International Journal of Computer Applications Technology and Research
Volume 4– Issue 6, 467 - 474, 2015, ISSN:- 2319–8656

Association Analysis of Pharmaceutical Imports in Kenya

Christopher Moturi
School of Computing and Informatics
University of Nairobi
Nairobi, Kenya

Victor Wahinya Kahonge
School of Computing and Informatics
University of Nairobi
Nairobi, Kenya

Abstract: The objective of this study was to apply Data Mining in the analysis of imports of pharmaceutical products in Kenya with the aim of discovering patterns of association and correlation among the various product groups. The RapidMiner Data Mining was used to analyze data obtained from the Pharmacy and Poison Board, the regulator of pharmaceutical products in the country, covering the period 2008 to 2010. The CRISP method was used to get a business understanding of the Board, understand the nature of the data held, prepare the data for analyze and actual analysis of the data. The results of the study discovered various patterns through correlation and association analysis of various product groups. The results were presented through various graphs and discussed with the domain experts. These patterns are similar to prescription patterns from studies in Ethiopia, Nigeria and India. The research will provide regulators of pharmaceutical products, not only in Kenya but other African countries, a better insight into the patterns of imports and exports of pharmaceutical products. This would result into better controls, not only in import and exports of the products, but also enforcement on their usage in order to avert negative effects to the citizens.

Keywords: Data Mining, Pharmaceutical Imports, Drug Discovery, Pharmacy and Poisons Board of Kenya

1.0.INTRODUCTION

A drug is any substance used to treat and or prevent disease. According to the Kenyan Pharmacy and Poisons Act CAP 244 a drug includes any medicine, medicinal preparation or therapeutic substance. A medicinal substance refers to any medicine, product, article, or substance which is claimed to be useful for purposes such as; treating, preventing, alleviating disease or symptoms of disease, diagnosing disease, preventing or interfering with the normal operation of a physiological function whether permanently or temporarily. The Pharmacy and Poisons Board (PPB) was created by an act of the Kenyan Parliament in 1957 to make better provision for the control of the profession of pharmacy and the trade in drugs and poisons. PPB is mandated by the Kenya Government to regulate the import and export of all medicines and their raw materials which are referred to as active pharmaceutical ingredients.

PPB has in place information systems that generate vast amounts of valuable data that, if exploited with the correct tools, will enable management unlock the relevant knowledge lying hidden in its databases and files. Currently PPB does not employ any tools relating to artificial intelligence in the extraction of knowledge from its data. However, with the current technological advances and the need to adapt to current trends will move it to utilizing the data they hold for more than what they are using it for.

Data Mining (DM), a relatively new field of analysis, is defined as the process of discovering several models, summaries and derived values from a given collection of data (Kantardzic, 2011). Data Mining which is often referred to as Knowledge Discovery in Databases (KDD) aims at the automatic interpretation of large datasets (Kriegel, Borgwardt, Kröger, Pryakhin, Matthias, &Zimek, 2007).

The potential of DM has been shown through application in various medical disciplines. Wilson et al (2003) discussed the potential use of data mining and knowledge discovery in databases for detection of adverse drug events in pharmacovigilance. They suggested the likely increase in importance of DM in the process of pharmacovigilance as they are likely to be able to detect signals earlier than using other methods. Ji et al (2013) discussed the effects of applying DM to the analysis of hierarchical nursing effects as an effective method to help hospitals improve service quality and strengthen clinical management. Other applications include Santosa et al (2013) and Deshpande (2010).

This research sought to analyze the data on imports of pharmaceutical products in Kenya with the aim of discovering patterns of association and correlation between the various pharmaceutical product groups. Pharmaceutical products are highly sensitive in terms of monetary value and social implications. Some of them are utilized in the black market to produce drugs which can be abused by the general public. Other pharmaceutical products such as antibiotics are prone to misuse leading to drug resistance (Afsan, Haque, Alam, & Noor, 2012). The benefits of obtaining patterns of imports or exports will be very valuable to pharmaceutical regulators in the process of making decisions and forecasting future needs. Control of such products will result in keeping citizens safe.

2.0.RELATED LITERATURE

2.1. Knowledge Discovery Process

Data Mining is a key phase in the process of Knowledge Discovery in Databases that is used in the creation of models from the mass data thus producing meaningful information. A review of the applications of data mining techniques to support knowledge management process has been done by (Silwattananusarn & Tuamsuk, 2012). The process of Data Mining has two primary goals which are prediction and description of a particular dataset under study. Prediction entails the use of some variables occurring in a data set so as to predict unknown values of other relevant variables Predictive data mining includes classification, regression, and anomaly detection. Description entails the discovery of human understandable patterns and sequences in the data. Descriptive data mining includes clustering, association rule learning, and summarization. Our study involved the use of descriptive data mining so as to bring out the patterns in the data they hold. This is because there is the need to know what information you hold first before you proceed to predict the future with the current data. The iterative nature of knowledge discovery process is represented by the model shown in fig 1.

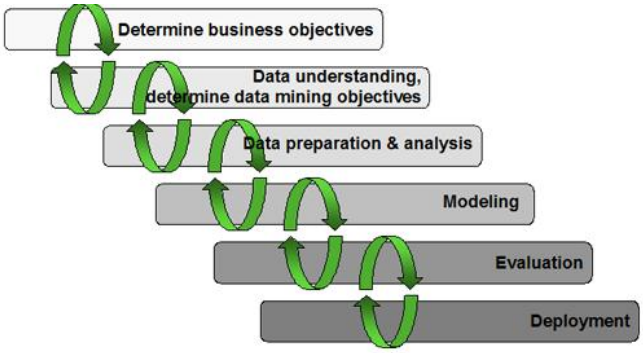

Figure 1: Process steps of Knowledge Discovery
Source: http://www.microsegment.hu/ms/methodology1.php

The steps are outlined below.

Business Understanding - This initial phase focuses on understanding the project objectives and requirements from a business perspective and then converting this knowledge into a data mining problem definition, and a preliminary project plan designed to achieve the objectives.

Data Understanding - The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data. Its aim is to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

Data Preparation - The data preparation phase covers all activities to construct the final dataset from the initial raw data. Raw data is usually noisy, incomplete or impure which is very likely to hamper the discovery useful patterns Current data mining tools require high quality data and high quality data results in intelligent patterns (Zhang, Zhang, & Yang, 2003). Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record, and attribute selection, data cleaning, construction of new attributes, and transformation of data for modelling tools.

Modelling - In this phase, various modelling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type as some techniques require specific data formats. There is a close link between Data Preparation and Modelling. Often, one realizes data problems while modelling or one gets ideas for constructing new data.

Evaluation - At this stage evaluation of models built are thoroughly evaluated in order to ascertain that it properly answers the business questions and satisfies the objectives set out at the beginning. Once this is complete, the model can now be used by the customer to solve the problem or satisfy the objectives set out.

Deployment - After the creation of the model, the usefulness of the model needs to be utilised by the customer (Džeroski, 2007), thereby justifying why so much time and resources have been spent on the project. The aim of this phase is to hand over the model to the customer in a way that he/she can utilise it on other similar scenarios with the same data source.

2.2. Knowledge Discovery in the Pharmaceutical Industry

The pharmaceutical industry mainly relies on decision oriented and systemic selection models which enable decision makers to evaluate the expected results of management decisions. A firm's competitive advantage and decision making ability can be greatly increased through the understanding of knowledge hidden in pharmaceutical data (Ranjan, 2007).This can be done through the application knowledge discovery to data repositories held by organizations. Data mining is currently being utilized in the development of solutions for the pharmaceutical industry in areas such as such as drug discovery technology, improved marketing strategies, and decision support (Yang, Adelstein, &Kassis, 2009). Yang et al (2009) propose data mining in the Target Discovery phase of the process of drug discovery as this phase utilizes biological data and information contained in vast data warehouses that usually double every two years or so.

Studies have been done to determine prescription patterns of the specific medicines within the pharmaceutical product groups. Studies done in Ethiopia show that the average number of drugs per prescription ranges from 1.98 to 2.24 (Angamo, Wabe, & Raju, 2011) and this goes to show that a visit to the hospital or pharmacy outlet will give a person more than one type of drug. Studies by (Kajungu, et al., 2012) shows that indeed there is irrational use of drugs in Tanzania which is quite similar to Kenya. Our study was aimed at understanding prescribing patterns in health facilities in order to reduce the rate of irrational use of drugs. The study utilized classification trees which are generally easy to represent information obtained from analysis. Kajungu et al (2012) propose that data mining is paramount in the process of the identification and control of polypharmacy. While polypharmacy may not necessarily be wrong, the practice may place patients at high risk of adverse drug reactions and increased bacterial resistance. The same phenomenon is replicated in a study in Nigeria by (Okoro & Shekari, 2013) where there is a high occurrence of polypharmacy.

2.3. Review of Knowledge Discovery Tools

2.3.1. WEKA

WEKA (Waikato Environment for Knowledge Analysis) is an open source data mining tool consisting of a wide array of state-of-the-art data mining and machine learning algorithms which are implemented in Java (Mikut&Reischl, 2011). The algorithms can either be applied directly to a dataset or called from Java code (Hall, Frank, Holmes, Pfahringer, Reutemann, & Witten, 2012). WEKA is recognized as a landmark system in data mining and machine learning. It has achieved widespread acceptance within academia and business circles, and has become a widely used tool for data mining research. WEKA has several tools incorporated into its structure namely: Data pre-processing, Classification, Regression, Clustering, Association rules and Visualization. WEKA has a modular and extensible architecture allowing users to test and compare different machine learning methods on new data sets while building sophisticated data mining processes from the wide collection of base algorithms.

2.3.2. KEEL

KEEL (Knowledge Extraction based on Evolutionary Learning) is an open source software that supports data management and a designer of experiments while paying special attention to the implementation of evolutionary learning and soft computing based techniques for Data Mining problems including regression, classification, clustering, and pattern mining (Alcalá-Fdez, et al., 2011). KEEL is implemented in Java and empowers the user to analyze the behaviour of evolutionary learning for different kinds of DM problems such as regression, classification, unsupervised learning. KEEL contains a library with evolutionary learning algorithms based on different paradigms and simplifies the integration of evolutionary learning algorithms with different pre-processing techniques thus reducing programming work. It requires less

technical work thus enabling researchers to focus more on analysis of new learning models in comparison with the existing ones. It contains a user friendly interface which is oriented to the analysis of algorithms.

2.3.3. KNIME

KNIME (Konstanz Information Miner) is also an open-source tool that allows a user to perform sophisticated statistics and data mining on data so as to analyze trends and predict potential results (Mikut&Reischl, 2011). It is a visual workbench that combines data access, data trans-formation, initial investigation, powerful predictive analytics and visualization. KNIME provides the ability to develop reports based on one's information or automate the application of new insight back into production systems. It can perform the following types of analysis; Regression, Classification, Clustering, Pattern Mining and Un-supervised learning. KNIME has a quite intuitive graphical user interface thus quite easy to operate. It contains an open integration platform which provides over 1000 modules including those on the KNIME community. It allows for parallel execution on multi core systems thus utilizing system resources efficiently. In addition, it is highly extensible.

2.3.4. RapidMiner

RapidMiner is one of the world's leading open source software for Data Mining that can be used as a standalone or integrated into other products. RapidMiner, developed in Java, is a complete analytics workbench with a strong focus on Data Mining, text mining and predictive analytics. The strength of RapidMiner includes: provides more than 400 operators, graphical user interface, several modes of access, XML process exchange, works on major operating systems and platforms, easy to learn and work with even for non –programmers, and has a high availability of materials on how to use it in the process of Data Mining. The software supports access to data sources like Excel, Access, Oracle, IBM DB2, Microsoft SQL, Sybase, Ingres, MySQL, Postgress, SPSS, dBase, Text files through a very simple process.

3.0.METHODOLOGY

3.1. Research Design

The study utilized quantitative research. RapidMiner and Microsoft SQL Server were used to analyze the PPB database containing data on imports of pharmaceutical products in order to understand and give out intelligent inferences from it.

3.2. Data Source

The study utilized data held in the import and export database of the Pharmacy and Poisons Board (PPB) of Kenya. This data originates from the Licit Control Department of PPB that is charged with the issuance of import/export permits regarding pharmaceutical products in the country. This database contains information pertaining to quantities of imports and exports products and the respective dates of permits issued. It also contains the business entities to which the permits have been given.

3.3. Data Collection and Analysis

The CRISP-DM (Cross Industry Standard Process for Data Mining) was used in the study. CRISP-DM is a comprehensive process model for carrying out Data Mining projects. The process model is independent of both the industry sector and the technology used (Wirth &Hipp, 2000). It is acknowledged and widely used especially in the fields of research and industrial communities (Kurgan &Musilek, 2006). The aim of CRISP-DM is to provide an efficient process that can be used by persons with lower technical skills in Data Mining to produce knowledge from their vast data repositories.

CRISP-DM can be integrated with a specific project management methodology complementing administrative and technical tasks. CRISP-DM defines a structure for Data Mining projects and provides orientation for their execution. It serves both as a reference model and a user guide (Chapman et al., 2000). The reference model gives a general view of a Data Mining project's life-cycle, containing each phase with its objective, the tasks, the relationships between them and the step-by-step instructions that must be carried out. CRISP-DM therefore guided the whole process from start to finish.

The CRISP-DM methodology has six phases: Business Understanding; Data Understanding; Data Preparation; Modelling; Evaluation; and Deployment (Fig. 2).

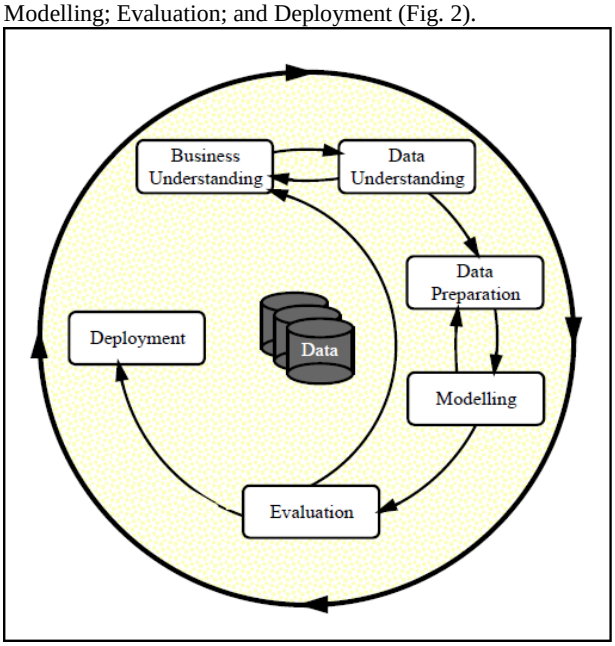

Figure 2: Phases of the CRISP-DM process model for the entire knowledge discovery process

3.3.1. Business Understanding

This phase was the key linking factor of this research to the business perspective of the board. The knowledge obtained was then converted into a Data Mining problem definition. The phase identified key persons in the import and export section, other sections that were likely to be impacted by the research, gathered user requirements and expectations and looked at expected benefits of the research. This phase related the business questions to Data Mining goals while specifying the Data Mining problem type and the criteria for model assessment. Examples of Data Mining problem types include classification, prediction, association and clustering. The business process for the issuance of import/export permits to agents is as shown in Fig. 3. Of major concern to PPB is obtaining valuable information from the data on approved permits. This is important to obtain yearly market trends of the various categories of pharmaceutical products, and to discover associations if any within the pharmaceutical product categories with time.

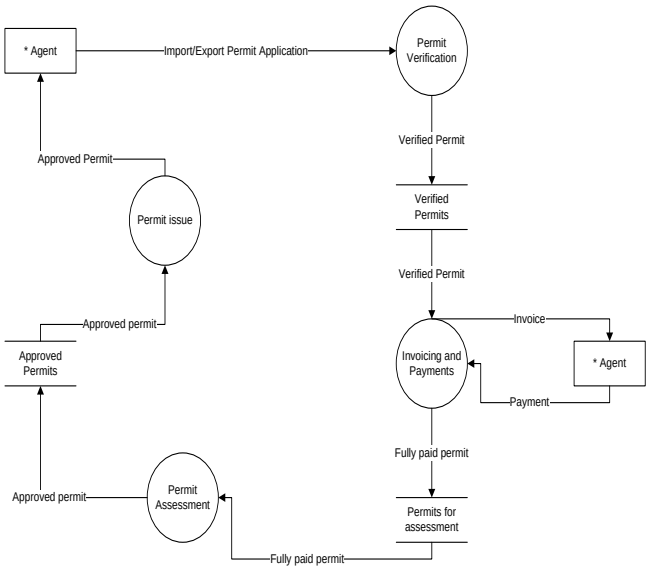

Figure 3: Data flow diagram representing the process of the application of a permit till issue

3.3.2. Data Understanding

Data on imports and exports permits was available from the electronic data repository from the data entry system. The database is housed in Microsoft Access and contains several tables and the tables have several attributes. The following are the tables as they appear in the repository of the data entry system.

Table 1: Summary of the PPB Import/Export Database

Table name	**Information contained**
dbo_dbo_Agent	Stores basic information of the agent such as City, Postal Address, Telephone Number and email
dbo_dbo_invoice	Stores information on the Proforma Invoice of each consignment, the manufacturer, the agent importing the consignment, the origin country and port, destination country and port, and the Invoice Value of the consignment
dbo_dbo_Manufacturer	Stores basic information of the Manufacturer such as City, Postal Address, Telephone Number and email
dbo_dbo_Country	Stores a list of all countries in the world
dbo_dbo_LineItem	Stores detailed information of the products imported/exported within each proforma Invoice. It is linked to dbo_dbo_invoice via a foreign key and is also linked to dbo_dbo_IDF via foreign key
dbo_dbo_IDF	Stores data from the IDF details such as the IDF total value, IDF number and the date of import

Fig. 4 shows an illustration of the various relationships that exist in the database that was used in the generation of queries to the database.

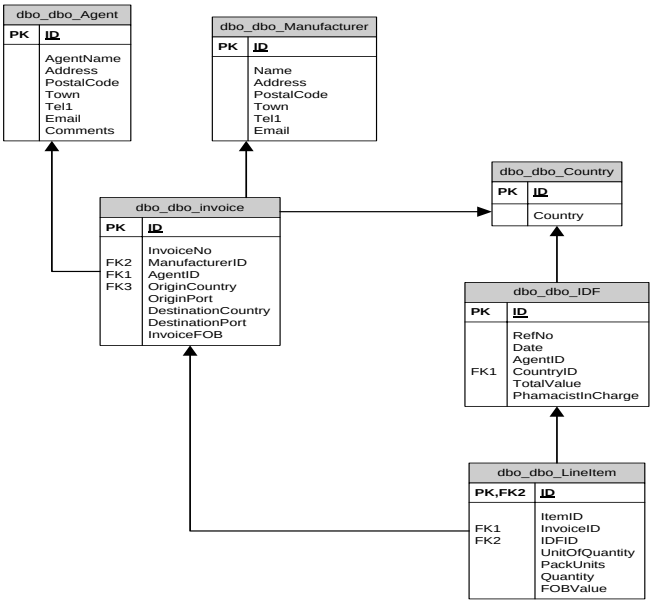

Figure 4: Entity relationship diagram of the PPB import / export database

3.3.3. Data Preparation

A full backup of the database was taken from PPB and migrated to Microsoft SQL Server through the Upsizing wizard present in Microsoft Access. Based on the relationships outlined in the data understanding phase, specific database scripts (see appendix 1) were developed to query the database of relevant data for the study. On running scripts missing data was identified and upon discussion with domain experts it was agreed to omit all missing values.

The process of data entry skipped the years 2009 and 2010.

Classification is usually done by the domain experts, who in this case are pharmacists and pharmaceutical technologists, who are conversant with the Kenyan pharmaceutical market. They usually match the names on the product to a list of existing group for example in the group 'Antibiotic' there are products having the name or part of their name containing 'Amoxil', a product having part of its name as 'Malarid' is likely to fall within the category if 'Anti-Malarials' and so on. Due to the large number of records in the database, it was close to impossible to classify the items one by one through a pharmaceutical expert.

With this knowledge of the technique of matching, a sample of all product names and their respective group names was obtained from PPB and this was used to generate a database script that was used to group all the products into their respective classifications. A new table was created to contain the names of all product categories and their category ID.

3.3.4. Modelling

The data obtained from the data preparation phase was fed into the RapidMiner for analysis. Since our data was in MSSQL Server, the first step was to create a database connection to be used to connect RapidMiner to MSSQL Server without having to specify the details over and over again each time a new query is to be run. The dates were changed appropriately so that the results could be obtained for the various years within the dataset. The next stage is the opening of a new 'process' which

shall open a new working space. The data is then loaded for modeling.

3.3.5. Correlation Analysis

Correlation analysis was done on the dataset obtained from the data preparation phase. A new process was created with the 'Read Database' operator and the 'Correlation Matrix' operator as illustrated in figure 5. The dataset range was modified to include the years 2008 through to 2013.

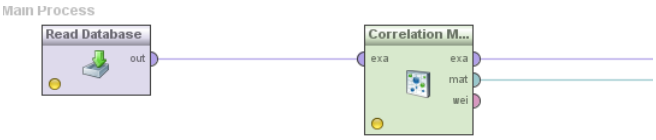

Figure 5: Correlation Process in Rapidminer

3.3.6. Association Analysis

The same data set was then subjected to a new process for association analysis. This process required the following operators illustrated in figure 6.

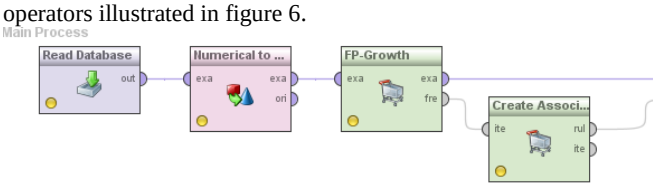

Figure 6: Process workflow for Association Analysis

4.0.RESULTS AND DISCUSSION

4.1. Correlation among Pharmaceutical Product groups

Through correlation analysis by RapidMiner it was found that there are product groups which are related based on their FOB values. A sample of the matrix indicating the level of correlation as shading of the coefficients obtained is as shown in Table 2. For better visualization of the correlations, we generated pair wise plots of the highly correlated pairs of pharmaceutical products. The following are the selected correlations done on the various groups of pharmaceutical products:

Drugs affecting Blood against Anti Diabetics
Respiratory Tract against Anti-Allergies
Respiratory Tract and Antibiotic drug
Drugs Affecting Blood and Cardiovascular Drugs
Drugs Affecting Blood and Hormones, Endocrine and Contraceptive Drugs
Drugs affecting Blood and Respiratory Tract Drugs
Drugs Affecting Blood and Anti-Allergies and Anaphylaxis
Respiratory Tract and Hormone, Endocrine and Contraceptives
Respiratory tract and Anti-Allergies, Anaphylaxis
Antiretrovirals and Antibiotics
Antiretrovirals and Cardiovascular Drugs
Antiretrovirals and Antimalarial Drugs
Antiretrovirals and Dermatological Drugs
Cardiovascular and Antineoplastics, Immunorepressive
Antibiotics and Dermatological Drugs
Antineoplastics_Immunorepressive and Antimalarials
Antibiotics and Dermatological Drugs
Antibiotics and Cardiovascular Drugs

Samples of the plots generated are shown in Fig 7 and Fig 8.

Table 2: Part of the Correlation Matrix showing the relationship between the various product groups

Attributes	DRUGS_AFFECTING_BLOOD	MUSCLE_RELAXANTS	ANTI_DIABETICS	BLOOD_PRODUCTS..	HORMONE..	RESPIRATORY_TR..	ANTI_A
DRUGS_AFFECTING_BLOOD	1	0.031	0.930	0.382	0.488	0.170	0.188
MUSCLE_RELAXANTS	0.031	1	-0.172	0.182	0.027	0.167	0.269
ANTI_DIABETICS	0.930	-0.172	1	0.342	0.489	0.311	0.252
BLOOD_PRODUCTS_BLOOD_SUBSTITUTES	0.382	0.182	0.342	1	0.413	0.251	-0.049
HORMONES_ENDOCRINE_and_CONTRACEPTIVES	0.488	0.027	0.489	0.413	1	0.038	0.241
RESPIRATORY_TRACT	0.170	0.167	0.311	0.251	0.038	1	0.712
ANTI_ALLERGICS_ANAPHYLAXIS	0.188	0.269	0.252	-0.049	0.241	0.712	1
PSYCHOTHERAPEUTIC_DRUGS	0.323	0.535	0.342	0.059	0.232	0.032	0.026
MISCELLANEOUS	-0.028	-0.137	-0.013	-0.035	-0.032	0.331	0.276
ANAESTHETICS	0.103	0.658	-0.112	0.112	0.239	0.306	0.026
VITAMIN_and_MINERALS	-0.302	0.122	-0.163	-0.023	-0.228	0.078	-0.215
ANTIEPILEPTICS	-0.132	-0.169	0.100	0.186	0.345	-0.009	-0.121
GASTROINTESTINAL	0.075	0.303	0.153	0.046	-0.017	0.003	-0.042
IMMUNOLOGICALS_VACCINES	-0.092	-0.276	-0.103	0.295	-0.083	0.147	0.065
ANTI_VIRALS	-0.092	-0.187	-0.087	-0.147	-0.132	0.250	0.208
ANALGESICS_ANTIPYRETICS_NSAIDs	-0.020	-0.121	0.091	0.117	0.018	0.102	-0.077
ANTIMIGRANE	-0.249	1	0.186	0.183	0.258	-0.071	-0.044
ANTI_RETROVIRALS	-0.006	-0.239	-0.004	-0.074	0.117	0.227	0.284
ANTINEOPLASTICS_and_IMMUNOSUPPRESSIVE	0.005	0.124	0.073	0.272	0.064	-0.035	-0.105
ANTI_MALARIALS	0.288	-0.127	0.110	-0.168	0.087	0.114	-0.019
OPHTALMOLOGICAL_ENT_PREPARATIONS	?	?	?	?	?	?	?

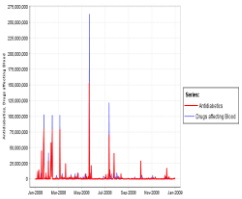

Figure 7: Correlation between Drugs affecting Blood and Anti-diabetics

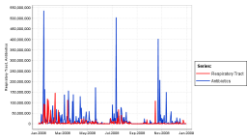

Figure 8: Correlation between Respiratory Tract against Anti-Allergics

From the selected plots of correlated product groups, the presence of correlation can be visually confirmed in addition to the correlation matrix. These correlations are indicative of the fact that importation of Respiratory Tract drugs and Anti-Allergics occurs simultaneously. The same case applies to Drugs Affecting Blood and Anti-diabetics. These correlations also indicate that an increase in the value of imports of one category is directly linked to an increase in the other.

4.2. Association Analysis

The occurrence of frequent item sets is common within transactions containing items. Frequent item sets are simply groups of items that often occur together in a data set. The data on imports of pharmaceuticals contains various associations. It was observed that the product group Antibiotics is associated to a number of other combinations of product groups which implies that within the various item sets, there exists an Antibiotic. It also implies that an antibiotic is a common conclusion to many product groups imported into the country.

Table 3: Association Rules obtained from Rapid Miner at 0.95 confidence level

PREMISES	**CONCLUSION**	**SU**	**ENCE**
DERMATOLOGICAL	ANTI_BIOTICS	0.6	
GASTROINTESTINAL, ANALGESICS ANTIPYRETICS_NSAIDs	ANTI_BIOTICS	0.573883	0.95977
GASTROINTESTINAL, RESPIRATORY_TRACT	ANTI_BIOTICS		0.965318
GASTROINTESTINAL, DERMATOLOGICAL	ANTI_BIOTICS	0.563574	0.982036
ANTI_BIOTICS, CARDIOVASCULAR	GASTROINTESTINAL	0.560137	0.958824
GASTROINTESTINAL, CARDIOVASCULAR	ANTI_BIOTICS	0.560137	0.970238
ANALGESICS_ANTIPYRETICS_NSAIDs, RESPIRATORY_TRACT	ANTI_BIOTICS	0.5	
ANALGESICS_ANTIPYRETICS_NSAIDs, RESPIRATORY_TRACT	GASTROINTESTINAL	0.532646	
ANALGESICS_ANTIPYRETICS_NSAIDs, DERMATOLOGICAL	ANTI_BIOTICS	0.52921	
ANALGESICS_ANTIPYRETICS_NSAIDs, CARDIOVASCULAR	ANTI_BIOTICS	0.522337	

Investigation of the associations conducted included the following groups of pharmaceutical products:
Dermatological and Antibiotics
Gastrointestinal, Analgesic, Antipyretic, NSAIDS and Antibiotics
Gastrointestinal, Antibiotics and Respiratory Tract
Gastrointestinal, Dermatological and Antibiotics
Antibiotics, Cardiovascular and Intestinal Drugs
Analgesics, Antipyretics, NSAIDS, Respiratory Tract and Antibiotics
Analgesics, Antipyretics, NSAIDS, Dermatological and Antibiotics
Analgesics, Antipyretics, NSAIDS, Cardiovascular and Antibiotics

Sample charts are shown in Fig 9 and Fig 10

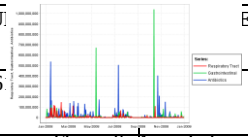

Figure 9: Association among three Products - Gastrointestinal, Respiratory Tract and Antibiotics

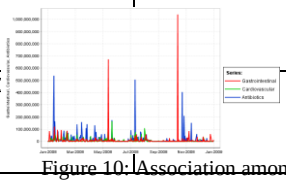

Figure 10: Association among three Products – Gastrointestinal, Cardiovascular, and Antibiotics

The following is a brief description of the functions of the mentioned categories:
Gastrointestinal – Drugs that are used to treat ailments of the digestive system which is also referred to as the gastrointestinal tract.
Antibiotics – Drugs that are used to treat diseases brought about by bacterial infections.
Respiratory Tract – drugs that are used in the treatment of diseases affecting the respiratory system.
Cardiovascular- Drugs used to control ailments related to the heart and blood circulation

4.3. Discussion

After performing correlation and association analysis on the data on imports of pharmaceutical products, some inferences can be made from it. From correlation analysis, it can be concluded that within the pharmaceutical product groups imported, there are related pairs based on the results of the coefficients in the correlation matrix. From association analysis, it can be shown that there are associations between several product groups which are represented as frequent item sets. The results also show that there are rules to which the instantaneous imports are done. For every instant purchase of Gastrointestinal and Dermatological drug, there is an Antibiotic drug as indicated by the rule Gastrointestinal, Dermatological → Antibiotics with a confidence of 0.982036. These are indicative of various product groups being imported at the same time and a particular order of occurrence. The common occurrence of Antibiotics as a conclusion to most premises is indicative of the fact that antibiotics are widely used to treat many infections which are mostly bacterial in nature. Graphs of combined product groups plotted against time, for example Gastrointestinal, Respiratory Tract, and Antibiotics, show a similar trend to confirm their association with periods of inactivity in June and September.
There is a major concern on the use of antibiotics, cough and cold medicines, painkillers and anti-diarrhœals in many developing countries (Le Grand, Hogerzeil, &Haaijer-Ruskam, 1999). The study shows that the sales of the aforementioned classes of medicines exceed the medical condition they are supposed to treat. From the association and correlation analysis done, it can be assumed that inflow of drugs into Kenya follows distinct patterns.
Through this research, we have demonstrated the existence of correlations and associations among the various groups of pharmaceutical products imported to Kenya. There correlations and associations are well described through the observation of trends of the particular product groups of interest. These trends are most likely due to the market responding to a particular need or situation. Research has shown that a visit to a hospital is likely to result in the prescription of a number of medicines to cure the ailment suffered (Adebayo &Hussain, 2010). Usually quite a number of Antibiotics are prescribed.
Research in Bangladesh on drug use and prescribing patterns of medical staff shows existence of a trend on issuance of multiple medicines or poly-pharmacy (Afsan et al, 2012). The same phenomenon is replicated in Ethiopia whereby the average number of drugs per prescription ranges from 1.98 to 2.24 (Angamo, Wabe, &Raju, 2011). The issuance of multiple medicines to a patient implies that there are associations between the various types of medicines available in the country and hence the importance of discovering these association patterns within them.

5.0.CONCLUSION

This research sought to demonstrate the application of Data Mining in the analysis of imports of pharmaceutical products in Kenya with the objective of discovering patterns of association and correlation among the product groups. Correlation analysis showed the product groups which are related. Association analysis showed the combination of product groups that are associated, for Antibiotics and several other groups.
It is hoped that this research will give regulators of pharmaceutical products in African countries a better insight into the usage of permits and licenses issued to agents for imports and exports of pharmaceutical products. This would result into better control mechanisms of the patterns of import and exports of the products in order to keep their citizen safe from abuse of these highly controlled products.
Associations between the various pharmaceutical product groups and the deduction that this is linked to the actual use, would require the regulators, such as the Pharmacy and Poison Board (PPB) of Kenya, to enforce its influence in the use of doctors' prescriptions in drug shops or liaise with partner regulatory boards such as those regulating doctors to look into the composition of their prescriptions of medicines to patients. The study can provide regulators with indications of excessive or limited amounts of a particular category of product so as to trigger the necessary regulatory action within its mandate to address the situation and avert negative effects to the country's citizens. The study might also trigger the investigation of the use of a particular product group, for example Antibiotics, with the aim of enforcing their proper use to reduce resistance and other issues related to their abuse.

6.0.REFERENCES

Angamo, M. T., Wabe, N. T., & Raju, N. J. (2011). Assessment of Patterns of Drug use by using World Health Organization's Prescribing,Patient Care and Health facility indicators inSelected Health Facilities in Southwest Ethiopia. *Journal of Applied Pharmaceutical Science, 7*, 62 - 66.

Džeroski, S. (2007). Towards a General Framework for Data Mining. *Knowledge Discovery in Inductive Databases, 4747*.

Kajungu, D. K., Selemani, M., Masanja, I., Baraka, A., Njozi, M., Khatib, R., . . . Speybroeck, N. (2012). Using classification tree modelling to investigate drug prescription practices at health facilities in rural Tanzania. *Malaria Journal*, 11: 311.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms.* New Jersey: John Wiley & Sons.

Kriegel, H.-P., Borgwardt, K. M., Kröger, P., Pryakhin, A., Matthias, & Zimek, A. (2007). Future trends in data mining. *Data Mining Knowledge Discovery*, 15:87–97.

Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining Process Models. *The Knowledge Engineering Review*, 1 - 24.

Le Grand, A., Hogerzeil, H. V., & Haaijer-Ruskam, F. M. (1999). Intervention Research in Rational use of

Drugs: a Review. *Health Policy and Planning*, 89 - 102.

Okoro, R. N., & Shekari, B. G. (2013). Physicians' drug prescribing patterns at the national health insurance scheme unit of a teaching hospital in the North Eastern Nigeria. *Archives of Pharmacy Practice, 4*(1), 3-8.

Ranjan, J. (2007). Application of Data Mining Techniques in the Pharmaceutical Industry. *Journal of Theoretical and Applied Information Technology*, 61 - 67.

Silwattananusarn, T., & Tuamsuk, K. (2012). Data Management and Its Application for Knowledge Management: A Literature Review from 2007 to 2012. *International Journal of Data Mining & Knowledge Management Process*, 13 - 24.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, 29 - 39.

Yang, Y., Adelstein, S. J., & Kassis, A. I. (2009, February). Target discovery from data mining approaches. *Drug Discovery Today*, *14*(3-4), 147-154.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data Preparation for Data Mining. *Applied Artificial Intelligence*, 375 - 381.