Implementing Big Data to Predict the Doctor Efficiency

K. Venkatesh[*1], Dr. S. Parthiban[*2], P. Santhosh Kumar[*3]

[*1, 2] Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
[*3] Department of Computer Science and Engineering, Siddharth Institute of Engineering and Technology, Puttur

Abstract **-** World is fully comprised of data. Now a days, data and information rule the world. Data analytics plays a major role in every sector. Predictive analytics a part of data analytics plays a vital role in different sectors of Medicine, Defense, Agriculture, Stock Exchange and many more. Machine learning algorithms are widely used to solve many of these problems. Medicine is essential for every Individual. Patient tracking and disease diagnosis is more important these days. One of the major problems is recommendation of doctors to patients according to Hospital Management System (HMS). Machine Learning algorithms are trained to solve this problem with higher accuracy. This not only helps Management but also Patients. Python Environment is used for Machine Learning Model whereas Django is used for User Interface (UI) and SQLite for backend Database.

Index Terms - Doctor, Django, Hospital Management System (HMS), Python, SQLite, User Interface (UI).

I. INTRODUCTION

Medicinal field is one of the important and emerging sectors in to- day's world. Hospitals play a vital role in saving many lives and increasing medicinal values. Due to the mechanical life of today's individual drastically change is observed in health issues when compared to past. For hospitals, the management of patients with upcoming health issues is huge task to be handled effectively. Hospital Management System is the Software approach used by majority in medical field for maintaining Patient records, Pharmaceutical Records, Doctor Records and many more. It is used by many hospitals for the records purpose. It not only works as database platform but also an efficient application for Statistical analysis of Data. HMS makes the management's work easier for maintenance of hospital data. Doctor efficiency is one of the important aspects to be calculated for hospital management to consider doctor for future reference because in a technological world of today people are choosing based upon reviews and successful cases solved by doctor or hospital management. So, it is necessary to predict the efficiency of doctor and rank them upon number of cases solved and complexity of each case. The Disease level complexity is defined by the previous study of diseases and references by many of doctors in all aspects.

II. LITERATURE REVIEW

The author [1] in this paper mainly discusses about mart Hospital Management System (SHS). It is an infrastructure component relies heavily on the actual resources made available to it for its proper functioning, operation, and maintenance. This is a solution aimed to present architecture Integration Framework using TOGAF's architecture development method. The TOGAF's Architecture consists mainly of 8 phases namely Architecture Vision, Business Architecture, Information Systems Architecture, Technology Architecture, Opportunities and Solutions, Migration Planning, Implementation Governance, Architecture change Management. The effectiveness of SHS is explained in terms of TOGAF with advantages and disadvantages.

The author [2] discusses about the benefits and problems that can be seen with the use of electronic versions of medical records. They constitute the permanent documentation of patient health, permitting the medical professional to evaluate symptoms and signs within a broader temporal perspective, contributing to improvements in making diagnoses and providing treatment. The hospital information systems constitute of three levels namely Operational level, Tactical level and Strategic level about three operational areas of Medical area, Administrative area and Hospitality. It also discusses about case study of testing made on HMS of three hospitals in city of Sa˜o Paulo, Brazil which includes visualization and testing results of HMS.

This author [3] in this discusses about an effective mechanism for chronic disease risk prediction by mining the data containing historical health records and personal lifestyle information. The method outperformed is the traditional mechanism in terms of accuracy, precision and sensitivity for predicting the risk of diabetes. Insightful observations show that the consideration of life-style information can effectively enhance whole performance for risk pre- diction. The framework described here is divided into three phases. In the first phase, health risk patterns of each item are found from a sequence dataset of each item in a health examination and a lifestyle dataset, respectively. The second phase, each health risk pattern is as a feature attribute and constructed a dynamic feature dataset for both static dynamic data respectively. The health risk prediction model is deployed using classification algorithm. Finally, in the third phase the patient's data to be predicted, the examinee's historical data related to health and lifestyle is converted to the suitable dynamic and static data according to the health risk patterns.

The author [4] explains about efficient patient management system based on Near Field Communication (NFC) technique. For registration, each NFC tag wrist band is identified by a Unique Identification Number (UIN) that can be programmed and can also be protected through password. NFC encoded device can be used to read this information from the patient at any time without causing any inconvenience to them. NFC is mainly introduced to reduce cost and paperwork for Hospital Management System (HMS). The system of the NFC based Hospital management system consists of six layers namely

physical layer, middle layer, process layer, data access layer, application layer and user interface layer.

In this paper author [5] explains about the Machine learning algorithms are used for effective prediction of chronic disease outbreak in disease-frequent communities. Latent factor model was used to re- construct the missing data. Convolution Neural Network-CNN based multimodal disease risk prediction algorithm is used for structured and unstructured data from hospital. The prediction accuracy was about98.4 when compared to other models of Big data analytics. The Data used was real-time hospital data stored in cloud from years of 2013-2015 which includes 31919 patients of 20320848 records overall. Stochastic Gradient Descent algorithm and CNN-based UNI model disease risk prediction algorithm are used for Structured and Unstructured data, respectively.

This author [6] speaks about the Artificial Intelligence (AI) based health physician system that would be able to interact with the patient, do the diagnosis and suggest quick remedy or treatment of their problem. The system uses a questionnaire-based approach to query the user (patient) about various Symptoms, based on which a decision is made, and a medicine is recommended. The usage of AI [10] is because it makes systems behave and work more like humans is gaining popularity. Natural Language Processing (NLP), smart agents, Machine Learning (ML) and knowledge base were used to make Rapid Application model for Smart Doctor. Decision Tree Algorithm is used here for every disease for the medicine prediction and doctor recommendation. This Software acts as efficient chatbot for the patient convenience.

The author [7] discusses about the importance of solutions for improving public health, healthcare providers are required to be fully equipped with appropriate infrastructure to systematically generate and analyze big data, whereas it need an efficient management, analysis, and interpretation. As the data is larger in size is not said to be big data because it should satisfy the conditions of velocity, veracity, volume, and variety. The collected data is stored in data warehouse and workflow is progressed by analytics 4.0 namely descriptive analytics (1.0), Diagnostic analytics (2.0), Predictive Analytics (3.0) and prescriptive analytics (4.0) for future usage. The healthcare information mainly consists of records mainly of Electronic health record, clinical information, and patient information. This brought a drastic change in journals publishing in health care associated with big data from 2000 to 2018. The IBM Watson is used here for Natural Language processing and deep learning process.

In this paper author [8] explains about impact of usage of Big Data analytics in health care [9]. One of the objectives in this paper is to address healthcare delivery problems and improve healthcare quality because Pharmaceutical-industry experts and shareholders have begun to routinely analyze big data to obtain insight, but these activities are still in the early stages. Predictive analytics plays a vital role in these analytics where it has been recognized as one of the major business intelligence approaches, but its real-world applications extend far beyond the business context. It does not include text analytics but also multimedia analytics whereas Hadoop architecture is used for health informatics.

III. MATERIALS AND RESEARCH METHODOLOGIES

A. Aim of the paper

The paper is designed for the benefit of hospital management. The objective is to find the efficiency of a doctor based on the number of solved successful cases. It is easy to predict the success rate of a doctor using number of cases he/she deals with, but success rate is not the important thing in today's world. The doctor solved cases is taken with the complexity of each case he solves. This approach not only involves doctor's success rate but also the complex rate of each disease. This prediction is made through Machine Learning algorithms easier for the problem to be solved. The analytics 4.0 is used where descriptive and diagnostic analytics is used for the disease level complexity prediction by Decision trees and Predictive and Perspective analytics are used for Doctor Efficiency prediction.

B. Paper Domain

To solve a real-time problem on a computer, we would like an algorithm. An algorithm is a sequence of instructions that should be carried out to transform the raw data input to required output. For example, one can devise an algorithm for sorting. The input may be a set of numbers and therefore the output is their ordered list. For an equivalent task, there could also be various algorithms and that we could also be curious about finding the foremost efficient one, requiring the smallest amount number of instructions or memory or both. Machine Learning is the way of training the algorithms with different types of data to get more accuracy in the resultant output. In this Paper we have used both Supervised Machine Learning techniques in order to get a clear and efficient output. Supervised learning makes training with known output variables with two prominent techniques namely classification and regression. It is study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of AI. Machine learning algorithms build a mathematical model supported sample data, mentioned as "training data", so on form predictions or decisions without being explicitly programmed to perform the task. Machine learning algorithms are utilized in a good sort of applications, like email filtering and computer vision, where it is difficult or infeasible to develop a standard algorithm for effectively performing the task.

C. Scope of the paper

This paper has future scope of deploying in server-based mechanism for better results. Treatment level complexity is one more case where different levels of treatment for same disease. The future scope of implementing the complexity levels uses Artificial Neural Networks. This implementation can be made as a qualitative approach as it needs high end discipline to work in. The proposed solution is developed depicts the outputs acquired at a higher rate of accuracy.

D. Methodology

The Main logic of code for prediction runs in Spyder console of anaconda environment written in python language at the backend. Data base acts as middleware where all the data is stored for the process. Database used here is My SQL workbench and SQLite3 as depends upon the system configurations. The User Interface is developed using Django framework of python in visual studio and UI screen appears in internet explorer using the static uniform resource locator of the Django framework. The data fed happens through the admin page for the patient record. The data is stored in database and retrieved in ML source code of anaconda framework. The training and testing of the data happens and validation is checked. The validated and predicted outputs are directly exported to the database. The results are retrieved and displayed in the User Interface Screen according to the user input and specification.

IV. PROPOSED SYSTEM

Hospital Management System (HMS) is an effective system of all records and statistics are stored explicitly. In this corporate world many IT companies have created many HMS applications which are being used by many hospitals for storing and analyzing their data explicitly. Other systems have many other features like medicine recommendation system, patient treatment re- minder system and many more. Many applications have no module or model of doctor recommendation system for management system or patient purpose.

The proposed system is an application which recommends doctor for disease with number of successful cases solved and efficiency too. This also explore about the top five doctors of specialization. This also has Pharmaceutical analysis according to the department wise of the concerned hospital. The recommendation system was developed under Python Language. Machine Learning Algorithms are developed using Anaconda environment in Spyder IDE. This is connected as backend system to middleware database. The Database is either of MySQL workbench or SQLite which depends upon the server configuration used. The front end is the main User Interface which was developed under Django framework in Visual Studio.

A. Feasibility study

The feasibility study of a paper is to study about the paper which fulfils the needs of today's market stage by considering different com- plications in development of the paper. This feasibility analysis of paper depicts a clear view to make paper as a business plan as it needs to be commissioned. This is to guarantee that the proposed framework is not a weight to the organization.

B. Module description

General Architecture

The simpler architecture has been designed of Prediction System. This explains about all the phases of Implementation System. Figure 1 shows the architecture diagram of the proposed system.

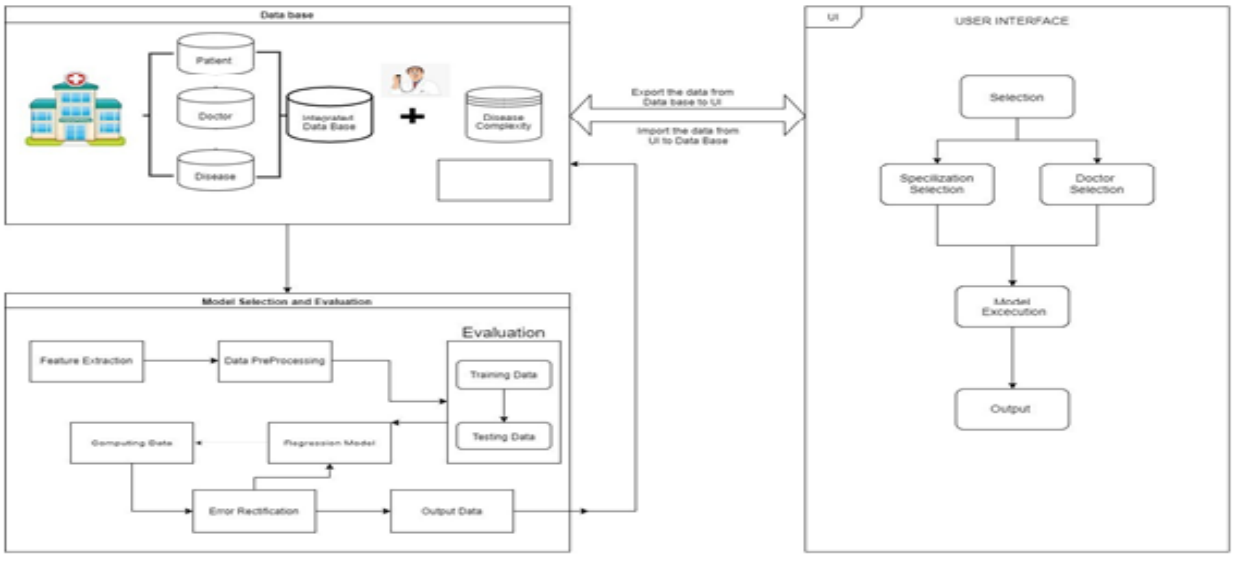

Figure 1: Architecture Diagram

The front end is User Interface (UI) which is connected to middle- ware Database integrated with the backend Machine Learning source code. The database is collective repository of Patient records, Doctor Records and Disease records. These three databases are collectively integrated to a single database with many tables. To this database disease complexity is added from doctor's perception and with the factors defining the disease complexity for new diseases is done through Decision Tree Algorithm. This module is for hospital management for the Disease Level Complexity (DLC) calculation purpose. The data from integrated database flows into the main code (or) source code. The complexity in raw data acts as tuple it is changed to attribute using pre-processing techniques. Feature Selection and Feature Extraction is done for this data for better algorithm induction. This data is splitter into training and testing data and Linear Regression Algorithm is applied. The predicted data from algorithm is compared with that of original data and checked for accuracy. This final data is stored into database. These final data is been stored into new table in database so that it is retrieved in UI screens. UI screens are created using Django framework in python language. There are four user screens in our UI. The first screen displays doctor efficiency according to the specialization selected as one need to select the specialization first and list comes accordingly and when doctor name is selected the efficiency is shown. The second screen displays efficiency of doctors using the name of doctor specifically. The third screen acts as ranking system where it ranks the topmost five doctors for a specific department. The final screen explores about the pharmaceutical data according to department wise. The admin page is used to check with raw data and adding new data to database. The crud operations in database can be applied from UI admin page itself. The number of users and groups can be checked using Admin Page itself.

V. RESULTS AND DISCUSSION

This paper exposes an application which recommends doctor for disease with number of successful cases solved and efficiency too. This also explore about the top five doctors of specialization. This also has Pharmaceutical analysis according to the department wise of the concerned hospital. The recommendation system was developed under Python Language. Machine Learning Algorithms are developed using Anaconda environment in Spyder IDE. This is connected as backend system to middleware database. The Database is either of MySQL work- bench and SQLite which depends upon the server configuration used. The front end is the main User Interface which was developed under Django framework in Visual Studio. The proposed system ML algorithm works more than 90% efficient with negligible error in prediction.

A. Implementation

Implementation phase is an essential and crucial phase in development of a paper. After all the feasibility studies of a paper are made and methodologies are studied it is implemented into a paper or a working module. The execution includes cautious arranging, examination of the current framework and its requirements on usage, planning of techniques to accomplish changeover and assessment of changeover strategies. All the strategies are taken under similar platforms where modified to better working model.

B. Input Design

The input to the model of is different according to system and algorithm performed. In the paper two algorithms are implemented.

Table 1: Table for DLC calculation

Disease Name	Cause	Genetic	Repetitive Occurrence	Success rate	Complexity
APPENDICITIS	General	No	No	High	One
BLEEDING GUMS	General	No	Yes	High	One
BODY PAINS	General	No	Yes	High	One
CAVITIES	Bacteria	No	Yes	High	One
COLD	Virus	No	Yes	High	One
DRY SKIN	General	No	Yes	High	One
HEADACHE	General	No	Yes	High	One
INSOMNIA	General	No	Yes	High	One
OBESITY	General	No	Yes	High	One
RICKETS	General	Yes	No	High	One
ROOT CANAL	General	No	Yes	High	One
SOLAR RETINOPATHY	General	No	Yes	High	One
SPRAINED ANKLE	Injury	No	Yes	High	One
STOMACHACHE	General	No	Yes	High	One
STRESS	General	No	Yes	High	One

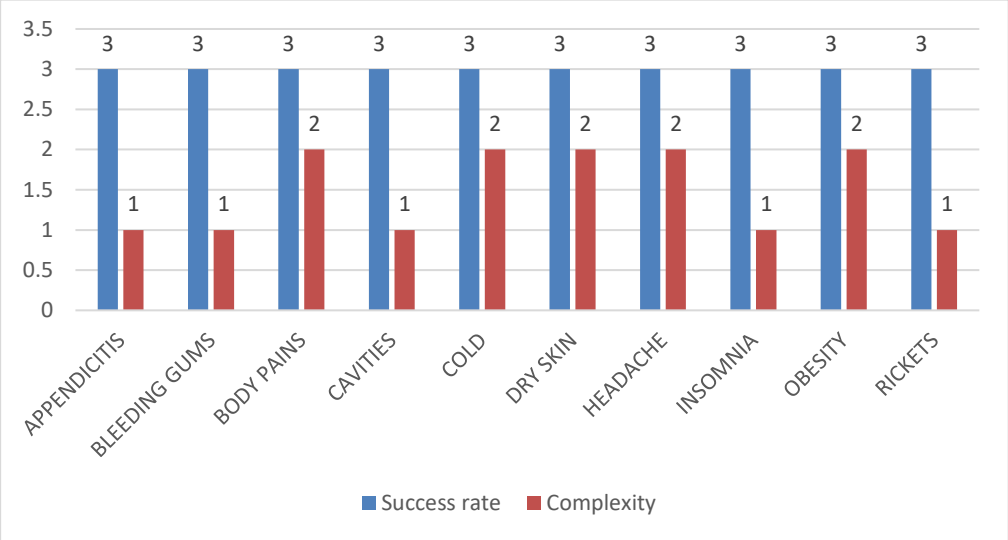

Figure 2: Sample graph

Table 2: Table for Doctor Efficiency calculation

Patient ID	Consultant	Specialization	Disease	Complexity
1592	Doctor 1	Urologist	Blasser prolapse	4
1593	Doctor 1	Urologist	Postate cancer	5
1594	Doctor 1	Urologist	Prostatis	1
1595	Doctor 1	Urologist	Hematuria	3
1596	Doctor 1	Urologist	Overactive bladder	2
1597	Doctor 1	Urologist	Blasser prolapse	4
1598	Doctor 1	Urologist	Prostatis	1
1599	Doctor 1	Urologist	Hematuria	3
1600	Doctor 1	Urologist	Overactive bladder	2
1601	Doctor 1	Urologist	Blasser prolapse	4
1602	Doctor 1	Urologist	Prostatis	1
1603	Doctor 1	Urologist	Postate cancer	5
1604	Doctor 1	Urologist	Overactive bladder	2
1605	Doctor 1	Urologist	Prostatis	1
1445	Doctor 2	Gynecologist	Amenorrhea	3
1446	Doctor 2	Gynecologist	Menstrual disorder	4
1447	Doctor 2	Gynecologist	Uterine fibroids	2
1448	Doctor 2	Gynecologist	Amenorrhea	3
1449	Doctor 2	Gynecologist	Pelvic pain	1
1450	Doctor 2	Gynecologist	Menstrual disorder	4
1451	Doctor 2	Gynecologist	Pelvic pain	1
1452	Doctor 2	Gynecologist	Endometriosis	5
1453	Doctor 2	Gynecologist	Uterine fibroids	2
1454	Doctor 2	Gynecologist	Amenorrhea	3
1455	Doctor 2	Gynecologist	Endometriosis	5

Firstly, data input contains attributes like disease name, genetic, repetitive, success rate and complexity which contribute to DLC calculation using decision tree algorithm. The second module is to predict the doctor efficiency

according to solved number of cases and their complexity. This data table contains attributes of Patient id, Consultant name, Specialization, Disease and Complexity which are used to predict the doctor efficiency using Linear Regression Algorithm. These both tables and output tables are stored in same database stored in DB SQLite3 version. This database is directly integrated with source code and UI code in Django framework.

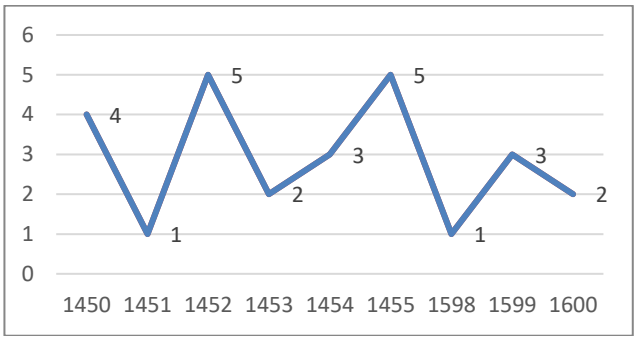

Fig: 3 Doctor Efficiency calculations

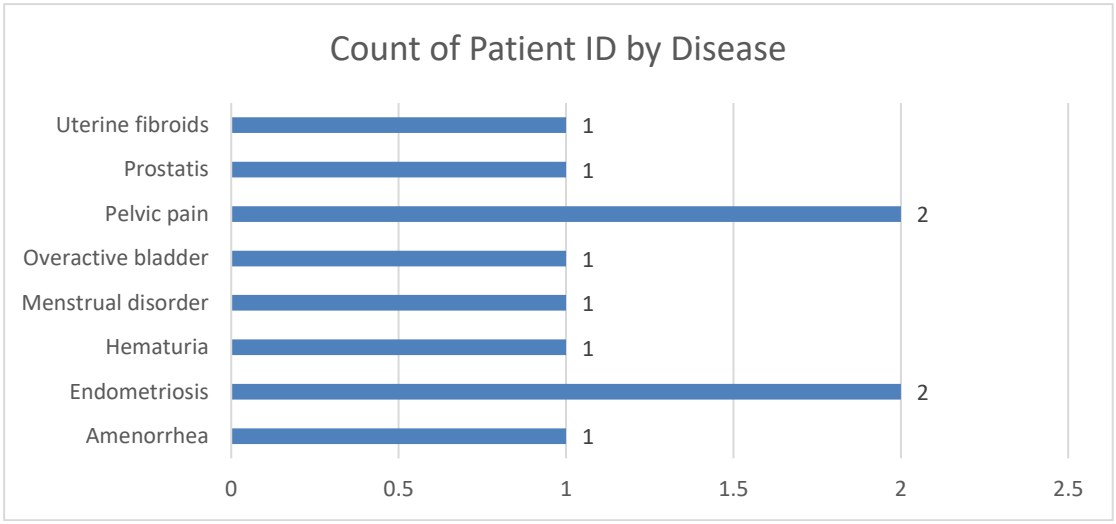

Fig: 4 Patient by disease

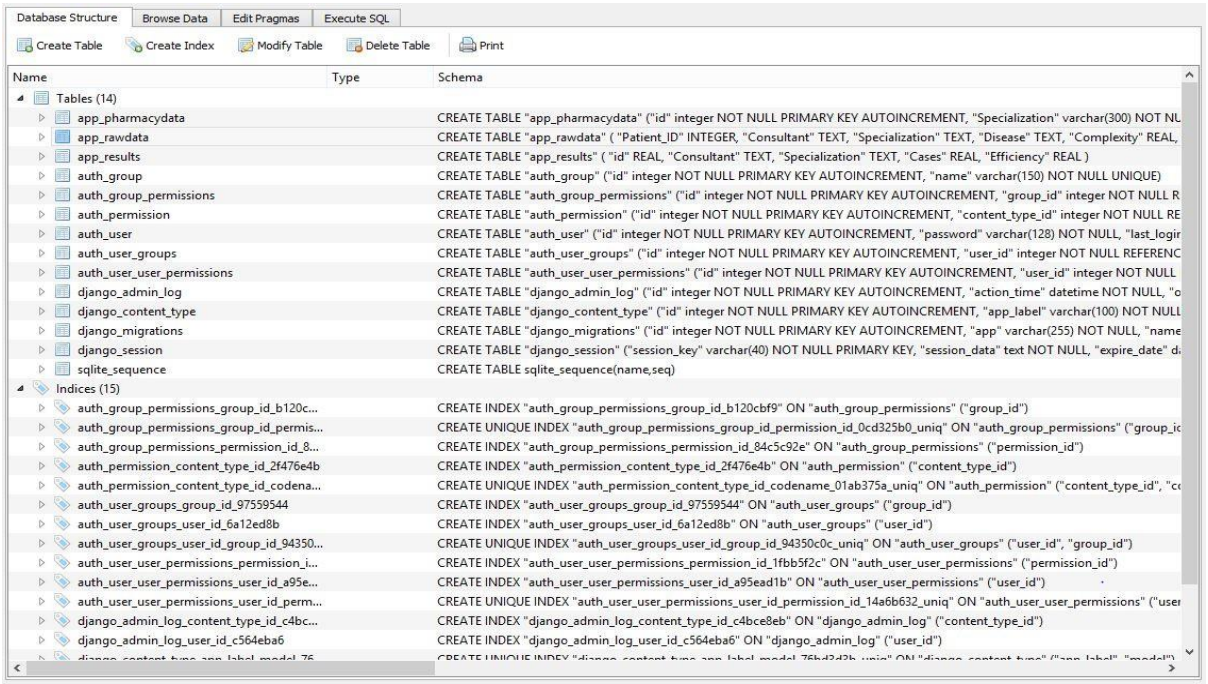

Figure 5 Database

Advantages of the Proposed System

The recommendation system is developed using the Regression Algorithms which depicts with least percentage of error. The UI screens are user friendly and easily can be interpreted by Management System. This approach of proposed solution is developed under an adaptive nature so that can be easily corrected for errors if occurs.

VI. CONCLUSION

This paper provides a clear, feasible, easily understandable view for any HMS in knowing an efficiency of doctor. According to Hospital management view it is more important to know how efficient their doctors in dealing with diseases. This paper helps to deal with the problem efficiently. This helps the hospital management in making better decision for future benefits of the hospital. This venture not only helps hospital management but also common people in selecting a better doctor for their problems to be dealt with.

REFERENCES

[1] Zenon Chaczko, Christopher chiu, Avtar Singh Kohli, Venkatesh Mahade- van, 2010, "Smart Hospital Management System: An Integration of Enterprise Level Solutions Utilising Open Group Architecture Framework (TOGAF)", 3rd International Conference on Computer Science and Information Technology, doi:10.1109/iccsit.2010.5564121.

[2] Pedro Luiz Coˆrtes, Eliana Golfette Paula Coˆrtes, 2011, "Hospital information systems: a study of electronic patient records", JISTEM - Journal of Information Systems and Technology Management, vol-8 no-1.

[3] Guo-Cheng Lan, Chao-Hui Lee, Yu-Yen Lee, Vincent S. Tseng, 2012, "Disease Risk Prediction by Mining Personalized Health Trend Patterns: A Case Study on Diabetes", Conference on Technologies and Applications of Artificial Intelli- gence.

[4] Atluri Venkata Gopi Krishna, CheerlaSreevardhan, S. Karun, S.Pranava Kumar, April 2013, "NFC-based Hospital Real-time Patient Management System" , In- ternational Journal of Engineering Trends and Technology (IJETT), Volume-4, Issue-4.

[5] MIN CHEN, YIXUE HAO, KAI HWANG, LU WANG, LIN WANG, 28 June 2017, "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities", SPECIAL SECTION ON HEALTHCARE BIG DATA, Digital Object Identifier11109/ACCESS.2017.2694446.

[6] Rida Sara Khan, Asad Ali Zardar, Zeeshan Bhatti, 2018, "Artificial Intelligence based Smart Doctor using Decision Tree Algorithm", Journal of Information Communication Technology, Vol. 11 Issue. 2.

[7] Sabyasachi Dash, Sushil Kumar Shakyawar, Mohit Sharma, Sandeep Kaushik, 2019, "Big data in healthcare: management, analysis and future prospects", Jour- nal of Big Data, Data Object Identifier 10.1186/s40537-019-0217-0.

[8] Sunil Kumar, Maninder Singh, March 2019, "Big Data Analytics for Healthcare Industry: Impact, Applications, and Tools", BIG DATA MINING AND ANA- LYTICS, ISSN222096-0654 l l05/06l lpp48–57 Volume 2, Number 1.

[9] K. Venkatesh, International Journal of Advanced Science and Technology (IJAST), "Big Data Analytics: Challenges, Issues, Applications and Research Opportunity", Vol. 29, No. 9s, (2020), pp. 2836-2848.

[10] K. Venkatesh, International conference on Fostering Interdisciplinary Research in health science (ICFIRHS) "A Survey on Machine Learning Algorithms and finding the best out there for the considered seven Medical Data Sets Scenario" Volume No. : 12, Issue No. : 6 Year: 2019, ISSN Print: 0974-3618, ISSN Online: 0974-360X. Impact Factor: 1.2106.

AUTHORS

First Author – K. Venkatesh, Assistant professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, email: venkiur@gmail.com.

Second Author – S. Parthiban, Assistant professor, Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, email: parthibans1983@gmail.com.

Third Author – P. Santhosh Kumar, Assistant Professor, Department of Computer Science and Engineering, Siddharth Institute of Engineering and Technology, Puttur – Andhra Pradesh. email: santhosh_cse07@yahoo.com