ANALYTICS FOR HOSPITALSHEALTH-CARE DATA

1Pravin A, 2Rimali S, 3Saran S, 4Sri Padma Priyadharshini S, 5Subha N

1,2,3,4Student, 5Guide and Assistant Professor
Knowledge Institute of Technology
Salem.

Abstract – Applying quantitative and qualitative methodologies to data analysis with the goal of gaining insightful knowledge is known as data analytics. With the use of data analytics, we can investigate data (exploratory data analysis) and even make inferences about our data (confirmatory data analysis). Go to the external API and retrieve the dataset. Check whether null values are found in the extracted dataset. The pre-processing of the data is carried out using the IBM Cog-nos Analytics tool, and a cleaned data module is produced. The module for cleansed data is used to do data exploration. These exploratory ideas are used to create a dashboard, report, and tale. Finally, the webpage includes links to the dashboard, report, and article. The prediction option can be used to forecast the length of the patient's stay. Machine learning algorithms are employed in the prediction process.

Keywords – Random forest, Predicting the length of stay, Data analytics, Dashboard.

1. INTRODUCTION
An extended hospital stay necessitates thinking about expenditures and suffering for patients. It also necessitates the requirement for improved planning by making better use of resources and facilities to meet upcoming resource demands. These factors inspire in-depth investigations that seek to shorten the length of stay. The dataset is compiled based on the requirements for the length of stay of the patients. The dataset includes the columns age, severity of disease, type of admission, and other hospital facilities. In this paper, we offer a concept in which hospital administration can estimate the exact length of the patient's stay. The model predicts and provides the duration of the patient's stay using the information gathered from the dataset. A random forest is used to train the model. The model's accuracy with RF is projected to be 99.71%.

1.1. PURPOSE
The purpose of conducting a length of stay in hospital prediction and data analysis is to forecast the amount of time a patient is likely to spend in the hospital based on various factors such as their medical condition, demographics, and treatment plan. The length of stay (LOS) is an important metric that impacts hospital operations, patient outcomes, and healthcare costs. Analyzing data related to LOS can help hospitals and healthcare providers. Hospitals can use LOS predictions to plan staffing levels and allocate resources more efficiently based on the expected number of patients and their estimated LOS. LOS predictions can help healthcare providers identify patients who are at higher risk of developing complications or requiring additional care, allowing them to provide more targeted interventions and improve patient outcomes.

2. PROBLEM STATEMENT
The goal is to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. The length of stay is divided into 11 different classes ranging from 0-10 days to more than 100 days. Understanding of key principles and familiarity with IBM Cog-nos Analytics are also required. Plotting multiple visualisations in order to offer the best solution, with which the user can quickly study the data and have a better understanding of the dataset. Dashboards, reports, and stories should be created.

3. PROPOSED SOLUTION
The proposed system includes both an analysis and a prediction process, and the user can select either an analytics or a prediction operation to run on the dataset. This process's analysis step entails acquiring and assessing data relating to the proposed system. During this phase, the user selects the analytics option and uploads the dataset to the analytics tool for an analytics or prediction operation to run on the dataset. This process's analysis step entails acquiring and assessing data relating to the proposed system. During this phase, the user selects the analytics option and uploads the dataset to the analytics tool. The user can visualise the dataset's contents and then use those visualisations to build a dashboard, report, or story. In addition, the user can examine previously created dashboards, reports, and stories. This phase's purpose is to identify the potential benefits and downsides of the proposed system, as well as any hurdles that must be overcome. Following the completion of the analysis phase, the prediction phase involves the use of machine learning algorithms to forecast future events based on historical data. In this phase, the user selects the prediction option, uploads the dataset using Python code, and forecasts the patients' length of stay using ML techniques such as the Random Forest algorithm. The expected output is saved in a CSV file. Finally, the CSV file will be displayed in the webpage.
4. DATA FLOW DIAGRAM (DFD)

Figure 4.1 Data Flow Diagram

SOLUTION ARCHITECTURE

Figure 5.1 Solution Architecture

5. TECHNICAL ARCHITECTURE

Figure 6.1 Technical Architecture

6. FEATURES

The model takes input from the dataset like their department, type of admission, severity of illness, and age. Then with the help of these inputs, the model predicts the length of stay of the patients.

7.1 FEATURE 1 – ANALYSING THE DATASET USING IBM Cognos ANALYTICS TOOL

Manually determining the number of spare rooms in each department of a hospital is a time-consuming task. However, once we have data on the availability of extra rooms in each department, we can upload it to IBM Cognos Analytics and construct comparison visualisations for departments and extra room availability. We can quickly examine the number of extra rooms available in each department using those visualisations. Cleaning data in IBM Cognos Analytics is very simple because it includes built-in
data processing solutions.

7.2 FEATURE 2 – CREATING DASHBOARD AND REPORT FOR THE EXTRACTED DATASET

The dashboard and report component of the IBM Cognos Analytics tool is used to build dashboards and reports for the pre-created visualisations. We can construct a large number of dashboards and reports using the IBM Cognos Analytics tool. The dashboards and reports must be saved after they are created, and an iframe link to the dashboard and report must be added to the webpage.

7.3 FEATURE 3 – CREATING STORY FOR THE EXTRACTED DATASET

The IBM Cognos Analytics tool's Story function is used to produce an animation effect (story) for the pre-created visualisations. The IBM Cognos Analytics tool has a number of animation effects that can be used to enhance the visualisations. Following the creation of the story, the file must be saved, and an iframe link of the story must be added to the webpage.
7.4 FEATURE 4 - PREDICTING THE LENGTH OF STAY OF THE PATIENTS AND DISPLAYING IN WEB PAGE

Predicting the length of stay for patients manually is a difficult endeavour. As a result, the prediction tool assists hospital administration in projecting the length of stay of patients based on information from their medical records. As a result, the hospital administration can easily manage and allocate resources to patients. The projected output is recorded as a CSV file and shown on the web page.

![Figure 7.4.1 Predicting Length of Stay](image)

7. ALGORITHM USED

8.1 RANDOM FOREST

A random forest is a machine learning technique for solving regression and classification problems. It makes use of ensemble learning, a technique that combines several classifiers to solve complex problems. A random forest algorithm is made up of several decision trees. The random forest algorithm’s ‘forest’ is trained via bagging or bootstrap aggregation. Bagging is a meta-algorithm that increases the accuracy of machine learning algorithms through an ensemble approach. The outcome is determined by the random forest algorithm based on the predictions of the decision trees. It forecasts by averaging the output of various trees. The precision of the outcome improves as the number of trees grows. A random forest algorithm overcomes the constraints of the decision tree technique. It reduces dataset overfitting and improves precision. It provides predictions without requiring additional packages.

![Figure 8.1 Model Training](image)

8. PERFORMANCE METRICS

The column chart visualization created for comparing the stay length and patientid gives the total number of patients in each stay period. Similarly, the hierarchy visualization for severity of illness and patientid gives the number of patients in each illness type. Radial visualization gives the percentage of patients in each admission type. Therefore IBM Cognos Analytics creates visualizations that gives the accurate count or percentage of the data.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
</table>
| 1.    | Metrics   | Classification Model:  
Confusion Matrix – [1840,0,0,407]  
Accuracy Score - 99.71%  
Classification Report – 90% |
| 2.    | Tune the Model | Hyper parameter Tuning – 89%  
Validation Method – Randomized Search CV |

Table 9.1 Performance Metrics
9. CONCLUSION
After the post-Covid-19 epidemic, one area of concern is healthcare. While there are many applications for healthcare management, patient length of stay is one of the most important variables to track and forecast if one wishes to increase the effectiveness of healthcare management at a hospital. In this project, we proposed application that helps to accurately predict the Length of Stay for each patient on case by case basis so that the Hospitals can use this information for optimal resource allocation and better functioning. This helps the hospital management to identify patients who will stay longer at the time of admission. The identified patients can have their treatment plan and it will lower the chance of staff/visitor infection.

The goal of this project is to analyze the healthcare dataset. The provided dataset is first checked for null values and the columns with null values are identified using python. The dataset is then uploaded to the database using IBM cloud and the database is connected to IBM Cognos Analytics tool. In IBM Cognos Analytics tool data module option is selected and the dataset is pre-processed (null values are replaced with valid values). Using this cleaned data module data exploration process is carried out, in which different visualizations are created. Interactive dashboard, report and story is created in IBM Cognos Analytics tool using the cleaned data module. Finally, the created dashboard, report and story is embedded to the webpage by fixing the iframe code in the html code.

10. FUTURE ENHANCEMENT
Data analytics techniques are being applied to improve research efforts in many health-related areas by gathering and analyzing clinical data. Here we created a dashboard for monitoring the length of stay of patients and also include the necessary data which were provided by the hospital to create an effective virtual dashboard using Cognos analytics. This dashboard supports clinicians and hospital managers in viewing and exploring data on processes and outcomes of care in an interactive manner. As heading for the future work right now, we are planning to create an app that works in an efficient manner which helps the hospital management for optimal resource allocation and to improve the accuracy to 100% for better functioning by predicting the Length of stay (LOS) of the patients by entering the age and other comorbidities related to the patient.

11. SAMPLE OUTPUT

![Figure 12.1 Dashboard](image1)

![Figure 12.2 Report](image2)
REFERENCES: