A Proposed Framework to Enhance Pharmaceutical Distribution Companies using Data Mining Techniques

Manal A. Abdel-Fattah[#1], Shimaa Ouf[*2], Esraa Ahmed[#3]

[#] *Faculty of computers and information, Helwan University, Egypt.*

[#] *Faculty of commerce &business administration, Business information systems department, Helwan University, Egypt*

[#] *Faculty of commerce and business administration, Postgraduate Business Information Systems, Helwan University. Egypt.*

[1]Manal_8@hotmail.com

[2]Shimaaouf@yahoo.com

[3]Israa.Ahmed.1811003@commerce.helwan.edu.eg

***Abstract*— This paper propose a framework to enhance the retail supply chain process of pharmaceutical distribution companies starting from forecasting the products of the next month ending with deliver the right amount of medicines to the right place and at the right time in addition to tracking the stock during the month by applying key performance indicator (KPIs).One of the challenges facing pharmaceutical distribution companies is how to control inventory levels in order to prevent costs of excessive inventory and prevent losing customers due to drug shortages. A literature review shows that there is a gap between the periodicity of deliveries of medicines to pharmacies, which may have many deliveries every day, and the purchase of medicine stock by distributors, which may take about two days. On the other hand, because of the short-term shelf-life of many medicines and the need to control stock levels, it is necessary for pharmaceutical distribution companies to have a reliable forecast of the needs of medicines as both surplus and shortage of medicines can cause problems to pharmaceutical distribution companies. Therefore, the paper focus on distribution phase the objective of paper is to improve the retail supply chain process by applying data mining techniques in pharmaceutical distribution companies. Data mining is the process of extracting useful information from a massive amount of data, as well as data mining may help identify uncover known facts, forecast unknown outcomes, increase revenue and reduce costs.**

***Keywords*--** *pharmaceutical distribution companies, retail supply chain process, Key performance indicators, Data mining, pharmacies.*

1-Introduction

Pharmaceutical distribution companies are the heart of the pharmaceutical industry because without it the medicines will not deliver to the patients and patient's life will be in danger.

Pharmaceutical distribution companies connecting the pharmaceutical companies with the pharmacies in an effective manner. And by using data mining the process will be more effective as data mining will solve the pharmaceutical distribution companies' problems. In Introduction section, we discuss the supply chain management, retail supply chain management, pharmaceutical supply chain, pharmaceutical distribution companies, KPIs.

1.1 Supply chain Management

Supply chain management (SCM) is a wide range of activities are needed to prepare, monitor and execute the flow of the product from obtaining the raw materials and manufacturing through distribution to final customers [11]. The concept of Supply Chain Management (SCM) is based on two ideas [25]:

1-The first is that every product deliver to the consumer show the cumulative effort that done to the product from the first phase until reaching to the final consumer.

2-The second is most of organizations paid attention to what happening in one phase only ignoring the phases of delivering the products to the final consumer resulting from that disjointed and ineffective supply chains.

1.2 Retail Supply Chain Management

Retail Supply Chain Management is a term refers to an end to end process in products from preparing the purchase order to the point of delivering the products to the customer

Retail supply chain also known as a method by which products reach the customers as it's making the products available to the customers whenever they need [27].

1.3 Pharmaceutical Supply Chain (PSC)

A pharmaceutical supply chain (PSC) can be defined as ‘‘the aggregation of all processes related to the transformation of drugs from raw materials until delivering to the consumer” [4]. The pharmaceutical supply chain consists of three main sides Producers, purchases, and pharmaceutical providers. Producers consist of: pharmaceutical companies, device manufacturers, Purchases consist of: pharmaceutical distributions companies, medical-surgical distributors while Providers consist of: Pharmacies and Hospitals [4]. Pharmaceutical Supply Chain considers a sensitive supply chain because anything less than a customer service level (CSL) of 100% is impermissible as its effect directly on customer’s health. It’s a very complicated supply chain and carries a huge responsibility to ensure that the right medicine reaches the right people at the right time and in the right situation to save human lives [4]. The main objective of the pharmaceutical supply chain (PSC) is to manufacture and distribute medicines among various areas to deliver medicines at the appropriate time with the quantity demanded at reasonable prices [18].

1.3.1 Benefits of Pharmaceutical Supply chain (PSC)

Supply chain management in the pharmaceutical field is important and critical as the pharmaceutical supply chain is sensitive and carries a huge responsibility whereas fights diseases and saves people’s lives. However, the pieces and parts of the pharmaceutical supply chain are typically out of sight but bring value to patients and the healthcare system overall.

Benefits of PSC include [14]:

-Increase profits.
-Fewer suppliers and shorter planning cycles.
-Response to customer’s demand inadequate time.
-Maximize productivity.
-Reduce costs.
-Control Inventory.

1.3.2 Pharmaceutical Supply Chain Issues

We mentioned before that the pharmaceutical supply chain is a critical and influential supply chain because it is affect directly on health and human lives, fights diseases, and illness. Consequently, the pharmaceutical supply chain faces some issues related to the flow of medicines, transportation, inventory, regulatory systems, suppliers, and raw materials of the medicines. From these issues [14]:

a) *Shortage or Surplus of Medicines.*

Both the shortage and surplus of medicines are problems faces the pharmaceutical supply chain. The shortage of medicines leads to customer dissatisfaction while the surplus of medicines caused excessive inventory and expired medicines [6].

b) *Hardness of estimating demand rates in the future.*

Due to unawareness of using accurate methods of forecasting and used traditional methods which lead to failure in expecting the upcoming demands from customers and consequently lose money and customers [6].

c) Order Management.

Issues in managing the order due to lack of transportation vehicle, the product/quantity demanded not available, destruction in medical devices, and lack of raw materials [35]

d) Transportation issues.

Due to the unprofessional behavior of the drivers and irresponsibility, malfunction in distribution vehicles, temperature control in the trunk, and damaged products during transporting [35].

e) Distribution Company issues.

Due to inaccurate temperature control which leads to damaged and expired drugs [35]. Besides the incapability of forecasting the right quantity of products, customer dissatisfaction, excessive products that have no place to store [10].

e) Storing and warehousing issues.

Caused by unawareness of drug type as every drug has its special type. There are drugs should keep in a cold room, Drugs should keep at a moderate temperature, other drugs which is in the category of cosmetics should keep in a special inventory and expensive drugs that should keep in a locker place [35].

1.4 Pharmaceutical Distribution Company

A pharmaceutical distribution company can be described as a middleman between pharmaceutical companies and pharmacies. **As shown in figure 1**. Pharmaceutical distribution companies purchase products (drugs & medical supplies) from pharmaceutical companies and sell them gradually to pharmacies. Without the existence of pharmaceutical distribution companies, pharmaceutical companies will spend large efforts on financial, transportation, and staff resources to do the function of pharmaceutical distribution companies.

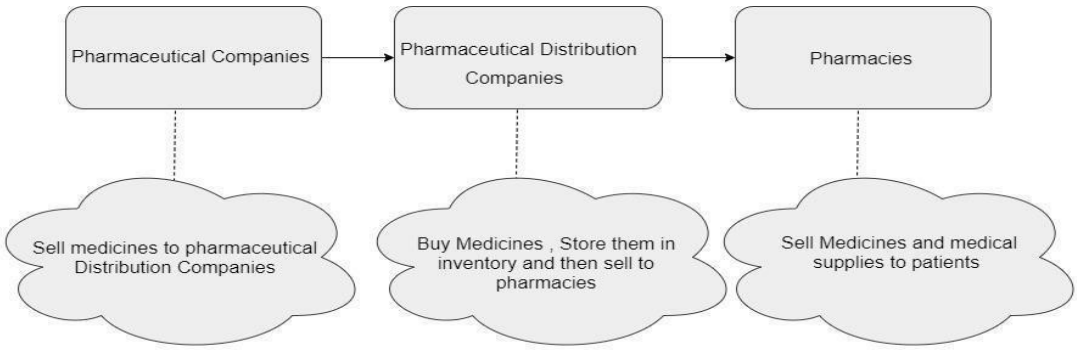

Figure 1: Pharmaceutical Distribution Companies

Pharmaceutical distribution companies face many business challenges and difficulties such as huge amounts of inventory, increased competition, and tough restrictions. They have to meet their customer's needs by delivering the right quantity of medicines to the right place and time [6]. Both the shortage and surplus of medicines can lead to loss of income for these companies. A shortage of medicines can lead to decreased profits and customer dissatisfaction while a surplus of medicines can lead to excessive expenses which cause damaged and expired items. So, one of the problems in pharmaceutical distribution companies is how much quantity of each drug should be kept in the inventory [10]. Existing inventory models that are used now not applicable to pharmaceutical products for several reasons. Special care should be taken in pharmaceutical inventory decisions to ensure 100% product availability at the appropriate time and to the right persons whereas the quality of health care industries strongly depends on the availability of pharmaceuticals products on time [4].

1.5 Key Performance Indicator (KPI)

A Key Performance Indicator is a measurable value that demonstrates how effectively a company is achieving key business objectives. Organizations use KPIs at multiple levels to evaluate their success at reaching targets. KPIs are a milestone that helps employees of the company and managers to understand the significance of their actions, work, and the results to be achieved. We can say that KPIs have great benefits that can help companies to achieve their goals, from these benefits [20]:

-Improving the time of response to market changes.
-Identifying costs that can be ignored
-Evaluation of the labor through dashboards.
-Improvement of labor force
-Forecasting future revenues

Accordingly to that applying KPIs in pharmaceutical distribution companies is very beneficial to monitor the inventory, follow up sales target also will be a quick response to customers if they need extra products or have a problem.

2. Data Mining

Data Mining is a process of extracting useful and usable information from a huge amount of data. Also, Data Mining is known as a knowledge discovery from data (KDD) because data mining plays an important role in the knowledge discovery process [2]. Data mining is part of a large process of finding knowledge that covers pre-treatment tasks like data extraction, data cleaning, fusion, data reduction and feature building, post treatment steps like pattern and model interpretation, confirmation of hypotheses, and production. This process of knowledge discovery and data mining is very iterative and interactive [1].

2.1 Data Mining Models

Generally, there are two types of data mining models predictive model and descriptive model. Predictive Models apply supervised learning functions to determine "What might happen in the future?" while descriptive models apply unsupervised learning functions to describe "What happened in the past?"[19]. the key differences between the two models are summarized in the following table [22]:

Points of comparison	Predictive Model	Descriptive Model
Usage	Use some variables to predict unknown and future values.	Finding patterns that describe the data that can explain to humans.
Accuracy	Results don't ensure accuracy	More accurate than a predictive model
Approach	Proactive	Reactive
Practical methods	Forecasting and Simulation	Ad-hoc and standard reporting

Table1: key differences between predictive and descriptive model

For each model of these two models (predictive and descriptive) models it's techniques. The predictive model has techniques and the descriptive models has also techniques. And for every technique, there are methods and algorithms **as shown in figure 2**.

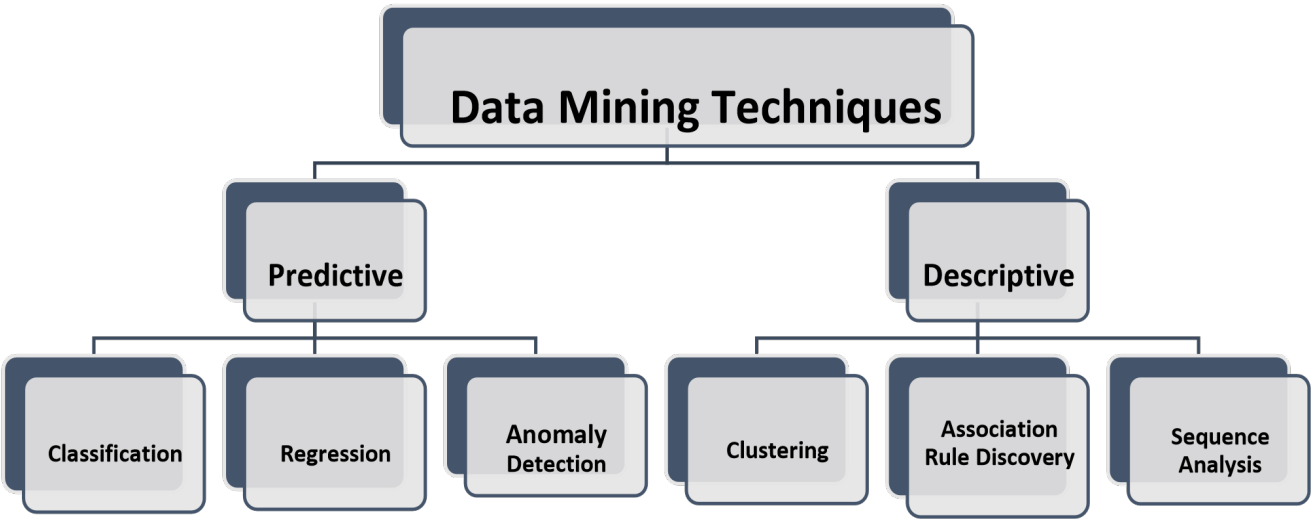

Figure 2: Data Mining Techniques

2.1.1 Predictive techniques

a) Classification Technique

It's a supervised learning method and an essential component of machine learning algorithms for extract patterns from data that could be used for prediction, retrieve important information about data and help you to classify data in different classes as data records are grouped in a set of predefined classes via classification method [23].Classification techniques are used in predicting unknown data, decision making, and pattern analysis. It is easily implementation and can handle numerical and categorical data [24].

b) Regression

Regression is a form of predictive technique that used to analyze the relationship between two factors dependent and independent. The dependent factor is the factor that predicted (target) while the independent factor is the (predictor) that used to predict the value of the dependent factor [30].

c) Anomaly Detection

Anomaly Detection is a technique of finding outliers in the dataset. And outliers refers to the observation of data items in the data set which don't match an expected behavior. There are several techniques in anomaly detection, and these techniques divided into two sections some of them based on machine learning and the other using statistical methods [31].

2.1.2 Descriptive Techniques

a) Clustering

Clustering is very similar to classification as it grouped data that are like each other and this helps to understand the difference and similarities between data. It's a process of grouping objects whose members are similar in some way into categories [23].Clustering is used in predicting unknown patterns, decision making, and document retrieval [24].

b) Association Rule Discovery

Association is one of the most popular techniques in data mining, It's applied to a large amount of data and help in finding the association between two or more items, discovers hidden patterns in the dataset. A pattern is discovered in association based on a relationship on specific items to other items in the same transaction. For example, Association is used to find the products that customers have to buy together in the supermarket. And based on this data supermarkets can sell more quantity from these products, or make an offer with these products together to increase profits [32].

c) Sequential Patterns

Help to discover sequential patterns in transaction data. Mainly it helps in finding the interesting subsequences in a set of sequences in a certain period.by providing a set of sequences and find a complete set of frequent [2].

2.2 Data mining in pharmaceutical distribution companies

The effectiveness of data mining techniques will be very useful and important to the pharmaceutical distribution company also will solve the problems and constraints that faced the company every month. In this case, we will use a predictive model to predict the future value that the company needs as we know that predictive models determine "what might happen in the future?" Thus by implementing data mining techniques, the company will reveal unknown facts, predict unknown outcomes, forecast the accurate amount of medicines and medical supplies then deliver it to customers (pharmacies) and consequently will achieve customer satisfaction and increase revenue[8]. Also by using data mining techniques company can classify their customers (pharmacies) based on their sales. At the end the company will be able to achieve its goals and objectives, solve constraints, increased revenues, and meet their customer's satisfaction.

3- Research Methodology

3.1 Methodology Approach.

The purpose of this study is to improve the retail supply chain process in pharmaceutical distribution companies by applying predictive data mining techniques to forecast the accurate amount of products for the next month, reduce costs of inventory, achieve customer satisfaction, increased revenues in addition to monitoring the inventory for the current month, follow up target and company's objectives through KPI. Research methodological is Quantitative because Quantitative methods are based on data that can be measured with numbers. The data is analyzed through numerical comparisons and statistical analysis.

3.2 Identifying the relevant scientific databases.

The following digital libraries were searched for the required articles; Research Gate, Science Direct, Springer, and IEEE. The main reason for selecting these digital libraries was they collect studies related to the fields of computer science and medical science, the index articles from different publication channels like journals, books, and workshops. In this paper, the searches were limited to journal articles. The searching process **as mentioned in table 2** was as follows:

Firstly we search by keyword "Data Mining and Supply Chain" in IEEE, Science Direct, Springers, and research gate. In IEEE from 2007 to 2020 found 19 articles, Science Direct from 2016 to 2020 found 8154, Springer from 2007 to 2020 found 15 while Research gate found 2 articles. Secondly, we search by using the keyword "Supply chain in pharmaceuticals" in IEEE from 2007 to 2020 found 17 articles, Science Direct from 2016 to 2020 found 12498, Springer from 2007 to 2020 found 9 while Research gate found 3 articles. Finally, we limited our search by using "Data Mining in pharmaceuticals" in IEEE from 2007 to 2020 found 14 articles, Science Direct from 2016 to 2020 found 617, Springer from 2007 to 2020 found 5 while Research gate found 2 articles. The articles that aren't related to our research already ignored and focus only on articles that actually related to our topic which are 4 articles from IEEE, 6 articles from science direct, 1 article from springers, and 5 articles from research gate.

Keywords	**IEEE**	**Science Direct**	**Springer**	**Research Gate**
	2007 - 2020	**2016 - 2020**	**2007-2020**	**2007-2020**
Data Mining and Supply Chain	19	8154	15	2
Supply Chain in Pharmaceuticals	17	12498	9	3
Data Mining in Pharmaceuticals	14	617	5	2
Articles actually related to research paper				
	4	6	1	5

Table 2: Inclusion and Exclusion of studies

4- Literature Review

An earlier study by Zadeh et al. in the field of pharmaceutical distribution companies are focused on the main problem which is how to control their Inventory level to prevent excessive costs also prevent losing customers by how much quantity of each drug should be kept in the inventory and to solve this problem, they propose a method to forecast sales of pharmaceutical distribution company, and the method is a combination of network analysis tools and time series forecasting methods as time series forecasting methods are constrained in their prediction by linear behavior so it isn't always acceptable while neural networks have been confirmed to make a prediction by dealing with non-linear [6].Subsequently, time series sales forecasting models were built with three different approaches including Autoregressive integrated moving average methodology (ARIMA), neural network, and an advanced hybrid neural network approach. They developed a hybrid method by applying each drug and its member's past records facilitate capturing both linear and nonlinear patterns of sales accurately. Their research demonstrated that for time series with a long memory, both ANNs and Box-Jenkins models had the same performance but at the same time they noticed that neural networks can't properly model the seasonal patterns in their data. However, the basic idea of the applied hybrid approach in this research is to let linear ANN model the linear parts and let nonlinear ANN model the nonlinear parts and then combine the results from both linear and nonlinear models [6]. Although they developed a method to forecast and control their inventory by how much quantity of each drug should be kept, they ignore to classify their customers (pharmacies) based on their sales history to make the prediction process more accurate.

While Guijuan et al. used data mining to illustrate the importance of data mining in customer relationship management (CRM) of the enterprise of the medical industry. And the problem which needs to solve is to find out the relationship between drug sales and drug attributes to realize what kind of drugs is better in the last period in order to decide the next purchase [7].The method that used is selection tree classifier to establish a relational model between sales volume and the role of the drug [7].

Uthayakumar et al. take a different direction and decided to implement OR model to solve the problem of the pharmaceutical supply chain, manage the inventory also deliver the right amount of drugs to the right people at the right time and in the right condition [9]. Although they decided to use OR Model to manage their inventory and solve the problems and of pharmaceutical supply chain, but they face some constraints, from these constraints they can't identify the total number of orders that placed per year, a number of delivers that can be accepted and missed prediction methods and segmentation by market or customer which is a very important step in developing an appropriate supply chain strategy for the target customer [9].

Ribeiro et al. detect two main issues in a pharmaceutical distribution company in Portugal. the first issue consisted of detecting customers (pharmacies) and products (medicines) which may be considered outliers and perform stock proration when these outliers are performed to avoid abnormal sales and out of stock in pharmacies, the second issue targeted the sales prediction of the pharmaceutical distribution company to control the inventory [10].They used Box-plot method for the detection of outliers (customers and products), they also used the SPSS statistical software. For sales prediction, the time series data mining method smoothed Pegals was used, while the implementation was done in SQL and the analyzed data was stored in an oracle database. [10] , but the methods they were used not to be acceptable and accurate 100% as the box plot method that used has many disadvantages exact values not retained, often difficult to locate, hide the multimodality, and other features of distribution[10].

While Sohrabi et al. adopted another theory and focus on doctors' prescription and pharmacies sales correlation using data mining. As they highlighted one of the major problems that face pharmaceutical companies which is the optimization of order size and inventory costs in order to prevent losing customers and reduce the expenses of unnecessary inventories. Thus they used data mining to examine the effect of the visit of doctors overlooking the pharmacies and the target was set for medical representatives on pharmaceutical sales. In that case, two types of classification rules were used decision tree and neural network [12].

Meanwhile, Janatyan et al propose a study to build a pharmaceutical distribution network based on the principles of sustainability that are economic, environmental, and social by using the NSGA-II algorithm. The goal of this study was to reduce the costs of transporting the products from main distribution centers to local distribution centers and from local distribution centers to customer zones. The second goal was to minimize the Co2 emissions and air pollution from the distribution network and the third goal was to increase the job rate in the pharmaceutical distribution network [13].

Kapoor et al. debated how to improve the supply chain in pharmaceuticals. And how to reduce the risks and gaps as pharmaceutical supply chains have become more complex because it includes the life-saving of people and needs the cooperation of different stakeholders such as pharmaceutical manufacturers, wholesalers, distributors, customers, information service providers, and regulatory agencies [14]. And they thought the benefits that will come from supply chain management in pharmaceuticals are quicker customer response and fulfillment rates, shorter lead time, greater productivity and lower costs and improved accuracy of forecasting [14].

Sousa et al. discussed the supply chain in pharmaceuticals, but globally as the aim of the research is the supply chain optimization of a large pharmaceutical company and consider the supply chain components as primary sites (manufactures) and secondary sites (final product market area)in order to locate the manufacture of primary and secondary products, the quantity of production, the level of inventory that will set for each manufacturing site, and how to establish the products flows between primary and secondary sites, whereas the distribution networks are out of the scope [15].

Malak et al. decided to implement a Kanban system to manage the multi-echelon inventory in a pharmaceutical supply chain. As the problem of the pharmaceutical supply chain is how to manage and control inventory levels by determining the size of orders, quantities, and meet customer expectations. [16] And for these reasons they decided to develop a Kanban system as a visual way to control production also provide the right and accurate information to all processes. But they found that Kanban system has prerequisites which are constant demand, constant cycle time, minimum quantity that often can't match with the pharmaceutical supply chain [16].

Pothitong et al. proposed a web-based information exchange system for the pharmaceutical industry to improve efficiency in the pharmaceutical supply chain, as supply chain operations in the pharmaceutical industry are one of the most complex operations to manage [17]. Electronic Data Interchange Systems (EDI) have been used mainly to exchange business information with their distributors. The exchanged electronic documents include monthly forecasts and purchase orders [17].

Abdulghani et al. explained the different challenges of pharmaceutical supply chain such as shortage or excess of drugs, customer dissatisfaction, and lack of coordination, so they decided to use the mixed integer programming (MLP) model, which describes inventory analysis of different ranges in the pharmaceutical supply chain, consisting of different facilities such as pharmaceutical suppliers, warehouses and pharmacies. The model considers multiple drugs supplied by a number of suppliers and manufacturers; therefore, the model involves a problem that chooses from which supplier a certain drug should be obtained. As the goal of the model is to reduce costs over the multiple stages of the pharmaceutical supply chain and conclude the order quantity of each product and the selection of the suppliers [18]. They decided to implement the model on CPLEX which is an automated program that can read problems interactively or from files, solves the problem, and delivers the solution interactively or in text files [18].

The following table summarizes all the studies that discussed above, identifying the research paper, methods used, and drawbacks found in their researches.

Research Paper	Year	Description	Methods	Drawbacks
A predictive analytics of physicians prescription and pharmacies sales correlation using data mining .	2019	Highlighted one of the major problems that face pharmaceutical companies which is optimization the order size and inventory costs in order to prevent losing customers and focus on doctor's prescription and pharmacies sales correlation using data mining to examine the relationship between doctor's Prescription and pharmacies sales.	Decision tree And neural network	Focus on the relationship between doctor's prescription and Pharmacies sales it's not an optimal solution to solve inventory costs and Customer satisfaction. Because sometimes patients buy medicines without a doctor's prescription.
Pharmaceutical Supply Chain Cost Optimization Model Considering Multiple Echelons and Multiple Drugs.	2019	Explain the different challenges of the pharmaceutical supply chain	Mixed-integer programming (MLP) Model and Implemented on CPLEX	While developing a model they assumed limited space for each product and ignore to consider the availability of floor space
Designing Sustainable Distribution Network In Pharmaceutical Supply Chain.	2018	Propose a study to build a pharmaceutical distribution network based on the principles of suitability	NSGA-II algorithm	Ignoring phase of distribution and how to manage inventory and focus on how to reduce the transportation costs from main centers, local centers to customer areas
An Overview on Pharmaceutical Supply Chain: A Next Step towards Good Manufacturing Practice	2018	Debated on how to improve the supply chain in pharmaceuticals and reduce the risks and gaps	No Methods mentioned	Authors talked in general, to improve the pharmaceutical supply chain and don't mentioned any methods used to improve.
Application of Data Mining Technology in the CRM of Pharmaceutical Industry.	2018	The problem which needs to solve is to find out the relationship between drug sales, the role of the drug, and drug attributes to realize what kind of drugs is better in the last period in order to decide the next purchase.	Selection tree classifier.	Ignore how to manage inventory after deciding the next purchase from customers also ignores to classify customers based on their purchases.
Detection of Outliers and Sales Prediction for a pharmaceutical Distribution Company	2017	Detecting customers (pharmacies) and products (medicines) which may be considered outliers, issue targeted the sales prediction of Pharmaceutical distribution company to control the inventory	Box-plot method	The Box-plot method it's exact values not retained, often difficult to locate, hide the multimodality and other features of distribution.
A Kanban based system for multi echelon inventory management The case of pharmaceutical supply chain	2016	Decided to manage inventory in pharmaceutical supply chain	Kanban System	Kanban system has prerequisites which are constant demand, constant cycle time, and minimum quantity.
Intelligent Sales Prediction for Pharmaceutical Distribution	2014	Focused on how to control their Inventory level to prevent excessive costs also prevent losing customers by how much	Combination methods of network analysis tools and time series forecasting methods.	Ignore to classify pharmacies based on their sales history.

Companies		quantity of each drug should be kept in the inventory.		
Pharmaceutical Supply Chain and Inventory Management Strategies: Optimization for a Pharmaceutical Distribution Company and a hospital.	2013	Decided to solve the problem of the pharmaceutical supply chain, deliver the right amount of drugs to the right people at the right time and in the right condition.	OR Model	they can't identify the total number of orders that placed, number of delivers also missed prediction methods and segmentation by market or customer
Global supply chain planning for pharmaceuticals.	2011	Discussed where to locate the manufacture of primary and secondary products, the quantity of production and the level of inventory that will set for each manufacturing site, How to establish the products flows between primary and secondary sites in the pharmaceutical supply chain	Mixed-integer linear programming	Mixed-integer linear programming doesn't take into consideration nonlinear effects and also require to determine all the time periods at once.
Improve Supply Chain Efficiency through a Web Based System: A Case Study on a Pharmaceutical Company in Thailand.	2011	proposed a web-based information exchange system for the pharmaceutical industry to improve the efficiency in the pharmaceutical supply chain	Electronic Data Interchange systems (EDI)	Electronic Data Interchange Systems are expensive, complex and small companies aren't able to implement it unlike data mining

Table 3: Systematic Literature review

While the second table identifies all the researches mentioned in the literature review and which researches focus on all stages of the supply chain and which focus on one or two phases only.

		Supply Chain Stages				
Research Paper	***Year***	***Raw Materials***	***Supplier***	***Manufacturing***	***Distribution***	***Customer***
A Predictive Analytics of Physicians Prescription and Pharmacies Sales correlation using data mining.	***2019***	✗	✗	✓	✓	✓
Pharmaceutical Supply Chain Cost Optimization Model Considering Multiple Echelons and Multiple Drugs.	***2019***	✓	✓	✓	✓	✓

Designing Sustainable Distribution Network in Pharmaceutical Supply Chain	***2018***					
An Overview on Pharmaceutical Supply Chain: A Next Step towards Good Manufacturing Practice.	***2018***					
Application of Data Mining Technology in the CRM of Pharmaceutical Industry.	***2018***					
Detection of outliers and sales prediction for Pharmaceutical Distribution on companies.	***2017***					
A Kanban based system for multi echelon inventory management the case of pharmaceutical supply chain.	***2016***					
Intelligent sales prediction for Pharmaceutical Distribution on Companies.	***2014***					
Pharmaceutical Supply Chain and Inventory Management Strategies: Optimization for a Pharmaceutical Distribution.	***2013***					

Global Supply Chain planning for Pharmaceuticals.	***2011***	✓	✓	✓	✓	✓
Improve Supply Chain Efficiency through a Web Based System : A Case study on a Pharmaceutical Company in Thailand.	***2011***	✓	✓	✓	✓	✓

.

Table 3: Supply chain stages matrix

The previous table pointed that there are rare researches used data mining to improve the pharmaceutical supply chain and few researches which focus on the distribution phase in the pharmaceutical supply chain even though it's a very important element in the pharmaceutical supply chain process because without it the drugs can't deliver to the pharmacy, and therefore can't deliver to the patient at the right time, and also pharmaceutical companies can't ever sell and market their products on a wider scale.

5- Proposed Framework

In this section, an overview of the proposed framework on how to apply data mining techniques in pharmaceutical distribution companies **as shown in figure 3.**

As we mentioned in the literature review that rare researches pointed out the gap that face pharmaceutical distribution companies and how to control the inventory and achieve customer satisfaction without affording huge expenses. So the proposed model discusses how to improve the retail supply chain in pharmaceutical distribution companies by applying data mining techniques to classify Customers based on their purchases and control the inventory by forecasting the accurate amount of products.

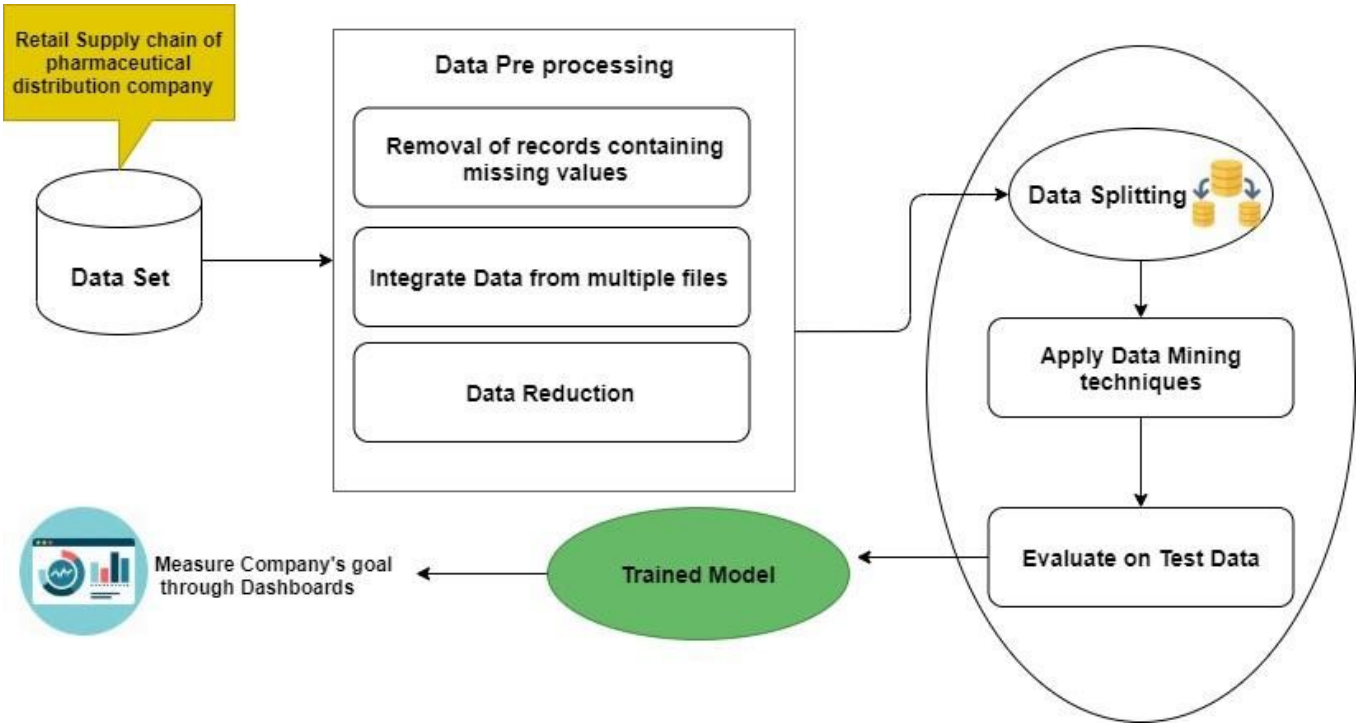

Figure 3: Proposed framework

5.1 Data preprocessing

Data preprocessing as we mentioned before consists of major tasks which are:

a) Data cleaning: Data cleaning helps to fill in missing values and noisy data. As data in the real world contains a lot of errors and mistakes resulting from human errors, computer errors, or transmission errors [23].

*b) Data Integration***:** the merging of data from multiple databases and files to be combined in a common source [23].

c) Data Reduction: using data reduction because while storing large amounts of data, it will be complex and takes a long time to run that data. [2].

5.2 Data Splitting

Data Splitting is the process in which the Data set is splitting and divided into subsets to train, validate, and test the models [36].

a) Training dataset: a subset to train different models, and check the efficiency and accuracy of the algorithms [36].

b) Validation dataset: The validation dataset records validation errors for each model. Used when comparing between two finals models and the optimal model set is determined based on the one with the lowest validation error [36].

c) Test dataset: a subset to test the trained model, used to provide an evaluation about the final model, and the accuracy that obtains at the test stage is the accuracy of the algorithm [36]. **As shown in figure 4**

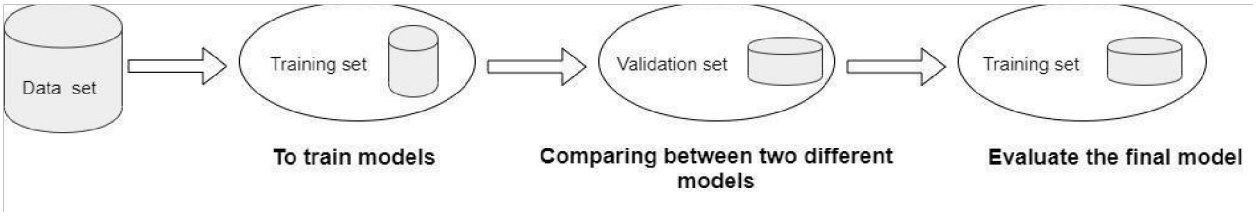

Figure 4: Data Splitting

5.3 Data Mining Techniques

To forecast the products for the coming month, classify customer (pharmacies) according to their sales history. Predictive techniques were selected.

5.4 Key performance indicators (KPIs)

KPIs will be used to monitor and track inventory, follow up the target, and detect if the company meets its target or not.

6- Conclusion

This paper will lead to better and effective success for pharmaceutical distribution companies.
And it's concluded that by using data mining techniques, the company will be able to manage its inventory, detecting customers (pharmacies) and products (medicines), and also detect its target and goals.so, The basic objective of this paper was to offer a precise sales prediction method to help pharmaceutical distribution companies to forecast product sales and tuning inventory management policies in order to prevent costs of excessive inventory and prevent losing customers due to drug shortage. In addition to that, the company will monitor the inventory process and products during the month through Dashboards. All will be beneficiary from this proposed model Pharmaceutical Companies, Pharmaceutical Distribution Companies, Pharmacies, also patients.

a) Pharmaceutical Companies will follow up with their products, customers, and ensure that all the products delivered to the customers on a wider scale. And if there are special or urgent cases pharmaceutical companies will provide pharmaceutical distribution companies as soon as possible.

b)Pharmaceutical Distribution Companies as we mentioned before that the main purpose of this paper is to solve all the constraints that face pharmaceutical distribution companies every month considering that an intermediate between pharmaceutical companies and pharmacies. And by applying data mining techniques pharmaceutical distribution companies will be able to forecast the accurate amount of products, avoid surplus and shortage of products as both of them caused loss to the company, rearrange the customers by their sales and finally monitor the inventory and company's target using KPIs.

c) Pharmacies the number of products that requested will be delivered at a right time thus customer satisfaction will be achieved.

d) Patients will find the medicines they need easily.

References

[1] ZAKI, Mohammed J.; MEIRA JR, Wagner; MEIRA, Wagner. Data mining and analysis: fundamental concepts and algorithms. Cambridge University Press, 2014.

[2] HAN, Jiawei; PEI, Jian; KAMBER, Micheline. Data mining: concepts and techniques. Elsevier, 2011.

[3] KOBERG, Esteban; LONGONI, Annachiara. A systematic review of sustainable supply chain management in global supply chains. Journal of cleaner production, 2019, 207: 1084-1098.

[4] UTHAYAKUMAR, R.; PRIYAN, S. Pharmaceutical supply chain and inventory management strategies: Optimization for a pharmaceutical company and a hospital. Operations Research for Health Care, 2013, 2.3: 52-64.

[5] RIBEIRO, Augusto; SERUCA, Isabel; DURÃO, Natércia. Improving organizational decision support: Detection of outliers and sales prediction for a pharmaceutical distribution company. Procedia computer science, 2017, 121: 282-290.

[6] KHALIL ZADEH, Neda; SEPEHRI, Mohammad Mehdi; FARVARESH, Hamid. Intelligent sales prediction for pharmaceutical distribution companies: A data mining based approach. *Mathematical Problems in Engineering*, 2014, 2014.

[7] SONG, Guijuan. Application of Data Mining Technology in the CRM of Pharmaceutical Industry. In: 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). IEEE, 2018. p. 61-64.

[8] THONGRAWD, Chairit; MEE-NGOEN, Benjabhon; JERMSITTIPARSERT, Kittisak. The supply chain innovation, supply chain transaction cost, supply chain risk and supply chain responsiveness and the supply base and its complexity. International Journal of Supply Chain Management, 2019, 8.4: 269-279.

[9] STAN, L., et al. KPI performance indicators for evaluating employees on industrial production lines. In: 8th internatonal DAAAM Baltic Conference. 2012. p. 1-6.

[10] AMOS, Daniel; AU-YONG, Cheong Peng; MUSA, Zairul Nisham. Developing key performance indicators for hospital facilities management services: a developing country perspective. Engineering, Construction and Architectural Management, 2020.

[11] MEE-NGOEN, Benjabhon; THONGRAWD, Chairit; JERMSITTIPARSERT, Kittisak. The Impact of Product and Process Strategies on the Service Performance: The Mediating Role of Supply Chain Integrations. International Journal of Supply Chain Management, 2019, 8.4: 291-301.

[12] SOHRABI, Babak, et al. A predictive analytics of physicians prescription and pharmacies sales correlation using data mining. International Journal of Pharmaceutical and Healthcare Marketing, 2019.

[13] ZANDIEH, Mostafa, et al. Designing sustainable distribution network in pharmaceutical supply chain: A case study. International journal of supply and operations management, 2018, 5.2: 122-133.

[14] KAPOOR, D., et al. An overview on pharmaceutical supply chain: a next step towards good manufacturing practice. Drug Designing & Intellectual Properties International Journal, 2018, 1.2: 49-54.

[15] SOUSA, Rui T., et al. Global supply chain planning for pharmaceuticals. chemical engineering research and design, 2011, 89.11: 2396-2409.

[16] MOUAKY, Malak; BERRADO, Abdelaziz; BENABBOU, Loubna. A Kanban based system for multi-echelon inventory management: The case of pharmaceutical supply chain. In: 2016 3rd International Conference on Logistics Operations Management (GOL). IEEE, 2016. p. 1-6.

[17] POTHITONG, Panthip; CHAROENSIRIWATH, Chayakrit. Improve supply chain efficiency through a web-based system: A case study on a pharmaceutical company in Thailand. In: 2011 IEEE International Conference on Quality and Reliability. IEEE, 2011. p. 566-569.

[18] ABDULGHANI, Ibrahim; BERMAMET, Hala; ALTARAZI, Safwan. Pharmaceutical Supply Chain Cost optimization Model Considering Multiple Echelons and Multiple Drugs. In: 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA). IEEE, 2019. p. 709-713.

[19] JOTHI, Neesha, et al. Data mining in healthcare–a review. Procedia computer science, 2015, 72: 306-313.

[20] STAN, L., et al. KPI performance indicators for evaluating employees on industrial production lines. In: 8th internatonal DAAAM Baltic Conference. 2012. p. 1-6.

[21] ARJI, Goli, et al. A systematic literature review and classification of knowledge discovery in traditional medicine. Computer methods and programs in biomedicine, 2019, 168: 39-57.

[22] JAIN, Nikita; SRIVASTAVA, Vishal. Data mining techniques: a survey paper. IJRET: International Journal of Research in Engineering and Technology, 2013, 2.11: 116-119.

[23] DOGRA, Ashish Kumar; WALA, Tanuj. A review paper on data mining techniques and algorithms. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2015, 4.5: 1976-1979.

[24] ALI, Faraaz, et al. A Brief Analysis of Data Mining Techniques. In: 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE). IEEE, 2019. p. 752-758.

[25] MOHAMAD, Mohd Fathi; UDIN, Zulkifli M.; SHARIF, Kamal Imran. Inventory, Warehousing and Transportation Management Impacts Towards Logistics Performance in Supply Chain Management. Int. J Sup. Chain. Mgt Vol, 2018, 7.6: 296.

[26] GUPTA, Aviral; SHARMA, Akshay; GOEL, Amita. Review of Regression Analysis Models.

[27] VALLANDINGHAM, Logan R., et al. Grocery retail supply chain planning and control: Impact of consumer trends and enabling technologies. IFAC-PapersOnLine, 2018, 51.11: 612-617.

[28] SINGH, Sonia; GUPTA, Priyanka. Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey. International Journal of Advanced Information Science and Technology (IJAIST), 2014, 27.27: 97-103.

[29] ÇIĞŞAR, Begüm; ÜNAL, Deniz. Comparison of data mining classification algorithms determining the default risk. Scientific Programming, 2019, 2019.

[30] GUPTA, Swati. A regression modeling technique on data mining. International Journal of Computer Applications, 2015, 116.9.

[31] KOTU, Vijay; DESHPANDE, Bala. Data science: concepts and practice. Morgan Kaufmann, 2018.

[32] DIWATE, Rahul B.; SAHU, Amit. Data Mining Techniques in Association Rule: A Review. International Journal of Computer Science and Information technology, 2014, 227-229.

[33] COLEMAN, Jamie J.; PONTEFRACT, Sarah K. Adverse drug reactions. Clinical Medicine, 2016, 16.5: 481.

[34] JEETU, Gangil; GIRISH, Thunga. Prescription drug labeling medication errors: a big deal for pharmacists. Journal of Young Pharmacists, 2010, 2.1: 107-111.

[35] SINGH, Rajesh Kr; KUMAR, Ravinder; KUMAR, Pravin. Strategic issues in pharmaceutical supply chains: a review. International Journal of Pharmaceutical and Healthcare Marketing, 2016.

[36] XU, Yun; GOODACRE, Royston. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. Journal of Analysis and Testing, 2018, 2.3: 249-262.