductKey]-[db0].[DimProduct].[ProductKey].MDD Messages Road Bikes Road 150 Red, 62 38 Bikes Mountain Bikes Mountain-100 Silver, 44 93 Bikes Mountain Bikes Mountain-100 Silver, 44 00 Bikes Road Bikes Road-50 Black, 62 01 Bikes Mountain Bikes Mountain-100 Silver, 44 02 Bikes Road Bikes Road-150 Red, 62 04 Bikes Road Bikes Road-150 Red, 62 05 Bikes Road Bikes Road-150 Red, 62 04 Bikes Road Bikes Road-150 Red, 62 05 Bikes Road Bikes Road-150 Red, 43	[db0].[DimProductSubcategoryKey]=[db0].[DimProductSubcategory].[ProductCategoryKey] [ProductSubcategoryKey] AND [db0].[DimProductSubcategory].[ProductCategoryKey] [l+MV32bits\admin (50)* × SELECT [SalesOrderNumber], [EnglishProductCategoryName], [EnglishProductSubcategory], FR0M [db0].[FactInternetSales], [db0].[DimProductCategory], [db0].[DimProductSubcategory], MHERE [db0].[FactInternetSales].[ProductKey]=[db0].[DimProduct].[ProductKey] AND [db0].[DimProductSubcategory], MHERE [db0].[FactInternetSales].[ProductKey]=[db0].[DimProduct].[ProductKey] AND [db0].[DimProduct].[ProductKey] AND [db0].[DimProductMen].[ProductMen].[DimProduct].[ProductKey] A	[dbo].[DimProduct].[ProductSubcategoryKey]=[dbo].[DimProductSubcategory]. [ProductSubcategoryKey] AND [dbo].[DimProductSubcategory].[ProductCategoryKey]= [dbo].[DimProductCategory].[ProductCategoryKey] [1-MV32bit\admin (50)* × SELECT [SalesOrderNamber], [EnglishProductCategoryName], [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProductSubcategory]. [dbo].[DimProduct].[ProductKey] AND [dbo].[DimPro

Source: The video tutorial created by the authors: y2u.be/2rW2wK77HD8



Figure no. 8. Results of applying the Microsoft Association Rules algorithms (the associate option of the addin)



Source: The video tutorial created by the authors: y2u.be/3_8E01hnSD0



5. Forecasting starting from aggregated historical data

For more historical data than in the previous example (**Figure no. 5**) we have considered to create a special forecasting scenario closer to reality. We have started from scratch with a new example involving data on 36 months in four calendar years, this time by using a simple SQL query on a single table but with ORDER BY and GROUP BY clauses for sorted results and aggregations meaning computing aggregated values as: sums, averages, total counts, counts for a specified condition and so on. In our case those were averages on every month of an year combined into a single numerical field derived / composed by passing from left to right in the specific order: years to months corresponding to larger to smaller units (**Figure no. 9** - just like in Microsoft's data sample which is provided when installing the Data Mining add-in).

Figure no. 9. Historical data aggregation (time stamp style from the Microsoft's sample) using a SQL Server query (GROUP BY clause) followed by simply copying results to use them in forecasting (Excel's Data Mining add-in)

		1	Sample data for I	orecasting.			_
		2					_
		3	Hist	ory of sales for the	M200 model in var	ious regions.	
		4		-			
		5	Year/Month	💌 Europe Amount 💌	NorthAmerica Amount 💌	Pacific Amount 🗾 💌	-
		6	20	0107 20324.94	20324.94	64424.81	-
		7	20	0108 20349.94	23724.93	60899.82	-
		8	20	0109 16949.95	16974.95	10174.97	-
		9	20	0110 16949.95	20299.94	54174.84	-
SQLQ	uery1.sql -	MV32bits\	admin (54))* ×				
	1 = SE	LECT LEFT	[OrderDateKey],6) AS	OrdYM, AVG(SalesAmount)) AS AvgSales		
	2 FR 3 GR	OUP BY LEF	T([OrderDateKey],6)	.[ractkesettersates]			
	4 OR	DER BY LEP	T([OrderDateKey],6)				
	Results 🗄	Messages					
1	OrdYM 200507	AvgSales					
2	200508	1959,7558					To to to
3	200509	1966.0996	Explore Clean Sample Cla	ssify Estimate Cluster Associate	Forecast Advanced Accuracy	Classification Profit Cross -	Browse Document Query
4	200510	1692.8276	Data Data - Data	10. Konstan (10. 1	Chart Chart	Matrix Chart Validation	Model
5	200511	2101.3886	Data Preparation	Data Modeling	م ا	ccuracy and Validation	Model Usage
6	200512	2120.7279	A1 • : ×	✓ 🛵 OrdYM			
7	200601	1831.5562					200 11 10 11 22
9	200602	2228 607	A B	C D E	F G H	I J	K L M
10	200604	1787.2468	1 OrdYM AvgSales	~		Browse	
11	200605	2040.5725	3 200508 1959 756	Charts Model			
12	200606	1223.2036	4 200509 1966.1	🔍 🔍 💱 Abs 🖹	Show historic predictions	Show Deviations	Prediction steps 6
13	200607	1389.2568	5 200510 1692.828				
14	200608	1214.9766	6 200511 2101.389				AvgSales 💌
15	200609	1320.5305	7 200512 2120.728	2500	1 1		
17	200611	1302.267	8 200601 1891.556				AvgSales
18	200612	1261.6704	9 200602 2194,906	2000			
19	200701	1340.3273	11 200604 1787 247		V		
20	200702	1433.201	12 200605 2040.573	1500			
21	200703	1287.2058	13 200606 1223.204		m.		
23	200704	1249 2422	14 200607 1389.257	1000	~		
24	200706	1170.7141	15 200608 1214.977				
25	200707	1219.4192	16 200609 1320.531	500			
26	200708	1122.2619	17 200610 1281.76				
27	200709	1076.4824	19 200612 1261.67	0			
28	200710	1260.8375	20 200701 1340.327				
30	200712	1159.1114	21 200702 1433.201	-500			
31	200801	1203.0009	22 200703 1287.206	200507	200605 200703	200801 200811	
32	200802	1300.321	23 200704 1287.287	1			
33	200803	1256.5921	24 200705 1249.242	Copy to Eyeod			
34	200804	1228.2024	25 200706 1170.714	Copy to Excer			
35	200805	1137 7345	201 200707 1215,415				
20	uery exercit	ed successfu	N. NV-14/5	1-32BITS (11.0 SP1) 014-1481-	32bits\admin (52) master 00/	0:00 36 rows	
	any execut	a succession	1010-000	1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	see to to an in the set of the set of the		

Source: The video tutorials created by the authors: y2u.be/RjTwGROD0TI and y2u.be/qHJ3Zm3JBT4



After the steps described above (Figure no. 9) and several other processing operations (Figure no. 10) we will get to a set of data suitable for forecasting implemented by using the Microsoft Time Series algorithms as a combination of ARIMA (Auto-Regressive Integrated Moving Average - optimized for improving accuracy in long-term predictions) and ARTXP (Auto-Regression Trees with Cross-Prediction - optimized for predicting the next likely value in a time series msdn.microsoft.com/.../bb677216.aspx) algorithms.

Figure no. 10. Deriving and explaining the correct time stamps as full dates internally stored (Excel) as numbers in right format data sources as support for undistorted forecasting

I =DATE(VALUE(LEFT(A2,4)),VALUE(RIGHT(A2,2)),DAY(EOMONTH(DATE(VALUE(LEFT(A2,4)),VALUE(RIGHT(A2,2)),1),0)))								
A B C D E F G	н	1	J	К	L	м	N	0
1 OrdYM AvgSales								
2 200507 1390.138 7/31/2005								
Format Cells ? 💌	1	A	В	_	C	D	E	F
	1	% from 1st	end of month	n (date)	AvgSales	OrdYM	AvgSales	% from 1st
Number Alignment Font Border Fill Protection	2	100.00%		38564	1390.138	200507	1390.138	100.0000 %
Category:	3	100.08%		38595	1959.756	200508	1959.756	100.0005 %
General Sample	4	100.16%		38625	1966.1	200509	1966.1	100.0010 %
Number 38564	5	100.24%		38656	1692.828	200510	1692.828	100.0015%
Accounting Decimal places:	6	100.32%		38686	2101.389	200511	2101.389	100.0020 %
Date Determini proces.	7	100.40%		38717	2120.728	200512	2120.728	100.0025 %
22 3/31/2007 1287.206 200703 1287.206	8	100.48%		38748	1891.556	200601	1891.556	100.0469%
23 4/30/2007 1287.287 200704 1287.287	9	100.55%		38776	2194.906	200602	2194.906	100.0474%
24 5/31/2007 1249.242 200705 1249.242	10	100.63%		38807	2228.607	200603	2228.607	100.0479 %

Source: The video tutorial created by the authors: y2u.be/e0SkDwG9mNY

We have also thought at automatically deriving the correct (the MM/DD/YYYY format translated into an integer number - **Figure no. 10**) time stamp labels and we presented more details about their comparative behavior when getting trend line functions and forecasting results with this Excel Data Mining add-in (last three tutorials in the aforementioned playlist).

6. Support for querying persistent Data Mining models

First of all, persistent in this context refers to a model defined the way it will be deployed and stored on the server (SQL Server Analysis Services – a module other than the Database Engine) and available for querying (Figure no. 11).

The Data Mining add-in available in Excel offers many advantages over the direct use of SQL Server Analysis Service. Among others, on can mention here: speed of use of Excel's tabular environment and formula language, possibility of many exports / imports as / from spreadsheets starting from different database formats and to indirectly involve multiple source tables by using the Structured Query Language (SQL), the possibilities of exploiting the resulting structures and models directly (the "copy to Excel" option), by using queries (SQL DMX extension - Figures no. 12 and 13) or programmatically (Figure no. 12). Last two are conditioned by activating persistency when defining models (use temporary model option unchecked - Figure no. 13 vs. Figures no. 1 and 2).



Figure no. 11. Data Mining eXtensions (DMX) sales prediction query examples (SQL Server Analysis Services) based on a DM time series model from a wrong format data source (text time stamp: 200815 /15th month in 2008)

Mining model: 1_ESELECT	
forecast AVG sales model	tTimeSeries([forecast_AVG_sales_model].[AvgSales],3) AS PredAvgSales
3 FROM L	forecast_AVG_sales_model]
Metadata J Functions 100 % - <	
Forecast_AVG_sales_model (Micro	III Results
AvgSales EredAvgSales	
	ae lae
\$TIME	AvoSales
200807	1443.24228516
200808	1454.63464828
200809	1450.83187874
Ouenverseted successfully	MV_W91_32BITS mv_w91_32bits/admin DMAddinsDB 00:00:01
Guery executed successionly.	
Object Explore	→ 0 ×
Connect -	स्थित र तो उ
E D WA-M	BI-32BITS (Microsoft Analysis Server 11.0.
E Dat	abases
= ()	DMAddinsDB
Œ	🛅 Data Sources
Ŧ	Data Source Views
The second se	Cuber
	Cabes
	Umensions
8	Mining Structures
	K forecast_AVG_sales_structure
	🖃 🥅 Mining Models
	K forecast AVG sales model
F	Roles
	Accomplian
	Assemblies
	emblies
Mining model:	1 -
	<pre>2 .es_model].[AvgSales],9)</pre>
lorecast_AVG_sales_model	3
Metadata 🛃 Functions	
A forecast_AVG_sales_mode	100 % - <
AvgSales	🔂 Messages 🛄 Results
TT OIDTM	PredAvgSales
	PredAvgSales
	ATTIME AugGales A
	200011 1452 00071500
	200011 1402.83271062.
	200812 1453.50726198.
	200813 1453.77358263.
	200814 1454.14759460.
	200815 1454.43486230.
Query executed successfully.	MV-W81-32BITS mv-w81-32bits\admin

Source: The video tutorial created by the authors: http://y2u.be/qHJ3Zm3JBT4



Figure no. 12. Rough example of how to programmatically query a well-defined Data Mining model by using a DMX query in Visual Basic (VB).NET preceded by testing most of it on SQL Server Analysis Services

	Select the files to reference
Formula de X Formulado (Decimal)	
(Form1.vo 4 × Porm1.vo [Design]	Organize Organize Vib329-12
Imports Microsoft.AnalysisServices.A	domdClient Desktop
Public Class Form1	Downloads
 Oreferences Private Sub Form1 Load(sender As 	Object, e As EventArgs) Handles MyBase.Load
Dim conn As AdomdConnection	
	Integrated Security=SSPI;Initial Catalog=DMAddinsDB")
conn.Open() Dim cmd As AdomdCommand	
<pre>cmd = New AdomdCommand("SELE" "SDOW [factorial command]</pre>	TT FLATTENED PredictTimeSeries([forecast_model_correct_TS].[AvgSales],1,1) AS PredAvgSales " &
cmd.Connection = conn	() X
Dim adr As AdomdDataReader adr = cmd.ExecuteReader()	
Dim matr(8, 1)	7/31/2008: 1449.64
i = 0	
While (adr.Read)	ateTime(adc.GetValue(0)) ToShortDateString
matr(i, 1) = Math.Round(adr.GetValue(1), 2, MidpointRounding.AwayFromZero).ToString
<pre>MessageBox.Show(matr(i, i = i + 1</pre>	<pre>3) & ": " & matr(i, 1))</pre>
End While	
End Sub	
End Class ObjectExplorer + 1 ×	forecast_structure_correct_TS[Browse] × • Mining Legend •
Connect* 말 밝 = 7 @ 3	Mining Model: forecast, model_correct, v Viewer: Microsoft Time Series Vie v (2) Timestamp: 3/31/2009 12:00
😑 📸 MV-W81-32BITS (Microsoft Analysis Server 11.0.3000.	Charts model Column Value
Databases DMAddinsDB	O C Star Show historic medictions O Show Devictions Deviction stars All AvgSales (predi 1457.31
Data Sources	
Data Source Views Data Cuber	AvgSales 🗸 🔘
Cubes Dimensions	
🖂 🛅 Mining Structures	2500 V AvgSales
🗈 🛅 Roles	
Assemblies Assemblies	
	1500
	1000
Mining model: 1 SELECT	I 500 +
forecast_model_correct_TS v 2 Predict FROM [*	:TimeSeries([forecast_model_correct_TS].[AvgSales],3) AS PredAvgSales Torecast_model_correct_TS]
Metadata J Functions 100 % • <	
AvgSales	Develop State Auto State
PredAvgSales.\$1/ME 7/31/2008 12:00:00/	птеминураавсяитураався M 1449.643217725663
8/31/2008 12:00:00/	M 1460.4264515793
Source and an and a second secon	RCmibb&MC mms/staffs-18w-vm 2TIBSE-18w-VM
and sector sectors and the sector an	
	1 //31/2005 5/31/2006 3/31/2007 1/31/2008 11/30/2008 I

Source: The authors' projection resulting after development attempts with VB and SQL Server



Figure no. 13. DMX prediction query example (houseowner) based on a persistent decision trees classification model (generic content view in background)

DT_s_HO [Browse] × DMXQuery1.dmxv	v81-32bits\admin)*	C	DT_s_HO [Browse] DMXQuery1.d	lmxw81-32bits\admin)* ×
Mining Model: DT_m_HO Node Caption (Unique ID)	Viewer: Microsoft Generic Node Details	Conter 🗸 👔	tining model: DT_m_HO V	1 ⊡SELECT 2 [houseowner], 3 PredictProbability([houseowner], 'Y') AS [HouseOwner = Yes]
 □· (0) □· All (00000000r) □· marital_status = 'M' (0000000000) □· marital_status = 'S' (00000000000) □· marital_status = 'S' (0000000000) □· marital_status = 'S' (0000000000) □· marital_status = 'S' (000000000) □· marital_status = 'S' (0000000000) □· marital_status = 'S' (000000000) □· marital_status = 'S' (0000000000) □· marital_status = 'S' (000000000) □· marital_status = 'S' (0000000000) □· marital_status = 'S' (00000000000) □· marital_status = 'S' (000000000000000) □· marital_status = 'S' (0000000000000000000000000000000000	MODEL_CATALOG MODEL_SCHEMA MODEL_NAME ATTRIBUTE_NAME NODE_NAME NODE_UNIQUE_NAME NODE_TYPE NODE_GUID NODE_GUID NODE_CAPTION CHILDREN_CARDINALITY DADBUM_UNICE_NAME	DMAddinsDB DT_m_HO houseowner 000000000107 000000000107 4 (Distribution) yearly_income = '\$150K +' 0	Metadata F Functions T DT_m_H0 (Microsoft Dec	4 PredictProbability([houseowner], 'N') AS [HouseOwner = No] 5 FROM [DT_m_HO] 6 NATURAL PREDICTION JOIN 7 (SELECT 'S' AS [marital_status], 8 '\$150K +' AS [yearly_income]) AS t 100 % • < > Messages Results houseowner HouseOwner = Yes 100.0805550755 0.19384449244
	NODE DESCRIPTION	marital status = 'S' and y	yearly income = '\$150K +'	
	NODE_RULE	<pre></pre>	nd"> =="S"> =="marital_status" /> =="\$150K +"> =="yearly_income" />	
	MARGINAL_RULE	<predicate op="eq" value="<br"><simple-attribute name="<br"></simple-attribute></predicate>	"\$150K +"> "yearly_income" />	
	NODE_PROBABILITY	0.0094483812699736		
	MARGINAL_PROBABILITY	0.019036954087346		
	NODE_DISTRIBUTION	ATTRIBUTE NAME ATTRIBUTE V houseowner Missing houseowner Y houseowner N	XALUE SUPPORT PROBABILIT 0 0 0 55 0.80615550755 13	YARIANCE VALUETYPE 0 1 (Missing) 59395 0 4 (Discrete) 40605 0 4 (Discrete)

Source: The authors' projection resulting from development attempts with SQL Server

This last advantage reminds us that the programmatic generation (Airinei and Homocianu, 2009) of Excel dashboards and scorecards by using suggestive representations, warning indicators and dynamic formatting with support for BI has been simplified a lot since the 2007 version of the Microsoft Office suite. Combining that with the ability to programmatically determine behavioral patterns and generate predicted values starting from high performance and easy to use tools such as this Data Mining add-in available for Office

2010, 2013 and 2016 promises much in terms of productivity. All these advances were defined after many years of using dedicated and now well-known technologies (e.g. SQL Server tested by authors since early 2000).

When it comes to spreadsheet products (dssresources.com/.../sshistory.html) such as: VisiCalc, Lotus 1-2-3, Microsoft Excel, Microsoft Works Spreadsheet, Sun Open Office Spreadsheets, Polaris Office Sheet, and Google Sheets the average

Anditorilor Financial

experience of final users is up to decades. Furthermore the easiness of using these applications even just as interface instruments to connect to data from databases and data warehouses and display it was an objective reason to continue with testing the Data Mining component that led to making this article.

By using a way of reporting which identifies itself with a sequence of steps which borrow their names from those eighteen support tutorials and also some techniques previously defined namely: E2P4CAFR (Homocianu, 2015), ACCORD / CADRE (Homocianu and Airinei, September 2014) and S-DOT (Homocianu and Airinei, August 2014) on can reach in stages, but with a minimal number of steps to follow some representations that are dynamic, interactive, suggestive, based on causality and rooted in the current reality and in the history defined by data stored in the organization's data sources.

Conclusions

We can conclude that the possibilities of the Excel Data Mining add-in component are above the expectations of a business analyst, offering the advantage of integrating identified classification patterns, association rules and predictions with the support for connectivity to various data formats, data validations, advanced graphical representations, geographical referencing, automatic conditional formatting and key performance indicators (KPI), pivot and power pivot tables and charts, automatic solving of optimization problems (solver) and the DAX (Data Analysis eXpressions) language together with the

- 1. Airinei, D. (2002), *Depozite de date*, Editura Polirom, Iaand.
- 2. Airinei, D. and Homocianu, D. (2009), The Geographical Dimension of DSS Applications, *Analele Ştiintifice ale Universității "Alexandru Ioan Cuza" din Iaand*, Tome LVI, p. 637-642.
- Chersan, I.C., Carp, M. and Mironiuc, M. (2013), Data mining – o provocare pentru auditorii financiari, *Audit Financiar*, vol. XI, no. 10, p.57-64.
- Cleland, D.I. and King, W.R. (1975), Competitive Business Intelligence Systems, *Business Horizons Journal*, vol. 18, no. 6, pp. 19-28, DOI 10.1016/0007-6813(75)90036-1.

traditional formula language thereby increasing the chances of defining dashboards based on simulations, analyzes and Data Mining models truly useful for audit staff interested in performance monitoring.

We hope we have identified many real motivations to choose this Microsoft add-in for the Office suite as a near real time Data Mining tool, beyond many other recommendations available in the specialized literature and practice.

Beyond effective examples of working with well-known software applications available for a considerable range of users and providing advanced methods for analysis, query and representation of current and historical data particular to support tools for Data Mining and Business Intelligence, the paper also provides a brief theoretical description necessary in order to understand a rapid way of generating complex dynamic reports as dashboards based on analyzes and Data Mining models starting especially from sales and financial data.

The video tutorials developed by the authors, integrated in a playlist, and successively referenced in the paper prove the attempts to enrich the aforementioned way of reporting and to ensure the minimization of the number of steps required when trying to implement similar examples.

Overall, the article is trying and we hope that it succeeds to convey by clear examples some desirable traits as speed, simplicity, capacity of synthesis, transparency, flexibility and availability in reporting with support in Data Mining, as key elements of performance in preparing financial statements and supporting audit activities.

REFERENCES

- Fraser, L.E. (1998), Public Sector Audit Business Integration and Causal Analysis, *Quality Audit Conference*, February 26-27, 1998, Louisville, KY, vol. 7.
- 6. Homocianu, D. (2015), Excel Power Pivot's Applications in Audit and Financial Reports, *Audit Financiar*, vol. XIII, no. 11, p.127-138.
- 7. Homocianu, D. and Airinei, D. (2014), Business Intelligence facilities with applications in audit and financial reporting, *Audit Financiar*, vol. XII, no. 9, p. 17-29.
- 8. Homocianu, D. and Airinei, D. (2014), Consolidating source data in audit reports, *Audit Financiar*, vol. XII, no. 8, p.10-19.



- 9. Inmon, W.H. and Linstedt, D. (2014), *Data* architecture: A primer for the data scientist. Big Data, Data Warehouse and Data Vault, Morgan Kaufmann, MA.
- Rasmussen, N.H., Goldy, P.S., Solli, P.O. (2002), Financial business intelligence – trends, technology, software selection and implementation, John Wiley and Sons, Inc., New York, p. 98-99.
- Sirikulvadhana, S. (2002), Data mining as a financial auditing tool (thesis), Swedish School of Economics and Business Administration, pp.49-57, available online at: https://pdfs.semanticscholar.org /2612/f764664796f911e9ff9a79b7bb9de84bf16c.pd f, accessed on 15.03.2017.
- Vintilescu Belciug, A., Creţu, D. and Gegea, C. (2010), Utilizarea tehnicilor de data mining ca metodă complementară în audit, *Audit Financiar*, vol. VIII, no. 7, p. 30-35.
- 13. Wang, J. and Yang, J.G.S. (2009), Data Mining Techniques for Auditing Attest Function and Fraud Detection, *Journal of Forensic & Investigative Accounting*, vol. 1, no. 1, pp. 1-24.
- 14. http://bi-insider.com/businessintelligence/operational-bi-vs-strategic-bi/
- 15. http://dssresources.com/faq/index.php?action=artik el&id=174
- 16. http://dssresources.com/faq/index.php?action=artik el&id=199
- 17. http://dssresources.com/history/dsshistory.html
- 18. http://dssresources.com/history/sshistory.html
- 19. http://inf.ucv.ro/documents/rstoean/5.%20Arbori%2 0de%20decizie.pdf

- 20. http://profs.info.uaic.ro/~val/statistica/StatWork_10. pdf
- 21. http://searchoracle.techtarget.com/tip/Optimizingdatabase-performance-part-2-Denormalizationand-clustering
- 22. http://searchsqlserver.techtarget.com/definition/dat a-mining
- 23. http://www.computerweekly.com/tip/Inmon-vs-Kimball-Which-approach-is-suitable-for-your-datawarehouse
- 24. http://www.techrepublic.com/article/10-tips-forusing-wildcard-characters-in-microsoft-accesscriteria-expressions/
- 25. http://www.w3schools.in/dbms/databasenormalization/
- 26. http://y2u.be/Xs2SWtBqdzI
- 27. https://deshpande.mit.edu/portfolio/project/hybriddbms-optimized-read-intensive-applications
- 28. https://developers.google.com/appsscript/guides/sheets
- 29. https://msdn.microsoft.com/enus/library/bb677216.aspx
- 30. https://msdn.microsoft.com/enus/library/dn282385.aspx
- 31. https://msdn.microsoft.com/enus/library/ms174828.aspx
- 32. https://msdn.microsoft.com/enus/library/office/ee814737.aspx
- 33. https://onlinecourses.science.psu.edu/stat504/node /149
- 34. https://sites.google.com/site/supp4excel2dataminin g2017af/d