IJCSN International Journal of Computer Science and Network, Volume 4, Issue 4, August 2015
ISSN (Online) : 2277-5420 www.IJCSN.org
Impact Factor: 0.417

Strategic Management System for Effective Health Care Planning (SMS-EHCP)

[1] **O. I. Omotoso,** [2] **I. A. Adeyanju ,** [3] **S. A. Ibraheem** [4] **K. S. Ibrahim**

[1,2,3,4] Department of Computer Science and Engineering, LadokeAkintola University of Technology, Ogbomoso, Oyo State.

Abstract - Ranging from primary health care institutions to the big health care centers, every healthcare organization uses an information system which stores, processes and retrieves healthcare data. Healthcare data stored electronically are used to improve healthcare planning. Healthcare centers and academic centers also use these data for education and research. Data Mining is the process of finding correlations or patterns among dozens of fields in large relational databases storing large set of transactional data. It basically has four modules: Classification, the grouping of data in predefined classes; , clustering, the association of data in classes; Regression, which models the data of the minimum error and Association, which find relationship among data objects.This research paper presents our findings on the strategic management of health care system for effective health care planning and advancement it has brought to the health care sector.

Keywords - *SMS-EHCP, strategic, Management, Health, Care, Planning, Data and Mining.*

1. Introduction

From primary health care institutions to the big health care centers, every healthcare organization uses an information system. These healthcare information systems (HCIS) store, process and retrieve healthcare data. Healthcare data is very valuable in today's world. By the help of rapidly developing healthcare informatics, there are efforts to use the valuable data stored electronically in HCIS databases to improve healthcare. Healthcare giving staff expects more than e-recording the data from HCIS. Besides using healthcare data for care giving, healthcare centers and academic centers use these data for education and research. Knowledge discovery in medical area is not limited to healthcare development such as developing new healing techniques and drugs, but critically important for medical planning for effective medical decisions (Ameraican Medical Association, 2010). When compared to human brain, computers are well suited to making rapid calculations and detecting hidden facts, facilitating decision networks that support near limitless complexity. Any educational (knowledge discovery) area, dealing with huge and valuable data, such as medical domain, requires creative techniques supported with computers and computer systems to utilize it. Data mining techniques are good examples for these required creative techniques (Ameraican Medical Association, 2010).

There are many ways to define a data warehouse (DW) due to its widespread adoption; a good working definition of a DW is a dedicated computer system or database that consolidates subject-oriented, time-variant, and non-volatile data from multiple sources to support decision-making processes. Recently, DWs have become invaluable resources in various domains, and they are used to analyze trends over time or to extract valuable information.

GoguSandeep defined data mining as discovering new patterns which are unknown before, statistically reliable and process able from data. Data mining is a field which is concerned to understanding data patterns from huge datasets. We can say that the aim is to find out new patterns in data. A number of data mining techniques are there like classification, clustering, advanced neural networks, prediction and regression models used for different data mining approaches in various areas.

2. Literature Review

A survey separately carried out by J Med (2001) and Kerkri E. et al (2010) indicated that the adoption rate of Data Warehouses (DWs) in Clinical and Translational Science Award (CTSA) institutions has increased from 64% (18 of 28 institutions) in 2008 to 86% (30 of 35) in 2010. DWs in hospitals, which are usually called clinical data warehouses (CDWs) are used for various purposes, including administration, management, clinical practice, and research. These can be categorized as either conventional usage or hospital-specific usage.

Conventional usage includes administration, operation, and management. Therefore, such a DW in a hospital is usually called an enterprise DW. It is an earlier type of DW in hospitals. The hospital-specific usage consists of clinical practice, quality improvementand biomedical research. However, research usage cannot be efficiently supported by conventional DW technology due to the complexity and heterogeneity of clinical and research data (Livne O E et al. Federated Querying Architecture with Clinical & Translational Health IT Application. (2011)).

Existing data stored in a hospital's transactional servers/records have enormous potential to improve performance measurement and health care quality. Accessing, organizing, and using these data to support research and quality improvement projects are evolving challenges for hospital systems.Development of a clinical data warehouse involved importing data from the information systems of at the selected hospital, mining the data and critically analyzing the data to discover hidden knowledge in the data stored by the hospital management. This knowledge, when implemented, helps the hospital management, their patients and even the government of the land.

3. Collection of Data

Electronic data sources do not encompass all the datanecessary for analysis. Data used for this work were manually collected from the aparticipating Health Care Center. The data were then stored into the database for easy access, which can be fetch and stored on the local system in an excel format for analysis using WEKA Analytic tool.

Criteria

Season-Ailment Criteria was used to analyze the data set. This enabled the discovery of how often (proportion by which) an ailment occurs in a particular season of the year.

Analysis

Figure 1 below is the WEKA Preprocess Window displayed when the data set is loaded to the WEKA analytic environment. It consists of the Dataset name which is denoted by ***Relation***, ***Instances*** denotes the numbers of instances of individual record considered.

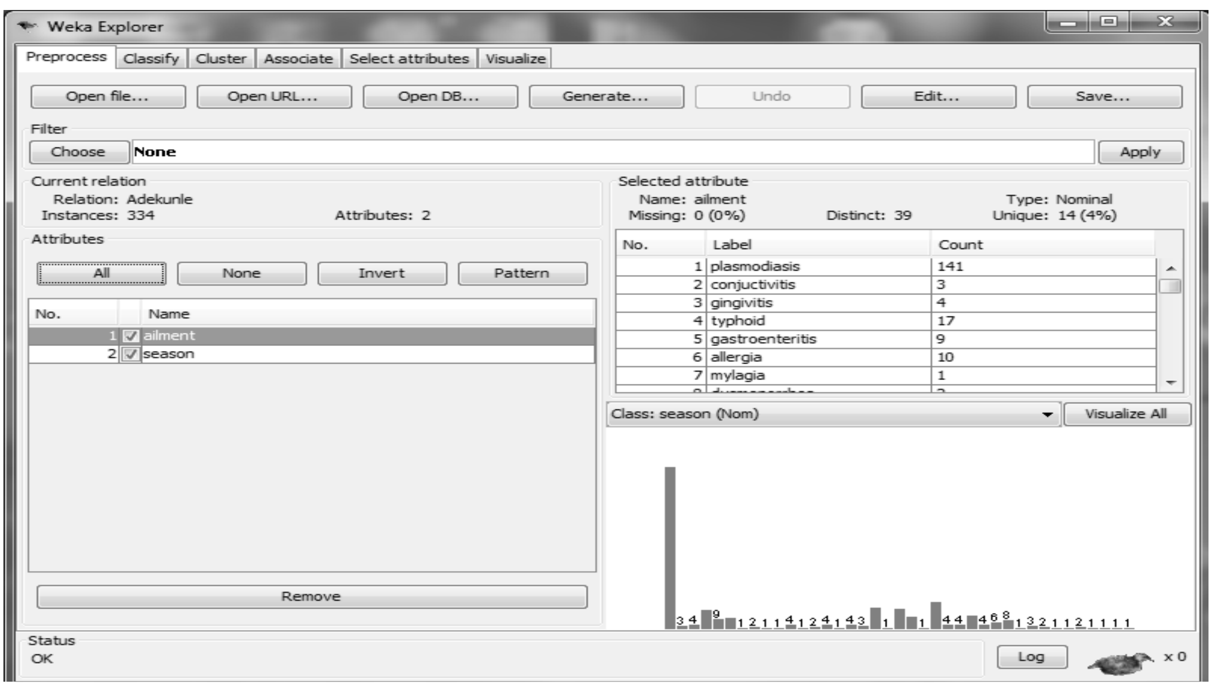

Figure 1: WEKA Preprocess Window

While ***Attributes*** denote the numbers of field making the column of the Record. The data were analyzed with ZeroR classification algorithm for obtaining the degree of occurrence of each ailment per season. After the analysis of the data set with the *ZeroR* classification algorithm of the WEKA analytic tool, the result obtained under different testing options are as shown in figure 2,figure 3, figure 4 andfigure 5.

When the dataset was analyzed with the training set test option,it was discovered that***Plasmodiasis***is the most occurring ailment for ***Rainy season*** with **59.58%** occurrence, as shown in figure 2.

```
=== Classifier model (full training set) ===

ZeroR predicts class value: Rainy

Time taken to build model: 0 seconds

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         199               59.5808 %
Incorrectly Classified Instances       135               40.4192 %
Kappa statistic                          0
Mean absolute error                      0.4818
Root mean squared error                  0.4907
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              334

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         1          0.596     1         0.747      0.5      Rainy
                 0         0          0         0         0          0.5      Dry
Weighted Avg.    0.596     0.596      0.355     0.596     0.445      0.5

=== Confusion Matrix ===

   a   b   <-- classified as
 199   0 |   a = Rainy
 135   0 |   b = Dry
```

Figure 2: ZeroR classification algorithm (Training test set)

To verified the accuracy of the analysis. The dataset was also analyzed with supplied test set option which also predicted ***Plasmodiasis*** as the most occurring ailment for ***Rainy Season*** with **59.57%** occurrence as depicted in figure 3.

```
=== Classifier model (full training set) ===

ZeroR predicts class value: Rainy

Time taken to build model: 0 seconds

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances         199               59.5808 %
Incorrectly Classified Instances       135               40.4192 %
Kappa statistic                          0
Mean absolute error                      0.4818
Root mean squared error                  0.4907
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              334

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         1          0.596     1         0.747      0.5      Rainy
                 0         0          0         0         0          0.5      Dry
Weighted Avg.    0.596     0.596      0.355     0.596     0.445      0.5

=== Confusion Matrix ===

   a   b   <-- classified as
 199   0 |   a = Rainy
 135   0 |   b = Dry
```

Figure 3: ZeroR classification algorithm (Supplied test set)

Also, Percentage Split (with 66%) was used to analyze the data set. *ZeroR* classification algorithm predicted ***Plasmodiasis*** as the most occurring ailment with **58.94%** occurrence in the ***Rainy season*** as depicted in figure 4.

```
=== Classifier model (full training set) ===

ZeroR predicts class value: Rainy

Time taken to build model: 0.01 seconds

=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances          73               64.0351 %
Incorrectly Classified Instances        41               35.9649 %
Kappa statistic                          0
Mean absolute error                      0.4798
Root mean squared error                  0.4847
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              114

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         1          0.64      1         0.781      0.5     Rainy
                 0         0          0         0         0          0.5     Dry
Weighted Avg.    0.64      0.64       0.41      0.64      0.5        0.5

=== Confusion Matrix ===

  a  b   <-- classified as
 73  0 |  a = Rainy
 41  0 |  b = Dry
```

Figure 4: ZeroR classification algorithm (Percentage split)

Cross Validation testing option predicted ***Plasmodiasis*** for ailment with **58.97 %**, while ***Rainy*** is predicted for season as depicted in figure 5.

```
=== Classifier model (full training set) ===

ZeroR predicts class value: Rainy

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         199               59.5808 %
Incorrectly Classified Instances       135               40.4192 %
Kappa statistic                          0
Mean absolute error                      0.4818
Root mean squared error                  0.4908
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              334

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                 1         1          0.596     1         0.747      0.488    Rainy
                 0         0          0         0         0          0.488    Dry
Weighted Avg.    0.596     0.596      0.355     0.596     0.445      0.488

=== Confusion Matrix ===

   a   b   <-- classified as
 199   0 |   a = Rainy
 135   0 |   b = Dry
```

Figure 5: ZeroR classification algorithm (Cross Validation)

The prediction is *perfect*in the sense that the analysis attained 59.58% classification accuracy. Even, this may seem unreliable to validate the correctness of the classifier, especially when the dataset is unbalanced. But the correctness of the prediction was upheld with the error values obtained. That is, all the errors (Mean Absolute Error, Relative absolute Error, and Root Relative Absolute Error) tend toward zero. The result of the analysis is summarized in table 3.1

Table 3.1: Ailment, Season and Percentage Relation table

Ailment	Season	% of occurrence in a season
Plasmodiasis	Rainy	59.58
Asthma	Dry	14.34
Pneumonia	Rainy	5.31
Cholera	Rainy	4.97
Gastroentritis	Rainy	3.04
Dysentry	Rainy	2.16
Sore Throat	Dry	2.11
Acne	Dry	2.05
Tines Vesicular	Dry	1.50
Boil	Dry	1.12
Anthritis	Dry	0.52
Anthics	Dry	0.26

The above table is therefore summarized with the chart shown in Figure 6

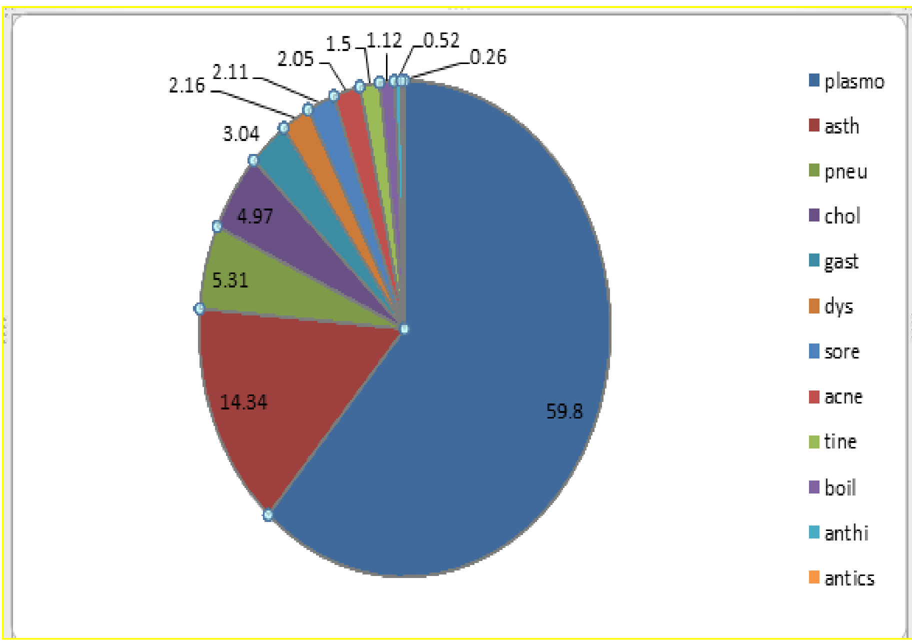

Figure 6: Ailment, Season and Percentage Relation Pie Chart

4. Results Implication

With the results obtained from the analysis, the following implications were derived;

Plasmodiasis is predicted as the most occurring ailment being treated in LadokeAkintola University Health center. It occurs highly in the *Rainy Season* with an accurate occurrence oat 59.58%. *Asthma*is predicted as second the most occurring ailment being treated in the health care center. It occurs much in the *dryseason.* It occurs at 14.34%.

5. Conclusion

For each characteristic, analysis of how the results vary whenever test mode is changed was noticed. Measure of interest includes the analysis of classifiers on the datasets, the results are described in value of prediction made by Zero R classification algorithm(for dataset with nominal class value), correlation coefficient (for dataset with numeric class value), mean absolute error, root mean squared error, relative absolute error, root relative squared error after applying the Training set test, Supplied test set, Cross-Validation, and Percentage split methods. A datasets (Medical Record) have nominal class value. All the testing methods predicted ***Plamodiasis*** for ailment and ***Rainy*** season was predicted as matching season for the predicted ailment.

References

[1] J Am Med Inform Assoc.2010 Mar-Apr;17(2):131–5.[PMC free article][PubMed].

[2] Verma R, Harper J. Life cycle of a data warehousing project in health care. J Health c In f Manag.2001 Summer;15(2):107–17.[PubMed].

[3] Kerkri EM, Quantin C, Allaert FA, et al. An approach for integrating heterogeneous information sources in a medical data warehouse.J Med Syst.2001 Jun;25(3):167–76.[PubMed]

[4] Zhang M, Zhang H, Tjandra D, Wong ST. DBMap: a space-conscious data visualization and knowledge discovery framework for biomedical data warehouse.IEEE Trans InfTechnol Biomed.2004 Sep;8(3):343–53.[PubMed]

[5] Livne OE, Schultz ND, Narus SP. Federated Querying Architecture with Clinical & Translational Health IT Application.J Med Syst.2011 Oct;35(5):1211–24.Epub 2011 May 3.[PubMed]