



A Proposed Hybrid Data Modelling Framework to Enhance Decision Making in Pharmaceutical Sector

Noura Mahmoud Abd Elazeem, Nevine Makram Labib and Aliaa Kamal Abdella

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 18, 2020

A Proposed Hybrid Data Modelling Framework to Enhance Decision Making in Pharmaceutical Sector

Noura Mahmoud Abd Elazeem, Nevine Makram Labib, Aliaa Kamal Abdella

Computer and Information Systems Department, Sadat Academy for Management Sciences,
Cairo, Egypt.

Tel: 02 23586819

Email: nouramahmoud1391@gmail.com, nevmakram@gmail.com, aliaaabdella@yahoo.com

Abstract

Pharmaceutical companies suffered from loss because of wrong decisions. At this point, data warehouses have been introduced as a solution for enhancing decision for those companies. Most recent trends of Business Intelligence (BI) techniques have taken attention of both communities and research in pharmaceutical sector. This paper aimed to propose a framework to enhance decisions of pharmaceutical distribution department by focusing on the sales department in Egyptian Company for Medicine Trade. High level managers often give critical decisions affects the future of the company. Large amounts of data need to be reached and modified to enhance these decisions. Creating a hybrid data modelling to help managers in decision support by analyzing data to make queries, reports using online analyzing processing (OLAP) based on multidimensional data model and predicting using artificial neural network (ANN) modelling . This leads to overcoming many problems facing them such as increasing the cost of inventory and expired medicines.

Keywords: Pharmaceutical Companies, Business Intelligence, Data Warehouse, Data Mining, Data Mart, OLAP, ANN.

1. Introduction

The value of pharmaceutical investments in Egypt hit 120 billion pounds, in 2017, up to 123 factories are operating in the domain and employing up to 300,000 workers. During a press conference at the Elysee Palace in Paris, President Abdel Fattah al-Sisi announced that there is a serious crisis in most notably medicines that greatly affect the Egyptian citizen. The medicine crisis includes several axes that overlap together, including the successive increases in prices, disappearance of vital medicines, and the problems of pharmaceutical industries and expired drugs in Egyptian markets. The Egyptian Pharmaceutical Trading Company was established in 1965,

since that date, it is considered to be the first distribution and importation company that is responsible for trading pharmaceutical products all over the governorates [1].

Data store and information retrieval is a very important subject nowadays. It is affecting a large number of people and organizations because of being a useful source for decision making and increasing business. This is usually completed by Business Intelligence (BI). BI helps companies to make a better understanding of themselves. [2].

Recent advances and success of data warehouse has offered an excellent solution toward right decision in the supply chain management of pharmaceutical companies. Data Warehouse (DWH) presents a solution prepared to help these companies to easily get up-to-date, relevant and accurate information about doctors, managed care organizations, wholesalers, distributors and consumers. Pharmaceutical companies need to collect large amounts of their downstream data such as inventory information to meet their needs. This data should be collected for a clear form of the supply chain to be integrated into a DWH so the company can take decisions based on this data [3].

Sales prediction in Pharmaceutical distribution companies plays an important role in organizations for making business plans more accurate and getting a competitive feature. These companies are in continues effort to increase their profits and reduce their costs. Forecast is a way that will be used to decrease the cost of inventory and trying not to lose their customers because of the shortage of medicines [4].

The researchers used data warehousing techniques to collect historical data of diabetic and hypertension drugs. Data mining techniques were applied for sales prediction to enhance decisions in pharmaceutical sector.

2. Hybrid data modelling Framework Description

A framework is a conceptual or actual structure prepared to serve as a conductor or support for building of something that extends the structure into something useful [5]. The main goal of this data warehouse framework (DWF) is to enhance decisions in the distributed pharmaceutical company. Data warehouses can be divided into data marts, which are small data warehouses focused on a specific area of interest, for improved performance in use [6]. Data mining can be applied to operational databases with individual transactions. OLAP cube was used to give the user an efficient way to access the data, make reports and queries to support decision making [6].

The framework (DWF) can be described through five phases. (as shown in figure 1) **Phase one** is data source phase which has relational DBs and flat files. **Phase two** is ETL phase which is selecting, extracting, transforming, integrating and loading data into data marts. **Phase three** is DWH phase which is including many data marts such as (finance, marketing, sales and employs). **Phase four** is Analysis phase which is divided in to two sub phases DM and OLAP cube techniques. **Phase five** is an evaluation phase of these techniques.

2.1 Phase One: (Data Collection)

The data was collected by gathering invoices about Diabetic and Hypertension drugs for the last ten years from 2008 to 2017 for Egyptian Company for Medicines Trade. The collected invoices were hard copy, so the researcher has to build two relational databases and one flat file as the following:

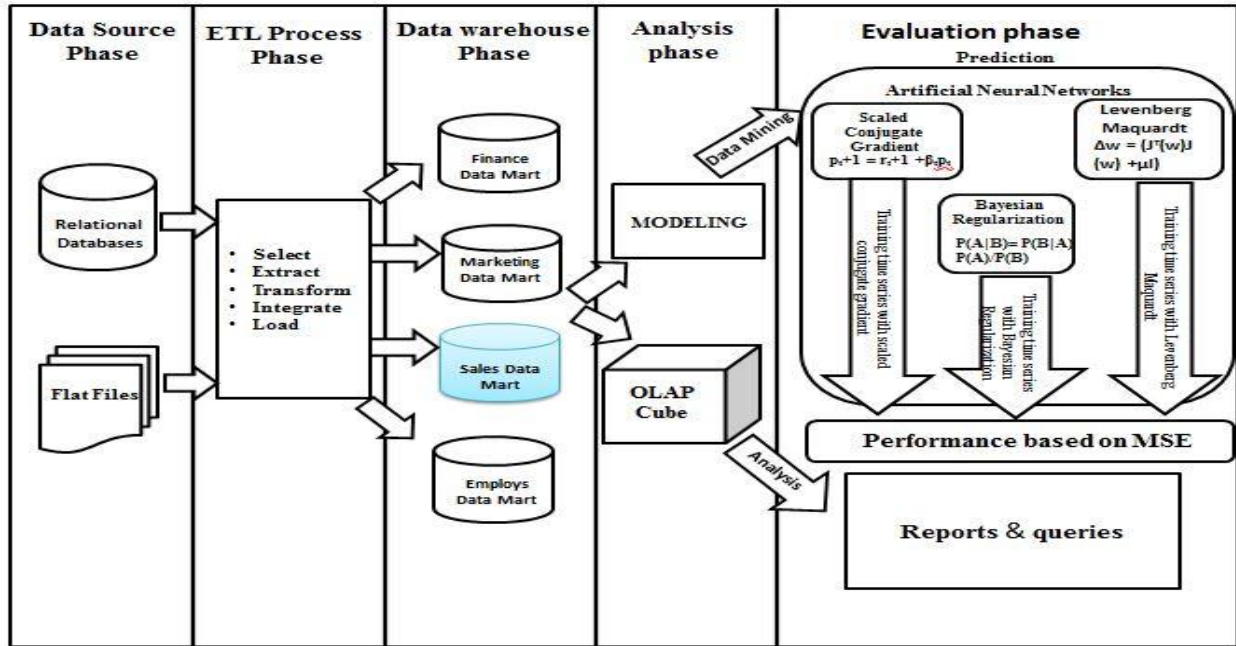


Figure 1 The proposed hybrid Data modelling Framework

2.1.1 Relational Databases Design:

Scalability is usually a very critical factor in data warehousing, so it is hard to predict how much a data warehouse will grow on the long term. Relational databases used to support scalability, high speed query processing, integration with OLAP servers and support of SQL.

A. Designing Product Database:

The product relational DB showed in figure (2) which represent five tables with their relationships as the following:

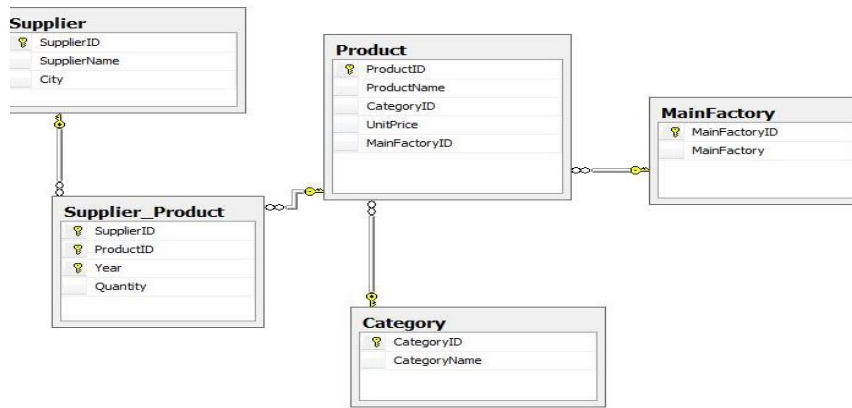


Figure 2 Product Relational Database

Manufactory table stores data about manufactory with a primary key ManFactoryID that has one to many relationships with Product table. Category table stores data about category of product with a primary key CategoryID that has one to many relationships with Product table. The supplier table stores data about suppliers with a primary key SupplierID that has indirect relationship with product table by supplier_product table that has a foreign key for both supplier and product tables.

B. Designing Customer Database:

The customer relational DB is shown in figure (3) which represented by five tables with their relationships as follows:

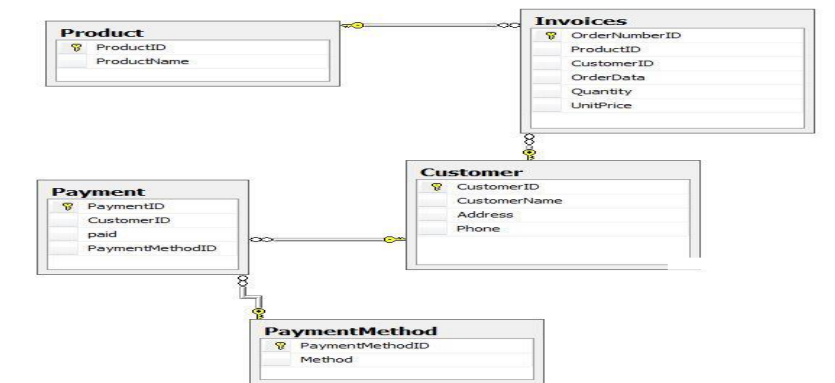


Figure 3 Customer Relational Database

The customer table stores data about customer with primary key CustomerID that has indirect relationship with product table by (invoices table) that has a foreign keys ProductID and CustomerID and contains a primay key called OrderNumberID. The payment method table stores data about methods for payment with a primary key PaymentMethodID and it has indirect relationship with customer table by (payment table) that contains a foreign keys PaymentMethodID and CustomerID and a primary key PaymentID.

2.1.2 Designing Time Flat File:

Data in flat file is simple and can port to any programming language so it is simple and portable. Data stored in plain text as shown in figure 4. Each line contains the values for different variables in the data set.

```
order number,order date,customer name,address ,phone,item,qty, price for item ,Total cost,Comment
1,1/1/08,El azeby Pharmacy ,el monakh district-portsaid,1-9600,Atelol,30, EGP 60 ," EGP 1,800 ",The
type of this medicine is Hypertension
2,1/1/08,Abd el malek pharmacy,el dwahy district-portsaid,066 3332566,Blokium,305, EGP 88 ," EGP 26,840
",The type of this medicine is Hypertension
3,1/5/08,Makah pharmacy,El arab district-portsaid,066 3366998,Tenormin,113, EGP 77 ," EGP 8,663 ",The
type of this medicine is Hypertension
4,1/5/08,19011pharmacy,El zohor district-portsaid,1-9011,Bisocord,63, EGP 50 ," EGP 3,125 ",The type of
this medicine is Hypertension
5,1/5/08,Delmar & Altallah pharmacy,Portfoad-portsaid,1-9955,Bisto1,60, EGP 200 ," EGP 12,000 ",The
type of this medicine is Hypertension
6,1/5/08,seif pharmacy,portfoad-portsaid,1-9199,Concor,1800, EGP 15 ," EGP 27,000 ",The type of this
medicine is Hypertension
7,1/9/08,Foda pharmacy,El shark district-portsaid,1-6832,Betacor,160, EGP 60 ," EGP 9,600 ",The type of
this medicine is Hypertension
8,1/13/08,El wakeel pharmacy,El ganob district-portsaid,1-9967,Alkapress,400, EGP 40 ," EGP 16,000
",The type of this medicine is Hypertension
9,1/17/08,El wakeel pharmacy,El ganob district-portsaid,1-9967,Myodura,50, EGP 50 ," EGP 2,500 ",The
type of this medicine is Hypertension
10,1/21/08,El sharkawy pharmacy,1st Compound-new cairo- cairo,0100 4333939,Ezapril co,463, EGP 50 ,"
EGP 23,125 ",The type of this medicine is Hypertension
11,1/25/08,George pharmacy,Portfoad-portsaid,0120 1000896,Zestril,160, EGP 40 ," EGP 6,400 ",The type
of this medicine is Hypertension
12,1/29/08,Dr. osama el tayeby pharmacy,el monakh district-portsaid,1-6840,Atacan1,375, EGP 40 ," EGP
15,000 ",The type of this medicine is Hypertension
13,2/2/08,Home pharmacy,el dwahy district- portsaid,1-6917,Norvasc,150, EGP 60 ," EGP 9,000 ",The type
of this medicine is Hypertension
14,2/6/08,Abo El Geet pharmacy ,El shark district-portsaid,066 3249333,Inderal,290, EGP 30 ," EGP 8,700
",The type of this medicine is Hypertension
15,2/10/08,Elwaly pharmacy,portfoad-portsaid,0101 6755553,Disartan,100, EGP 20 ," EGP 2,000 ",The type
of this medicine is Hypertension
16,2/14/08,omran pharmacy,El zohor district- portsaid,0100 0001564,Aldomet,125, EGP 70 ," EGP 8,750
```

Figure 4 Orders Flat File

Time flat file is about orders data in ten years from 2008 to 2017 and it contains many attributes such as (order number, order date, customer name, address, phone, items, quantity, price for item, total cost and comment).

2.2 Phase Two: (ETL Process)

This phase is containing three phases extracting, transforming, and loading [ETL]. ETL is the basic component of a BI system because the data quality of all other components builds on the ETL process. The data that will populate to the data mart will be selected to extract from the data sources SQL server database and text to the destination source MS SQL Server Data Transformation Services (DTS).

DTS can provide access to data and custom transformation specifications so the tables will be used for transforming. The data should be designed then clean it and converted into correct format and then put the corrected data into the data mart.

2.3 Phase Three: (DWH Phase)

Starting the DataMart first requires data loading using all the historical data. Fact table are updated every time according to data mart. Designing entity relationship for data mart diagram shown in figure (5)

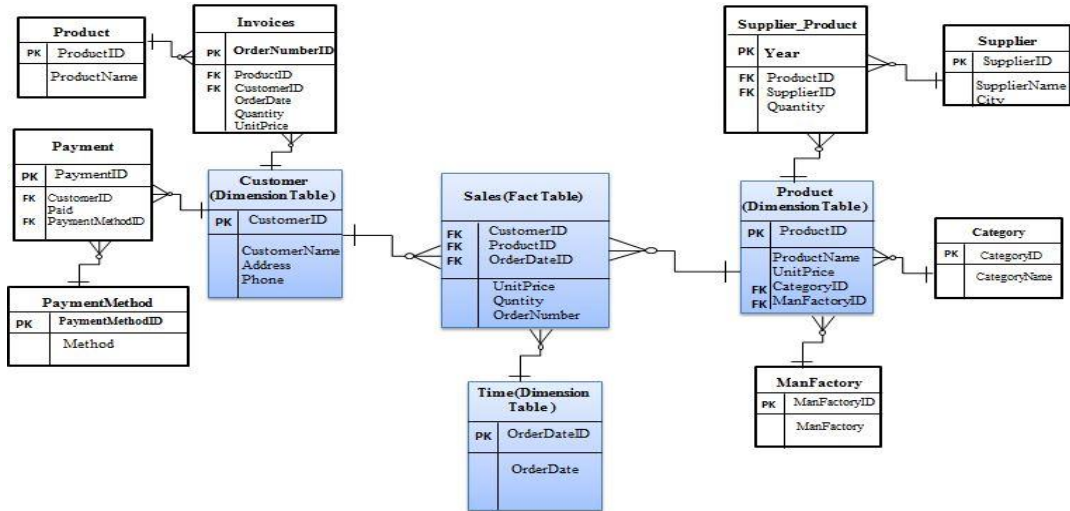


Figure 5 Operational Source Entity Relationships Diagram

2.3.1 Fact Table

A fact table must be defined for star schema shown in figure (6). The primary key in fact table are Customer ID , Product ID and OrderDateID that is made up of all of its foreign keys.

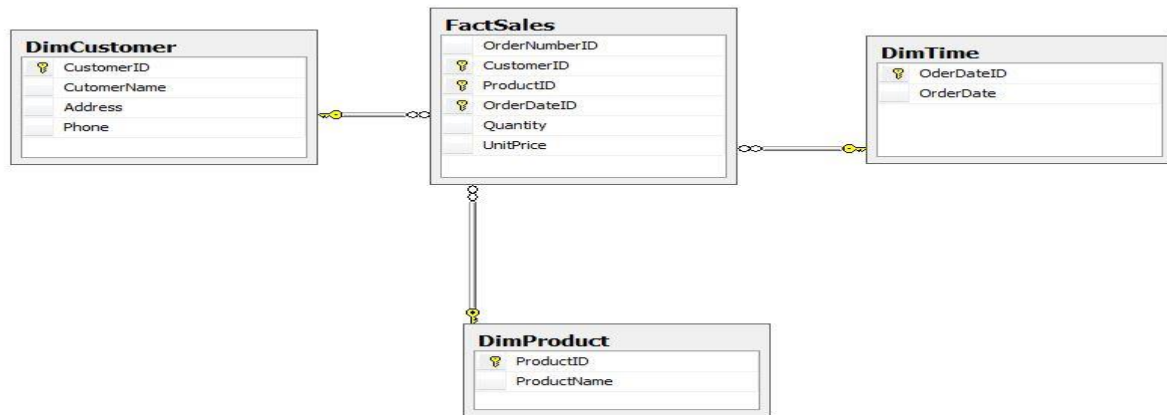


Figure 6. Star Schema

2.3.2 Dimension Tables

There are three dimensions tables as shown in figure (6) product dimension database, customer dimension database and time dimension flat file.

2.4 Phase Four: (Analysis Phase)

Analysis phase will go through two sub phases modeling using DM techniques and OLAP cube analysis. The two sub phases can be showed as the following:

2.4.1 Applying OLAP Cube:

Cubes model data in the dimension and fact tables in the DataMart to provide query and analysis to decision makers. The researcher creates OLAP cube for manipulating and deriving data for analysis purposes, storing data in a standard relational database and can be accessing by any SQL reporting tool and being more scalable in handling large data volumes. The researcher creates it by using Visual Studio2008 Tool and follows the next steps:

A. Creating Data Source:

The first step with a Cube is to connections to data sources. The Data Source Wizard helps to create data sources based on an existing data source so that a single connection can be shared by Analysis Services for multiple databases.

B. Creating a Data Source View:

Data Source Views (DSVs) enable the user to create a logical view of only the tables involved in the data warehouse or data mart design as shown in figure (7):

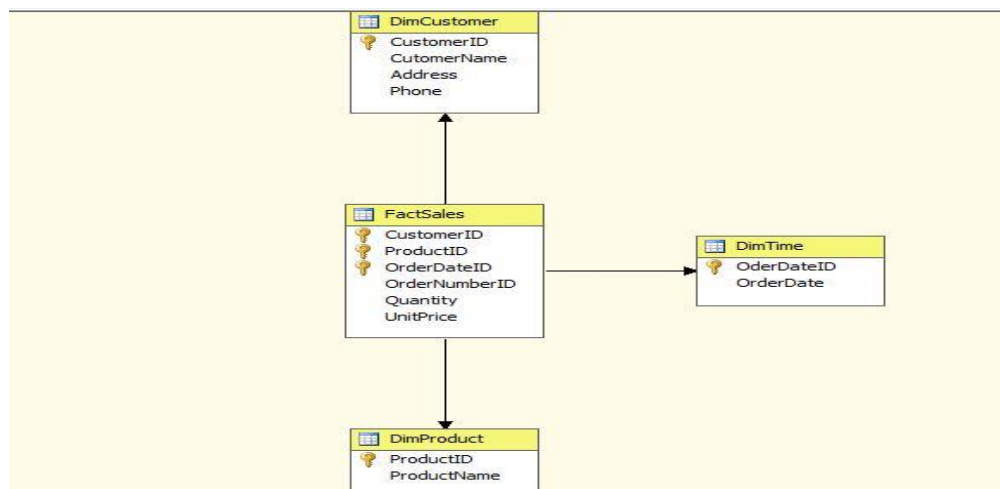


Figure 7 Data Source View for Sales Data Mart

C. Creating Product Dimension:

Dimensions are the collection of attributes based on the dimension tables in DSV. All these attributes are shown as attribute hierarchies. Analysis services help the user to create User Defined Hierarchies as well. The researcher creates product hierarchy by dragging attribute called “ProductID” so the result shows in figure (8):

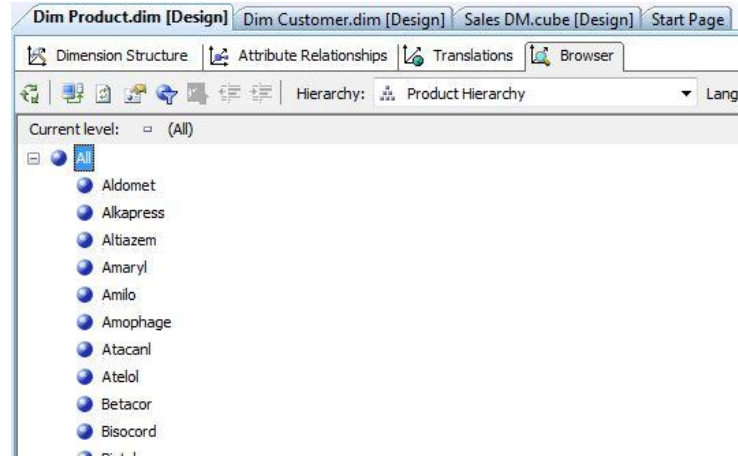


Figure 8 Product Dimension Browser

D. Creating Customer Dimension:

The researcher creates customer hierarchy by dragging attribute called “CustomerID” first and then “Address” and “Phone” last one so the result shows in figure (9) :

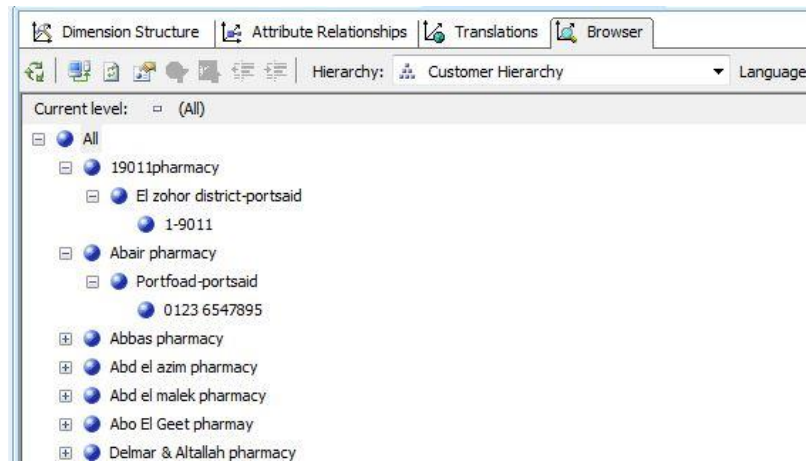


Figure 9 Customer Dimension Browser

E. Creating Time Dimension:

The researcher creates time hierarchy by dragging attribute called “OrderDateID” so the result shows in figure (10) :

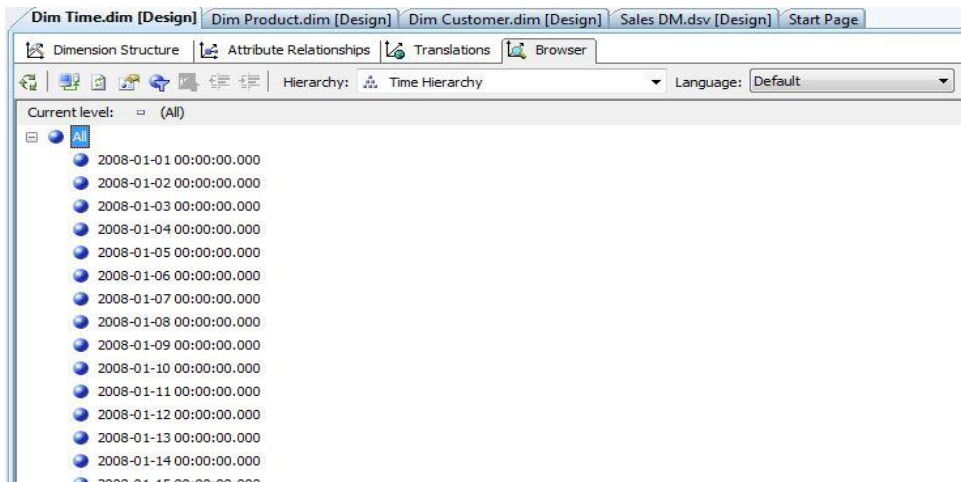


Figure 10

Time
Dimension

Browser

F. Creating a Cube:

The researcher creates a cube called “Sales Cube” and selects the measures “Quantity” and then selects the existing dimensions so that the results shows as the figure (11):

Order Date ID	Bin pharmacy	Omran pharmacy	Radana pharmacy	Seif pharmacy	Sheta pharmacy	Tamer pharmacy	Grand Total
2008-01-01 00:00:00.000	5409						10133
2008-01-02 00:00:00.000							4799
2008-01-03 00:00:00.000			3609				3609
2008-01-04 00:00:00.000					1479		12152
2008-01-05 00:00:00.000							4301
2008-01-06 00:00:00.000							8533
2008-01-07 00:00:00.000			3218				11546
2008-01-08 00:00:00.000	5457			1971		1448	16767
2008-01-09 00:00:00.000							15524
2008-01-10 00:00:00.000							2189
2008-01-11 00:00:00.000					1335		3347
2008-01-12 00:00:00.000							16264

Figure 11 Sales Cube

G. Experiment and Results:

After designing the aggregation of a new cube, the cube should be processed. Graphical charts are presented in figure(14)(15):

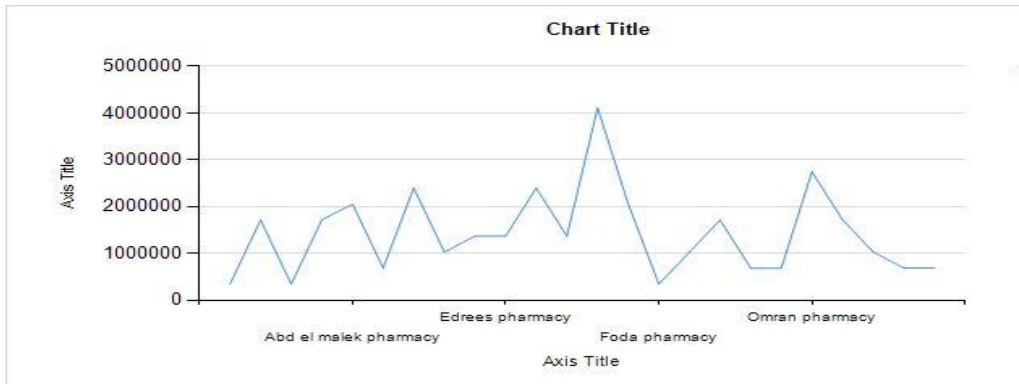


Figure 14 Top Customers at Last Ten Years

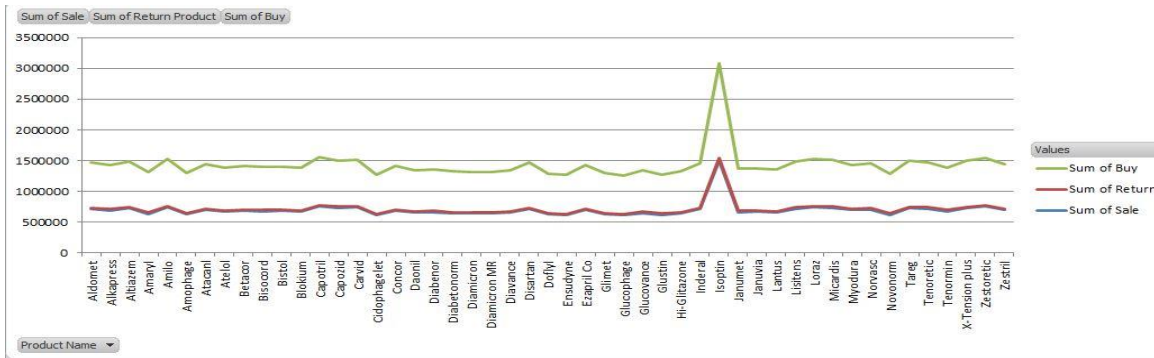


Figure 15 Sums of Purchases, Sales and Returns of Hypertension and Diabetic Drugs at Last Ten Years

2.4.2 Applying DM Technique

First, the researcher used a classifier model (WEKA) to know the correlation between variables that showed there is no correlation coefficient between variables in classifier model. Train future prediction shows that the data is non-linear and non-smoothly with very high mean square errors (MSE). time series technique was not efficient with non-linear and non-smooth data[7].

Second, DM is used to predict future trends, customer purchases and help in decision making. As ANN considered as machine learning techniques, so it is preferable to use MATLAB tool [8]

ANN has the ability to learn and model non-linear and complex relationships, many of the relationships between inputs and outputs are non-linear as well as complex. It can infer unseen relationships on unseen data as well, thus making the model generalizes and predict on unseen data. ANN does not impose any restrictions on the input variables [9]. There are Three Techniques to A High Order ANN:

A. The Levenberg-Marquardt (L|FM) is a numerical least-squares non-linear function minimization technique. When applied to the problem of minimizing the error function of a

Artificial Neural Network this method provides a compromise between Newton’s technique, which converges quickly in the neighborhood of a minimum but may otherwise diverge, and gradient descent, which converges slowly but is guaranteed to do so[10].

The LM technique computes the weight change according to

$$\Delta \mathbf{w} = (\mathbf{J}^T(\mathbf{w})\mathbf{J}(\mathbf{w}) + \mu\mathbf{I})^{-1}\mathbf{J}^T(\mathbf{w})\mathbf{e}(\mathbf{w})$$

B. Bayesian regularized ANNs (BRANNs) attempt to overcome these problems by incorporating Bayes’ theorem into the regularization scheme [11].

$$P(\mathbf{A}|\mathbf{B})= P(\mathbf{B}|\mathbf{A}) P(\mathbf{A})/P(\mathbf{B})$$

C. Scaled Conjugate Gradient adds to the complexity of the training procedure by performing a line search in each iteration to find the best step size along the conjugate direction. Instead of using a line search, the scaled conjugate gradient method uses a Levenberg-Marquardt approach to determine the optimal step size at each iteration [12].

$$\mathbf{p}_{t+1} = \mathbf{r}_{t+1} + \beta_t \mathbf{p}_t.$$

D. Experiment and Results:

The next table shows the comparison between Artificial Neural Network techniques that applied in Diabetic and Hypertension drugs data based on MSE:

Table 1 Comparison between Applied Techniques Based on MSE.

Applied Techniques	Artificial Neural Network		
	Levenberg Marquaradt	Bayesian Regularization	Scaled conjugate gradient
Diabetic drugs	149582.28294	306.10614	260490.75382
Hypertension drugs	5806.70695	426.43748	6515.51252

As shown in table 1 that applying Artificial Neural Network techniques in Diabetic drugs data to know the MSE for each technique. The results of applying Artificial Neural Network techniques present the MSE of LM technique was (149582.28294) ,the MSE of Bayesian Regularization technique was (**306.10614**) and the MSE of Scaled conjugate gradient technique was (260490.75382). By applying Artificial Neural Network techniques in Hypertension drugs data to knowing the MSE for each technique, there are many results. The results of applying Artificial Neural Network techniques present the MSE of LM technique was (5806.70695) ,the MSE of Bayesian Regularization technique was (**426.43748**) and the MSE of Scaled conjugate gradient

technique was (6515.51252). So the data declared that the BR was the best technique in Diabetic and Hypertension drugs data because it is searching for hidden correlation relationship that makes data smooth. BR training aims to minimize the sum of mean squared errors although it takes time to get results. So Artificial Neural Network are capable of fitting linear and nonlinear functions without the need for knowing the shape of the underlying function so that it is more suitable for nonlinear time series prediction.

Data mining is used in conjunction with data warehousing to enhance decisions. Data warehouse provides answers of many queries whereas datamining predicts future trends, customer purchase habits and help in decision making.

3. Conclusion and Recommendations

The data warehouse framework designed and developed in this paper to enhance decisions of distributed system in Egyptian Pharmaceutical Trading Company. The data mart designed and applied is used by sales department in this company.

Decision support was enhanced by using two techniques: first to answer the complex queries and the weekly or monthly reports, second to analyze relational data in the DWH to predict the future performance.

The conclusion can be summarized as follows:

1. Time series technique uses when data is smooth and lags were small that gives the best prediction in linear data.
2. Traditional statistical techniques can't predict in nonlinear data with no correlation and no smoothly so the researcher recommended to applying machine learning techniques such as ANNs that gives the best prediction in nonlinear data.

Accordingly, it is recommended to apply the proposed framework so as to enhance decisions of distribution systems in pharmaceutical companies to decrease the medicine industry cost and increase the productivity.

4. Future Work

The following are some research directions for future work:

1. Developing the enterprise DWH for all departments of the pharmaceutical company as the researcher applied on sales department only.
2. Developing a forecasting model by using ANNs technique in Egypt to decrease the expired medicines so it will also help in the sustainable development of the pharmaceutical sector.
3. Developing a proposed web based framework for feedback evaluation.

Summary

OLAP was selected as a development tool by creating data source, data source view and then creates product, customer and time dimensions. ANN and data mining technique was applied to the data mart for sales prediction. Finally training and comparing between the used techniques. By adopting the Levenberg-Marquardt, Bayesian regularized and Scaled Conjugate Gradient techniques of ANNs for data mining is more efficient with nonlinear data that have no correlation with non-smooth data. By implementing diabetic and hypertension drugs data, the results shown the best performance for both drugs data in Bayesian Regularization technique which gives best prediction of sales. Data mining is used in conjunction with data warehousing to enhance decisions. Data warehouse provides answers of many queries whereas datamining predicts future trends, customer purchase habits and help in decision making.

References:

- [1] Mostafa Gawish, "Egyptian Society 2017: Drug Crisis". Egyptian Institute for Studies.2018,[online] , Available: https://en.eipss-eg.org/egyptian-society-2017-drug-crisis/?fbclid=IwAR0v9lpN9_ohkBeavIgnuC8SW6F9zc49v52I6pkYUEM6y69bHraBW5S_wI
- [2] <http://www.holdipharma.com/en/home/Pages/EPTC.aspx>
- [3] Khalid Ibrahim Mohammed, "Data Warehouse Design And Implementation Based On Quality Requirements", International Journal of Advances in Engineering & Technology, 2014 [Online], Available: https://www.researchgate.net/publication/330666318_DATA_WAREHOUSE_DESIGN_AND_IMPLEMENTATION_BASED_ON_QUALITY_REQUIREMENTS
- [4] Matteo Golfarelli and Stefano Rizzi, "Data Warehouse Design: Modern Principles and Methodologies", McGraw-Hill Education; Italy, 2009, Available: http://cdn.ttgtmedia.com/searchDataManagement/downloads/Data_Warehouse_Design.pdf
- [5] Walid Qassim Qwaider, "Apply On-Line Analytical Processing (OLAP)With Data Mining For Clinical Decision Support", International Journal of Managing Information Technology (IJMIT) Vol,4, 2012, Available: <https://pdfs.semanticscholar.org/2e15/7943423a0f2f73f0af3e80f6f6d2c01dab3e.pdf>
- [6] G,Satyanarayana Reddy,Rallabandi Srinivasu,Poorna Chander Rao And Srikanth Reddy Rikkula,"Data Warehousing, Data Mining, OLAP And OLTP Technologies Are Essential Elements To Support Decision-Making Process In Industries", International Journal on Computer Science and Engineering,2010, Available: https://www.researchgate.net/publication/50235148_DATA_WAREHOUSING_DATA_MI

NING_OLAP_AND_OLTP_TECHNOLOGIES_ARE_ESSENTIAL_ELEMENTS_TO_SUPPORT_DECISION-MAKING_PROCESS_IN_INDUSTRIES

- [7] Noura Mahmoud Abd Elazeem, Nevine Makram Labib and Aliaa Kamal Abdella, “A Proposed Data Warehouse Framework to Enhance Decisions of Distribution System in Pharmaceutical Sector”, Egyptian Computer Science Journal, Volume 43, Issue 2, (May 2019) Available: http://ecsjournal.org/Archive/Volume43_Issue2.aspx
- [8] JEREMIAH GARDNER, “Brand Development Process Vs, Framework”, [Online], 2013, Available: <https://www.brandingmag.com/2013/09/05/brand-development-process-vs-framework/>
- [9] Ratnadip Adhikari, “An Introductory Study on Time Series Modeling and Forecasting”, Germany, 2013, Available: <https://arxiv.org/abs/1302.6613>
- [10] Montaña Moreno JJ, Palmer Pol and Muñoz Gracia "Artificial neural networks applied to forecasting time series", Spain, 2011, Available: <https://www.ncbi.nlm.nih.gov/pubmed/21504688>
- [11] Henri P, Gavin, “The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems”, 2011, Available: <https://www.semanticscholar.org/paper/The-Levenberg-Marquardt-method-for-nonlinear-least-Gavin/1d83aa91810162891c9f04e9a62da69197c78602>
- [12] Hayrettin Okut, “Bayesian Regularized Neural Networks for Small n Big p Data”, Artificial Neural Networks - Models and Applications, 2016, Available: <https://www.intechopen.com/books/artificial-neural-networks-models-and-applications/bayesian-regularized-neural-networks-for-small-n-big-p-data>