International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org



IJIEMR Transactions, online available on 13th Apr 2023. Link
:http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

**10.48047/IJIEMR/V12/ISSUE 04/134**

Title **PREDICTING PATIENT LENGTH OF STAY USING IN HOSPITAL**

Volume 12, ISSUE 04, Pages: 1048-1057

Paper Authors
**Mr. Anand Thota, Mr.Munagala Anudeep, Mr. Mallavolu jagadeesh,**

**Mr.Dlaia Venkata Vishnu Bhargava Teja, Mr. Chendika Haranadh Babu**

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

# Predicting Patient Length of Stay Using in Hospıtal

[1]Mr. Anand Thota, [2]Mr.Munagala Anudeep, [3]Mr. Mallavolu jagadeesh, [4]Mr.Dlaia Venkata Vishnu Bhargava Teja, [5]Mr. Chendika Haranadh Babu

[1]Associate Professor, Department of Computer Science and Engineering, Potti Sriramulu Chalavadi Mallikarjunarao College of Engineering and Technology, Vijayawada, AP, India

[2,3,4,5] Students, Department of Computer Science and Engineering, Potti Sriramulu Chalavadi Mallikarjunarao College of Engineering and Technology, Vijayawada, AP, India

[1]anand4uammu@gmail.com,[2]anudeepmunagala2002@gmail.com, [3]jagadeeshmallavolu@gmail.com, [5]harinadhchendikal1323@gmail.com

**Abstract**

Since modern machine learning (ML) techniques can use vast volumes of data to predict specific patient outcomes, predictive analytics is becoming an increasingly crucial tool in the healthcare industry. Predictions from machine learning, for instance, can help doctors make diagnostics, recommend treatments, and predict the future wellbeing of their patients. I decided to concentrate on a more practical healthcare indicator for this research, the length of hospital stays (LOS). The number of days between hospital admission and discharge is known as the length of stay, or LOS. As a result, A high risk of LOS is identified at the time of admission, and hospitals are encouraged to detect these patients. Patients who are at a high risk of LOS can have their treatment regimens changed once they have been identified to lessen LOS and lessen the likelihood that they will contract a hospital-acquired illness like staph infection. Another advantage is that prior understanding of LOS might help with planning practicalities as space and bed allotment. Hospital patient length of stay (LOS) is an important performance factor and control metric. The intensive care unit (ICU) may run out of supplies, personnel, and equipment if an extended length of stay (LOS) occurs there Furthermore, a precise prediction of patient LOS could help medical professionals decide what treatments to give patients and how to allocate staff and resources. Both the patient and the insurance company could plan their spending using this forecast. Based on this study's findings, it is possible to forecast how many days a patient would spend in the hospital by looking at certain personality features. The mathematical codification of patient information analysis makes it simpler to identify significant human qualities via conditioned space analysis. The number of hospital days is the objective variable.

**Keywords**

Length Of Stay (LOS), Clusterıng analysis, Regression, Conditional space, Hierarchical predictor.

## 1. INTRODUCTION

### 1.1 About Project

Hospitals must cope with two key uncertainties when managing resources and personnel: who will request the patient's admission and how long he would stay if hospitalised. These uncertainties significantly limit the optimal scheduling of the admission of elective patients, the considerable fluctuations in occupancy and demand for various services, and the effective use of labour and infrastructure.

Even though the hospital is unable to predict future admission requests, doing so may make it clearer how long each patient will be admitted for. Elective admissions may be scheduled in accordance with discharge dates if discharge dates can be precisely predicted. This lowers occupancy variance and creates the possibility of either lowering staff and facilities or raising average occupancy. The total amount of time a patient spends in the hospital between successive admissions and discharges during a predetermined period of time is referred to as "length of stay" (LOS).

Especially during the COVID-19 outbreak, most hospitals find it challenging to deliver prompt patient treatment while guaranteeing effective resource use. A yearly study by the American Hospital Association found that in 2019 hospitalised patients spent more than $1.16 trillion in all U.S. hospitals with registration (American Hospital Association, 2021). Inpatient deaths in the United States are modified to increase by 3% for every hour of transfer delay.

A significant amount of the gross domestic product of many countries goes toward paying for healthcare (GDP). For instance, healthcare spending in the UK nearly reached ten percent (9.3%) of GDP in 2012. In many countries, government funding has not kept up with the cost of patient care, forcing hospitals and other healthcare facilities to handle an expanding patient load. As a result, cutting healthcare costs has become one of the industry's top priorities today. Hospitalization is the main cost of patient care, hence healthcare administration pays a lot of attention to it.

For the patient and the hospital, the length of stay is crucial (LOS). It acts as a benchmark for how well hospitals are managed. Due to improved bed management, a decrease in inpatient days leads to a reduced risk of infection and medicine side effects, an improvement in the standard of treatment, and higher hospital profits. It is a complicated task, however, because it is impacted by a variety of significant factors, including age, sex, weight, diagnosis, sub-diagnosis, prescription, flora, active ingredient, etc.

### 1.1Scope

Predicting how long patients would stay in the hospital is the study's main objective. Determining how long patients would stay in the hospital is the study's main objective. In this process we need to collect the patient details and perform pre-processing for eliminating unwanted or unused data.

### 1.2 Purpose

In to figure out how many days a patient will reside in the hospital before being discharged, patient length of stay predictions are made. This project helps in assigning beds and rooms to the patients and also to predict the equipment that will be used to the patients.

### 1.3 Problem Statement

Predicting Patient length of stay mainly concentrates on early prediction of number of days that a patient stays in hospital so that to predict whether there is room for another patients to join in the hospital. During Corona period due to lack of accurate information many patients can't able to join in the hospitals dur to ambiguity in availability of beds in the hospital.

## 2. RELATED WORK

[1] Manuel Puentes Gutierrez et al. [] developed a technique for forecasting how long people will stay across all departments. The author of this paper implemented the suggested system using neural networks, support vector machines, and random forests.Of the above methods, the author found that the best results are obtained by the Radial Basis Function Kernel Algorithm. The issue that Jesus Manuel Puentes Gutierrez encountered was that Emergency pain management and rehabilitation facilities, which could not be examined owing to a lack of data, should also be taken into account.

[2] Leslie Mon Plaisir et al.[] a Forecast of COVID-19 Patients' Emergency Department Stay Length of Time. The author built the system using the procedures mentioned below: logistic regression, gradient boosting, decision trees, and random forests. In the study mentioned above, the author created a framework for prediction using the LR, GB, DT, and RF models. to foretell whether COVID-19 patients will spend more than or less than 4 hours in the emergency department. Only decision tree and gradient boosting were able to

outperform the other approaches in our investigation.

[3] Patrice Caulier et al.[] a machine learning technique for forecasting length of stay in a hospital context has been presented.The author here implemented the system in the proposed system using the Random Forest, Naive Bayes, Gradient Boosting, and K-Nearest Neighbour algorithms.

In the study mentioned above, As the most effective technique for LOS prediction, we built random forest. Lack of a valid data set was the issue Patrice Caulier encounteredOne method for improving the system's efficacy is to change the LOS from numerical data to category data based on the survey, current activities, and medical specialists.

[4] Stephanie L. Mekhaldi et al.[] employing machine learning, it was suggested that circulation failure might be predicted early in the critical care unit. Using the strategies detailed below, the author put the system into practise. Gradient boosting, recurrent neural network model based on LSTM, and logistic regression. The mean absolute SHAP values of the provide these on the validation set for each temporal split, based on the author's analysis of the aforementioned approach, were used to determine the relevance of certain qualities.

According to the study, the prognosis of death or LOS has little impact on subsequent treatment decisions following the first decision to admit a patient to the ICU.

[5] Gustavo A. Fernandez et al.[] It was proposed to compare statistical approaches for analysing tally data and applying them to calculate the duration of hospital stays. The following methods were employed by the author here to implement the system: Negative binomial, zero-inflated Poisson regression, and poisson regression. In several situations of both zero-inflation and overdispersion, the author observed that the NB model significantly improved than the ZINB model and had the closest fit for the over-dispersed data. The author found that the above technique does not consider the boundary restrictions present in the zero-infated data. More research should be done for scenarios with various data production systems in inpatient hospital LOS.

[6] James McNicholas et al.[] for patients in critical care units, a method for forecasting hospital mortality has been presented. The following methods were implemented by the author are: Bayesian Networks, Random Forest. The use of such tools for assisting actual decision choice is yet untested, according to the author, who observed that scoring systems have centred on giving ever-more-detailed ways of measuring ICU performance and have developed the structure for powerful equality control systems.

In our study there may be some missing data for each patient in our study because not all medical variables or tests are taken during the first few hours of arrival.

[7] Duncan Shillan et al.[] Machine learning was proposed as a method for evaluating regularly collected critical care unit data. The author created the system using decision trees, neural networks, and support vector machines.. In the study the author found that using random selections of the data with or without (44.1%) k-fold validation was the strategy that was most frequently utilised. Duncan Shillan ran into a problem where the study was descriptive and the results of the included studies did not account for the risk of bias.

[8] Eunbi Kim et al. [] offered a cost-benefit analysis of early hospitalisation prediction using machine learning. The author used support vector machines, logistic regression, XG boost, and NG boost in this study. According to our study's analysis of the predictive model's accuracy and the ROC curve's area under the curve, hospitalisation was more likely for patients who were older and had more urgent medical conditions. The issue that Eunbi Kim encountered was that the experiment throughout this research failed to show agreement in Values and

responsiveness to dataset amount. In future investigations, we may utilise additional data to observe the convergence of AUC and sensitivity by utilising conditional space and hierarchical predictors.

[9] Belal Alsinglaw et al. [] for cardiovascular hospitalisations in the critical care unit, a prediction of length of stay was made. The author implemented the following methods Gradient boosting, Logistic regression, LSTM based recurrent neural network model. The author identified that based on ICD-9, all unique hospitalisations for heart failure. Belal Alsinglaw's problem was that we failed to take into account the other medical comorbidities already present in HF inpatient patients. This is something that has to be explored more in a subsequent study in order to assess its effect on the LOS or the extended LOS.

[10] Didier Morel et al. [] suggested a method for predicting readmission to hospital in individuals with drug use or mental illnesses. The author here implemented the following approaches:

XG boost model, Boruta algorithmA recent study employing information collected from two hospital' medical record systems revealed that such ml approach produced readmission prediction models with higher AUROC than the baseline GLMNet (0.75-0.76 versus 0.68-0.70 for the general inpatient population). Didier Morel came into a dilemma since the dataset was based on claims: physicians' firsthand estimates of the severity of the sickness were missing.

Table 1: Existing System Analysis

| S.No | Author | Algorithm | Merits | De-Merits | Future Scope |
|---|---|---|---|---|---|
| 1. | Jesus Manuel Puentes Gutierrez | Decision tree C4.5, Random Forest, Support Vector Machines, Neural Networks | Our tests were the ones that used the methodology; prior to these studies, the techniques were not as effective at producing results. | Although these departments couldn't be studied owing to a lack of data, it is also important to take into account the limitations of the rehabilitation and pain management units previously discussed in emergency scenarios. | To determine how a severe pandemic like COVID-19 affects this feature and hospital departments, the length of stay should be investigated. |
| 2. | Leslie Monplaisir | Logistic Regression, Gradient Boosting, Decision Tree, Random Forest | Performance-wise, Gradient Boosting and Decision Tree were able to outperform the alternative approaches. | Several clinical values were missing as a result of administrative errors. If the missingness was not random, the matching mean value was used in place of the absent data points, which produced bias. | As deep learning technology develops, we will be able to collect more logical patient data pieces, increasing accuracy and decision-making speed. |
| 3. | Patrice Caulier | Random Forest , Gradient Boosting, Naive Bayes, K-nearest neighbour | In our study, random forest was employed since it had the best performance for LOS prediction in the literature. | Our investigation's validity could be compromised by the lack of a reliable dataset. As a matter of fact, we were forced to utilise the Microsoft dataset as an illustration since we had trouble getting access to the real data. | Depending on the survey, the ongoing work, and the opinions of medical professionals, the LOS might instead be transformed from numerical data to category data to improve the system's efficiency. |
| 4. | Stephanie L. Hyland | Gradient Boosting, Logistic Regression ,LSTM- | The model is simple, accurate, and practical for everyday use. Due to its outstanding | Once the decision has been taken to initially admit a patient to the ICU, the accuracy of the mortality or LOS | Our results do not corroborate our hypothesis, which asserts that lowering mortality through early diagnosis |

| | | | | | |
|---|---|---|---|---|---|
| | | based recurrent neural network model. | prediction accuracy, it performed better than the results of models employed in earlier experiments. | projection is not critical for future treatment choices. | and treatment of persons at risk for circulatory failure. This theory has to be verified in a follow-up prospective investigation, though. |
| 5. | Gustavo A. Fernandez | poisson Regression, negative Binomial ,Zero inflated poisson regression, zero-inflated negative binomial regression | The NB model outperformed the ZINB model in many scenarios of both zero-infation and overdispersion, and it offered the best performance for the over-dispersed data. | The boundary restrictions in the zero-infated data are not taken into account by this method. Another optimisation technique using zero-inflated data that is popular is Nelder-Mead Simplex Optimization for ZIP Regression Models (as in the case of R software). | More research should be done on inpatient hospital LOS scenarios with various data producing technologies. In this study, we also did not look at underdispersion. It would be interesting to examine how effectively count regression models describe under-dispersed data distributions, despite the fact that these data are rarely present in real-world datasets. |
| 6. | James McNicholas | Bayesian Networks, Random Forest | Hospitals can estimate the overall patient load using the prediction of cumulative hospital LOS over a specified time horizon. This makes it possible to arrange patient admissions more effectively, which reduces the fluctuation of hospital bed occupancies. | During the initial hours of admission, not all medical conditions or tests are assessed for every patient, therefore there may be some data that is missing for each patient. Missing values can be filled in in many ways, or they can be handled by simply ignoring the dataset's incomplete records. | We have shown this using a database with insufficient data. We propose that this signal may be strengthened in the future by modifications to the approach that we have utilised, to help physicians and patients in early outcome prediction. |
| 7. | Duncan Shillan | Neural Networks, Support Vector Machines, Classification/Decision trees | While we did a good job of searching the literature, it's likely that we overlooked studies that employed specialised machine learning methods or code repositories that weren't subjected to peer review and weren't present in the databases we searched. | Because the analysis was descriptive, the risk of bias in the findings of the included studies was not evaluated. It is therefore hard to draw firm conclusions about the causes of the variations in AUC between studies or the shortfalls between machine learning prediction models and those that use traditional statistical approaches. | The best technique to determine how well machine learning algorithms will work in actual clinical settings and to prevent overconfidence brought on by the selection of variables and parametrizations is to evaluate the algorithms using independent data.. |
| 8. | Eunbi Kim | Logistic Regression, | In our investigation, | AUC values and sensitivity to dataset size | The sensitivity of this indicator is significantly |

| | | | | | |
|---|---|---|---|---|---|
| | | XG Boost, NG Boost, Support Vector Machine, Decision tree models | XGBoost outperformed all other prediction models in terms of AUC. | do not converge in the experiments conducted for this study. The sensitivity of this indicator is significantly lower than that of other indicators due to the small number of hospitalised patients in the ED. | lower than that of other measures because there are so few ED patients who are hospitalised. To observe how AUC and sensitivity converge in future study, we might use more data. |
| 9. | Belal Alisinglaw | Gradient Boosting, Logistic Regression ,LSTM-based recurrent neural network model. | GBR finished the process in the shortest amount of time. The best R2 was discovered in GBR and stacking. In real-world settings, such as when getting ready for clinical and medical scenarios, model training and time assessment are essential. | The goal was to create a machine learning strategy that could estimate how long HF patients would need to stay in the hospital. The other current medical comorbidities linked to HF inpatient treatment were not considered in this investigation, and they may need to be further investigated in a follow-up study to ascertain how they would affect the LOS or the prolonged LOS. The impact of post-hospitalization intervention on heart failure LOS was also not covered. | To assess the effectiveness of the suggested model on various data sources, we will also verify our proposed predictive LOS framework on a real-world external dataset in next work. Our study adds new knowledge on how to anticipate patient outcomes in the future and identify pricey HF hospitalisations using artificial intelligence (AI) methods to clinical practitioners, hospital management systems, and clinical AI research. |
| 10. | Didier Morel | XGBoost model, Boruta algorithm | Compared to the GLMNet, the XGBoost method generates models with a markedly greater predictive value from a model discrimination standpoint. | Due to the dataset's claim-based foundation, clinical evaluations of the condition's severity by doctors were not readily available.The data source only contained information on patients who had commercial insurance. Although Medicare and Medicaid beneficiaries and those in the middle age range were not included in the database's age groups from adolescence to middle age, the majority of Americans received employer-sponsored health insurance. | Future study might investigate the social drivers of M/SUD and further validate our ML model using a more representative patient sample, particularly older patients and those with public health insurance, in addition to future scenarios. Researchers will be able to assess the model's utility in creating standardised hospitalisations for comparison and maybe enabling targeted demographic initiatives to reduce patient hospitalisations. |

## 3. PROPOSED MODEL

### 3.1 Dataset Description

The Data set was collected from the Figshare website which consists of details of the patients from the Lviv Public Hospital in Ukraine. The Dataset contains 11 Labels which almost cover every detail of the admitted patient in the Hospital and the size of the Data Set file is 4MB.

In the surgical department of Lviv Public Hospital, this data set was gathered (Ukraine). Postoperative problems in the abdomen were carefully managed in patients. Information as of June 28, 2021, may be found at https://doi.org/10.6084/m9.figshare.14865411.v1.

Table 2: Dataset Description

| Attribute | Description | Value |
|---|---|---|
| Id | It is the id of the person | String |
| Age | It provides us with the age group of each person | Categorical value |
| Sex | Male/Female | Categorical/ or can be interpreted as Boolean |
| Weight | It provides us with age group of each person | Categorical value |
| Date admission | Joining Date of a patient | Date |
| Diagnose | Main Problem | Categorical value |
| Sub Diagnose | Sub problem | Categorical value |
| Flora | Bacterial participation in this disease | Categorical value |
| Medicament | The substance used for medical treatment | Categorical Value |
| Active substance | The active substance used mostly in the treatment | Categorical value |
| Time in hospital | Time spent by the patient in hospital | Integer value |

```
mape(testY,predY)
0.14303748993178866
```

| METRICS | VALUES |
|---|---|
| Root mean square error | 0.9274502777137164 |
| Root squared error | 0.2553293450495254 |
| Mean absolute percentage error | 0.1430374899317886 |

Fig 3.1 Dataset

### 3.2 Methodology

There are several steps required for predicting the allotment of beds in hospital.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The steps required are:

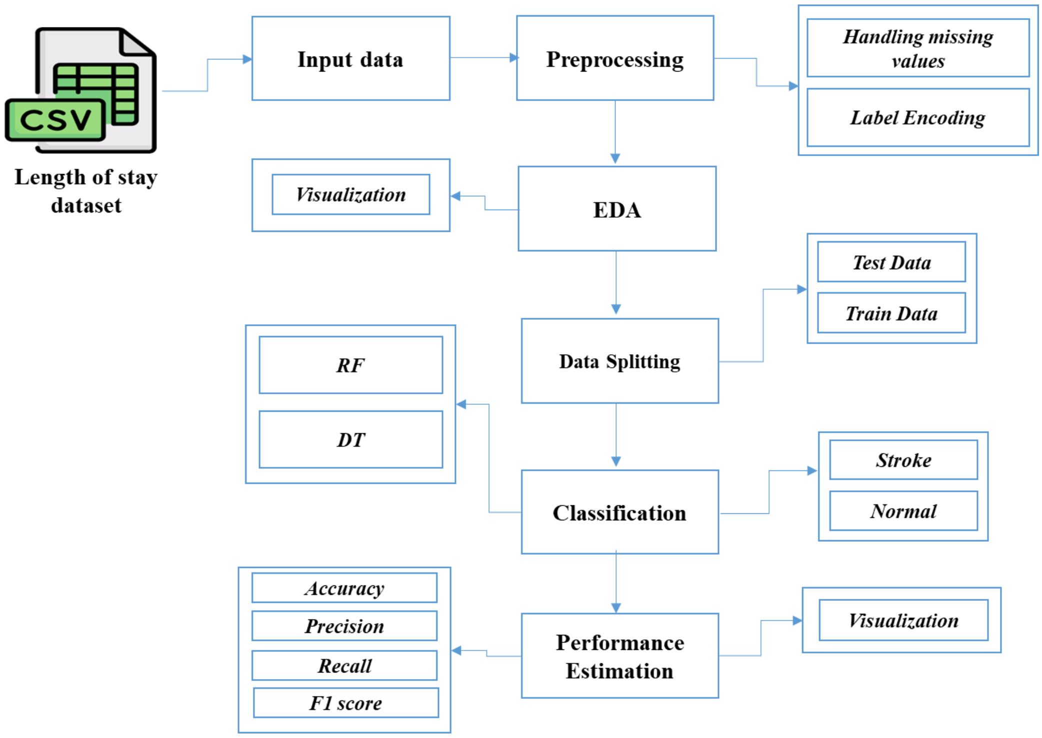


Fig 3.2: Proposed Methodology

**Conditional Space Analysis**

Firstly, we need to consider only time dependent parameters i.e.., we need to eliminate the parameters which are not dependent on time i.e.., for ex, if we use a drug for a patient that requires more time to stay should be included in data and Id of the patient will not affect the length of stay so, it will be discarded. To do this process we can follow two processes.

- Correlation matrix
- Boruta algorithm

**Correlation matrix**

Correlation matrix is used to find the dependency relation among parameters in the data set.

But the major problem with correlation matrix is it doesn't give the parameters which are required to analyze the result so that we use boruta algorithm.

**Boruta Algorithm**

The Boruta algorithm is a random forest classification wrapper. This method determines whether any of the actual traits are more important than others.

1. 1. In order to add unpredictability to the provided data set, it copies all features and shuffles them (which are called shadow features).
2. 2. Next, using the bigger data set, it analyses each feature's relevance using a feature importance measure (the default is Mean Decrease Accuracy), with higher means indicating more importance, and trains a random forest classifier.
3. It constantly eliminates attributes deemed to be very insignificant by determining whether a true quality is more significant than the finest of its ghost qualities (i.e., whether the feature has a higher Z score than the maximum Z score of its shadow features).
4. Either when all features are approved or rejected, or when the total number of random forest

runs hits a predefined limit, the method finishes.

After discarding the unimportant parameters, we need to give labels to the parameters as numbering so that we can easily compress the data by using labels. We give labels to the data by using label encoding.

**Label Encoding**

Label encoding is the process of converting labels into a numeric representation that machines can read.Then, using machine learning methods, the operation of those labels may be better understood. This phase of the structured dataset's supervised learning pre-processing is crucial.

**Clustering**

Clustering is the process of dividing the data points into several groups so that the data points within a group are more comparable to one another than to those within other groups. Now that the parameters have labels, we must decide how many clusters are required. Thus, using the K-Elbow approach, we may obtain the clustering result.

**K-Elbow**

To gauge how many clusters are there in a data set during a cluster analysis, the elbow technique is a heuristic utilised. The procedure involves charting the explained variance as a function of the number of clusters, then selecting the number of clusters to use at the elbow of the curve. The K-Elbow Visualizer uses the "elbow" method to determine the appropriate number of clusters for K-means clustering.

**Fuzzy C-mean Method**

Each data point in the dataset is partially assigned to each cluster using the fuzzy c-means (FCM) data clustering algorithm, which separates a data collection into N groups. A likelihood or probability score is assigned to each data point in the soft clustering technique known as fuzzy C-clustering Means to indicate how probable it is that the data point belongs to that cluster.

**Hierarchical Predictor development**

With the use of SVC with radial basis kernel, SVC with polynomial kernel, and linear regression random forest, each cluster is examined independently. On the findings received, an average vote is given. It will be used to determine the average value as the outcome.

**4. RESULTS**

**Root Mean Square Error Method:**

The residuals' standard deviation is known as Root Mean Square Error (RMSE) (prediction errors). Remainders are used to examine how far off the data points are from the regression line, and RMSE is used to gauge how far apart they are. In other words, it offers details on how closely the data are grouped around the line that fits it the best. In climatology, forecasting, and regression analysis, root mean square error is widely used to analyse experimental results.

**The final RMSE result is: 0.9274502777137164**

**Root Squared Error method**:

This root mean square approach has been normalized. The R squared error approach captures a higher percentage of the variation of the answer variable than the MSE, which only captures residual errorThe R-squared (R2) statistic in a regression model shows how much of a dependent variable's fluctuation can be explained by one or more independent variables. **The final R squared result is: 0.2553293450495254**

**Mean Absolute Percentage Error:**

The mean absolute percentage error of a forecasting system can be used to determine how accurate it is (MAPE). It is attainable by dividing the actual values by the average absolute percent inaccuracy for each time period, less genuine values. A percentage is used to represent this accuracy.

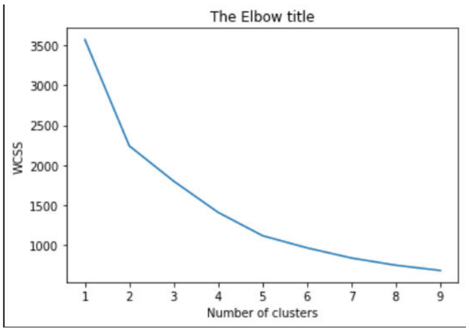


Fig 4.1 :- Knee Locator

**5. References**


1. JESÚS MANUEL PUENTES GUTIÉRREZ , MIGUEL-ÁNGEL SICILIA, SALVADOR SÁNCHEZ-ALONSO, AND ELENA GARCÍA-BARRIOCANAL , Predicting Length of Stay Across Hospital Departments , 2021 , DOI : r 10.1109/ACCESS.2021.3066562
2. EGBE-ETU ETU , LESLIE MONPLAISIR , SUZAN ARSLANTURK , SARA MASOUD , CELESTINE AGUWA , IHOR MARKEVYCH , JOSEPH MILLER , Prediction of Length of Stay in the Emergency Department for COVID-19 Patients: A Machine Learning Approach , 2022 , DOI : 10.1109/ACCESS.2022.3168045
3. Rachda Naila Mekhaldi, Patrice Caulier, Sondes Chaabane, Abdelahad Chraibi, and Sylvain Piechowiak , Using Machine Learning Models to Predict the Length of Stay in a Hospital Setting,2020 , DOI: https://www.researchgate.net/publication/341821304
4. Stephanie L. Hyland , Martin Faltys, Matthias Hüser , Xinrui Lyu, Thomas Gumbsch, Cristóbal Esteban, Christian Bock , Max Horn, Michael Moor, Bastian Rieck, Marc Zimmermann , Dean Bodenham, Karsten Borgwardt , Gunnar Rätsch and Tobias M. Merz , Early prediction of circulatory failure in the intensive care unit using machine learning , VOL 26 , 2020, DOI: https://doi.org/10.1038/s41591-020-0789-4
5. Gustavo A. Fernandez and Kristina P. Vatcheva , A comparison of statistical methods for modeling count data with an application to hospital length of stay , 2022, https://doi.org/10.1186/s12874-022-01685-8
6. Aya Awad, Mohamed Bader-El-Den, James McNicholas, Jim Briggs , Yasser El-Sonbaty , Predicting hospital mortality for intensive care unit patients: Timeseries analysis , Vol. 26(2) ,2020 ,DOI: 10.1177/1460458219850323
7. Duncan Shillan, Jonathan A. C. Sterne , Alan Champneys and Ben Gibbison , Use of machine learning to analyse routinely collected intensive care unit data: a systematic review , 2019, DOI : https://doi.org/10.1186/s13054-019-2564-9
8. EUNBI KIM , KAP SU HAN , TAESU CHEONG , SUNG WOO LEE , JOONYUP EUN , AND SU JIN KIM, Analysis on Benefits and Costs of Machine Learning-Based Early Hospitalization Prediction , VOLUME 10, 2022 , DOI : 10.1109/ACCESS.2022.3160742
9. Belal Alsinglawi , Fady Alnajjar , Omar Mubin , Mauricio Novoa , Mohammed Alorjani , Ola Karajeh , Omar Darwish , Predicting Length of Stay for Cardiovascular Hospitalizations in the Intensive Care Unit: Machine Learning Approach , 2020
10. Didier Morel, Kalvin C.Yu, Ann Liu-Ferrara, Ambiorix J. Caceres-Suriel, Stephan G. Kurtz, Ying P. Tabak, Predicting hospital readmission in patients with mental or substance use disorders: A machine learning approach , 2019 , https://doi.org/10.1016/j.ijmedinf.2020.104136