

Review

Extracting Healthcare Information from Retail Business Data Warehouses

Amer Al-Nassiri and Syed Imtiaz Ali Rizvi*

College of Information Technology, Ajman University of Science and Technology, United Arab Emirates (UAE)

Copyright: © 2016 Syed Imtiaz Ali Rizvi, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Healthcare information is traditionally collected through surveys. This traditional method cannot explore much of the real information. There should be an indirect way to collect the healthcare information about individuals and communities. Such information provides a real-time insight into the health situation of individuals or communities. Currently Data warehousing is a common source in Businesses to get information to plan and to know the current and future trends in business. The main source of business data are individuals and communities; so why not this huge reservoir of information is used for healthcare. In this article we describe two analytical techniques based on support vector machines (SVMs) for data analysis to extract and classify healthcare information from a typical Business data warehouse about individuals and communities. Working out on data collection of a local chain of retail market in UAE and from the purchase habits of consumers get their healthcare information. Different kernels are be used in Support of Vector Machines models. These include linear, polynomial, radial basis function (RBF) and sigmoid as part of mapping system and analysis of healthcare information. Finally this technique can be used to correlate the extracted information with the existing standards of International health organizations at a national and global level and suggest the change in purchase habits of individuals and communities in context of healthcare.

Keywords: Healthcare management, Data Wwarehouse, Business intelligence, Data mining, Retail business, Support Vector Machine (SVM), UAE health care

Introduction

Businesses need data from a variety of data sources and like to perform complex data analysis across these data sources, and create multidimensional views of data that represent the business analyst's perspective of the data [1]. This can be achieved through corporate knowledge repository, also called data warehouse.

Data warehousing in corporate businesses is a widely adopted technique in world. The data warehouse is the significant component of business intelligence. Successful Business Intelligence initiatives have been undertaken across major industries and for varied applications, including health care [2,3], security and event management [4], telecommunications [5], web analytics [6], and Six Sigma [7,8].

Application areas of datawarehouse technique has been limited to business benefits [9]. Although the healthcare industry is one of the world's largest, fastest-developing and most information-rich industry [11] but dataware house techniques was not used for health care as it can be. In health care industry it is used to increase the performance of healthcare centers or its business' benefits like drug analysis, diagnosis, quality control and epidemiological studies [9].

The idea we are presenting in this paper is to use the data warehouse of retail business for predicting unhealthy issues of a population. Literature overview on the topic shows that this concept is not discussed much in this context and the results are not flashy enough. For healthcare industry the datawarehouse technique was used for performance of pharmaceutical companies [10], to estimate the cost of treatment for cancer and other diseases, analysis of death rate in specific type of cancer and the impact of a particular drug on related cancer disease [12], to gather clinical information about influenza [13] and Malaria[19], to diagnose Lymphoma Cancer and support its treatment [14], and to improve the existing health care

statistics data warehouse design [15].

Our research objective is to find a solution from retail business data warehouse to lower disease occurrences, improve patient care, and hence lower the healthcare costs. Some work can be seen on disease management program using data warehouse techniques [16] but again they have not used the retail business data warehouse for this purpose. Disease Management Program is very important because considering only the chronic illnesses, it account for over 70% of the total healthcare costs in the United States.[17] Health insurance payers and health care providers are keen to identify strategies to manage number of patients involved in such diseases in order to reduce cost and improve patient care.[17] Health insurance payers have data from various sources; it is a data warehouse but all papers show that no one has suggested to use retail business data for this purpose [18].

Healthcare industry has not fully utilized the potential of retail business data for disease and healthcare management or strategic decision making. On the other hand retail business community has also ignored to use their retail transactional data for healthcare monitoring and management. Corporate and Business community used the retail business data for Sales Performance Analysis [20], Customer Management Analysis [21], and many such topics that has no relation with healthcare.

This research encourages using the retail business data warehouse for healthcare purposes. This research suggests using the data gathered from grocery shops, super and hyper markets and other outlets for understanding the necessary consumption of energy

*Corresponding author: Syed Imtiaz Ali Rizvi, College of Information Technology, Ajman University of Science and Technology, United Arab Emirates (UAE), Tel: E-mail: s.rizvi@ajman.ac.ae

Received: November 01 2016; **Accepted:** December 13, 2016; **Published:** December 16, 2016

intake of the population and compare it with the required standards of energy consumption of a healthy population and to take decisions to keep the population healthier. If any misbalance exists between standard consumptions of energy components and the consumed commodities, we can take decision according to the predicted health situation of the population e.g. if the trends are dangerous then the planners can plan something like launching an awareness program to control the situation and to get the required results.

Background

In a way healthcare of individuals and communities depends on planning, forecasting and managing diseases. Currently healthcare data is obtained from the health risk appraisals, general health assessments and satisfaction of patient and clinician but the retail business data which shows the purchases of energy ingredients of population is not considered. In this paper it is suggested to integrate retail business data with the other sources of healthcare sensitive data. This integrated data can determine health condition of a population. This data will help the authorities to forecast the disease treatment resources and can launch awareness programs to reduce the possibilities. This means that the responsible people or healthcare authorities can identify earlier the risk of population for getting a disease. Besides patients' risk of getting a disease, the rate of compliance with treatment protocols and severity of illness or patient's episodes of care, also depends on early prediction of disease [22]. Therefore the aim of this paper is to identify a less attended data source from where we can predict unhealthy situation and can plan to implement the programs to prevent or reduce disease progression or appearance which lead to public health [23]. The success of such healthcare program relates to application of "information technology". The data warehouse (DW) as a decision-making tool that can integrate different data sources and is the most reliable technology used by companies for planning, forecasting and management. It can be a key mechanism to detect population at risk. The correct application of data warehousing benefits for both individuals, healthcare organizations such as cost reducing and quality increasing in patient's life [24], and the government healthcare sector [25].

The retail business data systems; already installed in some form or another in grocery shops, super and hyper markets collect information of all purchases of population. This operational information collected for billing can be transformed to strategic information and can be used in Decision Support Systems (DSS) with specific tools. In start of 1990 Decision Support Systems were complemented with technologies like Data Warehouse and online analytical processing (OLAP). These technologies facilitate the analysis in DSS [26]. Data warehousing is the foundation of Business Intelligence. The design of data marts and tools for extraction, transformation, and load (ETL) are essential for converting and integrating enterprise specific data like retail business data with healthcare units. Database query, online analytical processing (OLAP), and advanced reporting tools are used to get important data characteristics. Business performance management (BPM) tools, such as scorecards and dashboards, are usually used in business intelligence to visualize and analyze various performance metrics. In addition to the business analytics functions inherent in BI software, advanced knowledge discovery using data and text mining are usually adopted for association rule mining, database segmentation and clustering, anomaly detection, and predictive modeling in various business applications. BI technologies provide historical, current, and predictive views of business operations to enhance the comprehension/understanding of fact based interrelationships and this makes BI an enabling technology for knowledge creation [27].

Proposed architecture of Retail Business Datawarehouse for extracting healthcare information

First of all we have to suggest architecture for proposed data warehouse that can bring all the components of a data warehouse together. This architecture should define the standards, measurements, general design, and support techniques. This architecture groups DW components into the three areas; data acquisition from different sources, data storage, and information delivery [28]. Building these areas require the components that are necessary to arrange properly in order to meet the proposed goals.

The components of proposed data warehouse architecture are described in the following:

Data Source:

In the proposed data warehouse the main source of data is operational data of purchases of all grocery stores of the target area like a city along with other data sources of healthcare importance like hospitals and clinics in the target area. Some other data will also be required like the data that refers to statistics published by governmental or non-governmental agencies about population census and healthcare facilities in the target area. Data about the basic nutrition of a healthy population is also required that can also be accessed by external sources.

Data staging:

Here four major functions need to be performed as follows:

Data extraction: This function has to deal with several formats and sources of data. That means if the source data from a grocery store is not in relational database, it can handle it. To extract data from different environment need to be used proper techniques to move the data into data warehouse.

Data cleansing: This function has to deal with cleaning the unnecessary data like the purchase data of items that have no relation with the healthcare. It also do correction of misspelling and solve conflicts among data. It eliminates duplicates that may be collected from different sources, and provides default values for missing data elements.

Data Standardization: This function deals with data transformation and other activities. Standardization of data is performed in two forms. First, the standardization of the data types and field length for the same data elements from multiple data sources. And second, the semantic standardization. The later means to resolve the problems related to synonyms and homonyms [29].

Data summarization: This function deals with purging source data that is not useful. It sorts and merges of data in the data staging area. When the data transformation function ends, a collection of integrated, cleaned, standardized, and summarized data is ready to load into data warehouse [30].

Data storage:

The data storage for data warehouse must be in structures suitable for analysis. The data structure on data warehouse depends on how it is used for complex analysis. In this research we are using the data for trend analysis, observation of standard reports and data screening, classification, comparing and verification using Support Vector Machine (SVM) which is an analytical techniques used to extract and classify healthcare information from a typical Business data warehouse about individuals and communities.

Analysis:

UAE Vision 2021 for healthcare:

The official brochure of UAE vision 2021 states:

“The UAE will promote long and healthy lives for all Emiratis by providing equitable access to world-class medical care while actively processing against health hazards through awareness and prevention” [31].

To help in achieving this vision our contribution is to prevent the population from health hazards by guiding them through balanced intake of energy ingredients in food consumption. UAE government can monitor their society by getting the food retail purchases from the proposed retail datawarehouse and analyze and compare it with the balance diet consumption data. In case of a misbalance; the UAE government can take steps to overcome the irregularities.

Data Analysis of UAE residents food consumption:

Data from the main food commodities provide a valuable insight into diets and their consumption by the masses. This can give a complete picture of utilization of food during a specified period. From this data, the average per capita supply of macronutrients (i.e. carbohydrates, protein, fats) can be derived for all food commodities [32].

Food consumption is expressed in kilocalories (kcal) per capita per day [32]. A more appropriate term for this variable would be “national average apparent food consumption” because it is assumed that what is purchased is also consumed, but always there is some wastage. We are ignoring the food wastage. We assume that the wastage is so low that it can be covered as precision in the result.

Table 1 and figure 1 shows that data from dietary energy measured in kcals per capita per day has been steadily increasing in UAE [33].

To keep a healthy, balanced diet, a man needs around 2,500Kcal/day to maintain his weight. For a woman, that figure is around 2,000Kcal/day [34]. We can say on average each person needs to take 2250 kcal per day to keep healthy.

According to the statistics provided in table 1, In UAE each person is taking 3170 kcal/day which is 920 kcal/day higher than the required or ideal consumption. This higher intake of daily calories results in unhealthy population (Figure 2).

In UAE; obesity is a well-known health burden and about 70 per cent of men and 67 per cent of women aged 15 years and older are considered overweight [36]. This can be controlled and planned by the government by understanding the food consumption and food purchasing habits of the population.

Table 2 provides data about the top ten commodities available in UAE and their respective Kcal per capital per day value used by the population. This table also shows that the energy available in each food item for a UAE resident to consume in a day is equivalent to 3160 Kcal/capita/day by assuming that the other food commodities; which are not in the top ten list; has 612 kcal energy for a person in a day.

Balanced food consumption and Human health:

Dieticians have suggested that calorie intake should be such that 50 to 60% of the total calorie intake is contributed by carbohydrates, 20% by proteins and 15 to 20% should come from fat [35]. The stated information is rounded and tabulated in Table 3 below. The same information can be visualized in Figure 3.

Table 4 has further split the information of Table 2 into the equivalent carbohydrates, proteins and fats percentages. This data can

be used to get the total carbohydrates, proteins and fats each person in UAE consumes per day.

UAE residents’ food consumption and health issues:

Table 5 shows the calculated values of carbohydrates, proteins and fats in Kcal/capita/day from the values from table 2 and 4 by applying the formula: $(E \cdot P / 100)$; where E is from table 2 which is total energy value consumed for a commodity and P is from table 4 which is the percentage of a category of energy in that particular commodity.

Figure 4 relates and compares the data of table 5 and the energy consumption values calculated for an ideal healthy person. It indicates that a UAE resident is consuming more carbohydrates and more fats, while less proteins.

This irregular consumption of carbohydrates, protein and fats can cause healthcare problems in population as we see that 70 percent of men and 67 per cent of women in UAE are obese [36] and obesity comes with excess consumption of carbohydrates and fats [37] (Figure 3).

Support Vector Machine (SVM) and Healthcare Information:

In this research two analytical techniques were used. Vector Machines (SVMs) technique was used to extract and classify healthcare information from a typical retail business data warehouse about individuals and communities. Different kernels are be used in Support of Vector Machines models. These include linear, polynomial, radial basis function (RBF) and sigmoid as part of mapping system and analysis of healthcare information.

The results show that using SVM method as analytical and classification tool for healthcare data is promising and comparable to other techniques like ANN (Figure 4 and 5).

Finally this technique can be used to correlate the extracted information with the existing standards of International health organizations at a national and global authorities and decision makers that can suggest tox`x` control it like the change in purchase habits of individuals or communities in context of healthcare.

Conclusions and future suggestions:

In this research it is proved that the data gathered by the retail outlets can be used smartly for healthcare planning using Data Mining and Data Warehousing techniques, as it is used for business planning.

In future a thorough and detailed research work in required to get the required results of healthcare from this technique as very less work has been done on this topic and researchers are ignoring this potential field.

In future to fully implement the proposed technique it is required to record all retail purchases and create a data warehouse to gather all retail business data and use SVM methods to classify data of food related commodities from non-food commodities, to explore the energy components from the purchased food commodities, to compare it with standard data for healthy population, to take decision to know whether the population is consuming the food abnormally and to verify it from the available healthcare data about general health of the population.

Table 1: Daily calorie intake per capita in UAE for last 10 years [33]

Years	1995-97	2000-2002	2005-07	2010-12	2015 data
Kcal/person/day	3130	3120	3140	3160	3170

Table 2: Top Ten commodities for consumption available in UAE (source: FAO Statistical Pocket Book 2015).

	Commodity	Quantity (Kcal/capita/day)
1	Wheat and products	769
2	Rice (Milled Equivalent)	496
3	Sugar (Raw Equivalent)	335
4	Milk - Excluding Butter	195
5	Pulses, and products	154
6	Soyabean Oil	146
7	Nuts and Products	144
8	Poultry Meat	132
9	Sunflower seed oil	97
10	Maize Germ Oil	80
11	Other Products	612
	Total	3160

Table 3: Healthy Calories intake

Healthy Calorie intake distribution		
1	Carbohydrates	60%
2	Proteins	20%
3	Fats	20%

Table 4: Top Ten Consumable food items in UAE with respective percentage in carbohydrates, proteins and fats.

Commodity	Carbohydrates percentage	Proteins percentage	Fats percentage
1 Wheat and products	71	12	2
2 Rice (Milled Equivalent)	80	8	1
3 Sugar (Raw Equivalent)	99	0	0
4 Milk-Excluding butter	6	4	4
5 Pulses, and products	60	1	26
6 Soabean oil	0	0	97
7 Nuts and products	21	20	54
8 Poultry Meat	0	11	25
9 Sunflower seed oil	0	0	99
10 Maize Germ Oil	0	0	99
11 Other Products	60	20	20

Table 5: carbohydrates, protein and fats consumption of each resident of UAE in Kcal per day.

Commodity	Carbohydrates Kcal/capita/day	Proteins Kcal/capita/day	Fats Kcal/capita/day
1 Wheat and products	545.99	92.28	15.38
2 Rice (Milled Equivalent)	396.8	39.68	4.96
3 Sugar (Raw Equivalent)	331.65	0	0
4 Milk-Excluding butter	11.7	7.8	7.8
5 Pulses, and products	92.4	1.54	40.04
6 Soabean oil	0	0	141.62
7 Nuts and products	30.24	28.8	77.76
8 Poultry Meat	0	14.52	33
9 Sunflower seed oil	0	0	96.03
10 Maize Germ Oil	0	0	79.2
11 Other Products	367.2	122.4	122.4
Total	1775.98	307.02	618.19

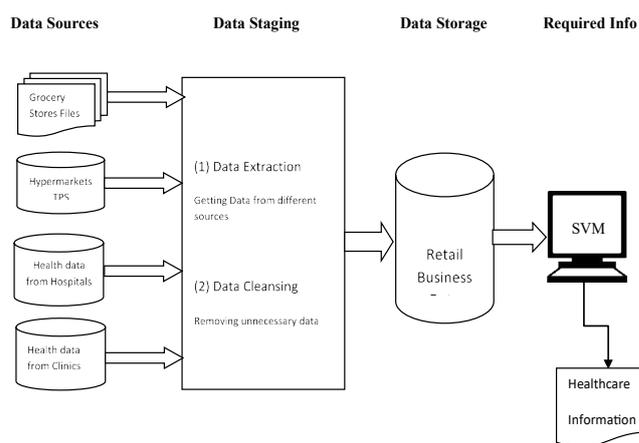


Figure 1: Architecture of Proposed Datawarehouse for Healthcare Information from Retail Business

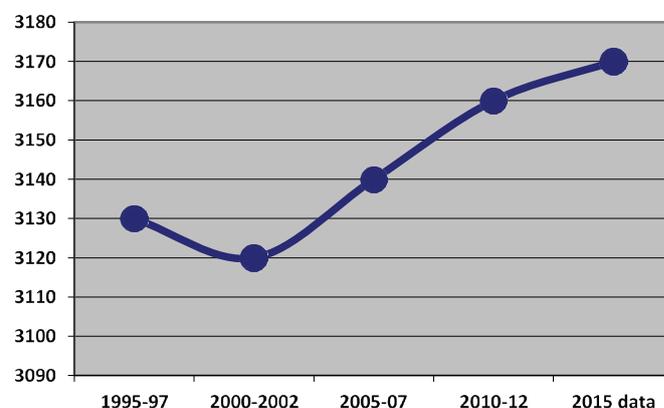


Figure 2: Showing gradual increase in daily Kcal intake of population in UAE

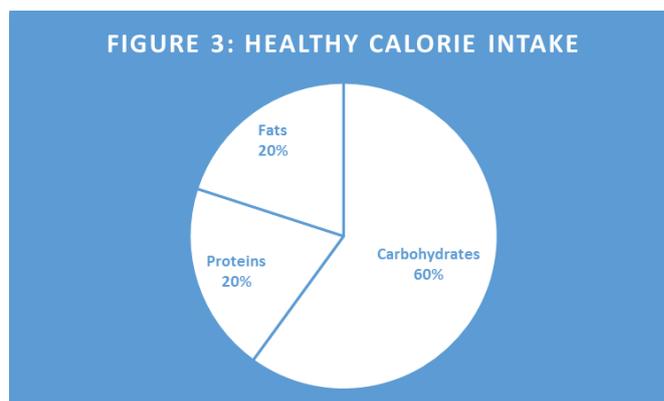


Figure 3: Healthy Calorie Intake

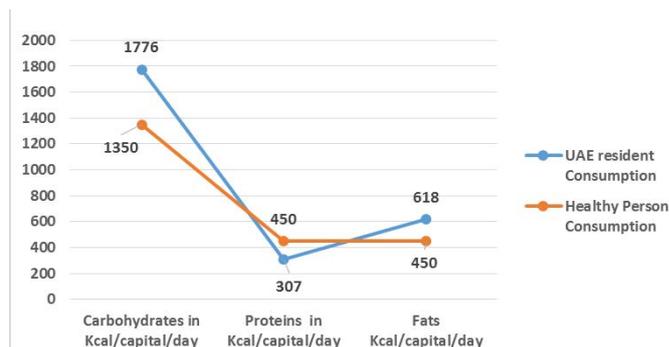


Figure 4: A Comparison Between UAE Residents consumption of Necessary Nutrition and an Healthy Person

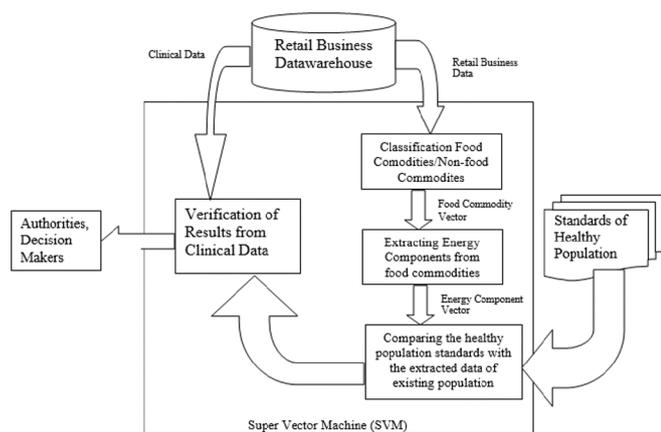


Figure 5: Proposed Framework of Analysis of data from retail business data warehouse using super Vector Machine (SVM)

References

- Bischoff, Joyce (1997) Introduction to Data Warehousing: Practical Advice From the Experts. New Jersey: Prentice-Hall, Inc.
- Carte TA, Schwarzkopf AB, TM Shaft, Zmud RW (2005) Advanced Business Intelligence at Cardinal Health. *MIS Quarterly Executive* 4: 413- 424.
- Olinsky A, Schumacher P (2010) Data Mining for Health Care Professionals: MBA Course Projects Resulting in Hospital Improvements. *International Journal of Business Intelligence Research* 1: 30-41.
- Lozito K (2011) Mitigating Risk: Analysis of Security Information and Event Management. *International Journal of Business Intelligence Research* 2: 67-75.
- Turban E, Sharda S, Aronson JE, King D (2008) Business Intelligence: A Managerial Approach. Pearson/Pren- tice Hall, Upper Saddle River.
- Iyer L, Raman R (2011) Intelligent Analytics: Integrating Business Intelligence and Web Analytics. *International Journal of Business Intelligence Research* 2: 31-45.
- D. Miller (2010) Improving Business Intelligence: The Six Sigma Way. *International Journal of Business Intelligence Research* 2: 31-45.
- Jayanthi Ranjan (2010) Business Intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology* 9: 60-70.
- Madhuri VJ (2011) Significance of Data Warehousing and Data Mining in Business Applications. *International Journal of Soft Computing and Engineering* 3: 329 -333.
- <http://www2.sas.com/proceedings/sugi24/Dat aware/p115-24.pdf>
- http://www.kff.org/insurance/snapshot/chcm0_50206oth2.cfm
- Osama ES, Ahmed NE (2009) Building a Health Care Data Warehouse for Cancer Diseases. *International Journal of Database Management Systems* 4: 39-46.
- <http://airccj.org/CSCP/vol3/csit3208.pdf>
- <http://www.wseas.us/e- library/transactions/information/2009/28- 906.pdf>
- <http://iacis.org/iis/2001/natek330.PDF>
- Denise CR (2000) Data Warehousing in Disease Management Programs. *Journal of Healthcare Information Management* 15: 98 -105.
- <http://support.sas.com/resources/papers/proce edings11/183-2011.pdf>
- Azadeh Nazari, Mahtab Karami, Reza Safdari, Majid, Yaghoubi Ashrafi (2013) Optimizing Disease Management with Data Warehousing. *Life Science Journal*.
- Nova Eka Diana, Aan Kardiana (2009) Comprehensive Centralized-Data Warehouse for Managing Malaria Cases. *International Journal of Advanced Computer Science and Applications* 10: 40 -46.
- Nasir JA, Shahzad MK, Pasha MA. Data Warehouse Design for Sales Performance Analysis. *Information Technology Journal* 5: 964-969.
- <http://crpit.com/confpapers/CRPITV83Cunni ngham.pdf>
- Perry T (2004) Technology-Driven Outcomes. *Health Management Technology* 25: 40.
- Ramick DC (2001) Data Warehousing in Disease Management Programs. *Journal of Healthcare Information Management* 15.
- Karami M (2006) The better management of disease with data warehousing. *The first international conference on telemedicine and electronic health*.
- http://usuaebusiness.org/wp- content/uploads/2014/06/HealthcareReport_ Update_June2014.pdf
- <http://articles.techrepublic.com>
- Steiger D (2010) Decision Support as Knowledge Creation: A Business Intelligence Design Theory. *International Jour nal of Business Intelligence Research* 1: 29-47.
- Ponniah P (2001) Data warehousing fundamentals.
- Watson H, Ariyachandra T, Matyska RJ (2001) Data warehousing stages of growth. *Information Systems Management* 18: 42-50.
- Mattison R (1996) Data warehousing: strategies technologies and techniques. MC Grow-Hill: Wiley.
- <https://www.vision2021.ae/en>
- <http://www.fao.org/docrep/005/AC911E/a c911e05.htm>
- <http://chartsbin.com/view/1150>
- <http://www.nhs.uk/chq/Pages/1126aspx?CategoryID=51&SubCategoryID=165>
- <http://www.iloveindia.com/nutrition/food-facts/calorie-requirement-ofbody.html#QAPmShKyikRphCvR.99>
- <http://gulfnnews.com/gn-focus/special-reports/health/the-state-of-the-uac-s- health- 2016-1.1658937>
- Swinburn BA, Caterson I, Seidell JC, James WPT (2000) Diet nutrition and the prevention of excess weight gain and obesity. *Journal of Public Health Nutrition* 7: 123-146.