International Journal of Computer Science, Communication & Information Technology (CSCIT)
Vol.6 pp.1-7

A Service-Oriented Data Mining Platform: System of Assistance to Epidemiological Research and Monitoring of the Diseases

SABRI Mohammed[#1], RAHAL Sidi Ahmed[*2]

[#]*Computer Sciences Department,*
National Polytechnic School of Oran, Algeria
[1]mohammed.sabri.u@gmail.com
[*]*Computer Sciences Department*
University of Sciences and Technology-Mohamed Boudiaf, Oran, Algeria
[2]rahalsa2001@yahoo.fr

***Abstract*— This article relates our contribution in the field of public health and epidemiology through the design of a System of Assistance to Epidemiologic Research and Monitoring of Diseases (SARESM). SARESM provides to epidemiologists assistance for the definition of medical policies, and more specifically the planning of acquisition of pharmaceutical products for a given disease according to different factors including for example the period of the year and the geographical distribution of their use. Our contribution in this field is to provide prediction models for chronic diseases such as diabetes and asthma. These models are based on Service-Oriented Data Mining techniques, and an approach of Boolean modeling of the induction graphs. The proposed platform is a flexible and scalable system for aid in decision-making by trades' experts after extracting the epidemiologic prediction rules.**

***Keywords*— Data Mining, Services Oriented Architecture (SOA), Data Warehouse, Epidemiology**

I. Introduction

The development of information systems and computer technologies has enabled the automation of the activities in every field of the real-world; this has induced a fast increase in the information availability, the development of high volume data ware-houses and finally, the emergence of Data Mining. The latter enables the extraction of available knowledge, until now hidden within the data, to be used in various fields such as trade, banking, public health, etc. Public health is the main concern for the world population and calls upon several disciplines, for the well-being of all.

However, the growing market draws attention to distributed Data Mining [8]; on the one hand, data and software are geographically distributed over a network instead of being located in a single site, and on the other hand, the cost is another reason for the distribution. In addition, thanks to the arrival of Web and Grid computing, distributed data is now much easier to access and distributed computing in heterogeneous environments became much more feasible [9]. At the same time, service-oriented architectures (SOA) are becoming one of the main paradigms for distributed computing [13]. Through an approach based on services, especially service-oriented architecture (SOA), integrated services can be defined to support the distributed data mining and knowledge discovery in databases (KDD) tasks in grids and the Web. The most important SOA implementation is represented by web services [6].

Let us recall that data mining has attracted a great deal of attention in the epidemiological information processing, as well as in area medical and public health as a whole. An important issue in the field is to relate a specific disease, for example asthma or diabetes, with physiological (e.g. age, gender) and environmental factors (temporal attributes such as period in the year, or spatial attributes such geographical locations). The available data in the field are complex, heterogeneous and uncertain, and it is not easy for medical doctors in the field to generate predictive rules linking a particular disease with these physiological and environmental factors. The main aim of our study will be to provide to experts in the field a platform that will enable them to generate such rules using data mining tools, and answer question related to the prevalence of a particular disease as a function of multiple factors such age or even geographical location.

Therefore, we propose here a platform for predicting models for the monitoring of chronic diseases (asthma and diabetes in the present case) guided by the service-oriented data mining which applied on real data related to drug sales in private pharmacies (that includes both physiological attributes such as age and gender, and environmental attributes). A new service-oriented data mining platform for the System of Assistance to Epidemiologic Research and Monitoring of Diseases (SARESM) [23] is thus proposed. This latter is part of an approach based on both (i) the storage and pre-processing of data and, and (ii) extraction of epidemiological prediction rules guided by the service-oriented data mining with the approach of Boolean modeling of the induction graphs [3].

The objective of our approach is to build a system that enables (i) extracting prediction rules, flexible and scalable for aid in decision-making by trades' experts, and (ii)

ISSN 1737-930X

reducing the knowledge management complexity and the response time.

II. Related Work

For our approach, we have performed researches simultaneously in the field of pharmaceutical sales and study oriented-service data mining platforms.

Data mining in the pharmaceutical field, thanks to the information collected (Data coming from the direct sales to customers and other sources such as hospitals and medical reports) from the data records, has targeted two potential entities to obtain data mining models answering various aspects: the first entity is related to the patient's behavior (or customer) and the second one is related to the study of the product. The various models, obtained from the application of various data mining techniques, can be classified according to the objectives of each approach:

- Public health: Often faced with the marketing pressures, it tries to follow the medical prescription, using the retail data obtained from pharmacies purchasing records and health insurance companies, to determine which drugs the doctors prefer for specific diagnoses. This will allow creating detailed portraits for prescription of each doctor [14].
- Management and monitoring of some pathologies: Several studies have been conducted to understand and monitor certain diseases. These studies are aimed at both (i) understanding the sale of certain products for selected patients (e.g. asthma [4]) and (ii) determining the long-term effects of the intervention program [5]. Moreover, the development of new pharmaceuticals and the advancement in cancer therapies [28], and finally, the development of the bio-monitoring systems that can be used to identify normal manifestations of the disease and events resulting from the bio-attacks [22] are also considered.
- Assistance system to the medical prescription: The objective is to improve the medical prescription among practitioners, based on the results of data mining on medical records and data from pharmacies. The contribution of this prescription assistance system is multiple: first, it helps medical doctors in setting up recommendations; second, it encourages the prescription of the products that can be cheaper and more efficient [19]; third it provides exploration of reactions for unfavourable drug to alert physicians about the potential adverse effects [7].
- Optimal management of stock and its modeling: Managing a multi-product stock means getting the right product, in terms of quantity and availability when needed, taking into account customer behavior. Several models are obtained, cited in [2] to examine the dependence of purchase in the sale.
- Profit making in the sales of drugs: Finally, the commercial side is a field that uses data mining techniques to make financial profits. The important business entities in the pharmaceutical field used data mining techniques to increase their turnover significantly and thus achieve remarkable profits [15].

There have been many studies published aimed at adopting data mining platforms to service-oriented architecture paradigm.

FAEHIM (Federated Analysis Environment for Heterogeneous Intelligent Mining) based on Web services-based toolkit for supporting distributed data [25]. Dynamic data mining process (DDMP) system based on service-oriented architecture (SOA) introduced in [10], where each Web service represents an activity in data mining process. Xu et al. [30] proposed a service-based architecture for data mining applications, including a set of services (configuration, engine, monitor...). Weka4WS described (design and implementation) in [29] uses the WSRF libraries and services provided by Globus Toolkit. SINDBAD SAILS (Service Architecture for Inductive Learning Schemes) is a web service interface implemented in PHP and dedicated to the SINDBAD platform; this interface allows a user to access SINDBAD data mining techniques [16]. Anteater: a service-oriented architecture for data mining was developed by Ferreira et al. [13]. This platform relies on Web services to achieve extensibility and interoperability. It offers simple abstractions for users, and supports computationally intensive processing on large amounts of data through massive parallelism. Latha et al. [18] proposed a novel method to develop service oriented architecture for a weather information system and forecast weather using data mining techniques. These authors mentioned that this method aims at developing a weather information system as a web service that can be used by any type of application and uses the prediction techniques of data mining for weather forecasting. In a different work [17], other authors present the design and realization of a new open source platform, WS4KDM: Web Services for Knowledge Discovery and Management. This platform regroups several web services dedicated to the extraction and knowledge management. Data mining techniques, implemented in the WS4KDM platform are based on induction graph and use a named Boolean modeling technique BML (Boolean Modeling Language), which is based on the cellular principle CASI. SOMiner proposed by Birant D. [6] is a flexible service-oriented data mining architecture that incorporates the main phases of knowledge discovery process. Overall, this architecture provides a large collection of machine learning algorithms written for knowledge discovery tasks. Zorrilla and Garcia-Saiz [33] proposed a model which joins both facets: data mining and SOA. It describes a data mining service addressed to non-expert data miners which can be delivered as Software-as-a-Service. Shelke et al. [26] have proposed an architecture to improve mobile data mining techniques so that data retrieval for mobile devices is faster and mobility management efficient using proper web services.

In this context, the SARESM platform's concept proposed in this paper focuses on service-oriented data mining applied to both Epidemiologic Research and Disease Monitoring, a field of investigation that necessitates important resources in

ISSN 1737-930X

terms of algorithm definition, life cycle management, and visualization reporting.

III. Approach

The approach that has been taken in the design and the implementation of SARESM (Fig. 1) results from the overall process of knowledge discovery from databases.

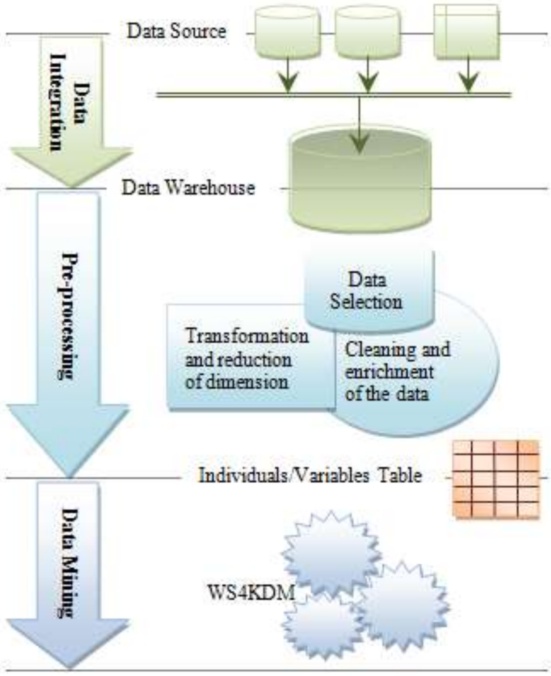

Fig. 1. Illustrating the selected approach for SARESM

A. Data warehousing

The first phase of our approach is the design of the data warehousing, with the objective of obtaining a unique source of data to carry out the data mining tasks.

Our data sources are based on records of sales from private Algerian pharmacies. The architecture of the SARESM data warehouse (Fig. 2) is articulated around three axes:

- Integration: This first step consists of extracting and gathering the data coming from the various databases of private pharmacies and the external sources [20]. These databases are supported by the same relational DBMS (Data Base Management System), they are identical in terms of structures, and are installed in different sites (where no connection exists between these sites). The source databases (files) are coded and stored in the file system.
- Building: It consists of extracting the relevant data and copying it in the warehouse [20]. Consequently, the SARESM warehouse will constitute a centralized collection of materialized and historical data, available for data mining. Whereas the data related to the drug sales and the characteristics related to the sold products are taken into account in our study, the other data such as the purchases are not considered.
- Structuring: This step consists in reorganizing the data, in data marts, to support the data mining [24]; a specific data mart is created, in relation to the information concerning the chronic diseases selected out of all the retail sales and the characteristics of the patients belonging to the essential variables for the data mining.

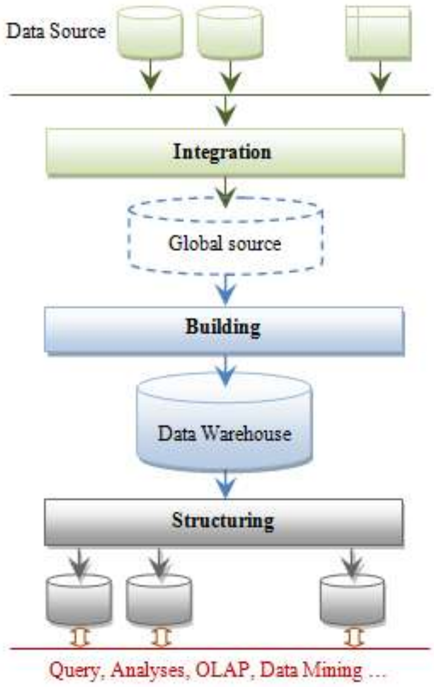

Fig. 2. Architecture of SARESM data warehouse.

SARESM data warehouse is based on the star model (Fig. 3) and contains all the information about retail sales, products and places where the pharmacies are located. The source data available is commercial data (basic sales records) to carry out medical research (epidemiologic), and according to the studies carried out by F. Ravat and al. [21], we chose a traditional multidimensional modeling for the SARESM data warehouse. The data of the warehouse are as follow:

The fact table “VENTES”; contains sold quantity (gross), selling price, etc.

The dimension tables:

- Localization of selected pharmacies “LOC_OFFI”.
- Dimension date “DATES”.
- Table of the handled products “PRODUITS” includes commercial name, etc.
- Specialties of the various existing products in database “SPECI_PDT”.
- “DCI” and corresponding diseases.
- Laboratories manufacturers’ products “LABORAT”.
- Patients “ASSURES”. Confidential details NOT downloaded in data warehouse.

ISSN 1737-930X

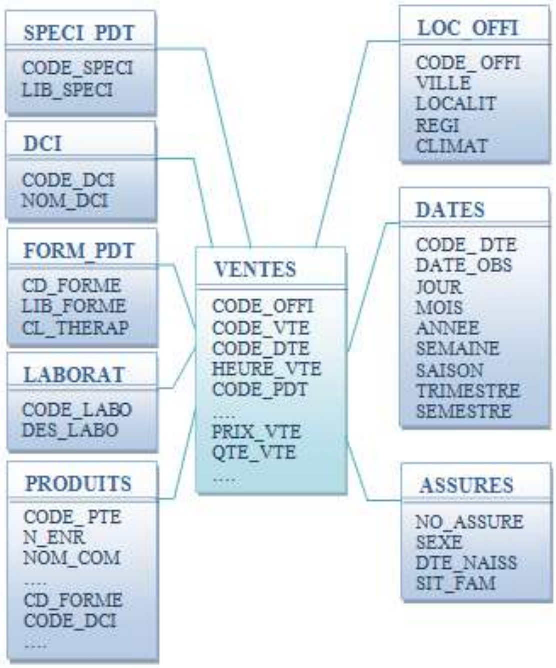

Fig. 3. The star model of SARESM.

B. *The pre-processing*

Data from the warehouse is very varied and is not necessarily all exploitable by the data mining techniques [27]. Most of the used techniques process only data tables in the traditional lines/columns. The objective is to prepare lines/columns tables, in other words, tables of individuals/variables (Table I), obtained by the following stages:

TABLE I
EXAMPLE OF A TABLE OF INDIVIDUALS/VARIABLES.

	Town	**Season**	**Age**	**Gender**	**Diseases**
ω1	Oran	Winter	Young	Male	Asthma
ω2	Oran	Winter	Young	Female	Asthma
ω3	Tlemcen	Winter	Young	Male	Diabetes
…	…	…	…	…	…

1) *Data selection*: It is carried out on the data which already exist in the data warehouse and which are in tabular form. Filters are then applied to select a subset of lines or columns [12]. Data selection is based on the following information:

- From the fact table “VENTES”, we will take the sold quantity, taken first in its “gross” state and aggregated according to the selected dimensions.
- From table “LOC_OFFI”, the attribute “LOCATION”.
- The date dimension “DATES” in order to carry out the data mining on a time interval. In our case, we proceed by the period “MONTH”.
- From the table “DCI”, we take information about present diseases. A filter is then applied to keep the records related to the selected diseases only.
- Finally, the patients, present in the table “ASSURES” and from which we take the gender and age attributes (recommendations of the experts).

2) *Cleaning and enrichment of the data:* A stage of cleaning of the data is essential in order to process the missing data (suppression of records). Besides, enrichment by external sources was carried out during the creation of the data warehouse [12].

3) *Transformation and reduction of dimension:* This is about transforming an attribute A into another A’ which would be more relevant to match the objectives of the study [12]. For example, patients’ dates of birth have been transformed to obtain age within intervals (see various discretization methods in [11]).

C. *Service-Oriented Data Mining (WS4KDM)*

After data storage and pre-processing, the phase of data mining may start. For producing epidemiological prediction rules, we chose WS4KDM [17] as the service-oriented data mining module. WS4KDM was developed in order to regroup several web services dedicated to the extraction and knowledge management. Data mining techniques, implemented in the WS4KDM platform are based on induction graph and use a named Boolean modeling which is based on the cellular principle [3] (Fig. 4). This approach is named CASI [1] and is based on the induction graphs (IG) methods produced by the SIPINA algorithm [31] [32].

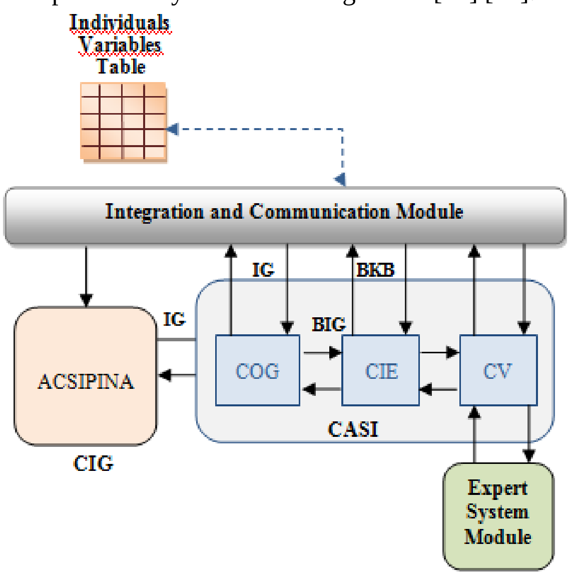

Fig. 4. Architecture of WS4KDM platform.

The architecture of WS4KDM is articulated around:

1) *CIG (Cellular Induction Graph) Module:* this module uses the SIPINA method that produces induction graphs.

2) *CASI Machine Module:* CASI [1] is a cellular automata which simulates the basic operating principle of an inference engine. It ensures the optimization of the induction graph (IG), the generation of the conjunctive rules (“If cond1 [AND cond2]…[AND condn] Then conclusion”), and the validation

ISSN 1737-930X

of the obtained model. Starting from a training sample, a symbolic processing starts for the IG building (e.g. ACSIPINA, ACID3, and ACJ48 algorithms). CASI module is composed by three sub-modules:

- COG (Cellular Optimization and Generation) consists of assisting the processing made by the ACSIPINA algorithm (or ACID3, or ACJ48) to generate the Boolean Induction Graph (BIG).
- CIE (Cellular Inference Engine) can, from the BIG, generate a Boolean Knowledge Base (BKB). This module simulates the operation of the basic cycle of an inference engine by using two finite layers of finite automata: CELFACT, for the fact base and CELRULE for the rule base.
- CV (Cellular Validation) is devoted to the validation process of the proposed model.

3) Integration and Communication Module: it ensures the communication between the various modules proposed in the WS4KDM platform. It presents the Input/Out layer of different operations proposed by the platform as web services.

4) *Expert System Module:* it operates the CASI machine principle and uses the CIE module as cellular inference engine to set off inferences in forward and backward. This module can be used for validation by the deduction of rules extracted from data, and it can be used independently of the KDD process to develop dedicated expert systems.

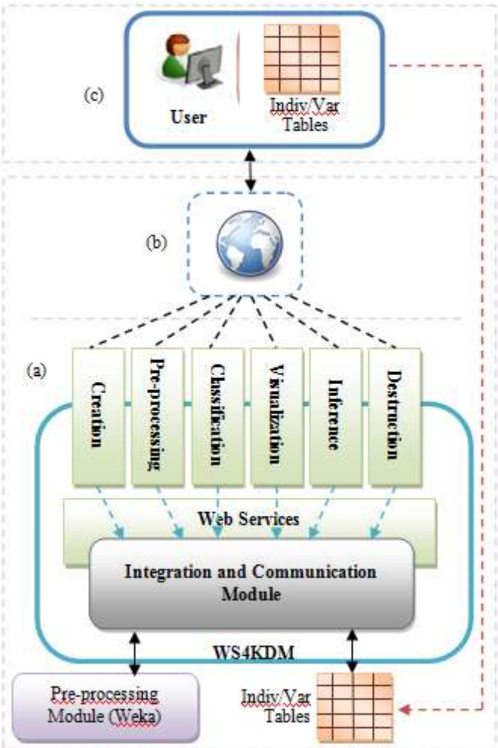

Fig. 5. Web services of WS4KDM platform.

Fig. 5 illustrates the implementation of the WS4KDM platform and the operating mode of its different web services (a). These latter are:

- Creation: This service allows a customer to create a session (account) or to access at its session via a login, to create its resources to ensure the services performance coherence.
- Pre-processing: This service is designed to make available to the customer various methods of filtering and discretization provided by the module pre-processing.
- Classification: A specification service of the desired method by the client to generate the model. It is characterized by the name and settings of the algorithm (the classification method), the learning base and the test database.
- Visualization: This service allows visualizing the generated model graphically (vertices and arcs). The graph is a Boolean matrix.
- Inference: This service renders available for the clients an inference engine and knowledge management. The service receives a set of rules and a test database (classification and class assignment) to validate the client model or affect the class for the individuals in the base.
- Destruction: Service for liberate the resources used by the client in the WS4KDM platform and to log out.

Via a network connection (local, Internet) (b), users can access to WS4KDM platform through an interface developed (c).

IV. Experiments

Before giving the results of the experimental phase, we want to highlight the fact that the data warehousing has represented a major task in the implementation of the project, especially the data collection. Data records of sales, spread out between 2003 and 2010, related to 4 departments in the west of Algeria. It should be noted that these records represent raw data of sales on which no form of aggregation was carried out. Therefore, in a first step, data were pre-processed by identifying diseases (Asthma and Diabetes), patients' characteristics (Gender and Age) and environmental attributes.

Moreover, after a second step of investigation, we realized that there was a significant number of redundant sales transactions. Because the presence of combinations of several products for the same disease exist in the same prescription (e.g. insulin and Glucophage in the same prescription to care diabetes), it was decided in collaboration with experts to analyse the data by medical prescription rather than by sale detail. Therefore, the generation of data in the DataMart has been improved, which allowed us to have a total of about 122,000 acts of prescriptions (combination of sales acts).

We then tested the service-oriented data mining platform SARESM with five different bases resulting from the chronic disease DataMart. We first applied service-oriented data mining on the entire database for the western region of Algeria (called BRO). Second, we passed to the other four departmental bases called TLM, SBA, ORN, and ATM referring, respectively, to the departments of Tlemcen,

ISSN 1737-930X

Belabbes, Oran and Temouchent. These bases are described by the same common attributes.

In order to carry out our experimentation, we imported the data in WS4KDM, and then selected the attributes and the variable to predict (Table II) (as stated above, induction was launched using SIPINA).

Fig. 6 presents the individual numbers of learning samples of the experimental bases.

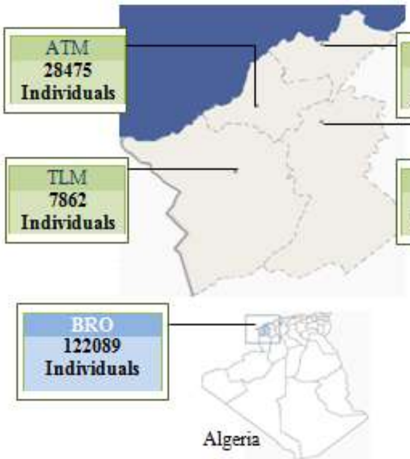

Fig. 6. Characteristics of the five bases of experimentation.

For the five selected bases, we used the same attributes mentioned in Table II (excluding the department attribute).

Note that we have added a diabetes-high blood pressure combination class since it was important for experts to dissociate between diabetes alone and diabetes accompany by high blood pressure.

TABLE III
REPRESENTATION OF ATTRIBUTES AND CLASS

Attributes	**Signification**	**Possible Values**
X1 : MONTH	The selected period is the calendar month	01 (January), .., 12 (December).
X2 : WILAYA	Department number	13, 22, 31, and 46
X3 : LOCATION	The locality or city relating to the chosen pharmacy	TLEMCEN, SENIA, ARZEW…
X4 : CLIMAT	The climate versus humidity mainly	High, Average, and Low.
X5 : PROXIM	Proximity of the location in relation to the sea	Inside, Coastal
X6 : GENDER	Patient's gender	M and F
X7 : AGE	Patient's age - Age intervals	Chi (<=16), Adt1 (16 - 40), Adt2 (41-65), and Old (>65).
Y : CLASS_THRP	Class of disease to be predicted	ast (Asthma), dbt (Diabetes) et dbt-hta (Diab-High blood pressure).

- "BRO" experimentation: firstly (case 1), we have used all attributes (Table II) and we can visualize 34 epidemiological prediction rules (conjunctive rules) produced by SARESM platform. In a second experiment (case 2), we did not select the attribute "LOCATION", and obtained 13 epidemiological prediction rules (Table III). In this case, the success rate was 86.7989% (105,972 correct instances for 122,089 acts).

TABLE IIIII
CONJUNCTIVE RULES PRODUCED IN THE EXPERIMENT OF BRO – CASE 2

1	if (age = old) then dbt-hta
2	if (age = adt1) then ast
3	if (age = chi) then ast
4	if (climat = high and age = adt2) then ast
5	if (wilaya = 46 and climat = aver and age = adt2) then ast
6	if (wilaya = 31 and climat = aver and gender = m and age = adt2) then dbt
7	if (wilaya = 31 and climat = aver and proxim = coastal and gender = f and age = adt2) then dbt
8	if (wilaya = 13 and climat = low and age = adt2) then dbt
9	if (wilaya = 46 and climat = low and age = adt2) then dbt
10	if (wilaya = 31 and climat = aver and proxim = inside and gender = f and age = adt2) then dbt
11	if (wilaya = 22 and climat = low and age = adt2) then dbt
12	if (wilaya = 31 and climat = low and proxim = inside and age = adt2) then ast
13	if (wilaya = 31 and climat = low and proxim = coastal and age = adt2) then dbt

V. CONCLUSIONS

In this article, we have used a service-oriented data mining approach applied to the field of epidemiology. Experts in the field are faced with the issue of relating a specific disease (asthma or diabetes in our case), with physiological and environmental factors (age, gender, period in the year, geographical locations). We provide here prediction models of such chronic diseases (asthma and diabetes), using Boolean modeling of the epidemiologic prediction rules. In the context of the analysis of the chronic diseases, the generated cellular induction graph enables us to determine the relations between the disease and the people exposed to it versus the physiognomic characteristics and the environment. These generated models will facilitate the identification of the diseases by experts so that the patients are better taken care of. In addition, it is worth noting that our platform is flexible at various levels. First, the user can decide which parameters are of interest in their analysis. For example, the "BRO" experiment (entire base) provided different classification rates (92.8% vs. 86.8%) and different number of predictive rules (34 vs. 13) as a function of including or not the variable "LOCATION". In sum, the user can opt between a very highly accurate but complex model (92.8% associate to 34 rules) and a highly accurate but simpler model (86.8% associated to 13 rules). Second, another form of flexibility is the possibility of enriching the bases with additional data

ISSN 1737-930X

resources (updating, new pharmacies, new patients, and new characteristics).

References

[1] B. Atmani and B. Beldjilali, “Knowledge Discovery in Database: Induction Graph and Cellular Automaton,” Computing and Informatics Journal, vol. 26, no. 2, pp. 171-197, 2007.

[2] P.K. Bala, “Data Mining for Retail Inventory Management,” Advances in Electrical Engineering and Computational Science, Sio-Iong Ao, Len Gelman, Springer Science & Business Media, pp. 587-598, 2009.

[3] M. Benamina and B. Atmani, “WCSS: un système cellulaire d'extraction et de gestion des connaissances,” Troisième atelier sur les systèmes décisionnels, 10 et 11 octobre 2008, Mohammadia, Maroc, pp. 223-234, 2008.

[4] B.J. Bereznicki, G.M. Peterson, S.L. Jackson, H. Walters, K. Fitzmaurice, and P. Gee, “Pharmacist-initiated general practitioner referral of patients with suboptimal asthma management,” Pharm World Sci, 2008.

[5] B.J. Bereznicki, G.M. Peterson, S.L. Jackson, H. Walters, and P. Gee, “The sustainability of a community pharmacy intervention to improve the quality use of asthma medication,” Journal of Clinical Pharmacy and Therapeutics, vol. 36, no. 2, pp. 144-151, 2011.

[6] D. Birant, “Service-Oriented Data Mining,” New Fundamental Technologies in Data Mining, K. Funatsu and K. Hasegawa, Published by InTech, Croatia, pp.3-18, 2011.

[7] J. Chen, H. He, J. Li, H. Jin, D. McAullay, G. Williams, R. Sparks, and C. Kelman, “Representing Association Classification Rules Mined from Health Data,” International Conference on Knowledge-Based Intelligent Information and Engineering Systems, 9th , KES 2005, Melbourne, Australia, Lecture Notes in Computer Science, Volume 3683, 2005, pp. 1225-1231, 2005.

[8] N. Chen, N.C. Marques, and N. Bolloju, “A Web Service-based approach for data mining in distributed environments,” in Proceeding of the 1th International Workshop on Web Services: Modeling, Architecture and Infrastructure (WSMAI-2003), Angers, France, pp. 74-81, 2003.

[9] W.K. Cheung, “Scalable and privacy preserving distributed data analysis over a service-oriented platform,” Data Mining Techniques in Grid Computing Environments, W. Dubitzky, University of Ulster, UK, pp. 105-118, 2008.

[10] C.C. Chiu and M.H. Tsai, “A Dynamic Web Service based Data Mining Process System,” in Proceeding of the 5th International Conference on Computer and Information Technology (CIT 2005), Shanghai, China, pp.1033-1039, 2005.

[11] J. Dougherty, R. Kohavi, and M. Sahami, “Supervised and unsupervised discretization of continuous attributes,” Machine Learning: Proceedings of the 12th International Conference (ICML-95), Morgan Kaufmann, pp. 194-202, 1995.

[12] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining, American Association for Artificial Intelligence Menlo Park, 1996.

[13] R.A. Ferreira, O.G. Dorgival, and W.Jr. Meira, “Anteater: A Service-Oriented Data Mining,” Data Mining Techniques in Grid Computing Environments, W. Dubitzky, University of Ulster, UK, pp. 179-198, 2008.

[14] A. Fugh-Berman, “Prescription Tracking and Public Health,” Journal of General Internal Medicine, vol. 23, no. 8, pp. 1277-1280, 2008.

[15] Y. Hamuro, N. Katoh, Y. Matsuda, and K. Yada, “Mining Pharmacy Data Helps to Make Profits,” Data Mining and Knowledge Discovery, vol. 2, no. 4, pp. 391-398, 1998.

[16] W. Jörg, C. Brosdau, L. Richter, and S. Kramer, “SINDBAD SAILS: A Service Architecture for Inductive Learning Schemes,” In Proceedings of the 1st Workshop on Third Generation Data Mining: Towards Service-oriented Knowledge Discovery, 2008.

[17] H. Kadem and B. Atmani, “Conception d’une Plateforme Cellulaire Open Source d’Extraction et de Gestion des Connaissances : WS4KDM,” 7ème Séminaire National en Informatique BISKRA (SNIB'2010), Université Mohamed Khider-Biskra, Algérie, 2010.

[18] C.B.C. Latha, P. Sujni, E. Kirubakaran, and S. Narayanan, “A Service Oriented Architecture for Weather Forecasting Using Data Mining,” The International Journal of Advanced Networking and Applications (IJANA), vol. 02, no. 02, pp. 608-613, 2010.

[19] K. Melley and K. Petersen, “Prescription Data Mining,” Fact Sheet, Pew Prescription Project, http://www.prescriptionproject.org, 2008.

[20] F. Ravat, O. Teste, and G. Zurfluh, “Modélisation et extraction de données pour un entrepôt objet,” Université Paul Sabatier (Toulouse III), IRIT, équipe SIG, 2000.

[21] F. Ravat, O. Teste, and G. Zurfluh, “Modélisation multidimensionnelle des systèmes décisionnels,” Extraction des connaissances et apprentissage, vol 1, no. 1, pp. 201-212, 2001.

[22] M.R. Sabhnani, D.B. Neill, and A.W. Moore, “Detecting Anomalous Patterns in Pharmacy Retail Data,” Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection, Chicago, Illinois, USA, pp. 132-137, 2005.

[23] M. Sabri and B. Atmani, “Système d’assistance aux recherches épidémiologiques et de surveillance des maladies : Modélisation Booléenne,” Colloque International “Veille Stratégique Scientifique et Technologique (VSST)”, VSST’10, Toulouse, France, 2010.

[24] N. Selmoune, S. Boukhedouma, and Z. Alimazighi, “Conception d’un outil décisionnel pour la gestion de la relation client dans un site de e-commerce,” SETIT 3rd International Conference, Tunisia, 2005.

[25] A.A. Shaikh, O.F. Rana, and I.J. Taylor, “Web services composition for distributed data mining,” in Proceeding of the 34th International Conference on Parallel Processing Workshops (ICPP 2005 Workshops), Oslo, Norway, pp. 11-18, 2005.

[26] R.R. Shelke, R.V. Dharaskar, and V.M. Thakare, “Data mining for mobile devices using web services,” in Proceeding of the International Conference on Industrial Automation And Computing (ICIAC), Lonara, Nagpur, 2014, Published in : International Journal of Engineering Research and Applications (IJERA), vol. 8, no. 2, pp. 7-9, 2014.

[27] L. Soibelman, M. Asce, and K. Hyunjoo, “Data Preparation Process for Construction Knowledge Generation through Knowledge Discovery in Databases,” Journal Of Computing In Civil Engineering, vol. 16, no. 1, pp. 39-48, 2002.

[28] S. Sumathi and S.N. Sivanandam, “Data Mining in Biomedicine and Science,” Introduction to Data Mining and its Applications, Springer, pp. 499-543, 2006.

[29] D. Talia, P. Trunfioy, and O. Verta, “The Weka4WS framework for distributed data mining in service-oriented Grids,” Concurrency and Computation: Practice and Experience, vol. 20, no. 16, pp.1933–1951, 2008.

[30] L. Xu, Y. Wang, G. Geng, X. Zhao, and N. Du, “SDMA: A Service based Architecture for Data Mining Applications,” IEEE International Conference on Services Computing, pp. 473-474, 2008.

[31] Zighed D.A., “Méthodes et outils pour les processus d'interrogation non arborescents,” PhD Thesis, Université Lyon 1, 1985.

[32] D.A. Zighed, J.P. Auray, and G. Duru, “SIPINA: Méthode et Logiciel,” Lacassagne, 1992.

[33] M. Zorrilla and D. García-Saiz, “A service oriented architecture to provide data mining services for non-expert data miners,” Decision Support Systems, Elsevier, vol. 55, no. 1, pp. 399-411, 2013.

ISSN 1737-930X