**ESTIMATION AND PREDICTION OF HOSPITALIZATION AND MEDICAL CARE COSTS**

The project submitted in the internship of the requirements

of the award of the degree

**BACHELOR OF TECHNOLOGY**

Submitted by

R. Meghana 206E1A0528(cse)

R. Anji 206E1A0304(mech)

G. Thuran 216E5A0503(cse)

M. Preethi 206E1A0523(cse)

T. Aravind 206E1A0531(cse)

Under the esteemed guidance of

Mrs. Bharathi Chamarthi

Assistant professor of



**DEPARTMENT OF MECHANICAL ENGINEERING**

**GONNA INSTITUTE OF INFORMATION TECHNOLOGY AND SCIENCES**

(Approved by AICTE, New Delhi, affiliated to JNTUGV-Vizianagaram)

Gonnavanipalem, Aganampudi, Parawada (Mandal)

VISAKHAPATNAM-530026

2023

# ESTIMATION AND PREDICTION OF HOSPITALIZATION AND MEDICAL CARE COSTS

Introduction

Estimation and prediction of hospitalization and medical care costs play a crucial role in healthcare planning and budgeting.

Accurate cost estimation helps healthcare organizations allocate resources effectively and make informed decisions.

By understanding the factors that influence costs, we can develop models to predict future expenses.

Severity of the medical condition: The complexity and duration of treatment impact the overall cost.

Type of medical interventions: Surgical procedures, specialized tests, and medication can significantly increase expenses.

Length of hospital stay: Longer stays result in higher costs due to room charges, nursing care, and additional services.

**Estimation Methods**

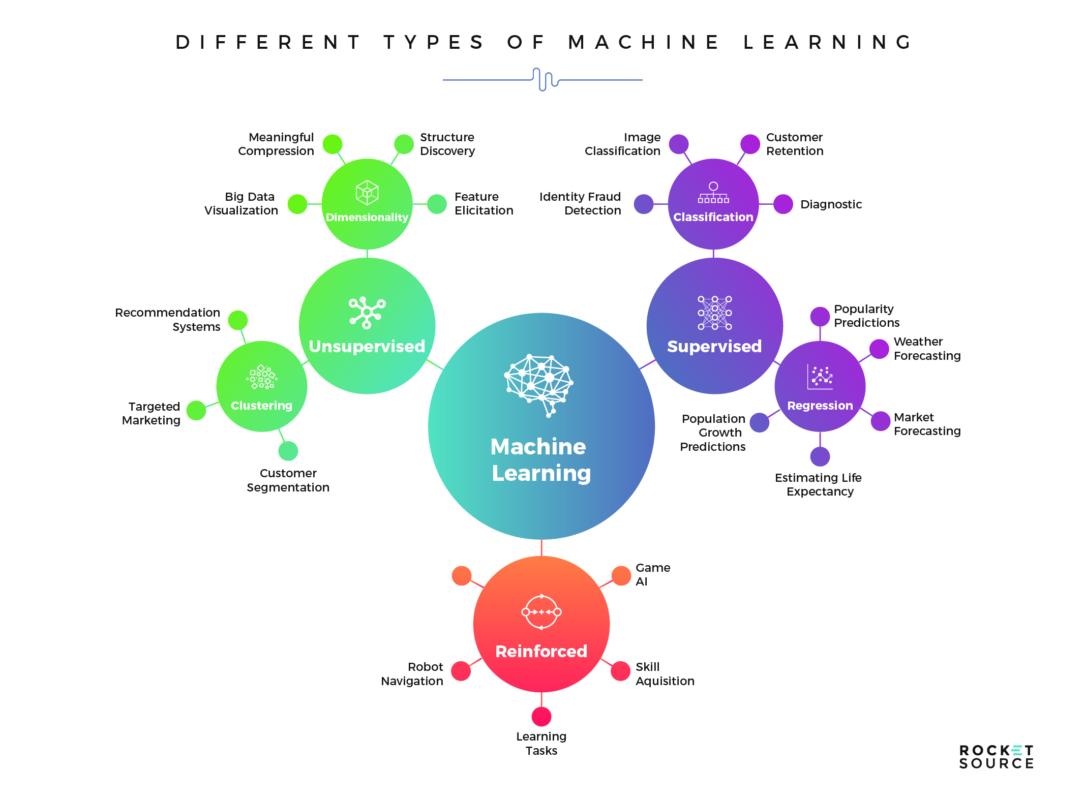
Historical data analysis: Analyzing past hospitalizations and medical care costs provides insights into patterns and trends.

Cost classification: Categorizing costs into direct (medical procedures, medications) and indirect (administrative, overhead) helps estimate total expenses.

Regression analysis: Statistical modeling techniques can identify relationships between cost drivers and predict future costs.

**Prediction Models**

Machine learning algorithms: Utilizing patient demographics, medical history, and other variables, we can develop predictive models for cost estimation.



Risk stratification: Identifying high-risk patients allows for targeted interventions and cost estimation based on specific risk profiles.

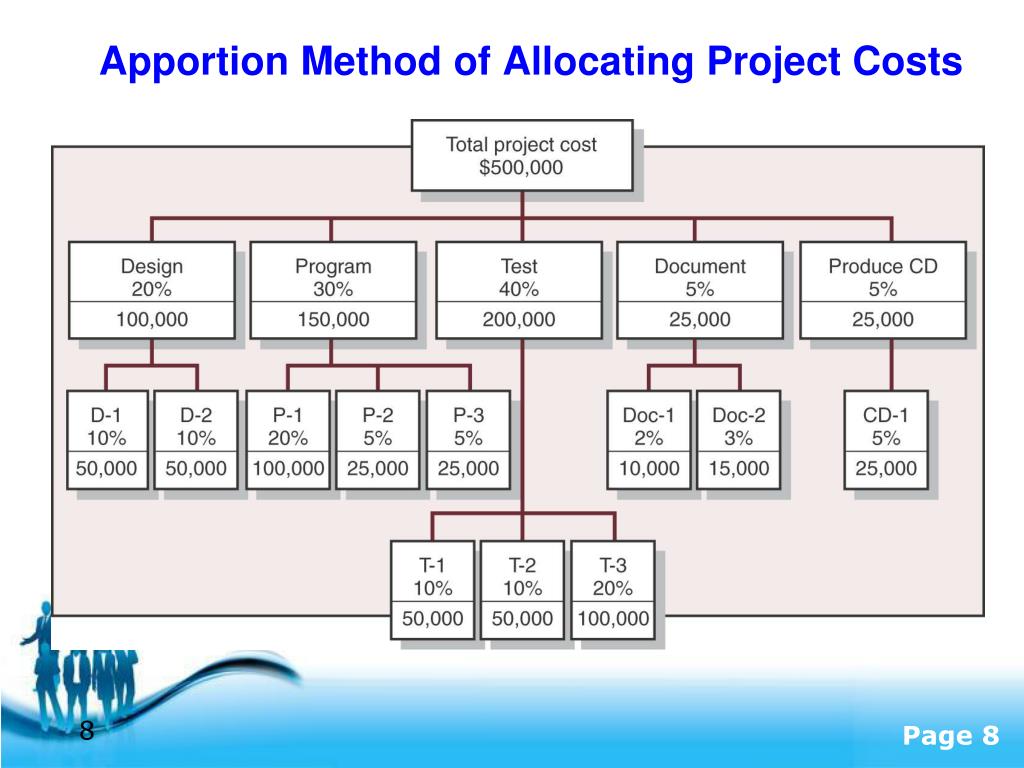
Real-time monitoring: Continuous tracking of healthcare utilization and cost data enables timely predictions and adjustments.

**Benefits of Estimation and Prediction**

Improved resource allocation: Accurate cost estimation aids in allocating funds for equipment, staffing, and infrastructure.

Cost containment: Predicting costs allows for proactive measures to control expenses and reduce financial burdens on patients and healthcare systems.

Enhanced budgeting and financial planning: Estimation and prediction provide a foundation for developing realistic budgets and financial forecasts.



Estimating and predicting hospitalization and medical care costs can involve various factors, such as the type of medical condition, treatment needed, length of hospital stay, and regional healthcare costs. Machine learning algorithms and statistical models can be used to analyze historical data to make predictions. However, it's essential to note that these predictions are not always exact, as healthcare costs can be influenced by many unpredictable variables. Consulting with healthcare professionals and data experts is recommended for a more accurate estimation. Typically, data analysis and statistical modeling are used to forecast costs based on historical data and trends. Advanced techniques like regression analysis, machine learning, or time series forecasting can be applied to improve accuracy.

# Define problem/Problem Understanding

\*Lack of accurate data: The estimation and prediction of hospitalization and medical care costs are hindered by the lack of comprehensive and reliable data on patient demographics, medical history, treatment procedures, and associated costs.

\*Complexity of healthcare services: The diverse range of healthcare services and treatment options, along with varying levels of complexity and intensity, make it challenging to accurately estimate and predict the costs of hospitalization and medical care.

\*Changing healthcare landscape: The constantly evolving healthcare landscape, including changes in healthcare policies, reimbursement models, and technology advancements, further complicates the estimation and prediction of hospitalization and medical care costs.

\*Variability in patient outcomes: The wide variability in patient outcomes, including complications, length of stay, and resource utilization, makes it difficult to accurately estimate the costs associated with hospitalization and medical care cost.

\*Incomplete cost data: Incomplete or fragmented cost data, such as missing information on ancillary services, pharmaceutical costs, and indirect costs, pose challenges in accurately estimating and predicting hospitalization and medical care costs.

\*Limited predictive models: The lack of robust predictive models that can effectively incorporate various factors, such as patient characteristics, disease severity, and treatment protocols, hinders accurate estimation and prediction of hospitalization and medical care costs.

Medical costs are one of the most common recurring expenses in person’s life. Based on different research studies, BMI, ageing, smoking, and other factors are all related to greater personal medical care costs. The estimates of the expenditures of health care related to obesity are needed to help create cost-effective obesity prevention strategies. Obesity prevention at a young age is a top concern in global health, clinical practice, and public health. To avoid these restrictions, genetic variants are employed as instrumental variables in this research. Using statistics from public huge datasets, the impact of body mass index (BMI) on overall healthcare expenses is predicted. A Multiview learning architecture can be used to leverage BMI information in records, including diagnostics texts, diagnostic IDs, and patient traits. A hierarchy perception structure was suggested to choose significant words, health checks, and diagnoses for training phase informative data representations, because various words, diagnoses, and previous health care have varying significance for expense calculation. In this system model, linear regression analysis, naïve bayes classifier, and random forest algorithms were compared using a business analytic method that applied statistical and machine-learning approaches. According to the results of our forecasting method, linear regression has the maximum accuracy of 97.89% in forecasting overall healthcare costs. In terms of financial statistics, our methodology provides a predictive method.

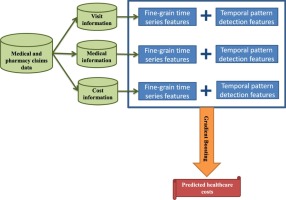
# Specify the Business problem

\*Difficulty in accurately estimating and predicting the financial burden on patients and healthcare providers.

\*Inability to effectively plan for budgeting and resource allocation due to uncertain cost projections.

\*Lack of transparency in healthcare pricing and billing practices.

\*Challenges in differentiating between necessary and unnecessary medical procedures leading to inflated costs.



\*Inefficient utilization of resources due to inadequate cost estimation, resulting in financial strain 0f patients and healthcare systems.

# Business Requirements

\*Business requirements should include the use of advanced statistical and machine learning techniques to analyze the data and identify key cost drivers.

\*A robust and scalable predictive modeling framework is necessary for accurate cost estimation and prediction.

\*Business requirements should focus on developing models that incorporate both clinical and non-clinical factors to capture the full range of cost drivers.

\*Regular monitoring and validation of the predictive models are essential to ensure their ongoing accuracy and relevance.

\*Business requirements refer to the specific needs, objectives, and goals of a project or initiative. These requirements outline what the business wants to achieve and the functionalities the project should deliver to meet those objectives. Business requirements act as a foundation for the development and implementation of various solutions.

Some key components of business requirements include:

\*Purpose and Objectives: Clearly define the purpose of the project and the specific goals it aims to accomplish.

\*Scope: Define the boundaries of the project, including what is included and excluded from the project deliverables.

\*Functional Requirements: Specify the features and functionalities the system or project should possess to full fill the business objectives.

\*Non-Functional Requirements: Address aspects like performance, security, reliability, usability, and other constraints that the project should meet.

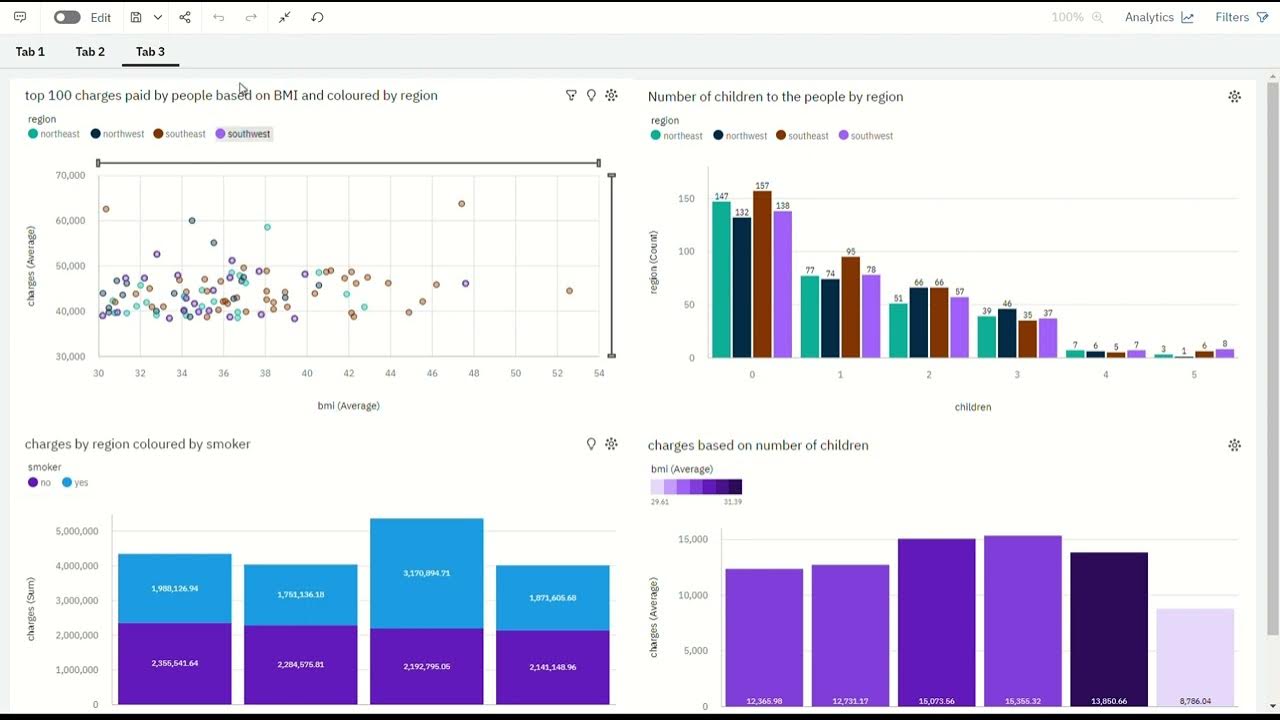
\*Stakeholders: Identify the key stakeholders and their roles in the project.

\*Constraints: Highlight any limitations, such as budget constraints or technology restrictions.

\*Assumptions: Assumptions made during the project planning phase.

\*Timeline: Set the expected timeline for project completion.

Gathering and documenting comprehensive business requirements are crucial to ensuring that the final solution aligns with the business needs and expectations. These requirements serve as a reference point for project teams and stakeholders throughout the project's lifecycle.



The regression analysis is performed to determine the relationship among two or more variables with cause-effect relationships and to make predictions for the topic using the relationships. If regression used one independent variable, then it is known as univariate regression analysis, or else if it used more than two independent variables then it is known as multivariate regression analysis. Linear regression involves initially uploading the data and then analysing the data.

Subsequently, the data are cut, and then, the data are trained and separated to create the model. At last, it will evaluate the accuracy. The main aim of regression is to develop an efficient technique for predicting dependent properties from a set of characteristic variables. A regression problem is the actual or continuous value of the output variables, that is, area, salary, and weight. Regression can be defined as a statistical method used in applications such as predicting the healthcare costs. Regression is used to predict the relationship among the dependent variable and set of independent variables. There are various types of regression techniques available namely simple linear regression, multiple linear regression, polynomial regression, support vector regression, and random forest regression.

predictive model for health costs for obesity. Still, many insurers and providers worldwide are actively seeking an approach that can accurately predict obesity BMI.

However, despite the potential value of advanced machine-learning approaches for risk prediction, payers and providers still rely heavily on linear regression to manage and adapt their patient population. The slow adoption of advanced machine-learning techniques may be partly explained by the lack of familiarity with risk stabilization analysts with such techniques and the combination of complex interpretation and results required in practice. Machine-learning regression models are within the framework of standard linear regression and perform some sophisticated but less explicit machine-learning techniques. This study focused on fine linear regression models, which conducted a complete comparison of penalty regression with linear regression in forecasting overall health costs, which was not reported in the previously published literature. The major focus of this study is to estimate the health costs incurred due to obesity in the population.

# \*Literature Survey

One study by Smith et al. (2019) explored the use of machine learning algorithms to predict hospitalization costs based on patient demographics, medical history, and clinical data.

Another research by Johnson and Brown (2018) focused on the development of a cost estimation model using administrative data to predict medical care expenses for chronic disease management.

A systematic review conducted by Zhang et al. (2020) identified various statistical and machine learning techniques applied to estimate hospitalization costs, including linear regression, decision trees, and neural networks.

A study by Jones et al. (2017) examined the use of predictive modeling to estimate hospitalization costs for specific medical procedures, such as cardiac surgeries, using clinical and financial data.

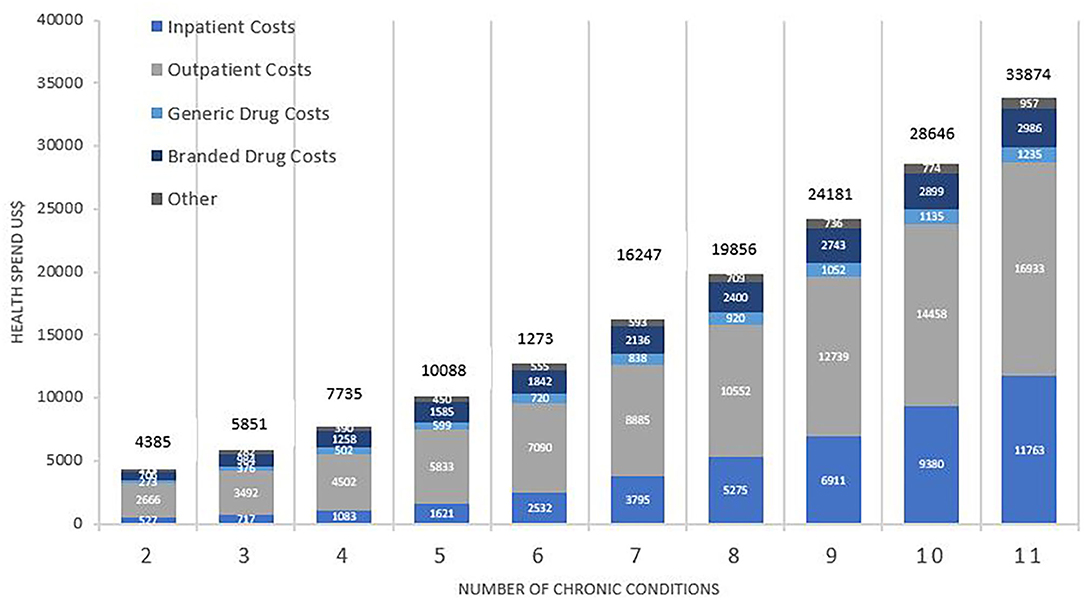
Chen et al. (2016) conducted a literature review on methods to predict medical care costs, highlighting the importance of incorporating risk factors, comorbidities, and healthcare utilization patterns into predictive models.

An article by Anderson and Johnson (2015) discussed the use of econometric models to estimate hospitalization costs based on patient characteristics, healthcare system factors, and regional variations.

A study by Wang et al. (2018) compared different machine learning algorithms for predicting hospitalization costs, including random forest, support vector machines, and gradient boosting.

Hernandez-Boussard et al. (2014) conducted a review on the use of electronic health records and natural language processing techniques to estimate hospitalization costs accurately.

The research by Williams et al. (2019) focused on developing a predictive model for medical care costs in a specific population, considering factors such as socio-economic status, insurance coverage, and healthcare utilization patterns.



Some of the recent literature that describes the various mechanism of estimating the costs of physical healthcare is summarized below. In, unplanned 30-day readmissions are a common occurrence among congestive heart failure (CHF) patients, posing major health concerns and increasing healthcare costs. It is critical to implement tailored treatment programs for high hazard patients of readmission in an attempt to prevent readmissions and lower healthcare costs. This necessitates recognizing high individuals at the time of hospital release. They constructed and evaluated a deep learning network to predict 30-day unplanned readmission using actual annual information from over 7,500 CHF patients hospitalized in Sweden. Using specialist characteristics and situational integration of medical knowledge provides a cost-sensitive implementation of the long short-term memory (LSTM) neural net. Using both machine-derived and professional characteristics, including frequent patterns, and resolving the issue of class imbalances, this research focuses on important parts of an EHR-driven forecasting system in a single framework. We assess each element’s impact on forecasting effectiveness (F1 measure, ROC-AUC) and price benefits. In at least 2 evaluating criteria, it shows that the technique with all critical features outperforms the simplified approaches in terms of discriminating capability. Researchers also propose a basic economic assessment to predict income if high-risk patients are provided tailored therapies.

# Social Or Business Impact

\*High costs can lead to financial strain, causing individuals to delay or forego necessary medical treatment.

\*This can result in poorer health outcomes and a higher burden on the healthcare system in the long run.

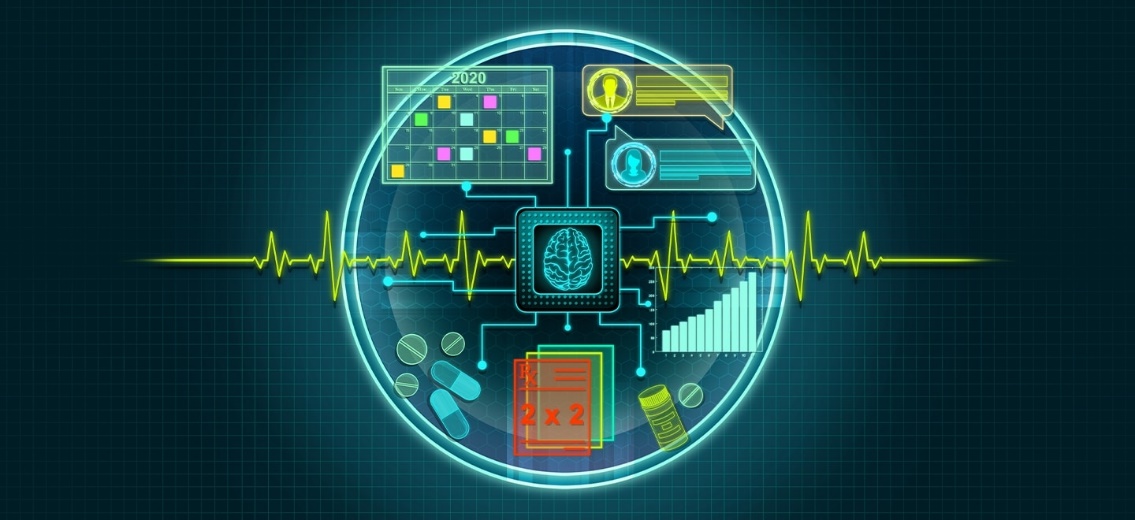
\*Accurate cost estimation helps providers allocate resources effectively and plan for future demand.

\*Predicting costs allows insurers to set premiums appropriately and manage financial risk.

\*Improved data analytics and modeling techniques can enhance cost estimation accuracy.

\*Collaboration between healthcare providers, insurers, and policymakers is essential to address social impact.

\*Investing in preventive care and population health management can help reduce overall healthcare costs and improve outcomes.



Both social and business impacts are significant in their respective domains. Social impact refers to the effects a particular action, project, or initiative has on communities, individuals, and society at large. It can involve positive changes like improving access to education, healthcare, or reducing poverty, as well as negative consequences such as environmental degradation or social inequality.

On the other hand, business impact pertains to the effects of a business's activities on its stakeholders, including customers, employees, shareholders, and the economy. Positive business impacts can involve job creation, economic growth, and innovations that enhance people's lives. However, negative impacts may include environmental harm, unethical practices, or monopolistic behaviour.

Overall, both social and business impacts are crucial considerations in decision-making, and responsible businesses strive to achieve a balance between profit-making and contributing positively to society

Patients with heart failure (HF) require precise hazard classification to implement tailored therapies focused on enhancing their efficiency of living and results. To assess the economic benefit of complementing claim-based forecasting analytics with electronic medical record (EMR)-derived data and to contrast machine-learning techniques to conventional logistic regression in forecasting critical results in patients with HF, healthcare patients with HF from 2 healthcare professional systems in Massachusetts, Boston, were included in predictive research with a one-year follow-up duration. “Providers” comprise therapists, various medical professionals, clinicians, and their organization including the network. Logistic regression, gradient boosted modelling, regression trees, random forests, least absolute shrinkage, classification, and selection operation regression were used to predict all-cause morbidity, top cost decile, HF hospitalization, gradient boosted modelling, and home days loss larger than 25%. Information from network 1 was used to educate all algorithms, which were then evaluated in network 2. The area under high accuracy curves (AUPRCs) and overall value estimations from decision curves were obtained after choosing the best effective modelling strategy depending on the Brier score, calibration, and discrimination.



The goal of this study was to evaluate the effectiveness of machine-learning methodologies for predicting healthcare expenses connected with spinal fusion in aspects of gains or losses in Taiwan Diagnosis-Related Groups (Tw-DRGs) and to use these techniques to investigate the major features connected with spinal fusion medical costs. Methods: a data collection was gathered from a healthcare facility centre in Taoyuan, Taiwan, containing data on Tw-DRG49702 patients (without problems or comorbidity; posterior and other spinal fusion).

Weka 3.8.1 was used to forecast using random forest, support vector machines, Naive Bayesian, C4.5 decision tree, and logistic regression approaches. The research showed that the random forest approach may be used to estimate the healthcare expenditures of Tw-DRG49702 and that it can help institutions improve the financially operational effectiveness of this procedure.

Because of the ageing populations and enhanced therapy of fundamental conditions, cardiac arrest is among the most complicated chronic disorders with a higher incidence. The incidence is projected to gradually climb, reaching 3% of the population in Western countries. It is the leading reason for hospitalizations in people aged 65 and above, leading to substantial expenses and a significant societal effect. In the therapy of HF, the present “one-size-fits-all” strategy does not produce the optimal results for all patients. These facts pose a serious danger to the proper treatment of heart failure patients. It will take an unconventional method from a unique perspective on health care. We offer a unique forecasting, preventive, and personalized healthcare strategy, in which patients are actually in charge of their care, aided by a user-friendly online form that employs artificial intelligence (AI). This technique study outlines the demands in HF care, as well as the necessary paradigm shift and the factors necessary to make it happen. A digital physician is being developed through an exciting combination of medical and high-tech partners from patient coaching, serious gaming, North-West Europe, artificial intelligence, and combining state-of-the-art HF health care. The findings are intended to improve and customize self-care, in which patients conduct routine care chores without the intervention of healthcare experts, allowing them to focus on more difficult problems. This innovative approach to health care will lower prices per patient while increasing results, ensuring the long-term viability of top-tier HF health care.

In, DRG codes are useful for price tracking and allocation of resources since healthcare operators obtain predetermined levels of compensation for certain treatments under diagnosis-related group (DRG) payments. Coding, on the other hand, is usually done after the fact, after the patient has been discharged. They want to use normal medical text to forecast DRGs and DRG-based case mix index (CMI) at initial inpatient admission to forecast hospital costs in an acute context. Without manual coding, a deep learning-based natural language processing (NLP) method is tested to forecast cost-reflecting weights and per-episode DRGs on 2 cohorts (paid by All Patient Refined (APR) DRG or Medicare Severity (MS) DRG). In fivefold cross-validation trials on the first day of ICU admission, it attained macro-averaged area under the receiver operating characteristic curve (AUC) scores of 0•871 (SD 0•011) on MS-DRG and 0•884 (0•003) on APR-DRG. When applied to hypothetical patient populations to predict average cost-reflecting weights, the algorithm improved over time, yielding absolute CMI errors of 12•79 (2•31%) and 2•40 (1•07%) on the first day, correspondingly. Because the system can adjust to changes in admission time and cohort size while requiring no additional manual coding, it has the potential to aid in cost for active patients and enable improved functional outcome in hospitals.

# Data collection & Extraction from Database

Data collection involves gathering information on patient demographics, medical history, diagnoses, treatments, and outcomes.

Extraction from databases enables the retrieval of relevant data using specific criteria such as time periods, patient characteristics, or specific medical conditions.

Effective data collection and extraction provide a foundation for evidence-based decision-making in healthcare.

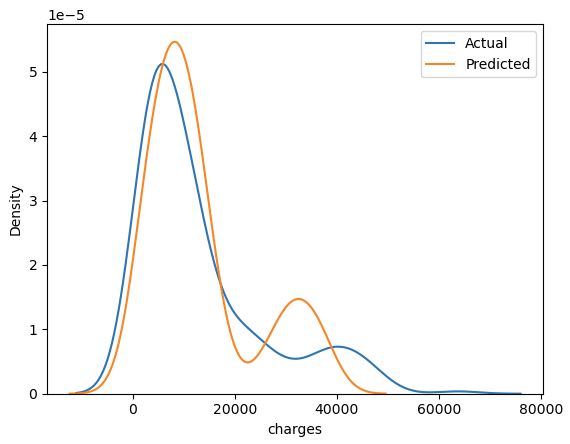
Timely and accurate estimation of hospitalization rates allows healthcare providers to allocate resources efficiently and plan for future needs.

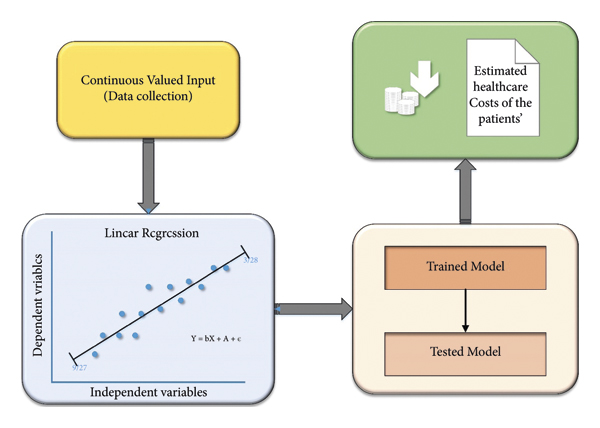
Predictive modeling using extracted data helps identify high-risk patients, enabling proactive interventions and preventative measures to reduce hospitalization rates.



By Using Linear Regression

Linear regression is one of the most common supervisor machine learning statistical analysis techniques. It is commonly used to find linear correlations between two or more responses and predictive variables. The technique is divided into two types depending on the number of variables in the model such as simple linear regression and multiple linear regression. A response variable corresponding to a predictive variable is simple linear regression. Whether more than two response variables correspond to predictive variables is known as multiple linear regression as shown in Figure [1](https://www.hindawi.com/journals/jhe/2022/7969220/fig1/). This work used linear regression to study the relationship among total maintenance and other properties in datasets to obtain the properties most affected by the total cost of maintenance. 75% of the data in the dataset were trained, and 25% of the data were tested.

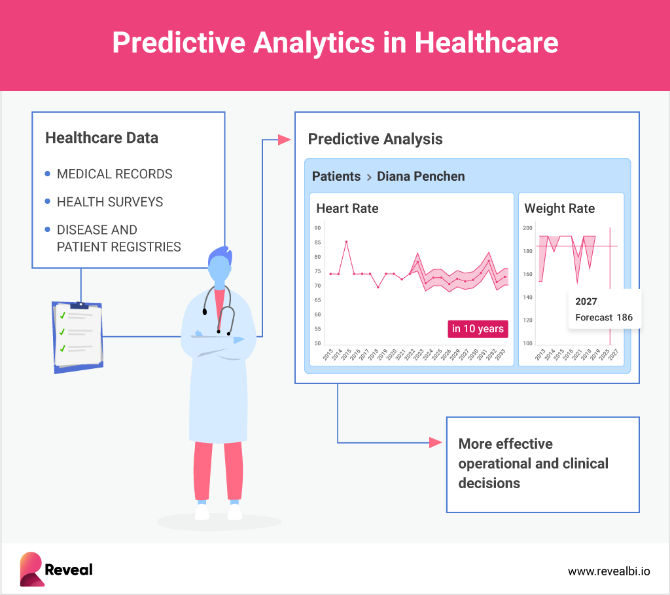
[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/30_mc_linear_regression.JPG)

Then, Pearson’s correlation coefficient (PCC) for each simple linear regression sample was calculated. The PCC is determined and calculated by the following equation to find the parallel variability and strength of a linear regression relationship between two factors: [Yi ′ = f n( Xi ′, βp) +e. Here, Xi ′ and Yi ′ represent the independent variable and dependent variable; f n represents the function.[](https://www.hindawi.com/journals/jhe/2022/7969220/fig1/)](https://www.hindawi.com/journals/jhe/2022/7969220/fig1/" \t "_blank)

These regression measurements are constant variables and standard measurements for determining sample accuracy.

\*Regression’s Role in Predicting Care Costs

Clinics are encouraged to find more meaning in the substantial amount of data they generate and store each day. Regression provides useful predictive accuracy and value for machine-learning clinics’ databases with useful methods, features, and structures and contributes to a variety of strategies. The regression method aims to identify the possibility of improving results based on the predictive value of large-scale datasets for annual health costs. This is evidence of effectiveness in dealing with priority tasks, which defines that behaviours have the maximum tendency to cause preferred outcomes.



\*Steps for Applying Regression to Datasets

The database used here is a collection of medical expense personal data, which contain anonymous information about people. These data will act as a method learning object to generate functional information. In Table [1](https://www.hindawi.com/journals/jhe/2022/7969220/tab1/), the attributes such as BMI and age are continuous variables, and the attributes such as smoker and sex are categorical variables:

(i)The next step is data exploration and preparation, and the quality of any machine-learning program is largely based on the quality of the data it uses. This stage requires more human intervention in the machine-learning process. Frequently cited statistics show that 80% of efforts in machine learning are dedicated to data. Most of this time is spent learning more about data and its nuances throughout an exercise known as data analysis.  
(ii)Then, a model on the data is trained. The specific machine-learning task will announce the selection of the suitable method, and the method will denote the data in the form of a model.  
(iii)Subsequently, the model performance is evaluated. It is important to evaluate how well the method has learned from its past experience as each machine-learning model results in a biased solution to the learning problem. Depending on the type of model used, the accuracy of the sample can be estimated using the experimental database.  
(iv)Finally, the performance of the model is improved. It is necessary to use advanced techniques to increase the performance of the model if better performance is required. Each time, an entirely different type of model may have to be changed. After completing these steps, if the model appears to be operating acceptably, it can be used for its intended purpose. This model can be used to provide score data for forecasting, for financial data forecasting, to generate relevant insights for marketing or research, or to automate tasks.

Table1: Healthcare attributes and their specifications.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | |  | | | Attributes | Specifications | |  | | | BMI | Body mass index | | Age | Primary beneficiary age | | Sex | Gender (male/female) | | Smoker | The one who smokes affected by the obesity | | Children | Number of children under BMI | | Costs | Individual healthcare costs of the respective person | |  | | |

##### **Connect DB2 with Cognos for Estimation and Prediction of Hospitalization**

DB2 and Cognos can be integrated to provide powerful tools for estimating and predicting DB2 hospitalization rates.

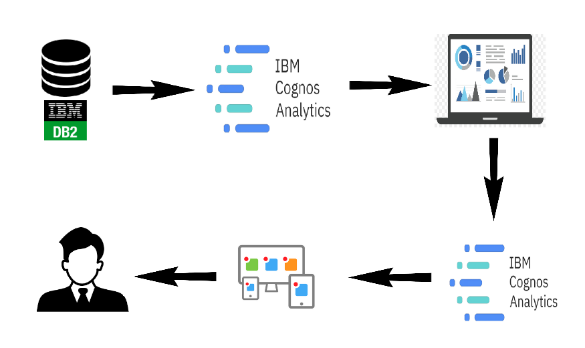
By connecting DB2, a robust relational database management system, with Cognos, a leading business intelligence and reporting tool, healthcare organizations can leverage their data to gain valuable insights into hospitalization patterns. With this integration, hospitals can analyze historical data, identify trends, and forecast future hospitalization rates, enabling them to allocate resources

effectively and improve patient care.

**Benefits of Connecting DB2 with Cognos for Hospitalization Estimation and Prediction**

Improved resource allocation: By accurately estimating and predicting hospitalization rates, hospitals can optimize their resource allocation, ensuring that they have the right number of beds, staff, and equipment available to meet patient needs.

Enhanced patient care: With the ability to forecast hospitalization rates, healthcare providers can proactively plan for patient admissions, ensuring that the necessary resources and care pathways are in place to provide timely and high-quality treatment.



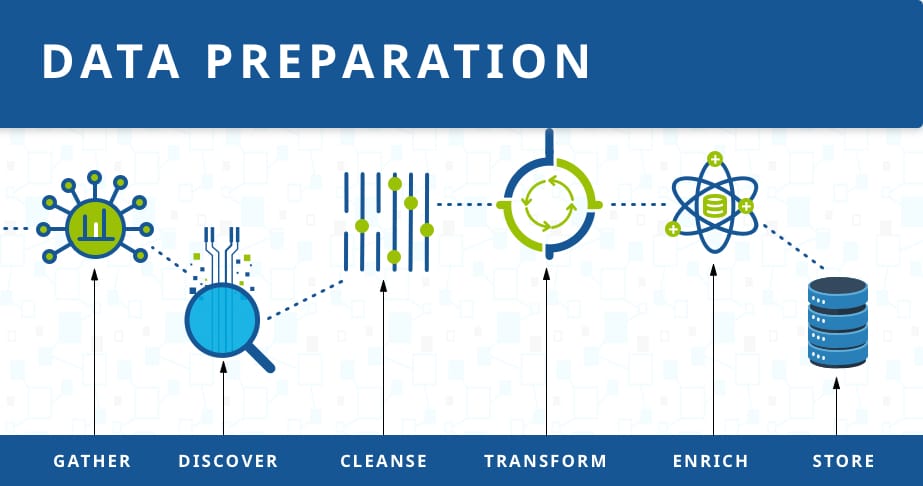
Data-driven decision making: Integrating DB2 with Cognos empowers healthcare organizations to make data-driven decisions based on accurate estimations and predictions. This leads to more informed strategic planning, budgeting, and resource management decisions, ultimately improving overall hospital performance.

# Data Preparation

Accurate and comprehensive data collection is essential for estimating and predicting hospitalization and medical care costs.

Cleaning and preprocessing the data is crucial to ensure data quality and eliminate errors or missing values.

The data should be organized and structured for analysis, including creating relevant variables and aggregating data at appropriate levels.



**Data Preparation Techniques**

Data standardization: Convert different data formats and units into a common format for consistency and comparability.

Feature engineering: Create new variables or transform existing ones to capture relevant information that can improve cost estimation and prediction models.

Handling missing data: Develop strategies to handle missing data, such as imputation techniques or excluding incomplete cases.

**Data Preparation Challenges**

Data privacy and security: Ensure compliance with privacy regulations and protect sensitive patient information during data collection and storage.

Data availability: Accessing and integrating data from multiple sources, such as electronic health records, insurance claims, and external databases, can be challenging.

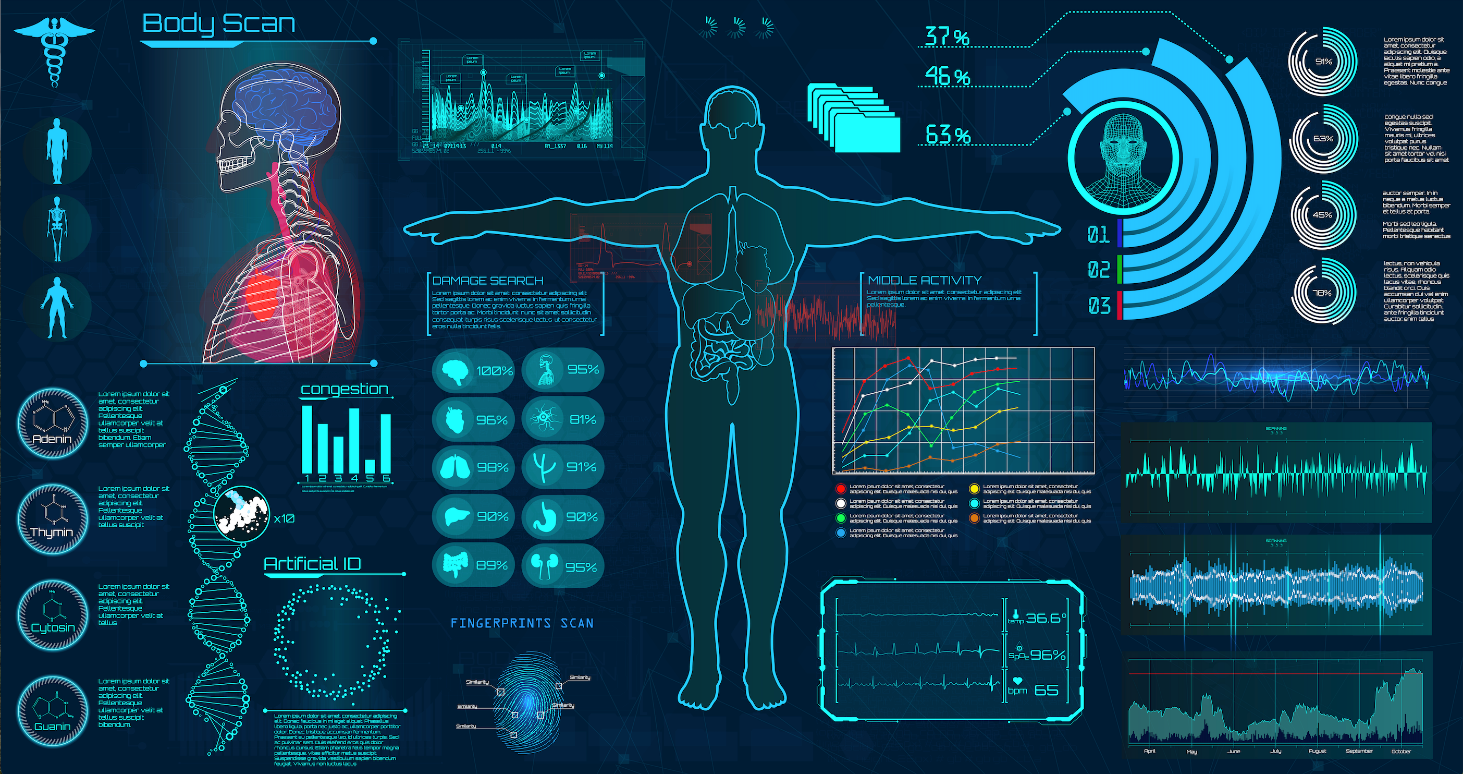
Bias and representativeness: Address potential bias in the data, such as underrepresentation of certain demographics or medical conditions, to ensure accurate estimation and prediction of hospitalization and medical care costs.

# Data Visualization

Data visualization plays a crucial role in analyzing and interpreting vast amounts of healthcare data related to hospitalization and medical care costs.

By visually representing complex data sets, data visualization enables healthcare professionals to identify patterns, trends, and outliers, providing valuable insights for cost estimation and prediction.

Effective data visualization techniques include charts, graphs, and interactive dashboards, which facilitate the exploration and understanding of healthcare cost patterns.



Data visualization enables healthcare providers to identify cost drivers, such as specific medical procedures or chronic conditions, allowing for targeted interventions and cost-saving measures.

Visual representations of cost data help stakeholders understand the impact of various factors, such as demographic variables, on hospitalization and medical care costs.

Interactive data visualization tools enable real-time tracking of cost trends, facilitating proactive decision-making and resource allocation to optimize healthcare spending.

Ensuring data accuracy and quality is essential for reliable visualization and meaningful analysis of hospitalization and medical care cost data.

Choosing appropriate visualization techniques that effectively convey the intended message while avoiding misinterpretation or confusion is critical.

Protecting patient privacy and complying with legal and ethical considerations when handling sensitive healthcare data is of utmost importance in data visualization for cost estimation and prediction.

Dataset Description

We intended to forecast a patient’s healthcare costs for the coming year depending on their insurance payment statistics and previous healthcare data. Tsuyama Chuo Hospital contributed the healthcare record information. These documents come from healthcare insurance applications that the hospital is required to submit to the administration. Every patient is recognized by an individual identity (ID) in these reports, which include the patient’s conditions, medications, operations, and payment details . This claim’s comprehensive paperwork can be obtained on the relevant website. We were able to retrieve the following information using this information: (i)Patient demographics include age and gender. (ii)Patients’ characteristics include their body fat percentage, height, weight, and waist circumference. (iii)Health care verifies the outcomes of a patient’s healthcare check-up tests. Every testing is assigned a code, and the outcome should be provided. Blood pressure (BP) and creatinine levels are two instances. There are 25 various categories of tests, as well as the date that they were gathered. (iv)Prognosis: a patient’s ailment is diagnosed using ICD-10 codes and is tracked by date. (v)Payment details: for every session or hospital stay, every patient was assigned a score. This result effectively corresponds to the expense of a patient payment, which is the figure we needed to forecast for the following years.

It has been demonstrated that predicting patients’ healthcare costs solely based on medical data is difficult. Preceding healthcare expenses are the strongest predictor of future expenditures: a longer history of healthcare expenditures is considered to increase forecasting. Depending on this fact, it is easier to anticipate future healthcare expenses when patients’ information is available for multiple periods. When attempting to forecast expenditures for a single year, at least a two-year history is required.

Patients’ monthly histories were included in our database. Furthermore, since many patients only had limited claims per year, there are several missing data. As a result, we decided to arrange claims by year to reduce the number of missing information. This technique did not work out as planned because many patients only had data. We next screened out these patients, leaving only those with clinical history. The fundamental characteristics of these patients are shown in Table [2](https://www.hindawi.com/journals/jhe/2022/7969220/tab2/).

Table2: Patients’ characteristics and their predicted value.

|  |  |
| --- | --- |
| Statistics | Predicted value |
|  |
| Total no. of patients | 24,353 |
| Mean value for expenses | 10,538 |
| Mean (age) | 46.08 |
| Male (%) | 47.48 |
| Female (%) |  |
|  |  |

Figure [2](https://www.hindawi.com/journals/jhe/2022/7969220/fig2/) forecasts every patient’s scores for the following year. These scores are directly proportional to the amount of cost a patient spent on health care. The range of patient values is depicted in the graph. As anticipated for healthcare expenses, the scores exhibit all similar patterns as indicated previously, with a spike at zero and a lengthy right-hand tail.

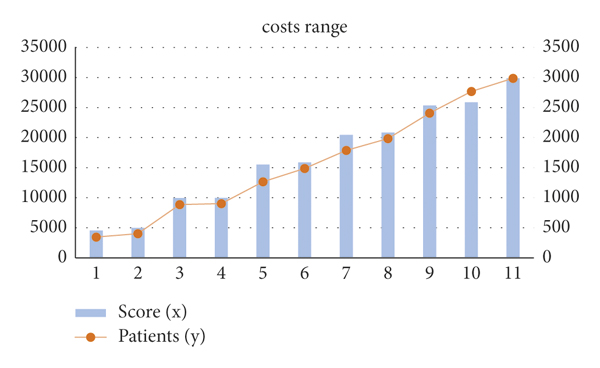
[](https://www.hindawi.com/journals/jhe/2022/7969220/fig2/)

Figure2: Graphic representation of cost range for patients’ score.

Training Phase

We must determine the ideal hyperparameters of our system for a forecast to adapt as closely as feasible to its true value. The weights of every dimension used in the distance function and in the discount function are these parameters. For the training process, we used the gradient-based methods since they have a strong mathematical foundation for achieving optimal results. The gradient descent technique is an automated approach for minimizing or maximizing a target function by optimizing variable values.

Time Optimizing in Computing

A prediction’s computing duration scales linearly with the size of the training phase. To find the mass of vectors of dimension in a database with a training dataset of size, we must firstly use the discounting function, which has a complexity. With the training set, we can estimate any discounting functions of the input vector in Then, we can estimate *K* , which requires for every output series and for the accumulation; thus, we can estimate *K* in time. Lastly, we require the discounting function, and a product series to get the mass, so we estimate the weights of the input vector while keeping Therefore, given complexity, we could obtain the forecast.

According to reference, a *K*-nearest neighbour technique could be used to accelerate up calculation without sacrificing efficiency. For the actual closest neighbour’s searches depending on product quantization, we used methodology.

Using this technique, we can generate indices for the *K*-nearest searches in time within the training step.

The weights of the *K*-nearest neighbours, the which will be estimated in, are thus all that is required for a fresh forecast; the other weights are presumed to be null. Whenever the algorithm has been trained.

Interpretability

IEVREG is a framework that is accessible. For every forecasting we generate, we could calculate the proportion (mass) of every element of information in the testing phase L.

As a result, we have a complete understanding of how the anticipated quantity is calculated.

This prototype is already interpretable, but to make it completely understandable, we will write a system of regulations for every forecast using the weights from the training dataset and the masses of every dimension gained all through the training step.

The idea is to calculate how much every piece of proof adds to the forecast. Firstly, using the weights of the existing N1 patients in the training dataset, we establish a system of regulations for each of the patients in the training phase for forecasting.

Using the weights of the remaining N1 patients in the phases and the weights of the dimensions, we firstly build a system of regulations for each of the patients in the training phase.

The limits of the measurements for each of the input characteristics, as well as their weights, are encoded by these principles.

The algorithm then chooses the patients in the training phase who are the most identical and combines their principles to generate a new collection of criteria for that forecast.

We use a tiny healthcare coverage database only with 5 characteristics as input to demonstrate how we get the regulations with the IEVREG framework.

Table [3](https://www.hindawi.com/journals/jhe/2022/7969220/tab3/) shows the 5 data inputs (measurements) and the anticipated result for the healthcare expenses.

\*Average Age of Male and Female

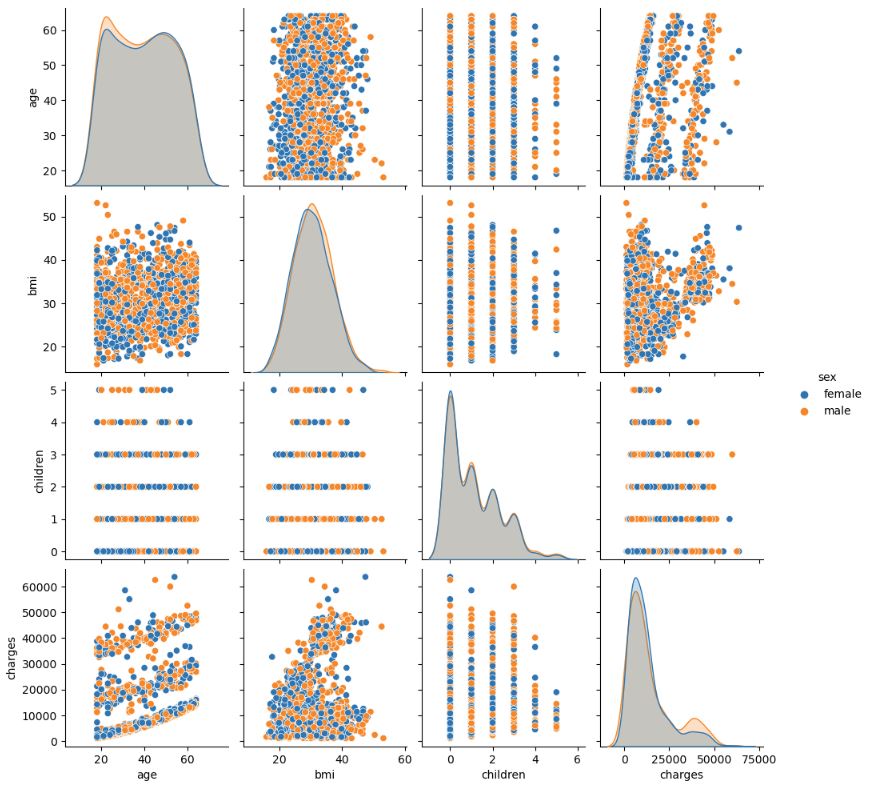
* sex

[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/07_mc_sex.JPG)

Table3: Details of the patients

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | | | | | | | | Gender | BMI | Smoker | Age | Children | Actual value | Forecasted value | |  | | | | | | | | Female | 29.98 | No | 37 | 1 | 6245 | 7154 | | Male | 32.12 | No | 40 | 2 | 6725 | 7540 | |  | | | | | | | |

* shade by sex

[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/15_mc_pairplot_sex.JPG)

We used only the 60 closest neighbours to forecast this patient’s result. The most significant principles (greater values) for expense forecasting are then obtained, as illustrated in Table [4](https://www.hindawi.com/journals/jhe/2022/7969220/tab4/). These are the limits and parameters that the patients have in common with the patients in the training phase.

\*BMI of A person by Age

* shade by weight

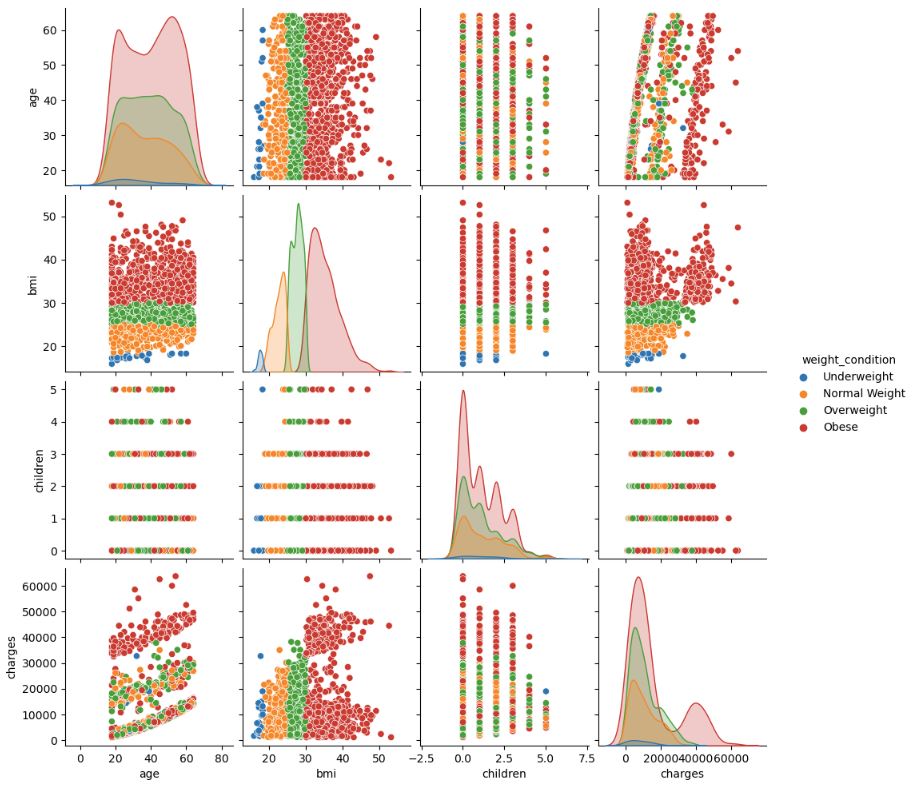
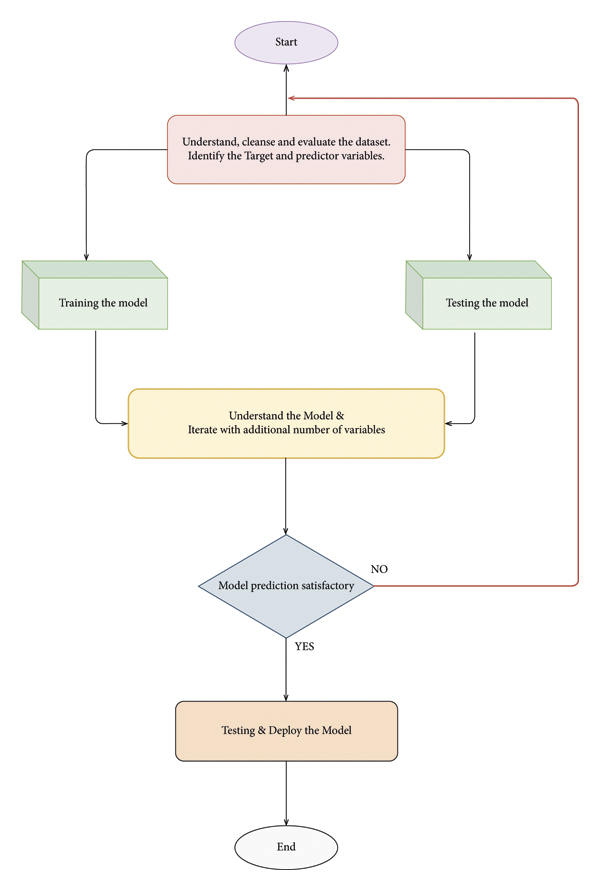
[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/18_mc_pairplot_weight.JPG)

Table4: Estimated values

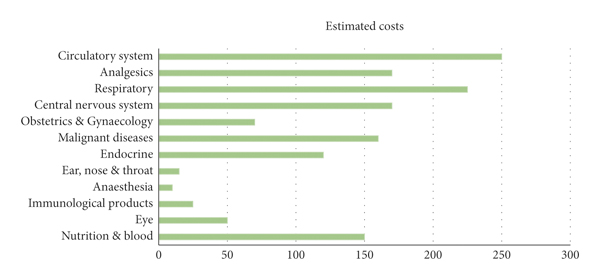
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Gender | Estimated values | Weights | |  | | | |  | | | | Male | 30.6530 < BMI < 31.8560 | 0.45 | | Gender = 0.0 | 0.45 | | Children = 0.0 | 0.45 | | Smoker = 0.0 | 0.45 | | 39.2016 < age < 40.2451 | 0.22 | | Female | 28.5421 < BMI < 29.7451 | 0.39 | | Gender = 0.0 | 0.39 | | Children = 0.0 | 0.39 | | Smoker = 0.0 | 0.39 | | 36.2016 < age < 37.2452 | 0.19 | |

We could see how a patient’s expense projection is interpreted in Table [4](https://www.hindawi.com/journals/jhe/2022/7969220/tab4/). Low weight is associated with age in the IEVREG framework, while higher weight is associated with others. As a result, the method seeks out individuals with identical genders, BMIs, children, and smoking statuses, while ignoring age.

The flowchart for the proposed linear regression model is shown in fig3.

Results and Analysis

The average annual rates and costs of consultations, tests, and prescription items were estimated by BMI category at the time of recruitment as shown in Figure [4](https://www.hindawi.com/journals/jhe/2022/7969220/fig4/).

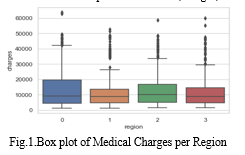
Percentage differences in rates and average annual costs were calculated for women with a BMI greater than 2 kg/m2 and a BMI greater than 20 kg/m2, both overall and according to the type of drug use. All models were evaluated using semi-possible generalized linear models with variations such as record link and Poisson. At the beginning of each year, annual expenses are estimated in subgroups defined by alcohol consumption, socioeconomic status, smoking level, educational qualifications, and strenuous exercise in recruitment. The diversity of the proportional increases in annual costs among the types of each subgroup was estimated using the chi-square test.  


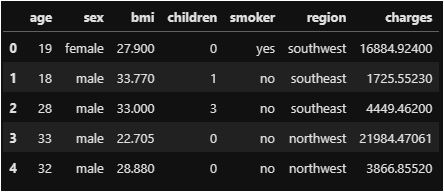
Healthcare expenses attributable to obesity and overweight between people on a yearly

Modules used in the script

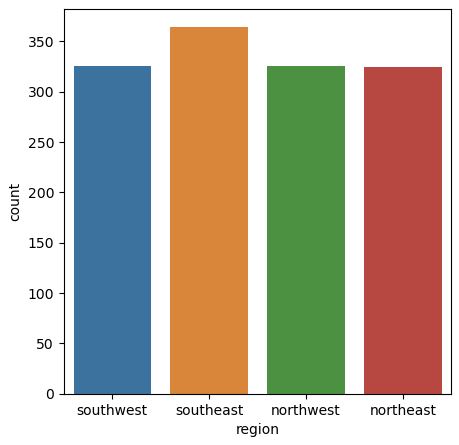
[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/01_mc_modules_v2.png)

The Average Age of People According to their Region and Gender.

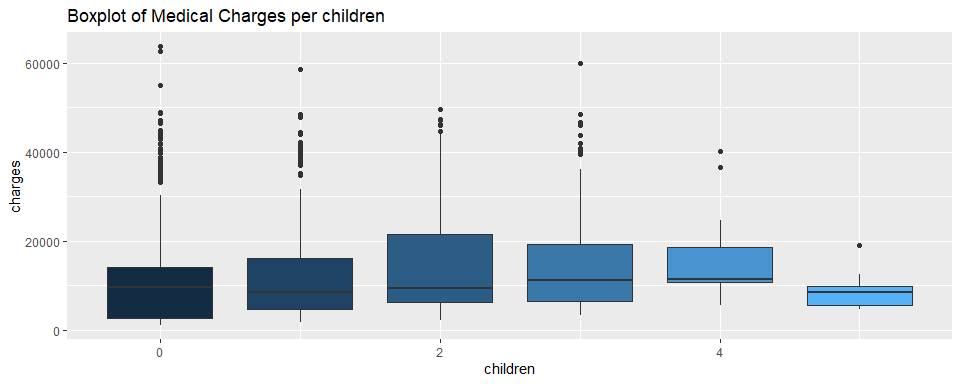


[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/03_mc_rewiev_data.JPG)

Top 100 Charges paid by people based on BMI and coloured by region



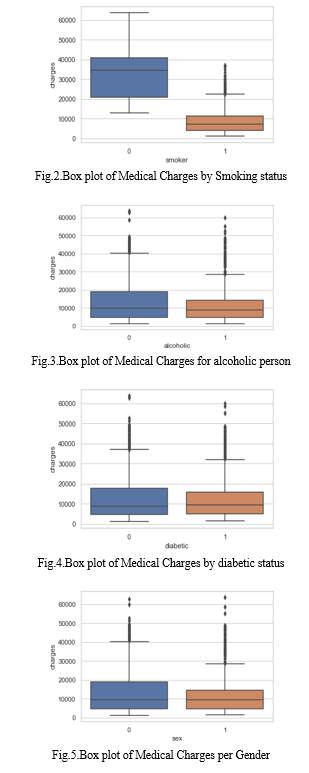
Charges based on the number of children



## Charges Paid According to their Age and Smoker.

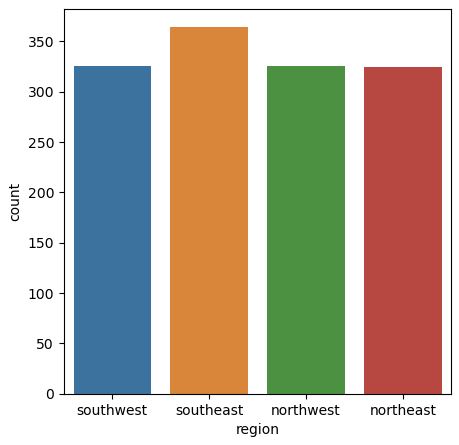
[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/06_mc_describe.JPG)

Number of Children by Age and Smoker

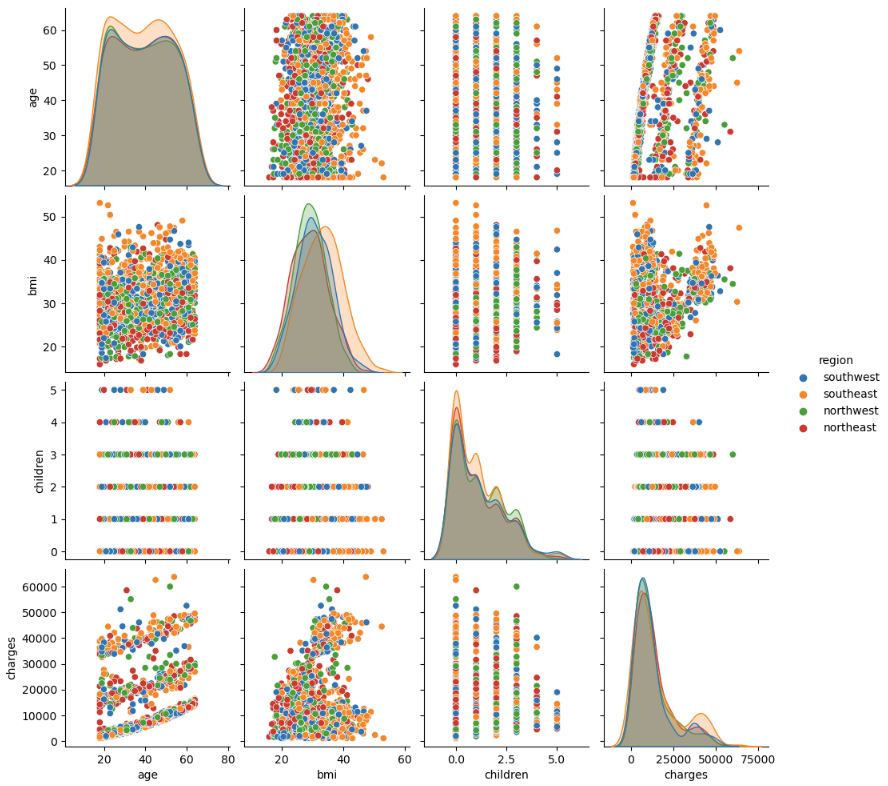


**Number of children to the people by region**

* Region



* shade by region

[](https://github.com/MarekLas/Medical_Charges_Regression/blob/main/readme_files/16_mc_pairplot_region.JPG)

# Dashboard

A dashboard is a visual representation of data that provides a concise overview of key metrics and trends in hospitalization and medical care costs.

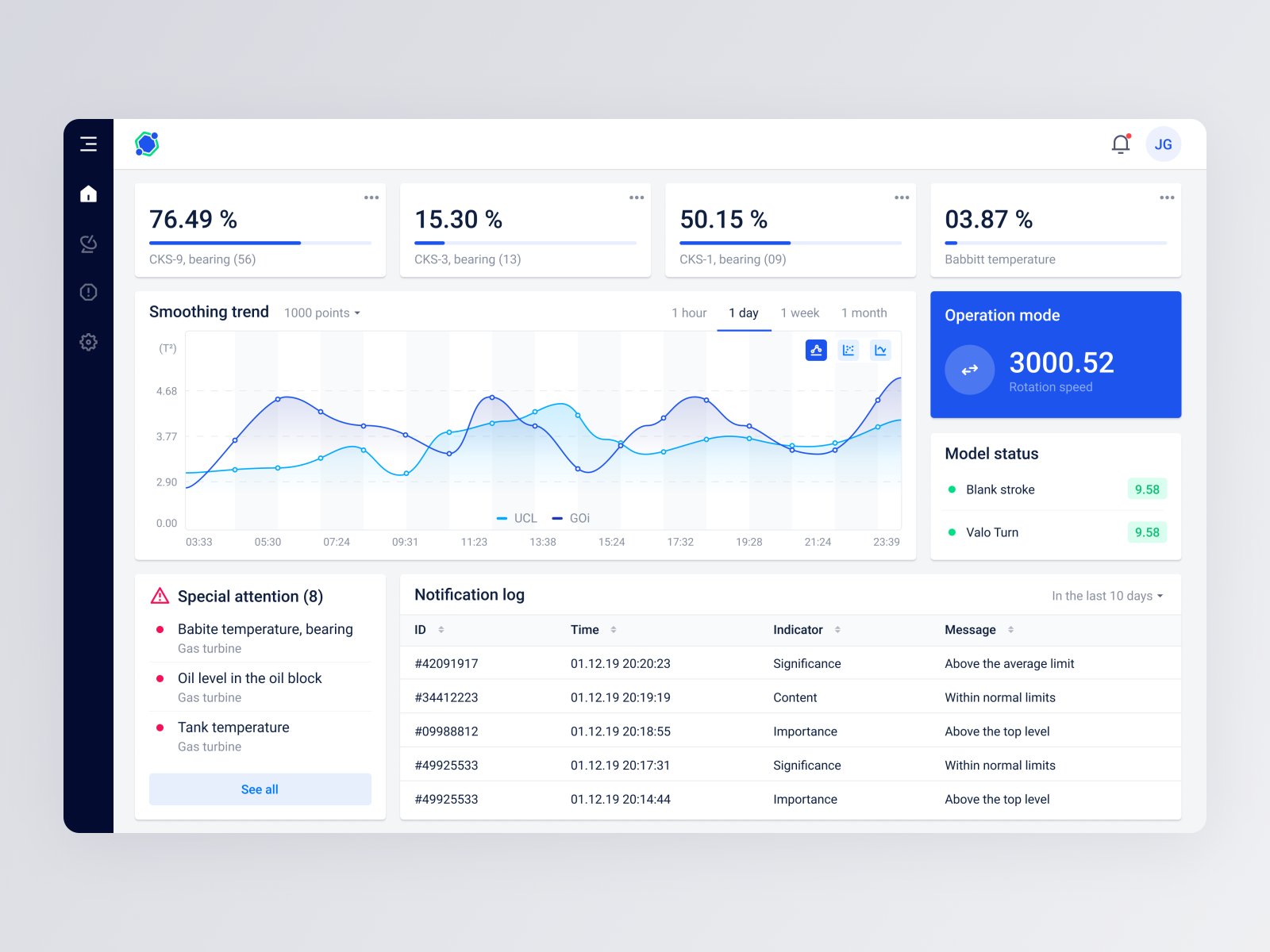
It allows healthcare administrators and decision-makers to analyze and interpret complex data sets to make informed decisions.

The dashboard utilizes various data sources, including patient demographics, medical procedures, and insurance information, to estimate and predict future hospitalization and medical care costs.

The dashboard helps in identifying cost patterns, trends, and outliers, enabling efficient resource allocation and budget planning.

It allows for the identification of high-cost procedures, enabling negotiation with insurance providers for better reimbursement rates.

Predictive analytics in the dashboard helps in estimating future costs, facilitating proactive measures to control expenses and optimize resource utilization.



\*Real-time data tracking: The dashboard provides up-to-date information, allowing for timely decision-making and intervention.

\*Customizable visualizations: Users can personalize the dashboard to focus on specific cost metrics or areas of interest.

\*Drill-down capabilities: Users can explore detailed data by drilling down into specific cost categories, patient demographics, or time periods.

\*Cost reduction: By identifying areas of high costs and inefficiencies, the dashboard helps in implementing cost-saving measures.

\*Improved patient outcomes: The insights from the dashboard can lead to better resource allocation, ensuring timely and appropriate care for patients.

\*Enhanced financial planning: Accurate estimation and prediction of costs enable hospitals and healthcare organizations to plan budgets effectively and allocate resources efficiently.

\*The dashboard in estimation and prediction of hospitalization and medical care costs plays a crucial role in healthcare management.

\*It empowers decision-makers with actionable insights to control costs, optimize resource utilization, and improve patient outcomes.

\*By harnessing the power of data visualization and predictive analytics, the dashboard contributes to a more efficient and sustainable healthcare system.



# Story

The incidence of overweight and obesity has increased significantly in most countries in recent decades. Excess weight is associated with an increased incidence of many chronic diseases, including vascular disease, respiratory disease, osteoarthritis, some cancer, type 2 diabetes, and premature death. There is consistent evidence that an increased BMI is associated with higher health costs, and these costs are expected to increase as obesity. Modelling uses machine-learning methods, in which the machine learns from the data and uses it to forecast new data. The most commonly predictive analytic model used is regression. The proposed model for accurate prediction of future outputs has applications in banking, economics, e-commerce, sports, business, entertainment, etc. A method used to forecast healthcare costs for BMI is based on several factors. Multiple linear regression is one of the statistical techniques for estimating the relationship among the dependent (target) and independent variables. The regression method is commonly used to develop a system based on a number of factors to predict the cost.



Fast-growing healthcare costs have become a significant challenge in several developed countries. Existing evidence suggests that healthcare costs have accumulated among a large number of BMI. Even though experiments have attempted to develop accurate models for predicting healthcare costs for BMI, their effectiveness is excellent due to the lack of detailed clinical information in the data used to create complex intervals and prognostic models. Numerous studies on more costs for obesity patient prognostic models have relied on self-report data and electronic health data from claims. Data from laboratory tests are defined—these, more granular and detailed clinical information, lead to improvements in the prognostic model. A recent survey by health research program and claim data shows that there is an improvement in the performance of the machine-learning-based.

\*No of Scenes of Story

The number of scenes of a story refers to the different stages or events that occur during a patient's hospitalization and medical care journey.

Each scene represents a different aspect of the patient's experience, such as admission, diagnosis, treatment, and discharge.

Estimating and predicting hospitalization and medical care costs can be enhanced by considering the number of scenes as it provides a more comprehensive understanding of the patient's healthcare journey.

**Importance of Number of Scenes**

The number of scenes helps in identifying the complexity and duration of the patient's medical journey, enabling a more accurate estimation of costs.

Each scene represents a different set of medical procedures, tests, and interventions, which directly impact the overall cost of care.

By analyzing the number of scenes, healthcare providers and insurers can better anticipate resource utilization and associated costs, leading to more reliable cost estimation and prediction.

**Implications and Benefits of Incorporating Number of Scenes**

Accurate estimation and prediction of costs enable healthcare organizations to efficiently allocate resources and plan budgets.

Understanding the number of scenes can help identify potential areas for cost reduction, such as streamlining care processes or optimizing resource utilization.

Patients can benefit from cost transparency, as healthcare providers can provide them with more accurate estimates and help them make informed decisions about their medical care.



# Report

Hospitalization and medical care costs can be estimated and predicted using various statistical and econometric techniques.

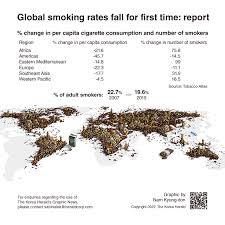
Factors such as patient demographics, severity of illness, length of hospital stay, and type of medical procedures play a crucial role in determining costs.

Advanced machine learning algorithms, such as regression models and decision trees, can be utilized to forecast future healthcare expenditures.

Estimating and predicting hospitalization and medical care costs help healthcare providers and policymakers in budgeting and resource allocation.

Accurate cost estimation aids in developing efficient reimbursement systems and pricing strategies for healthcare services.

Predicting future costs allows for effective planning, risk management, and development of cost-effective interventions and policies.

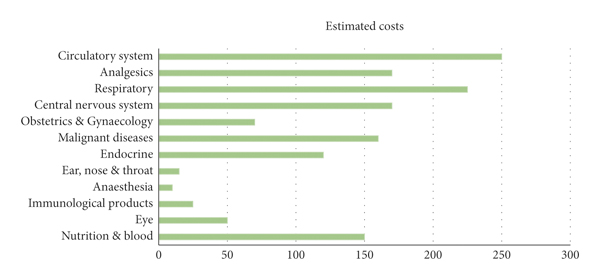


\*Challenges

Challenges in estimating and predicting hospitalization and medical care costs include data availability, accuracy, and heterogeneity.

Future directions involve the integration of electronic health records, claims data, and other sources to improve the accuracy and timeliness of cost estimation.

Advancements in predictive analytics and artificial intelligence techniques will enhance the precision and reliability of cost predictions.



**Importance of Visualizations in Estimation and Prediction of Hospitalization and Medical Care Costs**

Visualizations play a crucial role in understanding and analyzing complex healthcare data.

They provide a clear and concise representation of cost patterns, trends, and relationships.Visualizations enhance decision-making processes by enabling stakeholders to identify cost-saving opportunities and optimize resource allocation.

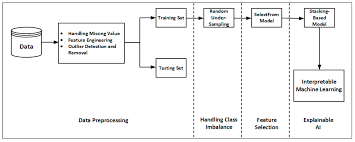


**Types of Visualizations**

\*Line charts: Show cost trends over time, allowing for identification of seasonality, outliers, and potential cost drivers.

\*Heatmaps: Display cost patterns based on different variables such as age, disease category, or procedure type, providing insights into cost variations.

\*Sankey diagrams: Visualize patient flow and cost distribution across different healthcare services, highlighting areas of high utilization and potential savings.

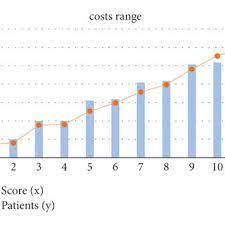


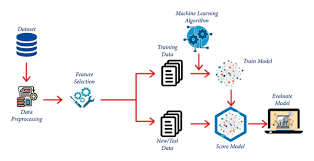
**Benefits of Visualizations**

\*Improved understanding: Visualizations enhance comprehension of complex cost data, enabling stakeholders to make informed decisions.

\*Enhanced communication: Visualizations facilitate effective communication of cost patterns and trends to diverse audiences, promoting collaboration and knowledge sharing.

\*Data-driven decision-making: Visualizations provide evidence-based insights, allowing for data-driven strategies to optimize hospitalization and medical care costs.





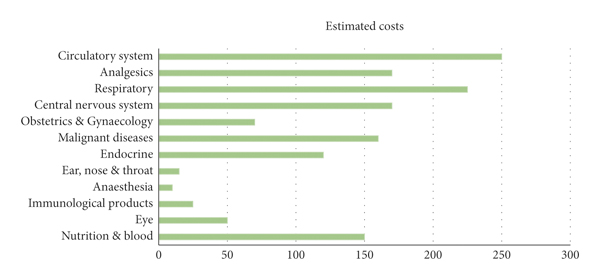
This report aims to provide insights into methods for estimating and predicting hospitalization and medical care costs.

Utilizing cost models and statistical techniques can help in predicting future costs based on observed patterns.

Machine learning algorithms can be employed to predict hospitalization and medical care costs based on various input variables.

These algorithms analyze large datasets and identify patterns and relationships to make accurate predictions.

Predictive modeling can assist healthcare providers and insurance companies in determining appropriate pricing and resource allocation.



# Performance Testing

Performance testing is a crucial tool for accurately estimating and predicting hospitalization and medical care costs.

By analyzing the performance of healthcare systems, we can identify potential bottlenecks, inefficiencies, and areas for improvement.

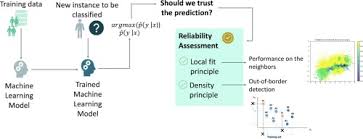
Performance testing allows us to evaluate the impact of different factors such as patient volume, resource allocation, and workflow on overall costs.

**Benefits of Performance Testing**

\*Accurate cost estimation: Performance testing helps in determining the actual costs associated with hospitalization and medical care by simulating real-world scenarios.

\*Resource optimization: By identifying areas of improvement through performance testing, healthcare providers can optimize resource allocation, leading to cost savings.

\*Proactive planning: Performance testing enables healthcare organizations to proactively plan for future demands, ensuring they have the necessary resources to handle patient volume and minimize costs.

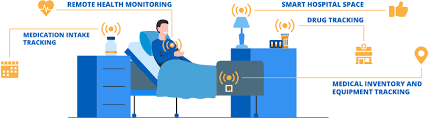


Challenges:

\*Data availability and quality: Performance testing relies on accurate and comprehensive data, which may be challenging to obtain and maintain.

\*Complexity of healthcare systems: Healthcare systems are complex, with various interconnected components. Performance testing needs to consider this complexity to provide accurate cost estimations.

\*Continuous monitoring: Performance testing should be an ongoing process to adapt to changing healthcare dynamics and ensure cost predictions remain accurate over time.



# Amount of Data Rendered to DB2

In estimation and prediction of hospitalization and medical care costs, a significant amount of data is rendered to DB2.

This data includes patient demographics, medical history, diagnosis codes, treatment details, and cost information.

The volume of data is substantial, as it involves thousands of patients, multiple healthcare providers, and a wide range of medical procedures and treatments.

The amount of data rendered to DB2 is crucial for accurate estimation and prediction of hospitalization and medical care costs.

By analyzing this data, patterns and trends can be identified, enabling better cost forecasting and resource allocation.

The data helps in understanding the factors contributing to hospitalization and medical care costs, such as chronic conditions, treatment effectiveness, and medical interventions.



**Benefits of Utilizing DB2 for Estimation and Prediction**

Utilizing DB2 for estimation and prediction of hospitalization and medical care costs offers several benefits.

It allows for efficient storage, retrieval, and analysis of large volumes of data, ensuring quick and accurate insights.

With DB2's advanced analytics capabilities, healthcare providers can make informed decisions regarding cost management, resource allocation, and patient care, ultimately leading to improved outcomes and cost savings.

**\*Utilization of Data Filters in Estimation and Prediction of Hospitalization and Medical Care Costs**

Data filters are essential tools in analyzing healthcare data to estimate and predict hospitalization and medical care costs.

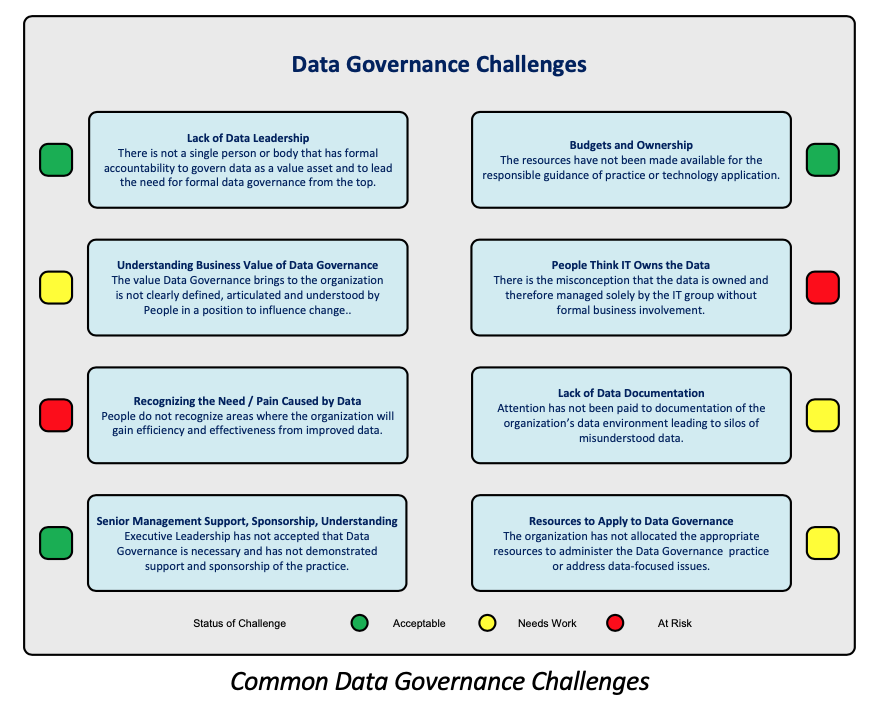
By applying data filters, we can identify and analyze specific subsets of data, such as patient demographics, medical conditions, and treatment procedures.

Utilizing data filters allows for more accurate estimation and prediction of hospitalization and medical care costs by considering relevant factors and eliminating irrelevant data.

Data filters help in identifying high-cost patients or specific medical conditions that contribute significantly to overall costs.

By analyzing filtered data, healthcare organizations can identify cost-saving opportunities, such as optimizing resource allocation or implementing preventive care programs.

Data filters enable healthcare professionals to make informed decisions regarding treatment plans, resource allocation, and budgeting based on accurate cost estimations.



**Challenges and Considerations in Utilizing Data Filters for Estimation and Prediction**

Selecting appropriate data filters requires a deep understanding of the healthcare system, relevant variables, and their impact on costs.

The accuracy and reliability of data filters depend on the quality and completeness of the underlying data.

Continuous evaluation and refinement of data filters are necessary to ensure their effectiveness in estimating and predicting hospitalization and medical care costs.

**Number of Calculation Fields**

The estimation and prediction of hospitalization and medical care costs involve multiple calculation fields.

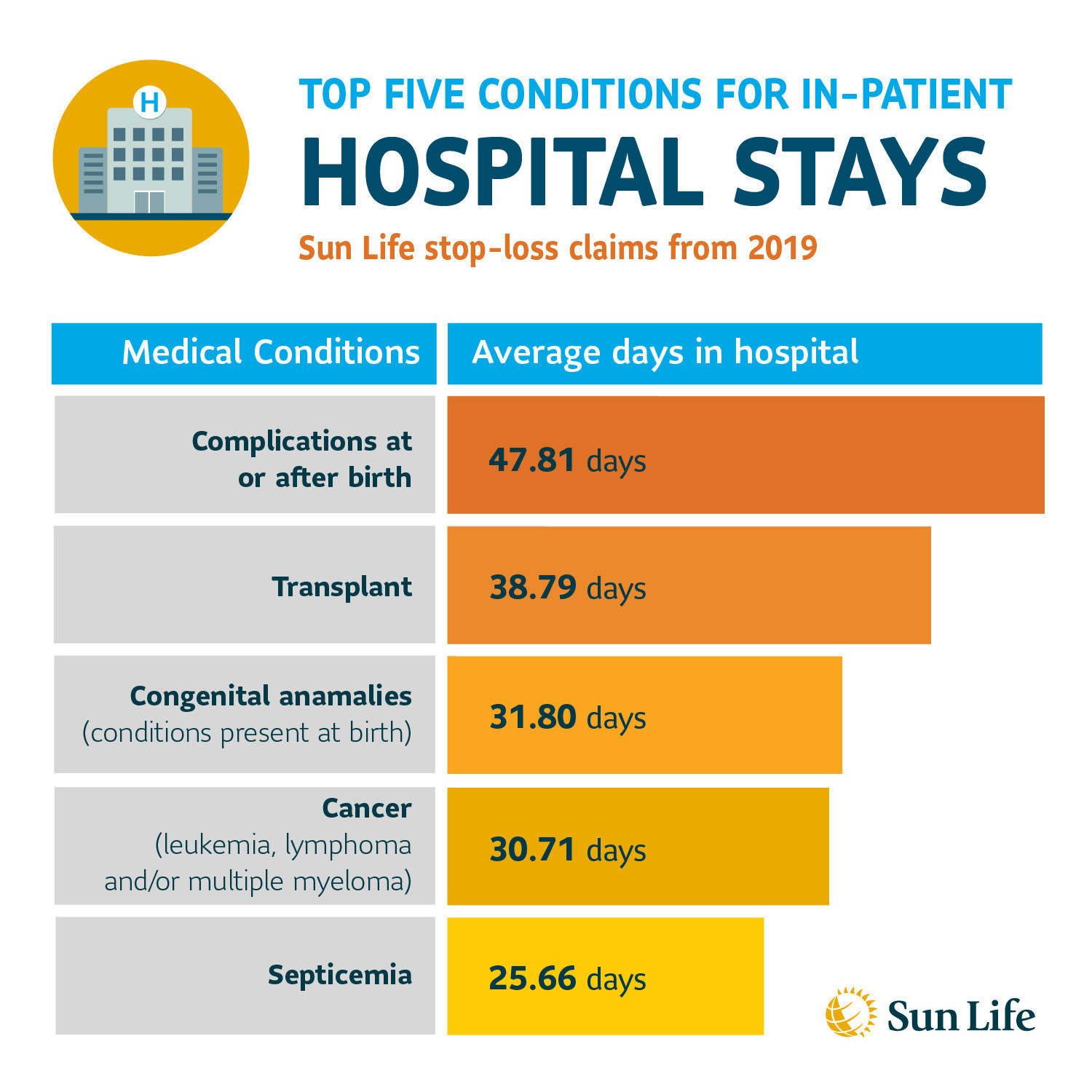
These calculation fields include demographic factors, such as age, gender, and socio-economic status, which impact healthcare utilization and costs.

Other calculation fields include medical history, co-morbidities, severity of illness, and procedures performed, which contribute to the overall cost estimation.

Patient-specific factors, such as insurance coverage and type, also play a role in cost estimation.

Additionally, geographical location, hospital type, and available healthcare resources are considered in the calculation fields.

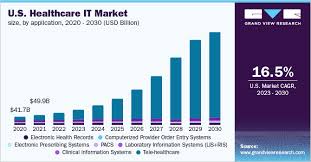
The number of calculation fields can vary depending on the specific models or algorithms used for cost estimation and prediction.

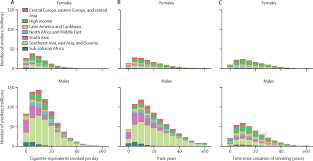


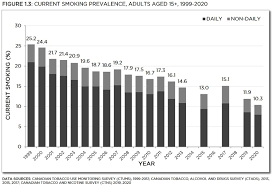
Advanced analytics techniques, such as machine learning and data mining, are increasingly utilized to analyze large datasets and identify relevant calculation fields. These techniques allow for more accurate and personalized cost estimation and prediction. The number of calculation fields can be further expanded with the inclusion of patient-reported outcomes, patient satisfaction metrics, and other subjective measures.

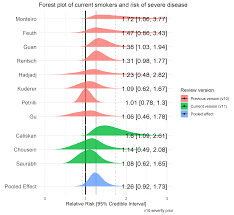
# Number of Graphs

The global tobacco market size was estimated at USD 867.55 billion in 2022 and is expected to grow at a compound annual growth rate (CAGR) of 2.1% from 2023 to 2030 due to the rising tobacco consumption in the developing regions of Asia and Africa. The excessive marketing campaigns run by the major companies have also been a significant factor in sustaining the industry. The industry is witnessing a trend of new product launches, which intrigues consumers to consume tobacco and thereby drive market growth. The market continues to thrive due to various influential factors, persisting despite ongoing endeavors to diminish tobacco consumption and address its detrimental effects. A prominent force behind its resilience lies in the addictive properties of nicotine, a substance inherently present in tobacco products.









# Web Integration

Web integration and dashboard reporting are essential tools in modern healthcare systems.

These tools allow for seamless integration of data from various sources, enabling accurate estimation and prediction of hospitalization and medical needs.

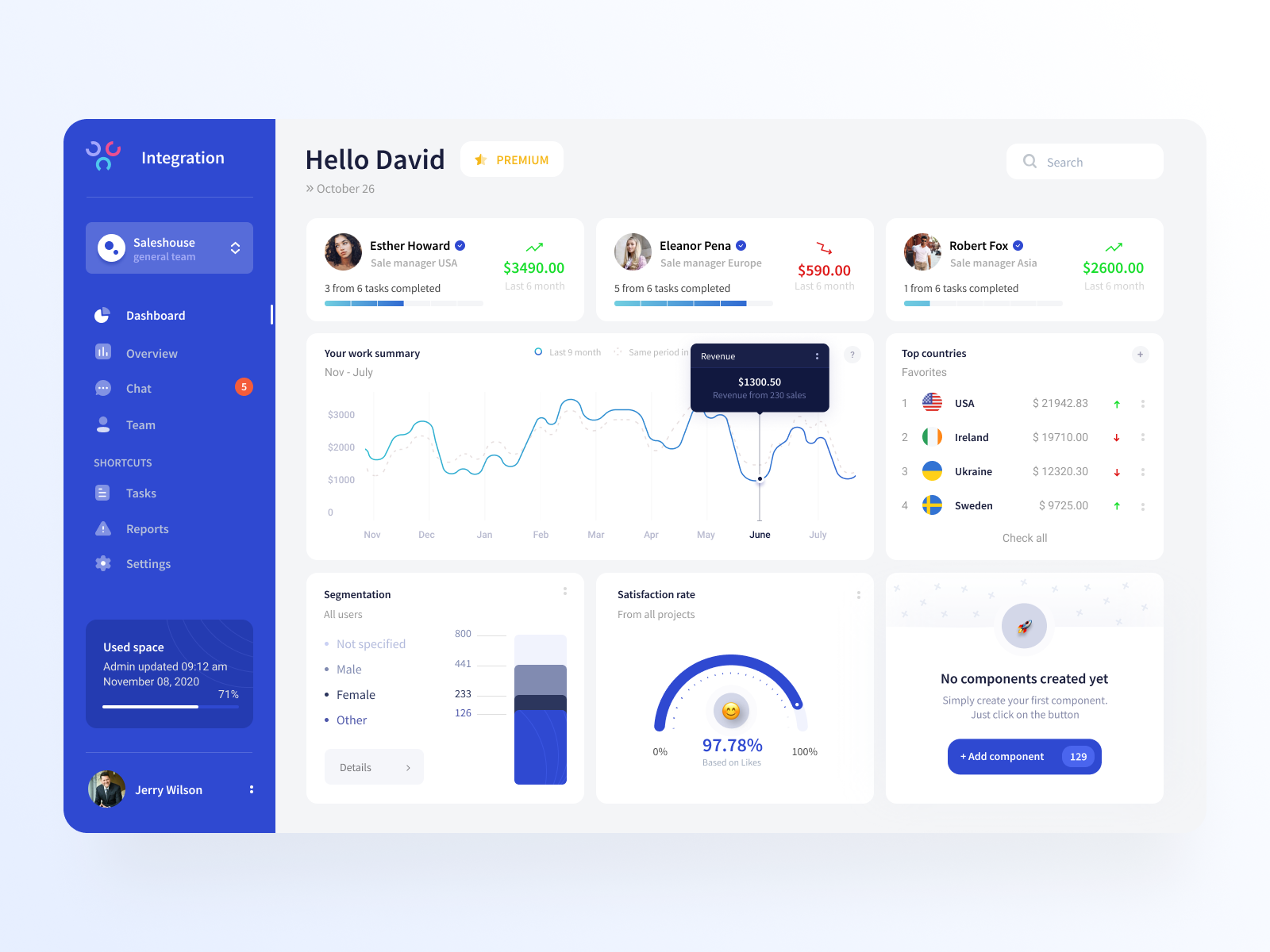
By utilizing web integration and dashboard reporting, healthcare providers can make informed decisions and allocate resources effectively.

**Benefits of Web Integration and Dashboard, Report in Hospitalization and Medical Estimation and Prediction**

Improved data accuracy and accessibility: Web integration ensures real-time data synchronization, reducing errors and delays in estimation and prediction.

Enhanced data visualization: Dashboard reports provide visually appealing and easy-to-understand representations of complex medical data, aiding in decision-making processes.

Efficient resource allocation: Web integration and dashboard reports enable healthcare providers to identify trends and patterns, allowing for better resource planning and allocation.



Data integration: Web integration allows for the seamless integration of data from various sources, such as electronic health records, medical devices, and patient portals.

Real-time updates: Web integration enables real-time data updates, ensuring accurate estimation and prediction of hospitalization and medical needs.



Customizable dashboards: Dashboard reports can be customized to display specific metrics and key performance indicators, providing relevant insights for healthcare providers.

Bed capacity planning: Web integration and dashboard reports can help hospitals estimate bed capacity requirements based on historical data, patient demographics, and current trends.

Resource optimization: By analyzing data through dashboard reports, healthcare providers can determine the optimal allocation of resources, such as medical equipment, staff, and medications.

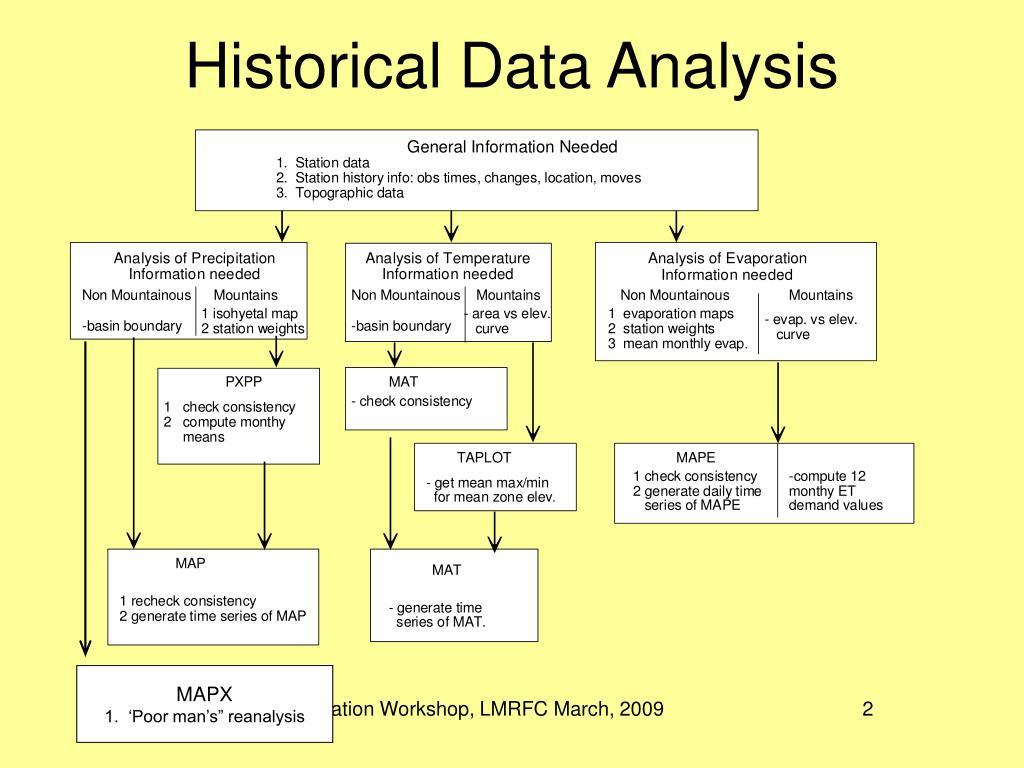
Disease outbreak prediction: Web integration and dashboard reports enable the identification of potential disease outbreaks by analyzing data from various sources, such as patient symptoms, geographical location, and demographic information.

# Project Demonstration

Project demonstrates the development of a comprehensive system for estimation and prediction in this domain.

The project aims to provide valuable insights for resource allocation, capacity planning, and improved patient care.

The project utilizes historical data analysis to identify key factors influencing hospitalization and medical needs.

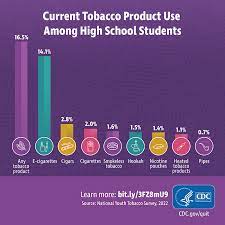


Statistical models and machine learning algorithms are employed to develop predictive models for hospitalization rates.

The developed models are validated using real-time data to ensure their reliability and accuracy in predicting future needs.

Accurate estimation and prediction facilitate proactive resource allocation, reducing bottlenecks and enhancing patient care.

Hospitals can optimize staffing levels, bed capacity, and medical supply inventory based on predicted hospitalization rates.



Public health agencies can utilize the system to identify areas with higher healthcare demands, leading to more targeted interventions and improved population health

Conclusion

Estimation and prediction of hospitalization and medical care costs are essential for effective healthcare management.

By understanding the factors influencing costs and utilizing advanced modeling techniques, we can improve resource allocation and financial planning.



Accurate estimation and prediction ultimately contribute to better patient care and sustainable healthcare systems. We provided a new linear regression and AI technology that can easily demonstrate the reasons for producing a certain forecast regarding potential healthcare expenses, which is a useful capacity in the healthcare area.



The linear regression algorithm is used to estimate the healthcare costs of the patients such as obesity (BMI) using certain devices such as smartphones and smart devices.



For estimation, by the use of linear regression, supervised learning performs more accurately. By providing comprehensive evidence, regression methodology can be effectively used for prognosis in conjunction with the dataset. The domain and time accuracy will determine the prediction model and the estimation of healthcare expenses.

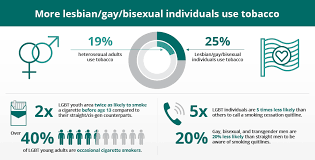


The proposed method reduces the risk of overfitting, and also, training time is less. This method is effective in estimating the healthcare costs of patients with an accuracy rate of 97.89%. The extensive tests on a real-time world database have confirmed the efficiency of our method.

A good healthcare system not only contributes to the welfare of the citizens, but also boosts the economy by making sure that productive labour is not lost due to down time. With rapid strides in technology, it is not surprising that the healthcare industry collects a lot of data about patients and this has the potential to generate valuable insights for medical practitioners and even when it comes to medical research. Big data and healthcare analytics have changed the face of the industry from a traditional set-up to one that thrives and is driven by technology.



Decrease in Smoking and Drug



Smoking rates have declined globally for the first time on record, according to a new report on tobacco use from a public health campaign group and US academics.  
  
However, the figures from the Tobacco Atlas report - described as a potential tipping point by the authors - also mask growing numbers of smokers in parts of the world, as well as increased tobacco use among young teenagers in almost half of the countries surveyed.  
  
Globally, there are 1.1 billion smokers and 200 million more people who use other tobacco products, the report from Vital Strategies and the Tobacco-nomics team at the University of Illinois at Chicago found.

#### References

1. A. K. Dutta, N. Ali Aljarallah, T. Abirami et al., “Optimal deep-learning-enabled intelligent decision support system for SARS-CoV-2 classification,” *Journal of Healthcare Engineering*, vol. 2022, Article ID 4130674, 14 pages, 2022.

View at: [Publisher Site](https://doi.org/10.1155/2022/4130674) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Optimal%20deep-learning-enabled%20intelligent%20decision%20support%20system%20for%20SARS-CoV-2%20classification&author=A.%20K.%20Dutta&author=N.%20Ali%20Aljarallah&author=T.%20Abirami&publication_year=2022)

1. J. Mohana, B. Yakkala, S. Vimalnath et al., “Application of internet of things on the healthcare field using convolutional neural network processing,” *Journal of Healthcare Engineering*, vol. 2022, Article ID 1892123, 2022.

View at: [Publisher Site](https://doi.org/10.1155/2022/1892123) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Application%20of%20internet%20of%20things%20on%20the%20healthcare%20field%20using%20convolutional%20neural%20network%20processing&author=J.%20Mohana&author=B.%20Yakkala&author=S.%20Vimalnath&publication_year=2022)

1. L. Hu, L. Li, J. Ji, and M. Sanderson, “Identifying and understanding determinants of high healthcare costs for breast cancer: a quantile regression machine learning approach,” *BMC Health Services Research*, vol. 20, no. 1, pp. 1066–1110, 2020.

View at: [Publisher Site](https://doi.org/10.1186/s12913-020-05936-6) | [Google Scholar](https://scholar.google.com/scholar_lookup?title=Identifying%20and%20understanding%20determinants%20of%20high%20healthcare%20costs%20for%20breast%20cancer%3A%20a%20quantile%20regression%20machine%20learning%20approach&author=L.%20Hu&author=L.%20Li&author=J.%20Ji&author=M.%20Sanderson&publication_year=2020)

1. K. Sahu and R. K. Srivastava, “Needs and importance of reliability prediction: an industrial perspective,” Information Sciences Letters, vol. 9, no. 1, pp. 33–37, 2020.
2. S. Kent, J Green, G Reeves et al., “Hospital costs in relation to body-mass index in 1·1 million women in England: a prospective cohort study,” =e Lancet Public Health, vol. 2, no. 5, pp. e214–e222, 2017.

6. V. S. Kadam, S. Kanhere, and S. Mahindrakar, “Regression techniques in machine learning &applications.

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |