مجلة الدلتا الدولية للعلوم التجارية ونظم المعلومات

Journal Home page https://dhic.edu.eg/DIJCSIS/

A Proposed Framework to Enhance Healthcare Systems Utilizing Data Analytics for IDSS

[1,2]Asaad Elsaadany Bedeir, and [3]Mohamed M. El Hadi

[1]Ph.D. Candidate in Sadat Academy for Management Sciences (SAMS), Department of Computer Sciences and Information Systems Sadat Academy for Management Sciences, Cairo, Egypt

[2]Department of Computer and Information Systems, Delta Higher Institute for Computers, Egypt

[3]Professor of Computer and Information Systems Department, Sadat Academy for Management Sciences (SAMS), Cairo, Egypt

Email: [1]asaad@dhic.edu.eg

Email: [3]mohamed.m.elhadi@gmail.com

ABSTRACT

Health Informatics is one of the quickly growing fields that focus on using Computer Science and Information Technology for health data. The fast digitization of these vast volumes of data is the present trend. These massive amounts of data hold the promise of supporting a wide range of medical and healthcare functions, allowing us to derive hidden patterns and insights from data to address many new and important questions. These massive quantities of data are driven by mandatory requirements and the potential to improve the quality of healthcare delivery while reducing costs. As depending on data analytics (DA) has become critical in the age of digital transformation, the Egyptian health sector still uses descriptive reporting and traditional statistical analysis for planning.

Therefore, to plan and make decisions, integrated healthcare systems must rely on DA and intelligent decision support systems (IDSS). A clear approach to developing an intelligent healthcare system is to go toward DA. The use of DA not only supports cost-saving, but also offers more accurate and fast information to help with decision-making and to achieve strategic goals. The main goal of this thesis is to build a proposed framework for healthcare data analytics to help healthcare planners use data analytics and data-driven conclusions for Intelligent Decision Support Systems (IDSS).

In this thesis, a descriptive and analytical method has been used to study and analyze the existing healthcare systems for a successful move towards applying the proposed data analytics framework. The proposed framework of healthcare data analytics for IDSS consists of five layers: data layer; data aggregation layer; data analytics layer; presentation layer; and decision-making layer. First, the proposed framework was applied to predictive analytics for cirrhosis of the liver using two machine learning methods: “logistics regression” and random forest classification. The two methods' performance results were evaluated and compared. The random forest classification was found to be the best fit for predicting cirrhosis stage. second, the proposed framework was applied to descriptive and predictive analytics for Dakahlia's healthcare system. Third, the proposed framework was applied to predictive analytics for diabetes prediction using machine active learning with selection methods. The four methods' performance results were evaluated and compared. The random selection method was found to be the best fit for predicting diabetes.

Keywords: Data analytics - Predictive Analytics – Healthcare Data Analytics Framework- Descriptive Analytics – Intelligent Decision Support System (IDSS) – Machine Learning – Active Learning - Healthcare Systems – Diabetes Prediction – Cirrhosis Prediction.

1.1 Research Introduction and Motivation of the Research

One of the fields with the fastest rate of growth is health informatics, which focuses on using computer science and information technology to analyze medical and health data. It gives managers added information about the health sector that could improve patient outcomes, advance personalized care, and cut down on wasteful spending. Big Data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and hardware; nor can they be easily managed with traditional or common data management tools and methods (Oyelade et al., 2015).

By discovering associations and understanding patterns and trends within the data, Big Data analytics has the potential to improve care, save lives and lower costs. Thus, Big Data analytics in healthcare takes advantage of the explosion in data to extract insights for making better informed decisions.

The new generation of decision-makers is moving away from relying on reports to make responsive decisions and toward using big data analytics to forecast outcomes. The foundation of Predictive Analytics (PA) is building a data-driven arithmetic relationship between the plans and goals of the business function and the achievement or failure of the enterprise's strategic goals. With the aid of this relationship, managers can develop a long-term strategy and analyze the outcomes of their decisions (Holsapple et al., 2014).

There are different methodologies and algorithms to predict different diseases called the predictive learning algorithms. The main purpose of these algorithms is to classify, predict, and analyze the data. The utilization of ML-based analyses is to discover the relationships and correlation of the data for necessary training and testing. Therefore, when the information is abundant, and the labels are complex with maximum elapsed time, it is more appreciated to use AL algorithms for this purpose. The evolution of depending on Artificial Intelligence (AI) in the decision-making process requires changing the decision- making environment of the healthcare sector to enhance analytics in all activities and using AI tools and techniques. these are considered the main motivations of this research to try to help decision-makers (Siegel, 2010). The aim of this thesis is to build a framework for healthcare systems that includes predictive analytics of diabetes and cirrhosis as well as descriptive analytics of healthcare data, and to present them in the right way to the person making the decision.

1.2 Research Problem and Questions

The healthcare system has generated enormous amounts of data generated from electronic records. To improve the quality of healthcare, it is necessary that large volumes of data generated should be analyzed effectively to answer new challenges. But dealing with this huge amount of data requires analytical embracing ideas that go beyond the traditional statistical methods already in use in the Egyptian context. Although nowadays countries have moved from using decision support systems to using intelligent decision support systems to improve their decisions, improving the services, and reducing costs.

Big Data Analytics helps in discovering valuable decisions by understanding the data patterns and the relationship between them with the help of machine learning algorithms, Data Mining and Artificial Intelligence. It requires a technology that helps to perform a real time analysis on the enormous data set, which is not utilized in the Egyptian healthcare information system.

Therefore, the following problems exist within the Egyptian healthcare system:

a) Modern analytics of the Egyptian healthcare system are still missing to help national decision making.
b) The lack of studies on the healthcare systems utilizing advanced data analytics for improving Egyptian healthcare policies and decision-making process generates a gap in the knowledge of the benefits, tools, factors, limits, and framework of healthcare data analytics.
c) There is a weakness in exploiting information and communication technology in the field of saving and spreading of information.

Hence this research tries to help those managers and planners to improve their decisions and get the greatest value from their data and information delivery systems by using data analytics techniques to improve services by trying to answer the following questions:

- What are the standards that can be used in healthcare data analytics framework for IDSS?
- What is a dynamic simulation framework of healthcare data analytics for IDSS?
- How can a Data Analytics framework help healthcare manager?
  - What are the common characteristics of hospitals according to occupancy rate whether higher or lower?
  - What are the critical features that affect public health and patient satisfaction?
  - What are the types of hospitals through daily census of inpatients, bed turnover rate, average length of stay and medical services?

- What are the necessary recommendations for the development of future policies and take the necessary decisions to improve health services that meet the needs of citizens?
- What is the accuracy of disease prediction if different machine learning algorithms are used, such as logistic regression and random forest regression?

1.3 Research Limitations

Limitations of the study include the following:

- The proposed Data Analytics Framework for IDSS to be applied for the Dakahlia Healthcare system (Curative care hospitals and public data centralization, basic care units in the maternal and child data).
- The data of the four years of Dakahlia healthcare system will be used.
- Machine Learning algorithms with Spark Libraries are used for cirrhosis prediction.
- Active learning algorithms with python programming language are used for diabetes prediction.

1.4 Research Objectives and Importance

The main objective of this study is to propose a data analytics framework for diabetes prediction, cirrhosis prediction, and descriptive analytics for healthcare systems to aid planners and decision makers in moving towards relying on data analytics for IDSS.

1.4.1 Research Objective

This study aims to fulfil the following objectives:

1. To determine the attributes of healthcare data analytics framework for IDSS, by a review of literature.
2. To survey the analyses of existing healthcare system in the Dakahlia governorate.
3. To determine the most important technologies and algorithms for predictive and descriptive analytics for diabetes prediction and cirrhosis prediction.
4. To create a dynamic system framework for healthcare data analytics for critical success factors.

1.4.2 The importance of the research

1. The importance of the study comes from its added value to healthcare in Egypt. Disease prediction is considered at the heart of healthcare policymaking and the success of healthcare systems because it affects the services and reduces costs.
2. Scarcity of Egyptian studies that deal with Big Data Analytics in the Egyptian healthcare system in general means it may contribute to the study of the enrichment of decision support system in this area. The proposed Predictive Analytics framework for IDSS in healthcare is applied to the Dakahlia governorate. This framework can be used to predict diabetes and cirrhosis for all members of the public.
3. The study supports the trend towards the improvement of data analytics and the decision-making process in healthcare through predictive and descriptive analytics for diabetes prediction and cirrhosis prediction.

1.5 Research Methodology

To achieve the goal of this study, the following research strategies and techniques are applied as follows:

1.5.1 Scientific Procedures

The methodology used in the study's context is the multi-paradigm approach, using a combination of techniques including survey method, framework building, and simulation that applies to the framework.

1. **Survey method**

   The best way to gather the data needed to answer the research questions for this dissertation is through a survey, which is a quantitative method.

2. **Framework building**

   Framework development is regarded as a successful research strategy. It helps researchers and scientists relate to reality more precisely, and it also helps them explain, forecast, test, or comprehend complicated systems or processes. Frameworks can be physical items or abstract forms like sketches, formulas, or diagrams, and they frequently serve as a foundation for conducting research. The study compares classification algorithms to see which algorithm delivers higher accuracy, using authorized criteria.

3. Simulation method

Simulation is an efficient way of analyzing the different dynamic issues in recent years. Using mathematical representations of the operational relations in the system over a predetermined time, a simulation represents the real system. A framework has been created to assess various design and control scenarios (Demir, 2018). The simulation applies the proposed framework to cirrhosis of the liver, Dakahlia's healthcare system, and diabetes.

1.5.2 Techniques for Data Collection

1. Collecting published literature regarding "the healthcare data analytics framework for IDSS" from various libraries and practical experience
2. Collecting data was from information centers, hospitals, and health units in Dakahlia governorate
3. Distributing the questionnaire among the doctors using Google Forms

1.5.3 Research Preparation Procedures

The study was executed in several phases as follows:

Step 1: Diagnostic Investigation

A diagnostic study of healthcare systems in Dakahlia governorate was conducted to indicate the already existing problems.

Step 2: Conducting a Preliminary Literature review.

A review of published books, research papers, and published Ph.D. dissertations in these disciplines was undertaken to become familiar with the prior research work on predictive analytics framework and intelligent decision support systems.

Step 3: Questionnaire and Checklist Criteria

The questionnaire and a letter outlining the motivation of this research were distributed and analyzed. The steps included preparing the checklist criteria questionnaire and assessing it, Choosing the appropriate sample, collecting the data relative to the study, drawing the appropriate criteria underlying the proposed data analytics framework, refereeing it, and testing it.

Step 4: Experimental Study

The tools that can be utilized for exploratory data analysis were reexamined to select one that is quick and simple to use for developing the suggested framework. The framework was developed in an experimental study utilizing PySpark with the Python programming language, which is excellent for performing large-scale exploratory data analytics and creating pipelines for machine learning and active learning. And simulating the framework of healthcare data analytics for IDSS.

Step 5: Drafting the research report; conclusions, recommendations, and future research.

1.6 Research Results

- The researcher found through a survey of healthcare in Dakahlia governorate that there is a relative importance of big data analytics in improving healthcare.

- Requirements for digital transformation (like making a digital transformation strategy, spreading the digital transformation culture, and designing programs) have a significant impact on improving healthcare.
- There is a significant impact of digital transformation on the decision-making process.
- There is importance to the determinants of big data analytics (administrative determinants, human determinants, technological determinants, strategic determinants) for improving healthcare systems and decision making.
- There is the importance of administrative, human, technological, and strategic determinants in improving healthcare systems.
- There is a Positive impact of intelligent decision support systems on improving healthcare.
- The architecture of healthcare data analytics framework consists of five layers; 1) data layer; 2) data aggregation layer; 3) data analytics layer; 4) presentation layer; and 5) decision making layer.
- The Random Forest algorithm was chosen for Predictive Analytics for Cirrhosis of the Liver because it has the highest prediction accuracy and is the best fit for predicting.
- The Random Selection Method with Recall Measure was chosen for Diabetes Prediction Using Active Learning because it has the highest prediction accuracy and is best fit for predicting.
- Healthcare administrative needs to rely on data analytics for intelligent decision support systems to improve their services and reduce costs.

1.7 Terminology Definitions

Listed below are definitions of terms used in this study:

Data Analytics: Data analytics is the process of employing specialized computer systems to extract meaning from unstructured data. These systems alter, arrange, and model the data to make judgments and spot patterns (Pathak et al., 2018).

Big Data: Everything around us constantly produces large amounts of data. It is created by every digital procedure and social media conversation. It is transmitted through systems, sensors, and mobile devices. Big Data is being made by various places and people at an alarming rate, size, and range (Sagiroglu & Sinanc, 2013).

Predictive Analytics

It is a type of data mining that deals with prediction probabilities and trends for the future. Predictive Analytics has now entered the mainstream, according to Halper (2014), who defines it as "a statistical or data mining solution that includes algorithms and methodologies that may be applied to both structured and unstructured data to determine outcomes."

Descriptive Analytics: The analysis of data or information to determine "What happened?" (or "What is happening?") is known as Descriptive Analytics. It is characterized by conventional business intelligence (BI) and visualizations such as pie charts, bar charts, line graphs, tables, or created narratives. (MOKRAN, 2020).

Machine Learning: Machine learning explores the study and construction of algorithms that can be learned from historical data to make predictions. Data-driven algorithms operate by building a model from example inputs to make predictions or decisions. (Cheng et al., 2020)

Health informatics: a broad term for the application of information science, computing, networking, and communications methodologies and techniques to support health and health-related disciplines such as medicine, nursing, pharmacy, dentistry, and so on (Music, 2020).

Healthcare Analytics

When discussing healthcare analytics, it is important to ask how the statistics numbers are regarding the usage of analytics in healthcare and how this affects the end user's knowledge (Russom, 2011).

Intelligent decision support systems (IDSS): IDSSs are interactive computer-based systems that enable decision-makers in businesses by using artificial intelligence approaches to handle complicated, uncertain, and ill-structured problems (Tamimi et al., 2019).

Active Learning (AL)

Active Learning approaches are becoming new and interesting ways to look at healthcare data because they work well in many fields, and their methods are quickly improving. Active Learning research is driven by two main concepts: Unlabeled data is typically readily available or easy to collect, and learners should be allowed to ask questions. Active learning tries to reduce the amount of time and money needed for labelling so that a prediction model that can make accurate predictions can be created (Monarch, 2021).

Predictive Analytics Using Active Learning Algorithm

Predictive Analytics is the process of modeling historical electronic health records for predicting future events. Predictive modeling involves multiple steps. First, define the prediction target. The second step is defining all the potentially relevant patients for this study, and the third is selecting which features that are relevant for predicting the target. The fourth step is computing the model that maps the input features of the patient to the output target, and the fifth is evaluating the model (Wang & Byrd, 2018).

Apache Spark: Apache Spark is an open source, distributed computing framework and a collection of tools for real-time, huge data processing, and PySpark is its Python API. If you are already familiar with Python and tools like Pandas, PySpark is a useful language to learn to build more scalable analytics and pipelines (Gupta et al., 2017).

References

- Cheng, J. C., Chen, W., Chen, K., & Wang, Q. (2020). Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms. *Automation in Construction*, *112*, 103087.
- Gupta, A., Thakur, H. K., Shrivastava, R., Kumar, P., & Nag, S. (2017). A big data analysis framework using apache spark and deep learning. In *2017 IEEE international conference on data mining workshops (ICDMW)* (pp. 9-16). IEEE.

- Halper, J., & Kjaer, M. (2014). Basic components of connective tissues and extracellular matrix: elastin, fibrillin, fibulins, fibrinogen, fibronectin, laminin, tenascins, and thrombospondins. *Progress in heritable soft connective tissue diseases*, 31-47.

- Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. Decision support systems, 64, 130-141.
- Jeble, S., Kumari, S., & Patil, Y. (2016). Role of big data and predictive analytics. *International Journal of Automation and Logistics*, *2*(4), 307-331.

- Monarch, R. M. (2021). *Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI*. Simon and Schuster.

- Oyelade, J., Soyemi, J., Isewon, I., & Obembe, O. (2015). Bioinformatics, healthcare informatics and analytics: an imperative for improved healthcare system. *International Journal of Applied Information System*, *13*(5), 1-6.
- Pathak, A. R., Pandey, M., & Rautaray, S. (2018). Construing the big data based on taxonomy, analytics, and approaches. *Iran Journal of Computer Science*, *1*(4), 237-259.
- Phillips-Wren, G. (2013). Intelligent decision support systems. *Multicriteria decision aid and artificial intelligence: links, theory, and applications*, 25-43.

- Rahul, A. E., & Narukulla, S. (2018). Introduction to Data Mining and Machine Learning Algorithms. *Int. J. Res. Eng. Sci. Manag*, *1*.

- Russom, P. (2011). Big data analytics. TDWI best practices report, fourth quarter, 19, p.40. Available online at: https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf
- Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 international conference on collaboration technologies and systems (CTS)* (pp. 42-47). IEEE.
- Siegel, D. J. (2010). The Mindful Therapist: A Clinician's Guide to Mindsight and Neural Integration (Norton Series on Interpersonal Neurobiology). WW Norton & Company.
- Tamimi, N., Samani, S., Minaei, M., & Harirchi, F. (2019, July). An artificial intelligence decision support system for unconventional field development design. In *SPE/AAPG/SEG Unconventional Resources Technology Conference*. OnePetro.
- Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological forecasting and social change*, *126*, 3-13.