UNITECH '12 GABROVO

INTERNATIONAL SCIENTIFIC CONFERENCE
16 – 17 November 2012, GABROVO

ELECTRICITY END USERS' CONSUMPTION PATTERNS IN DIFFERENT WEST BULGARIAN LOCATIONS[1]

Andrey Bachvarov
BICA Ltd.

Petko Ruskov
Sofia University

Kaloyan Haralampiev
Sofia University

Abstract

In recent years, electricity consumption climbed steadily, especially in developing countries like Bulgaria mainly due to increased access to electrical appliances, dependence on electrical power for climate control, introduction of expensive renewable energy sources and the same time artificially low prices and related. Governments and electricity power providers find it increasingly difficult to respond to the "new challenges" by developing new consumption model but due to the overwhelming complexity of data, most of those models are basically at the macro level and it is almost impossible to identify individual consumption patterns that could generate useful business process innovation. End consumers (mostly households) decide intuitively on their consumption due to lack of access to necessary tools or guidance to plan accordingly. Our research demonstrates that hidden patterns exist in the clients' consumption data depending on specific parameters of their respective environment. Our approach relies on data processing of large scale sample data base provided by the Operations department of "CEZ Electro Bulgaria" AD. and data mining using IBM SPSS Modeler predictive models and tools. This paper identifies and highlights typical consumption patterns within a database of randomly selected clients provided by CEZ. The results include identification and visualization of electricity consumption patterns based on Auto Cluster node method. The generated cluster models, key patterns recognition and predictions are useful for researchers as they can use those findings for informed decision making and argumented business process innovation versus "blind" and "intuitive" actions. Electric distribution companies can further use the results specific business transformations and process innovations including planning of delivery, network optimization and maintenance, proper cost allocation to most problematic areas, demand forecast based on improved supply predictions and related.

Keywords: electricity consumption, cluster analysis, predictive models, process innovation, renewable energy, energy efficiency

[1] The paper presents preliminary results of project "Методи и модели за подобряване на ефективността при хибридни електроснабдителни системи от възобновяеми енергийни източници с малка мощност" applied for „Финансиране на научни и научноприложени изследвания в приоритетните области", Вх. Номер FFNNIPO_12_00914

INTRODUCTION

New technologies and today's research methods enables shifts in the strategy of the energy sector in many different ways, allowing it to deliver enhanced services, to become more efficient, and to respond to environmental concerns such as local air pollution and global climate change [4, 8, 10, 17, 18, 20]. Those shifts would be very time consuming and challenging to analyze without predictive analytics [11, 12]. This paper discusses aspects of data processing of large scale sample data base provided by the Operations department of "CEZ Electro Bulgaria" AD [2]. The research is based on 4582 cases about CEZ electricity user consumption for a period of 17 months from January 2011 to May 2012.

The generated cluster models, important patterns recognition and predictions are useful for researchers and electric utility distributors for electric energy business transformation and efficiency processes innovations such as modeling and optimization of processes and services in the renewable energy sources with low power [1, 13, 14, 15].

Electric distribution companies can further use the results specific business transformations and process innovations including planning of delivery, network optimization and maintenance, proper cost allocation to most problematic areas, demand forecast based on improved supply predictions and related [3, 5, 6, 16].

METHODOLOGY AND RESEARCH DESIGN

The research is based on the CRISP – DM model, data mining methodology that reduce the time required for large data mining projects, improves efficiency and helps identify new, successful patterns [7, 9]. The methodology consists of the following steps - figure 1.

1. Business understanding.
2. Data understanding.
3. Data preparation.
4. Modeling.
5. Evaluation.
6. Deployment.

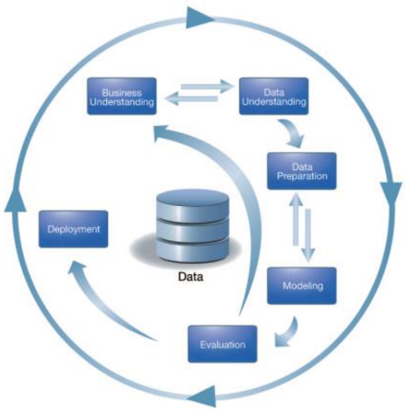

Fig. 1. *The methodology steps [7]*

Another significant part of our work is devoted to the Auto Cluster node. It is statistical method that illustrates and defines the problematical areas of the research.

Below we have adopted key questions to assess the above proposed objectives based on a previous similar study from Teerlink [19].

1. Do we understand what drives our target consumer's consumption?

2. Do we understand in what areas we are being commoditized?

3. How do we communicate information with our customers and how do we measure the effectiveness of our communications?

4. How well can we integrate channels and business partners?

5. How well are we geared up to adapt or change our business models?

6. Do we have a roadmap for customer analytics?

We will use analytics methodology and tools to [7]:

• Develop models to identify patterns in energy consumption

• Identify and where possible predict high and energy consumption patterns

• Classifying customers into groups with distinct usage or need patterns

• Connect different clusters with different type of locations

IBM SPSS Modeler has many predictive modeling nodes available, some of which are popular data mining methods while others come from classic statistics. Segmentation, known as "clustering" models divides the data into segments, or clusters, of records that have similar patterns of input fields. As they are only interested in the input fields, segmentation models have no concept of output or target fields. The Auto Cluster node works in the same manner as other automated modeling nodes, allowing you to experiment with multiple combinations of options in a single modeling pass. Models can be compared using basic measures with which to attempt to filter and rank the usefulness of the cluster models, and provide a measure based on the importance of particular fields. Supported model types include TwoStep, K-Means, and Kohonen – figure 2. [7].

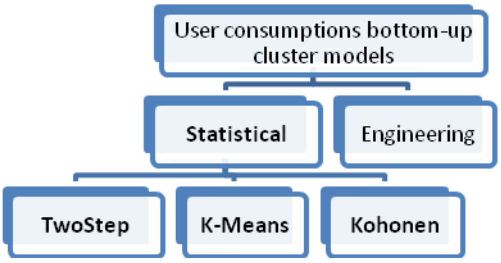

Fig. 2. *Bottom-up model classifications and techniques*

The use of the Auto Cluster node can reduce the time to build models for classification substantially. Once we find one or more models that look promising, the node can create a model node in the model manager that you can then use to explore and investigate the models in more detail.

CEZ ELECTRO BULGARIA AD CASE STUDY AND DISCUSSIONS

Overview: CEZ Electro Bulgaria AD is a company, registered at Sofia City Court on August 29, 2006. The capital of the company is 50 000 BGN, 67% of which property of CEZ a.s. Czech Republic, and the rest 33% of the Republic of Bulgaria in the person of the Ministry of Economy and Energy (http://www.cez.bg/en/home.html) [2]. CEZ Electro Bulgaria allows clients to pay the electricity bills cash or non-cash.

Business need: To create new electricity consumption business models to be able to plan their network expansion, predict customer demand, plan maintenance operations, calculate associated costs and related. On the level of end users, CEZ has certain obligations to Bulgaria as “obliged person” in the context of the law on energy efficiency. The team recognizes that to help CEZ achieve the above it has to use business analytics and value co-creation paradigms. We believe that better visibility of processes and identification of useful patterns will also help change end clients mindset and values, key for meeting successfully the challenges of the changing environment.

Solution: The team decided to use IBM SPSS Statistics and IBM SPSS Modeler as the latter implementation was easy to learn and model. The software seamlessly integrated with entire data bases and families of analytics products. This analytics and decision management software helped us discover patterns, identify critical thresholds and monitor for problem areas.

Benefits: Data-driven statistical analyses of consumer behavior and company analytics provide all stakeholders with a valuable basis for understanding energy efficiency and for process innovation.

We used Auto Cluster node for identification and definition of end users consumption patterns. We implemented the customer’s postal codes into a new variable “Type of location”, which includes the following categories – table 1.

Table 1. *Type of location coding*

Code	Type of location	Postal code
1.	Capital	1XYZ
2.	City	XY00
3.	Town	XYZ0
4.	Village	XYZT

For the purpose of the investigation we averaged the CEZ input data into 20 new variables – table 2.

Table 2. *Averaged consumption new variables*

	Winter	Spring	Summer	Autumn	Total
Active day	X	X	X	x	X
Active night	X	X	X	x	X
Whole day	X	X	X	x	X
Total	X	X	X	x	x

RESULTS AND DISCUSSIONS

After series of experimentation we select to present two types of results:

- Visualisation and graphics
- Auto cluster models

The “stream”, an instrument from the IBM SPSS Modeler used for the above analyses is visualized in Fig. 3.

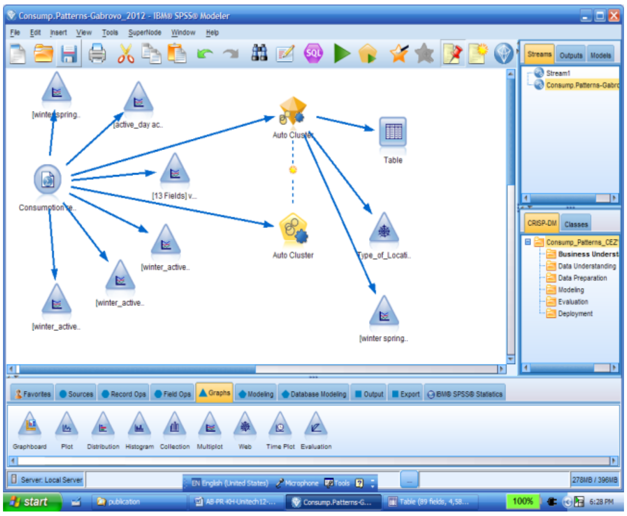

Fig. 3. *Entire SPSS Modeler stream*

Visualization and graphics

From the whole set of available Graphs in the SPSS Modeler the best visualization for our research offers the so called Multiplot – figures 4, 5 and 6. In all following figures on the X axis there are different types of location, according to Table 1. and on the Y axis – average energy consumption according to table 2.

In figure 4 as variable we have used consumption distinguished by seasons:

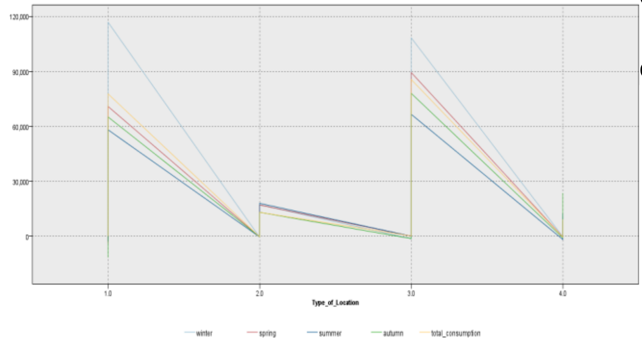

Fig.4. *Average consumption distinguished by season - Multiplot visualisation*

The largest consumption is in the capital and in the towns, followed by the consumption in the cities. The least total consumption is in the villages. The figure indicates that the seasonal consumption regardless of tariffs is comparable within cities and villages and differs quite significantly within capital and towns. The consumption for the last two is the biggest during winter, declining over spring and summer and slightly increasing during autumn however autumn consumption is still less than spring values.

In figure 5 as variable we use consumption distinguished by tariffs.

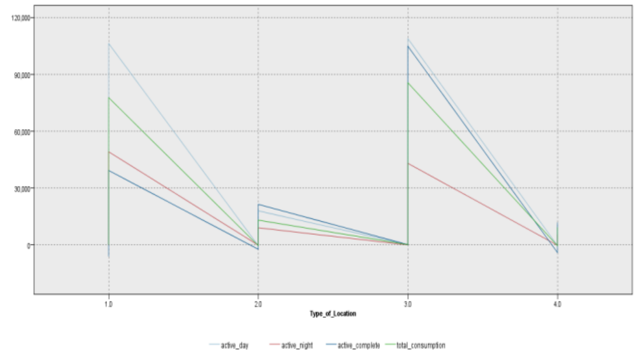

Fig. 5. *Average consumption distinguished by tariff - Multiplot visualization*

In Fig. 5 the largest consumption is again in the capital and in the towns, followed by the consumption in the cities. The least total consumption is again in the villages. The figure indicates that the tariff consumption regardless of seasons is comparable within cities and villages and differs quite significantly within capital and towns. In the capital most consumption is using active day tariff, less consumers use active night and whole day tariffs ranked by respective values from bigger to smaller. The consumption for the towns is the biggest during the days, declining over whole day and night.

In figure 6 as variable we use the combined consumption (by tariffs and seasons).

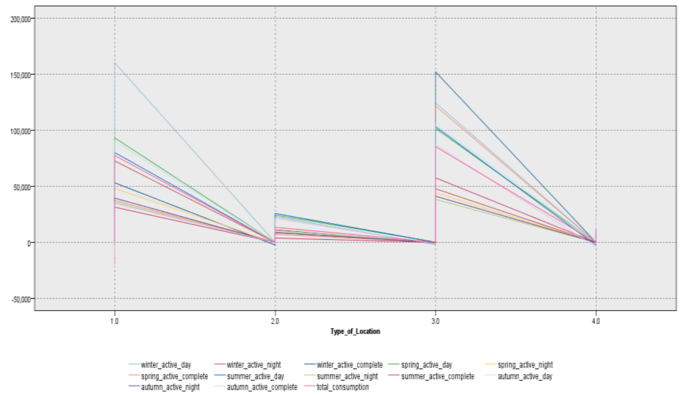

Fig. 6. *Season and tariff average consumption Multiplot visualisation*

In Fig. 6 the largest total consumption is again in the capital and in the towns, followed by the consumption in the cities. The least total consumption is again in the villages. The figure indicates that the consumption is comparable within cities and villages and differs quite significantly within capital and towns. The combined consumption for the capital is the biggest during the winter days, and lowest during the summer whole day tariffs. The consumption for the towns is the biggest during the winter whole day and lowest during summer at active night tariffs.

Auto cluster node models

As variable for Auto cluster node we have used consumption distinguished by seasons and total consumption.

In the figure 7 are shown all three models.

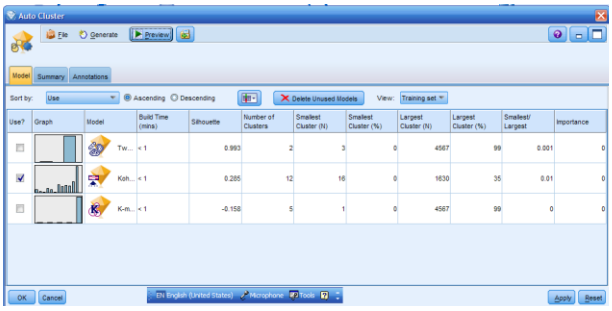

***Fig. 7.** The three best Auto cluster models*

We choose to explain the second model – Kohonen – because the first and the third models are quite disproportional – there is one big cluster which includes over 99.5% of cases. In figure 8 are shown seasonal consumption patterns of all 12 clusters obtained by Kohonen method and the mean consumptions patterns are illustrated in figure 9.

Clusters

Input (Predictor) Importance: 1.0 0.8 0.6 0.4 0.2 0.0

Cluster	Label	Description	Size	Inputs				
X=3, Y=2			35.6% (1630)	winter 131.91	spring 104.95	summer 61.50	autumn 33.46	total_consumption 82.16
X=3, Y=1			14.5% (663)	winter 373.47	spring 378.07	summer 105.39	autumn 89.32	total_consumption 244.09
X=2, Y=0			10.5% (480)	winter 770.76	spring 728.88	summer 288.57	autumn 331.11	total_consumption 606.99
X=2, Y=2			9.1% (419)	winter 144.82	spring 91.13	summer 159.86	autumn 453.82	total_consumption 212.82
X=3, Y=0			8.2% (377)	winter 717.50	spring 596.89	summer 126.05	autumn 115.31	total_consumption 405.65
X=2, Y=1			7.8% (358)	winter 383.05	spring 331.38	summer 278.43	autumn 401.36	total_consumption 354.63
X=1, Y=0			5.3% (242)	winter 1,572.5	spring 1,116.0	summer 433.46	autumn 489.48	total_consumption 998.34
X=0, Y=0			3.8% (174)	winter 3,090.5	spring 2,341.2	summer 1,474.0	autumn 1,763.0	total_consumption 2,216.8
X=1, Y=1			3.6% (164)	winter 592.79	spring 483.37	summer 705.45	autumn 1,151.9	total_consumption 752.52
X=0, Y=1			0.9% (40)	winter 6,462.4	spring 5,057.1	summer 5,444.7	autumn 6,616.6	total_consumption 5,832.5
X=0, Y=2			0.4% (19)	winter 17,585	spring 14,842	summer 16,228	autumn 36,810	total_consumption 14,408
X=1, Y=2			0.3% (16)	winter 721.08	spring 1,007.1	summer 1,347.1	autumn 3,832.7	total_consumption 1,727.0

***Fig. 8.** Kohonen Cluster description*

In the figure we can see the descriptive statistics of obtained clusters. In the fourth column there are clusters' sizes. In the next five columns there are mean consumptions of all clusters.

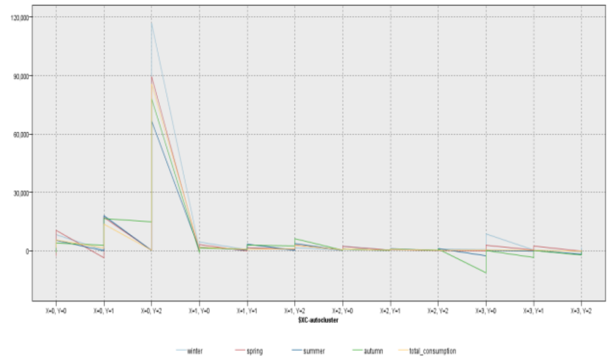

***Fig. 9.** Kohonen Cluster average consumption Multiplot visualisation*

The biggest cluster (X=3, Y=2) includes 35.6% of cases, but its mean consumption are relatively small. On the other hand the cluster with largest consumption (X=0, Y=2) includes only 0.4% of cases – it is next to the last according to the clusters' size.

The next step after the clusters' description is to determine cluster membership of each case. The fragment of table with cluster membership (last column) is shown in the figure 10.

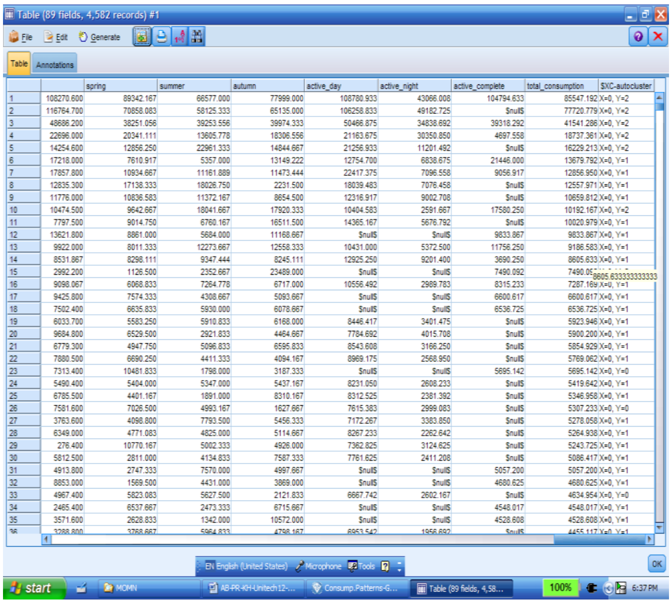

***Fig. 10.** Fragment of table with Kohonen Cluster membership*

These results could assist decision making and innovation processes of each household about investment in the private, small power hybrid electro generator.

To integrate and observe the results about the type of location and clusters from the available Graphs in the SPSS Modeler we used so called Web – figure 11.

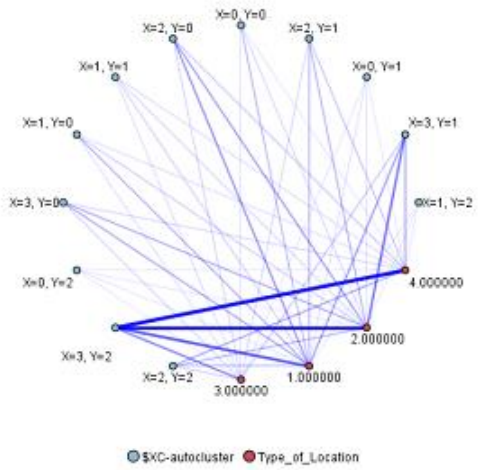

Fig. 11. *Relationship between Kohonen Cluster membership and type of location - web visualization*

The figure shows that normally cities and villages are connected with cluster (X=3, Y=2), i.e. the biggest cluster with relatively low consumptions. On the opposite – the cluster with largest consumption (X=0, Y=2) is connected with all four types of locations, but the relationships are weakest.

CONCLUSION AND FUTURE WORK

The results include electricity consumption patterns in west Bulgaria regions based on Auto Cluster node. In conclusion, this paper investigates the concept of utility company's client consumption patterns and moves it towards the broader concept of eco system in which not only utility, but also other players and factors involved in the energy supply process - client's characteristics and processes are considered. It is trying to illustrate how deep electricity consumption pattern understanding and future innovation efforts be required to work together to increase the energy efficiency and effectiveness.

The analyzed models, consumption patterns recognition and results can be applied by utility companies, their partners and consultants for future energy processes innovation and electric business transformation.

Our future work will integrate other variables such as statistical climate data, demographic data, specific social and economic parameters and related to create a powerful mechanism aimed to identify and forecast changes in consumption based on identified triggers in the surrounding environment.

Acknowledgement

The authors are grateful to the Marketing & Sales support department of "CEZ Electro Bulgaria" AD for the explicit sample data base and helpful information.

The work on this paper has been sponsored by IBS Bulgaria, IBM Premier Business partner.

REFERENCE

[1]. Anthony S., Johnson, Mark W., Sinfield, Joseph V., Altman, Elizabeth J. Innovator's Guide to Growth - Putting Disruptive Innovation to Work - Foreword by Clayton M. Christensen. Harvard Business School Press. 2008, ISBN 13-9781591398462.

[2]. CEZ Electro Bulgaria AD http://www.cez.bg/en/home.html, http://www.cez.bg/en/customer-service/ways-of-payment.html.

[3]. Christensen C. M., Scott D. Anthony, and Erik A. Roth, Seeing What's Next: Using the Theories of Innovation to Predict Industry Change, Harvard Business Review Press, 2004.

[4]. A. Druckman, T. Jackson, Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model, Energy Policy, - Elsevier, 2008.

[5]. Geels, F.W., Processes and patterns in transitions and system innovations: Refining the co-evolutionary multi-level perspective, Technological Forecasting & Social Change 72 (2005) 681–696.

[6]. Carl Haigney, Customer Relationships and the Economy, Shaping a New Era in Energy, www.UtilitiesProject.com.

[7]. IBM Corp., CRISP-DM 1.0, Step-by-step data mining guide, © Copyright IBM Corporation 2010.

[8]. IBM Corp., Switching perspectives, creating new business models for a changing world of energy, 2010.

[9]. LaValle S., Michael Hopkins, Eric Lesser, Rebecca Shockley and Nina Kruschwitz, Analytics: The new path to value, How the smartest organizations are embedding analytics to transform insights into action, IBM Institute for Business Value, MIT Sloan Management Review, 2010.

[10]. A-L Lindén and A Carlsson-Kanyama, Energy Efficiency in Residences - A challenge for women and men, Energy Policy, 35, 2007, pp. 2163-2172.

[11]. Miles I., Patterns of innovation in service industries, IBM Systems journal, vol.47, No 1, 2008, pp.115-128.

[12]. Morelli T., Shearer C., Buecker A., IBM SPSS predictive analytics: Optimizing decisions at the point of impact, IBM redpaper 4710, Copyright IBM Corp. 2010.

[13]. Pearson N., Eric Lesser and Joel Sapp, IBM Global Business Services, Executive Report, IBM Institute for Business Value, A new way of working, Insights from global leaders, 2010.

[14]. Peter Skarzynski, Rowan Gibson, Innovation to the Core A Blueprint for Transforming the Way Your Company Innovates, Harvard Business School Publishing Corporation, 2008.

[15]. C. K. Prahalad and M. S. Krishnan, The new age of innovation, McGraw-Hill, 2008.

[16]. Ruskov P., A. Bachvarov, Disruptive innovations for the energy system, Anniversary Scientific Conference with International Participation "40 Years Department of Industrial Automation", 18 March 2011, Sofia, Bulgaria, pp. 141-144.

[17]. Ambuj D. Sagara, Bob van der Zwaana, Technological innovation in the energy sector: R&D, deployment, and learning-by-doing, Energy Policy 34 (2006) 2601–2608.

[18]. Lukas G. Swan, V. Ismet Ugursal, Modeling of end-use energy consumption in the residential sector: A review of modeling techniques, Renewable and Sustainable Energy Reviews 13 (2009) 1819–1835, 2009.

[19]. Marc Teerlink and Michael Haydock, Customer analytics pay off, Driving top-line growth by bringing science to the art of marketing, Copyright IBM Corporation 2011.

[20]. Timothy J. Foxona, Geoffrey P. Hammondb and Peter J.G. (2010), Developing transition pathways for a low carbon electricity system in the UK, Pearsonc Technological Forecasting and Social Change, Issue includes a Special Section on "Infrastructures and Transitions", Volume 77, Issue 8, October 2010, Elsevier Inc., pp. 1203-1213.

.