**Data Mining Project**

**Prof Tony**

**INFO 317**

**03/14/2021**

**Team Members**

1. **Ayaz Rathod**
2. **Devam Patel**
3. **Junaid Asim**
4. **Parth Patel**
5. **SM Pantho**

**Introduction**

The physical and emotional fitness of individuals is one of the most important indices of general wellbeing and quality of life in the country. In every area of Ghana, public and private health facilities can be located. Ghana finances and administers public health services, while private health facilities are financed and run privately. Health authorities have consistently tried to provide all people with health insurance with an equal access program to ensure the correct treatment is provided in the right place and at the right time. There are several issues (challenges) that make it impossible to enforce such a health program. One of these problems is the connection between the demand for health facilities and their distance from services, financial stability, time limits, social disadvantages, and the psychological burden of commuting to health services. (Wang & Sun, 2017).

Healthcare managers must make daily choices on service quality without understanding what will come next. Predictions will allow them to forecast and prepare for the future. Market forecasting is a key business feature for suppliers, distributors, and service providers. Good projections are the basis for short, medium, and long-term planning in all forms of services. Forecasting has two purposes: assisting supervisors in the forecasting and preparation of the system. The preparation of the system is long-term in nature: what resources are offered and how many of them are available, what infrastructure and equipment are required, where patients are better served, etc. (Milovic, 2012).

Life is lost because the market for vital health goods lacks correct and reliable statistics. Deficiencies and market predictions tend to a discrepancy between production and demand, leading to unnecessarily high costs and lack of supply. Demand forecasting for both consumers and manufacturers is a vital policy, but it is only one step through a lengthy and sometimes complex supply chain. The remainder of the supply chain cannot be effectively mobilized to provide care due to the inherent complexities in the demand projection, for younger markets (Moore, 2002).

The medical companies today generate huge volumes of sophisticated data, including information on patients, hospital resources, illness identification, patient electronic records and medical devices. Large quantities of data are a vital resource that must be collected and analyzed to gather information to cut costs and make decisions. Data mining offers several methods and techniques to be used to detect secret trends for these data, providing healthcare professionals with another source of evidence for making prediction decisions on demand (Sun& Wang, 2017).

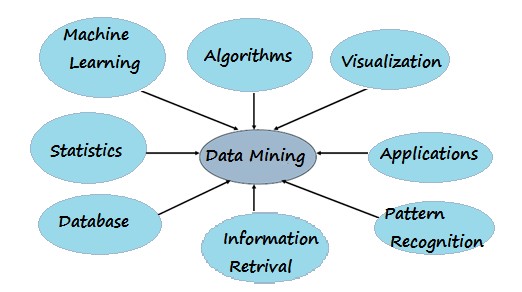
Data mining is an automated way to detect or to infer from vast quantities of data obscure significant patterns. The expression "hidden" refers to patterns which cannot be seen by plain observation. Data mining is an artificial intelligence, computing, machine-learning, information management, data viewing, statistical algorithms, and statistics. This technology offers various methodologies for decision making, troubleshooting, research, integration, mitigation, studying, planning, diagnosis, detection, creativity, forecasts, and estimates. With the need for healthcare organizations to decide on clinical and financial data intensified financial pressure, data mining and review became ever more relevant (Lin, 2020).

Many factors that cause diseases are basically spatial since their distribution and concentration differ depending on their location. Geospatial data mining is a subfield of data mining which aims to find spatial patterns in databases that are interesting and potentially helpful. Geospatial data sets are owned, produced, and managed by organizations that need powerful GIS software to extract data from massive geospatial data sets. The present research project proposes a new advanced data mining approach for private healthcare providers in Ghana (Moore, 2002).

**Data Description**

There is already a huge quantity of data contained in real-world databases. It is growing exponentially, providing an incentive for and the need for semi-automatic methods to find secret information in these databases. If a research operation has been fruitful, the information gathered will support the decision-making process of an organization (Sun& Wang, 2017).

The area of data mining encompasses artificial intelligence, computational science, machine learning, database administration, visualized data, statistical algorithms, and statistics, and is cross-disciplinary. It applies to a range of machine learning, mathematical processing, simulation, and database techniques applicable to different industries. A combination of those techniques is used to find different types of structures and relations in data and to derive rules and models that allow prediction and decision-making in new situations. Classification, estimate, projections, grouping of affinities, classification, definition, and display are all possible. (Lin, 2020).



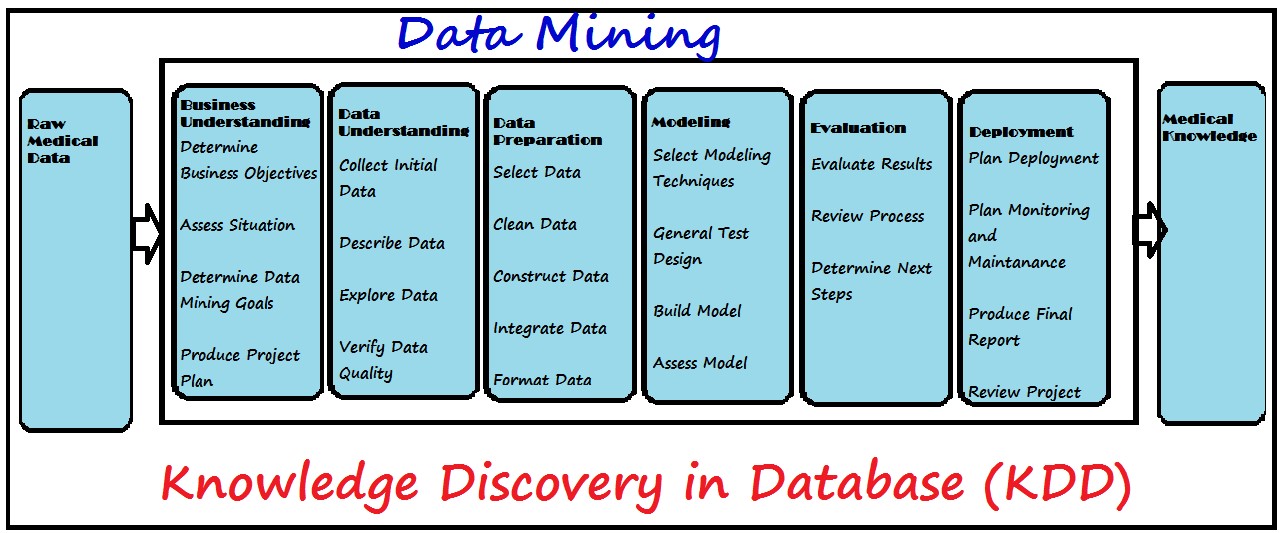
**Figure 1.0 Data Mining Architecture**

Insights into data mining will impact prices, sales and operating performance while ensuring a high degree of service. Data extraction helps healthcare institutions to fulfill their long-term needs; data can be a vital tool for healthcare organizations, but first and foremost it needs to be converted. Data mining tools can support healthcare providers such as hospitals, clinics, doctors, and patients by predicting healthcare demand and preparing growth programs (Lin, 2020).

**Data mining process**

Data mining is a task for pre-treatment, real data mining and post-treatment. During preprocessing, the data mining problems and all data sources are defined, and a subset of data are generated from the accumulated data. The data collection is analyzed to ensure that the noise is eliminated, handled, and transformed properly. The extracted data collection employs a methodology for data mining or a mixture of techniques appropriate for the form of information to be sought. The final stage of the processing is the evaluation and analysis of the information found (Milovic, 2012)

Information discovery is a challenging process for identifying useful and potential advantages between results. It is understood from the figure below that data collection is one of the main processes of exploration. The description of the scholars shows that the use of data mining in several findings is a research method. The word "data mining" steadily substitutes "finding information" for time (Moore, 2002).



**Figure 1.1 Knowledge Discovery**

**DATA MINING STRATEGIES**

Data mining is used to draw on data and two forms of data mining techniques exist: supervised and unattended learning. Supervised learning is the values of the variables (inputs) used to assume another variable (target) with known values, supervised learning methods (or strategically positioned for use in IT context) are applied (Morea, 2019).

In related ways, unsupervised learning methods can be used but are used most often in data that does not have a target of known values. The models and characteristics are interpreted and added to the data for the prediction and discovery of information in supervised methods. Uncontrolled simulation does not know the properties and model of fraud, but the patterns and clusters of data found in data mining contribute to findings (Sun& Wang, 2017).

Demand forecasters may use a range of existing forecasting techniques and processes (Kestin & Armstrong, 2012). These methods can be categorized in 17 distinct areas. The decision is dependent on the unaided judgments, breakups, expert polls, four systematic analogies, philosophy of games, judgmental bootstrapping’s, surveys of intentions and expectations, simulated involvement, collaborative research, experimentation, prediction markets and expert systems. For the remaining five methods, quantitative proof is required. The following section highlights many methods including quantitative data and data mining.

**Judgmental Forecasts**

The forecasts for judgments are based on executive decisions, contracts/insurance/POS industry numbers, consumer settings, mental market projections, insights, outside (consultant) opinions and managers and staff opinion. A health manager can ask his employees for help in creating a judgment forecast or several forecasts to pick (Lin, 2020). **Methods requiring quantitative data**

**Causal Models**

Both models are examples of causal models, originating from segmentation, regression analysis and the index method. If knowledge and evidence are available on variables that may affect the situation of concern, these methods may be useful. Extrapolation of the dependent variable is more accurate than causal model predictions where major deviations are predicted. Theory, prior research, and specialist domain knowledge offer an insight into the relationship between the explanatory variables and the predicted variable. Causal versions are useful when Theory, past research and experience in the field provide insights into the links between explanatory and forecast variables. Causal variants are most useful when:

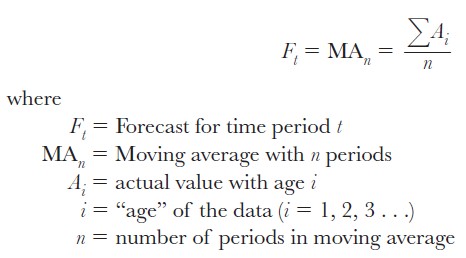
* There are clear causal relationships.
* Significant changes in the causal variables are anticipated over the forecast horizon,
* the causal variables can be reliably forecasted or monitored, especially in terms of direction.

**Segmentation** involves breaking down a problem into individual components, predicting each component by using knowledge and data and combining component forecasts. For example, a hardware company will forecast market sales for each type of product and then aggregate the forecasts. Determining important causal variables to be used to assess divisions and their goals through segmentation. Determine cut-off points for each variable to increase the relationship to the dependent variable, the higher the non-linearity of the relationship, and the more data available, the more cutting points can be used. Provide each segment demographic and population behavior for each segment and incorporate and summarize population and behavioral projections for each segment. As variables interact, variables on demand have non-linear effects and simplistic causal biases exist, segmentation has benefits over study of regression. When errors are likely to be in opposite directions in segment expectations, segmentation is especially useful (Moore, 2002).

**Time-Series Approach**

A time series is a set of measurement over a consistently spaced timeframe at regular intervals (such as daily, hourly, weekly, monthly, or yearly). Monthly hospital visits are an example of a time series. Prognoses based on time series data assume that the future values of the series can be calculated based on previous values. An overview of the time series will show the behavior of the series in terms of design, seasonality, cycles, irregular variance, and spontaneous variations. A trend is a steady and long-term upward or downward transition of results. Seasonality is a phenomenon that explains variations in gene expression at short notice and relatively regularly. Seasonality means relatively frequent, short-term fluctuations in factors such as weather, holidays, and holidays; healthcare facilities also experience "seasonal" differences weekly and often day by day. Cycles are recurrent patterns in data, often linked to current economic conditions, that occur every few years. These systems also have wave-like properties like the business cycle. Irregular deviations are data "spikes" that occur due to fortuitous or irregular conditions (e.g. extreme weather, work strike, the use of new high-tech health services); they do not constitute natural conduct and should be observed and disposed of as soon as possible. Random anomalies are those that remain after analysis of all such behaviors. The data can be graphed to assist a health care manager in choosing the right prediction process. Methodologies Averaging Noise (random variation), which distorts patterns in data, is popular in historical data. Randomness is the product of a slew of minor factors that are not certain to be expected. The optimal scenario would be to rule out all randomness, leaving only "true" variations (for example, changes in the level of patient demand). Unfortunately, it is usually impossible to differentiate between these two kinds of combinations. Minor variations are used as random variations which are essentially 'smoothed' out of the data set during the use of averaging techniques. Whilst greater differences represent "real" changes, they are smoothed, in a lesser way. In this section we will consider three related average techniques: average changes, exponential smoothing, and naive forecasts (Milovic, 2012).

**Moving Averages** (MA). A rolling average forecast uses various the most recent real data values, whereas a naive forecast uses data from the last time. The following equation is used to calculate the moving average projection.



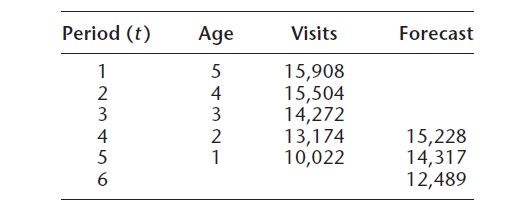
For budgeting purposes, an OB/GYN office has the following yearly patient visits and would like to forecast the operation rate for the next year.



*Solution:* Using formula [2.1], the three-period moving average (MA3) for period 6 is



A health care strategist uses the available data to retrieve forecasts previously; this is a valuable tool for calculating prediction precision. It is how 3-period moving averages are determined for OB/GYN visits.



The name of this technique derives from a revision of the forecast, by adding latest values, subtracting the oldest and recalculating the average if new actual values become available.

As a result, the projection "moves" with only the latest prices. For example, visits from periods 1 to 3 were paired with an expected value of 15,228 for period 4 (F4); visits from time one was reduced to F5, but period 4 was added to the average. Moving averages would be smoother with certain data points but would be less likely to be "absolute" moves. The chance of a slower response to data changes must be balanced against the expense of reacting to random variations.

**METHODOLOGY**

This section provides a brief overview of the research design that was chosen. It explains exactly what is to be done and how it can be done.

**Research Site**

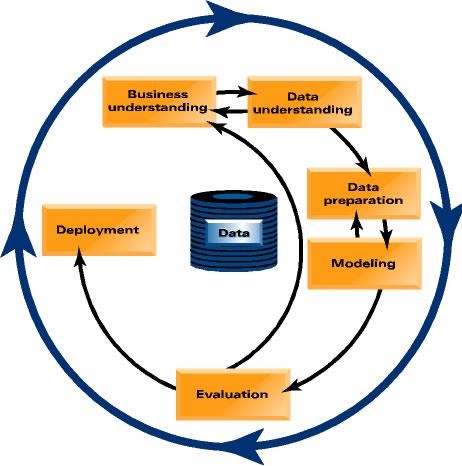
Ghana is a western African nation located on the Gulf of Guinea's coast. Despite its limited size and population, Ghana is one of Africa's most powerful nations, partly due to its vast natural resources and partly because it was the first black African nation south of the Sahara to gain independence from colonial rule.

Different factors were considered when selecting this country as the study site, including access to information on health demand and health resources offered in the healthcare facilities. In addition, all planning issues raised in this country are relevant to the hospitals of the country.

**Research Design**

Data mining is a groundbreaking approach that requires a variety of skills, experiences, and resources. It includes a standard approach that helps convert industry problems into data mining operations, proposes appropriate data transitions and data mining techniques and provides tools to evaluate the viability of results and to document experience.

The CRISP-DM describes a process model that provides basis for the implementation of data mining programs, regardless of industry and technology. The CRISP DM method model aims to reduce the expense of large data mining programs, their reliability, their replicability, their management and their speed. This is the method of preference in this initiative. The data mining reference model of CRISP-DM provides an overview of the life cycle of a data mining project. Which covers the phases of a project and its activities and outputs. The life cycle of a data mining project is split into six phases, as seen in the diagram below. Arrows only indicate the most important and routine interphase dependencies; in addition, in each procedure, each stage outcome determines which stages, or which particular operation of a phase, must be performed next. The outer circle of the figure reflects the cyclical nature of data mining. The data mining method does not stop until a solution is implemented. The lessons learned from the method and the solution applied will lead to modern, more market-oriented problems. The previous data mining processes will help future process for data mining (Sun& Wang, 2017.)



**Figure 1.3 The life cycle of a data mining project**

In the following, I outline each phase briefly:

This method is designed to achieve a business awareness of the project objectives and requirements and translate this insight into a definition of data mining and a tentative project plan to reach its targets.

* Recognize the company's goals.
* Assessing Current Systems, Assessing Current Systems,
* Task Decomposition,
* Identifying Constraints,
* Creating a Project Schedule

**Data Understanding**

The data sensitization process starts with the compilation of data. It works to familiarize itself with the data, find problems of data consistency, gain early insights into the data, or recognize interesting sub-sets to form hidden intelligence theories. Inextricably related are the two principles of market and data comprehension. At least to formulate the topic of data mining and the project approach, some understanding of the data available is needed.

* Collect Data
* Data Description
* Data Exploration

**Data Preparation**

The data collection includes all activities leading to the final data set (data being fed to the simulation tool(s)) being built from the raw data. Data processing tasks would probably be replicated several times and in no specific order. All tasks are table, record, and attribute compilation, data cleaning, construction of attributes, and modeling data transformation

* Data Transformation
* Clean Data
* Data Construction
* Integrate Data
* Select Data

**Modeling**

During this method, various modelling techniques are selected and applied and their parameters are optimized. There are typically several strategies for the same form of data mining query. Any approach requires the use of dynamic data formats. Inextricably related are data processing and modelling. Sometimes, you find data problems while modeling or get ideas for data creation.

* Build Model
* Assess the model
* Select the appropriate modeling technique
* Develop a testing regime

**Evaluation**

From the data interpretation point of view, we have built one or two models at this point in the project that seem to be of high quality. Until proceeding with the final roll-out of the framework, it is important to assess the concept more in depth and review the steps to build it to ensure it meets market targets. One of the key objectives is to see if a critical business challenge has been solved insufficiently. A decision should be taken at the conclusion of this phase about how to use the results of data mining.

* Review Process
* Determine next steps
* Validate Model

**Deployment**

The completion of the concept is not necessarily the end of the project. The knowledge collected must typically be arranged and distributed in a way that the customer understands. Depending on the specifications, the implementation stage can be as simple as reporting or as difficult as using repeatable data mining methods. In certain cases, the client will be implemented rather than the data scientist. In any event, it is necessary to know in advance what steps need to be taken to use the templates created.

. • Process deployment/production

• Produce final project report

• Knowledge Deployment is specific to objectives

**Overview of WEKA**

Waikato Environment for Knowledge Analysis is a free and open-source platform for machine learning algorithms and data mining activities.

WEKA includes data pre-processing, sorting, Decision Trees, Artificial Neural Networks, data mining, statistical analytics, and logic regression, as well as a collection of visualization methods. This would be the project's preferred tech tool.

**How WEKA Works**

The GUI user can choose from four different ways to work with WEKA. Explorer, Experimenter, Knowledge Flow, or a basic CLI are the four options. Explorer would be the tool of choice to work within this project. WEKA accepts data in the ARFF format (Attribute-Relation File Format), an ASCII text file that represents a list of instances that share a collection of attributes. This file is divided into two sections: header information and data information. The header includes the relation's name and a list of attributes and their data types. Users can access data from WEKA using JDBC from Windows databases like MS Access or other databases in general (MySQL, PostgreSQL, MS SQL Server, Oracle, and so on). WEKA can be accessed from any of the three interfaces mentioned below: command line, graphical user interface, or java programs.

**Data Collection Methods**

The various methods to be used includes

1. Interviews

Online interviews were used to gather market participants' in-depth perspectives, viewpoints, and anecdotal proof. The interviewer will take time and time to conduct interviews and equipment to record and transcribe interviews. There were also extensive interviews in which the interviewer did not hold to several questions. The interviewer encouraged transparent and truthful responses and the compromise could exist between a wide range of topics and a thorough review of a smaller set of issues. His interviews will also enable participants to articulate a highly desirable approach to quality data collection in their language. This would allow the evaluator to present the importance of the encounter from the point of view of the respondent. This can be applied to medical practitioners in the field.

1. Document Studies

Existing documents can reveal something about a location and a group of persons that cannot be observed or noted in any other way. This material would be available in the form of text. Established sources' utility can vary based on how available and reliable they are.

**Analysis**

The data mining and information discovery processes used in this study aim to predict the demand for health services in Ghana to plan health administrators and planners. This chapter focuses on evaluating and discussing the model's outcomes. The model's effects are tested and checked. They're often compared to the findings of other market forecasting methods used in the healthcare industry.

**System Evaluation**

The evaluation of the system involved testing of the model that was implemented and the results from WEKA:

1. Unit testing

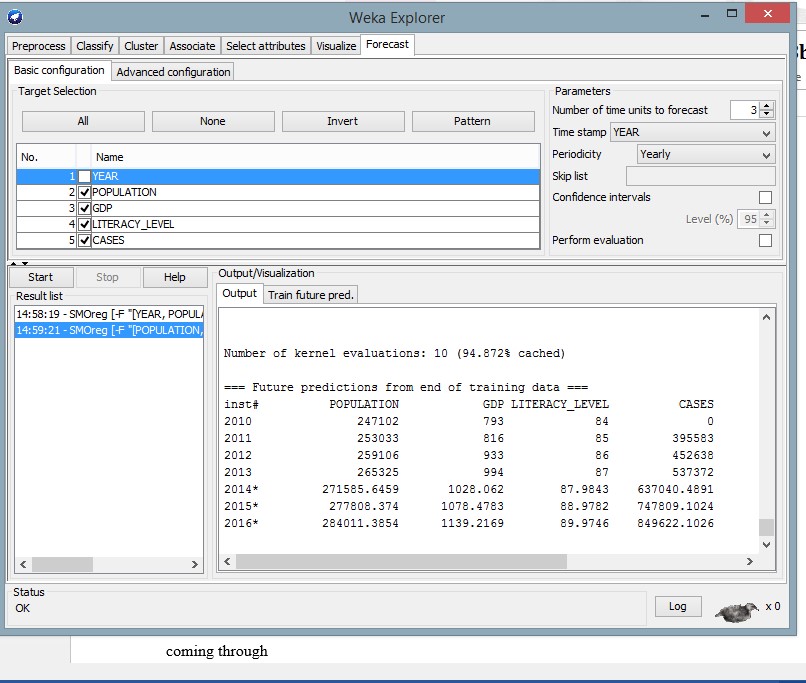
2. System testing.

**Unit Testing**

This testing level was designed to ensure that various subsystems could bind to databases and read and manipulate data.

|  |  |
| --- | --- |
| Task no. | Description |
| 1 | Running WEKA and Forecasting Package |
| 2 | Connection SQL Server Database Server from the model and WEKA through ODBC |
| 3 | Reading Data into WEKA from the database |
| 4 | Performing a Forecast |
| 6 | Running of Decision Support Reports |

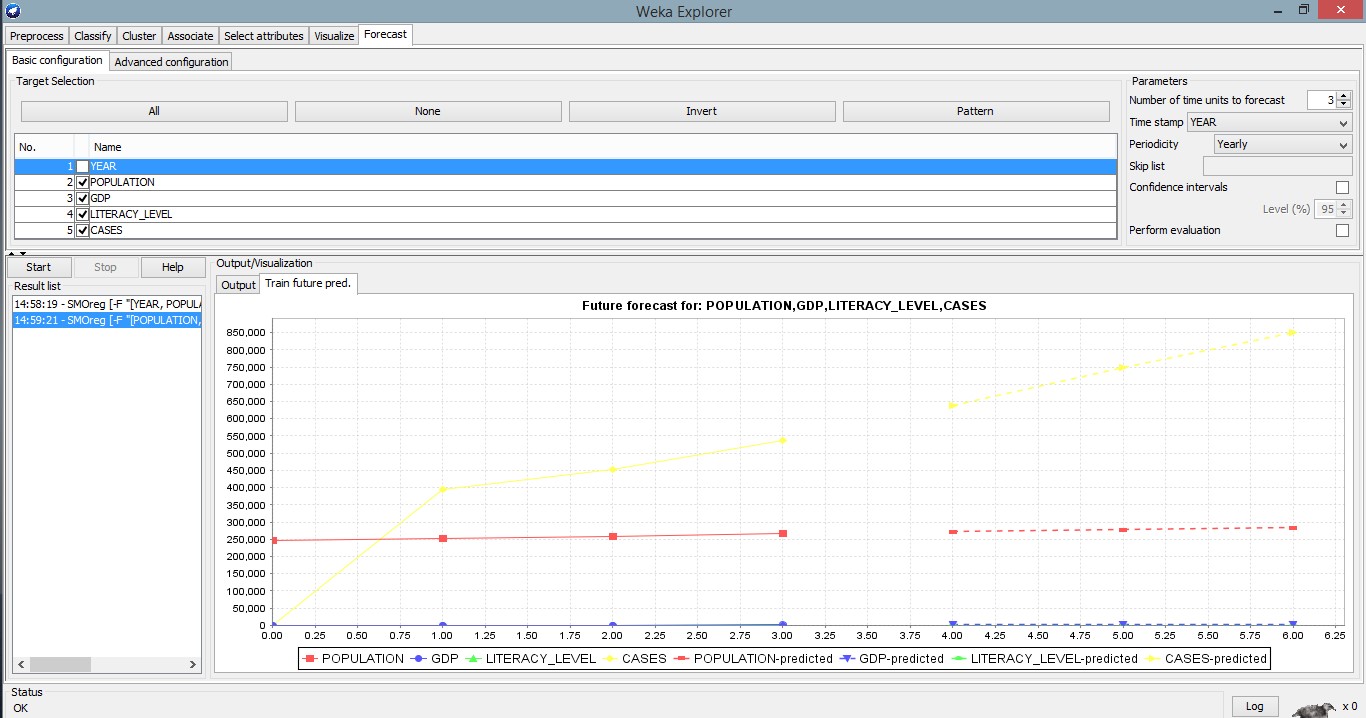
**Tasks in Unit Testing**



##### Figure 1.4: WEKA Results explorer System testing

**System testing**

Device maintenance was performed to ensure that the system was performing as expected and producing verifiable outcomes. The predictions were being tested against the data to ensure that the correct production was being generated.



**Figure 1.5 WEKA Forecast Graphs**

**Deployment**

Any private hospital in Ghana that wants to use the demand forecasting feature and decision support mechanism in this project will use this solution. The deployment phase will range from a basic report generation to a sophisticated application of a repeatable data mining method and real-time integration of data sources in specialized servers and network networks, depending on the hospital's goal. However, for this study's purposes, the prototype was installed on a single laptop for presentation and to design the final project report.

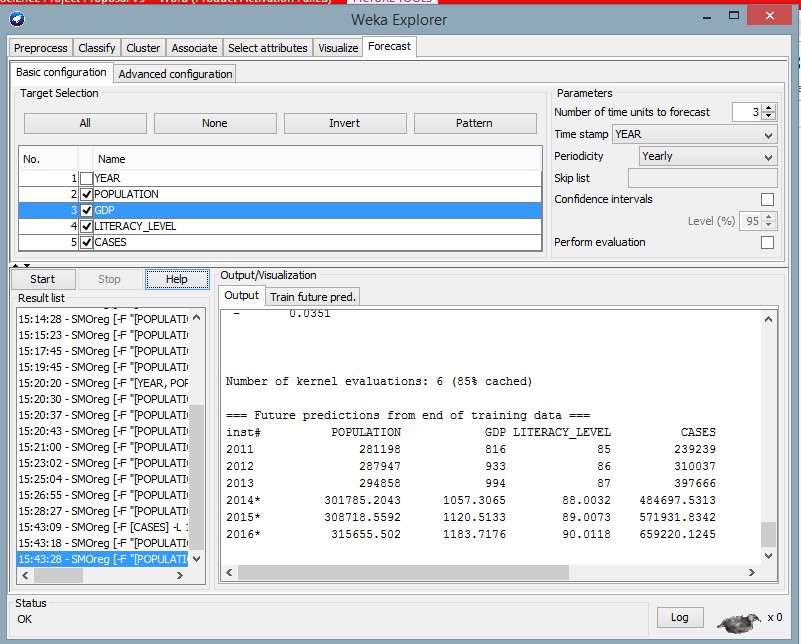
**Demand Forecasting Results and Discussions**

The effects of the prototype's development are described and discussed in this section. The aim was to see how the prototype met the system's functional specifications and if the findings could be trusted to make a decision. The outcomes are divided into three categories: WEKA's Forecasting Package allowed for forecasting several variables. Compared to standard linear forecasting methodologies used in other tools, this vastly increased the results' dependability.

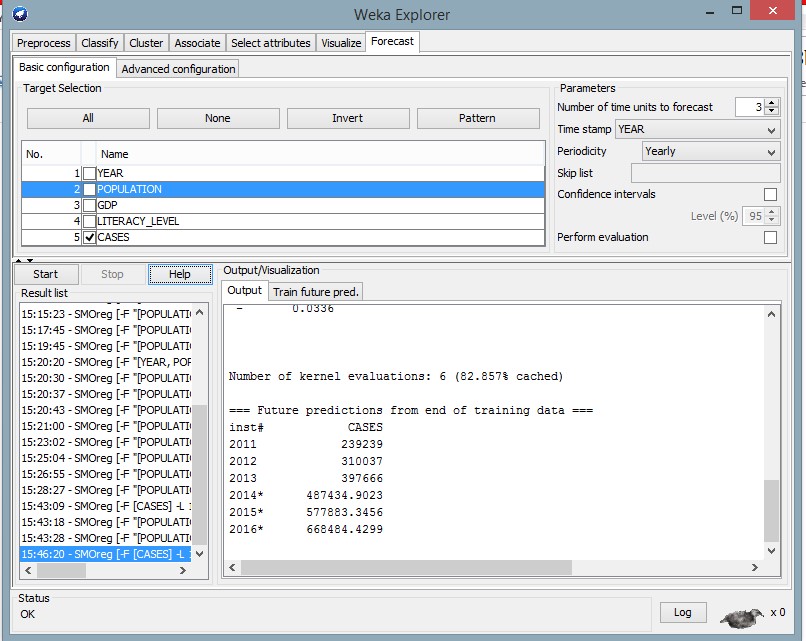
**Factors affecting the forecast results**

* Income – The outcomes vary if the GDP stays stable versus if power increases marginally. This demonstrated that the forecasting process accounted for several factors in the forecasted outcomes.
* Population and population growth – The demand for overall outpatients was influenced by population fluctuations and growth, as seen in the forecast results.
* Average Literacy Levels - This affects the findings because it dictates the average ability to go to private clinics for treatment.

Suppose the prediction is made without taking into account the values of these parameters. In that case, the forecast seems to be based on a linear forecasting model, which does not provide reliable forecast results because it only considers the time factor.



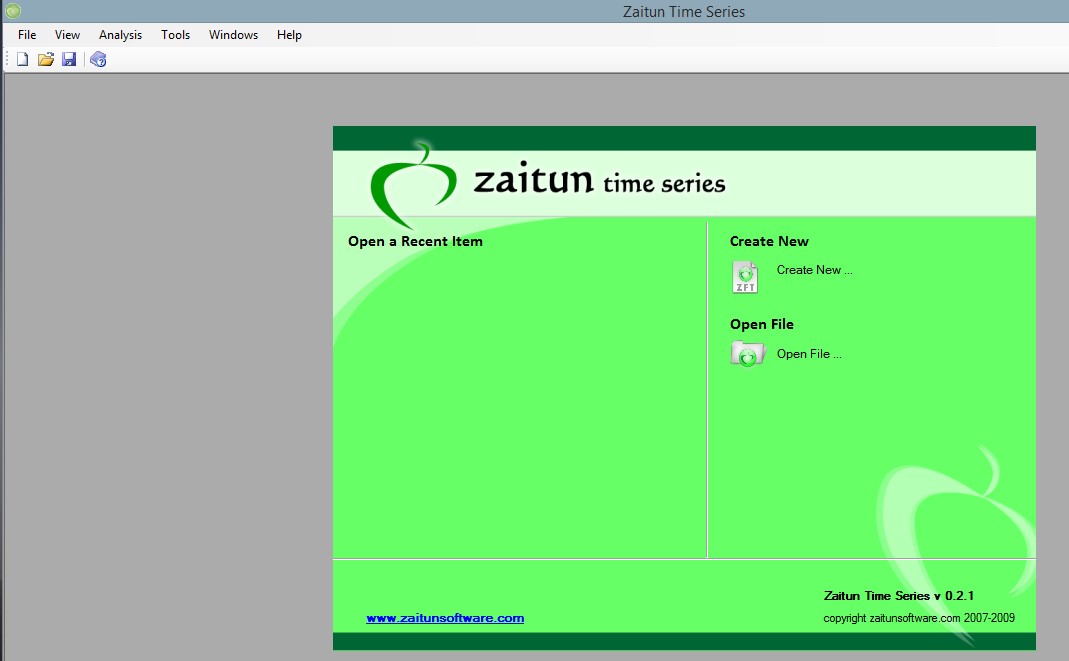
**Figure 1.5 OP Results with all demand factors variations**



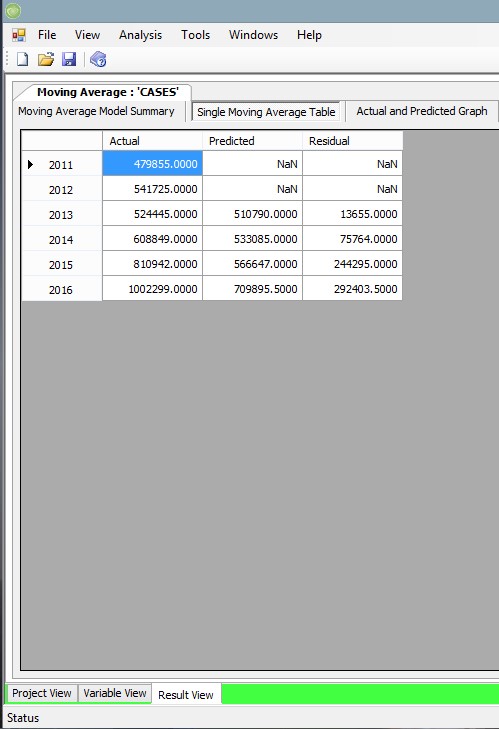
**Figure 1.6 Results without Demand Factors Variations**

**Performing comparison with other forecasting methods**

The results were compared to the performance of time series forecasts performed with the ZAITUN Time Series method, an open-source tool for performing linear forecasts with Time Series data.



**Zaitun Time Series Software Interface**



**Zaitun Forecasts**

**CONCLUSIONS & RECOMMENDATIONS**

Today's healthcare administrators and planners in Ghana must make future healthcare service delivery decisions without understanding what will happen. This is due to the lack of an insightful and sophisticated infrastructure capable of mining big data and generating analytical patterns on healthcare demand forecasting in Ghana. Forecasts will enable managers to plan. Forecasts will enable managers to forecast potential demand and make appropriate plans.

The large and complex data in Ghana's private medical facilities have not been properly used by market analysis and forecasting facilities, thus assisting in decision-making. The majority of hospitals in the country do not employ any forecasting system to help them schedule their care instead of opening their doors to the sick. Many who use forecasting do so using a linear mechanism mostly rooted in Excel, which does not consider the multiple variables influencing customer demand for their services. They look at the figures and predict future numbers, which are rarely accurate forecasts, but they do not use them to make decisions. They do not have a mechanism to assist with planning and make decisions on potential market forecasts.

This project's key goal was to look at data mining to forecast demand for private health services in Ghana. We set out to investigate the different forecasting methods in healthcare services demand, determine the most appropriate healthcare services demand forecasting methods for Ghana and then build and validate a data mining model for forecasting private healthcare services demand in Ghana.

In the real world, successful information discovery methods and data mining methods are at the heart of constructing any strategic research and forecasting models from big data in various industries. WEKA, a data mining and forecasting method used in this project and prototype development, has proven to be extremely effective in forecasting from big data in the medical field. The method can perform complex tasks such as forecasting future patterns by combining several variables and calculating them appropriately to arrive at forecast estimates. Aside from predicting practices, the method has a slew of other features that were not included in this study, such as clustering, classifications, and comparisons, all of which will help discover new medical significant data trends and aid decision-making.

WEKA's Forecasting Package allowed for the forecasting of several variables. Compared to standard linear forecasting methodologies used in other tools, this vastly increased the results' dependability. The factors that influence demand for private health care services in Ghana were examined, and they were found to affect the forecasted outcomes. Income, population, and literacy rate are three of these variables. When forecasting without considering these variables' values, the forecast seemed to use a linear forecasting model with a greater margin error than when all factors were used.

The predicted results must seem to be precise and practical to be trusted. Forecasting of real known data was used to validate this model. We set out to predict the outcome about 2019, for which the results were known, and we had data from 2016 to 2019. The model predicted the actual outcomes with a marginal error of 7.9%, projected to decrease as more training data is collected. Testing if the scheme considered several variables – Since, as mentioned above, many factors influence the demand for health care in private hospitals, this was also put to the test. The findings revealed that if these variables were removed, the results differed from what was anticipated. The findings found that when these variables were removed from the equation, they differed from when they were added.

As a result, using data mining to predict demand for health care is the most reliable and optimal method of creating a forecasting model with multiple variables. Using a data mining method to develop a healthcare demand forecasting model in Ghana was the project's fundamental goal. The project proposed and adopted a prototype for forecasting demand for health care in private health facilities focused on data mining. Demand for public facilities services is influenced by conditions separate from those that influence demand in private facilities. As a result, the study was narrowed to only look at private facilities. This is where all of the variables that influence demand, such as buying capacity, population increase, expenditures and equipment placed in place, time limits, and the class effect, will play a role.

As a result, this technology is a very useful invention that private healthcare planners, suppliers, and administrators will use to predict potential demand and accordingly prepare their facilities in terms of equipment, new locations, workforce training and management, and cash flow.

**Recommendations**

This model's accurate operation is highly dependent on the central availability of accurate records for all medical problems handled by all hospitals in the country. My advice to the government and the Ghana Health Information System implementers in the Ministry of Health is to collect all data from all hospitals around the country to improve forecasting models' performance based on those data. This would help the further developments in this model and provide more complex situations with Ghana's well-being. Data exchange among all involved parties should be improved so that data used to help any potential works can be obtained.

This project's prototype was created with Microsoft Visual Studio 2012 and a SQL Server database, while WEKA was introduced on the JAVA platform. As a result, the WEKA forecasting classes could not be viewed directly from the prototype, and forecast data had to be manually entered into the database. This limits the number of possibilities that can be tried inside the model outside of WEKA if such data is not available. To facilitate a smooth integration of WEKA with the model, I suggest that this model be developed in JAVA in the future. Alternatively, you can use JAVA bridges to run JAVA classes from Visual Studio.

# **APPENDICES**

## **Appendix A – Contribution of each group member**

**Ayaz Rathod**

This member was responsible for conducting online interviews with various stakeholders in the healthcare industry. This included ministry executives, doctors, and patients. Hershel also conducted document studies from multiple sources, majorly from the Ghanaian library, on the working of the Ghanaian health system

**Devam Patel**

This member was involved in data cleaning, which involved removing outliers from the data that the previous member gathered through online interviews and document studies

**Junaid Asim**

After data cleaning by the previous member, this member was responsible for data representation, which involved coming up with relevant data that could be loaded into our training models.

**Parth Patel**

This member was responsible for loading the data gathered and cleaned into our modeling tool, which was WEKA; he was also responsible for training the models to come up with a predictive forecast for the private health care demands in Ghana

**SM Pantho**

This member was responsible for the documentation of this project report at every stage.

**References**

Cios, K. J., & Moore, G. W. (2002). The uniqueness of medical data mining. *Artificial intelligence in medicine*, *26*(1-2), 1-24. <https://www.sciencedirect.com/science/article/pii/S0933365702000490>

Lin, A. X., Ho, A. F. W., Cheong, K. H., Li, Z., Cai, W., Chee, M. L., ... & Ong, M. E. H. (2020). Leveraging machine learning techniques and engineering of multi-nature features for national daily regional ambulance demand prediction. *International journal of environmental research and public health*, *17*(11), 4179. <https://www.mdpi.com/1660-4601/17/11/4179>

Milovic, B., & Milovic, M. (2012). Prediction and decision making in health care using data mining. *International Journal of Public Health Science (IJPHS)*, *1*(2), 69-78. <https://www.academia.edu/download/33572666/1380-4163-1-PB_(1).pdf>

Moro Visconti, R., & Morea, D. (2019). Big data for the sustainability of healthcare project financing. *Sustainability*, *11*(13), 3748.

<https://www.mdpi.com/2071-1050/11/13/3748>

Sun, W., Cai, Z., Liu, F., Fang, S., & Wang, G. (2017, October). A survey of data mining technology on electronic medical records. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1-6). IEEE. <https://ieeexplore.ieee.org/abstract/document/8210774/>