Predictive Modeling of In Hospital Mortality followed by Elective Surgery

|  |
| --- |
| **Citation:** To be added by editorial staff during production.  Academic Editor: Firstname Lastname  Received: date  Revised: date  Accepted: date  Published: date    **Copyright:** © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). |

Jahnavi Jammula, Jithendra Uppalapati and Navanth Velagaturi

1 Affiliation 1: Department of Data Science, Saint Peters University  
jjammula@saintpeters.edu

2 Affiliation 2: Department of Data Science, Saint Peters University  
juppalapati@saintpeters.edu

**\*** Correspondence: svelagaturi@saintpeters.edu; Tel.:

**Featured Application: This study introduces a machine learning based framework for predicting in-hospital mortality risk following elective surgeries. The findings can be used to identify the underlying causes, help clinical decision making & post operative care strategies.**

**Abstract:** Identifying patients at high risk of in-hospital mortality is crucial for improving patient outcomes.  
This study developed and evaluated predictive models using logistic regression, random forest classifier, support vector machine and k nearest neighbors on a dataset containing Facility address, state, what complications, surgery-related features. The models were applied on this selected dataset with those features.   
Among the models, Random Forest Classifier demonstrated the highest performance with an, followed by logistic regression. Key measure ID including Score, Lower and higher estimate.   
The proposed machine learning framework effectively predicts mortality risk, enabling targeted interventions.

**Keywords:** In-Hospital Mortality; Elective surgery; predictive modeling

1. Introduction

Elective surgeries are common procedures that require precise preoperative assessments to minimize risks, Nathaniel R. Smilowitz[1] research says that Heart failure (HR) is associated with increased risk of complications after non-cardiac surgery. Miyake et al.[2]study found that patients with pre-operative AKI (Acute Kidney Injury) had significantly increased mortality regardless of CKD (chronic kidney disease) factors. AKI was associated with increased use of vasoactive drugs, mechanical ventilation, blood transfusions. Castro-Dominguez et al. [3] research highlighted the factors like procedural urgency, cardiovascular instability, level of consciousness after cardiac arrest were some of the factors contributing to in hospital mortality. Our dataset consists of deaths and their complications in respective facilities, with this data we can identify which complications has the highest deaths, predicting mortality rate for new patients and this can provide clinicians with actionable insights.

Our dataset contains records of death and complications at many facilities across different states in US. With this data we can identify the highest mortality rates and predict the mortality for the new patients.

2. Materials and Methods

Data Collection  
 The dataset used in this study was obtained from <https://data.cms.gov/provider-data/dataset/ynj2-r877> encompassing 90802 patient records from July 2020 to June 2023.

Each record includes

* Facility ID, Facility Name, Address, ZIP Code, State, Telephone Number, City, County
* Measure ID (Complication), Measure Name
* Compared to National, Denominator, Score, Footnote, Lower estimate, Higher Estimate
* Start Date, End Date.

1. Data Preprocessing

* Converted columns to appropriate data types
* Dropped irrelevant columns like facility telephone number
* Created visualizations for exploratory analysis.

2. Imputation of Missing Values

* We found missing values in denominator, score, lower estimate, higher estimate and replaced them with their respective mean

3. Feature Engineering

* Measure duration (from start and end date)
* Complication severity - score per denominator
* Surgical season (categorized by months)

Model Development  
 We have implemented the following machine learning models   
 1. Logistic Regression  
 2. Random Forest Classifier  
 3. Support Vector Machine  
 4. K-Nearest Neighbors

Evaluation Metrics  
 AUC / ROC : Measures model discrimination capability  
 Precision and Accuracy: Evaluate predictive reliability

Error Metrics

* Mean Absolute Error (MAE)
* Mean Square Error (MSE)
* Root Mean Square Error (RMSE)
* Mean Absolute Percentage Error (MAPE)

3. Results

3.1.1 Model Performance

1. Random Forest Classifier: Has the highest accuracy and precision

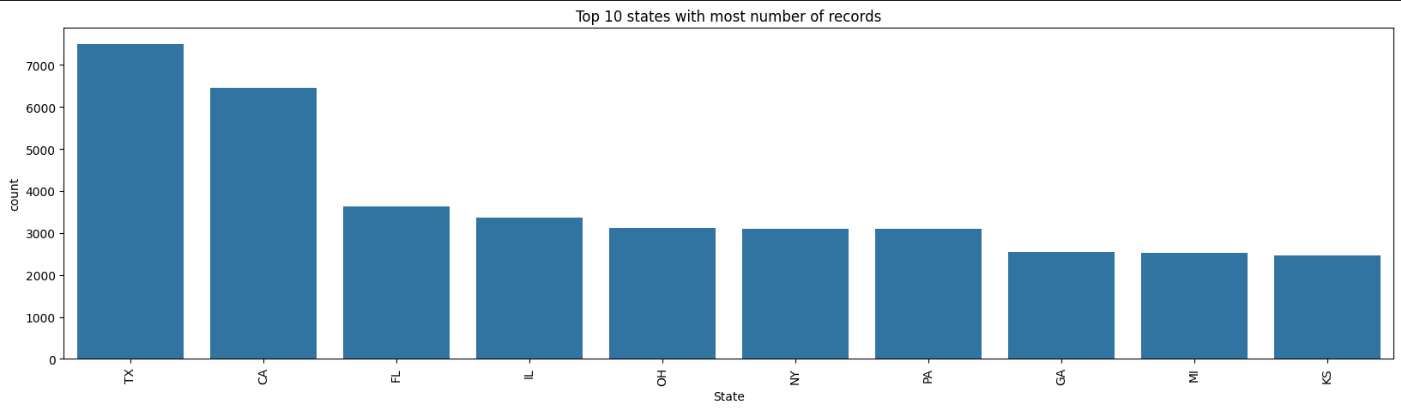
2. Logistic Regression: Has the next best accuracy, precision

3. SVM and KNN: Had low performance when compared to above models

3.1.2 Observations

1. Complications: Heart attack patients, hip & knee surgery patients have the most mortality rates. Memorial hospital and Community memorial hospital has the highest complications and deaths.

3.2. Figures, Tables and Schemes

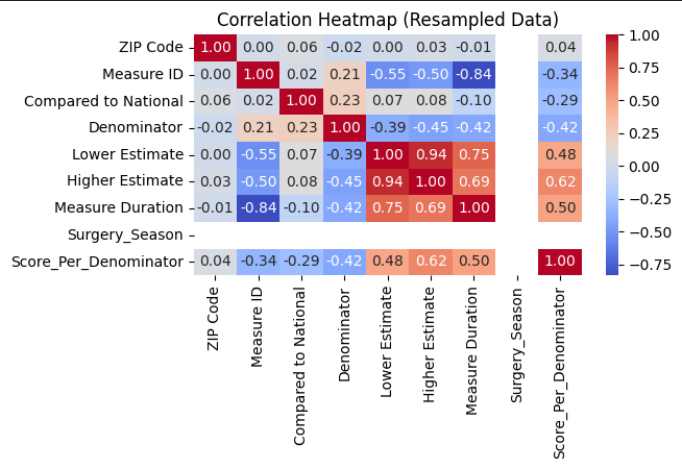
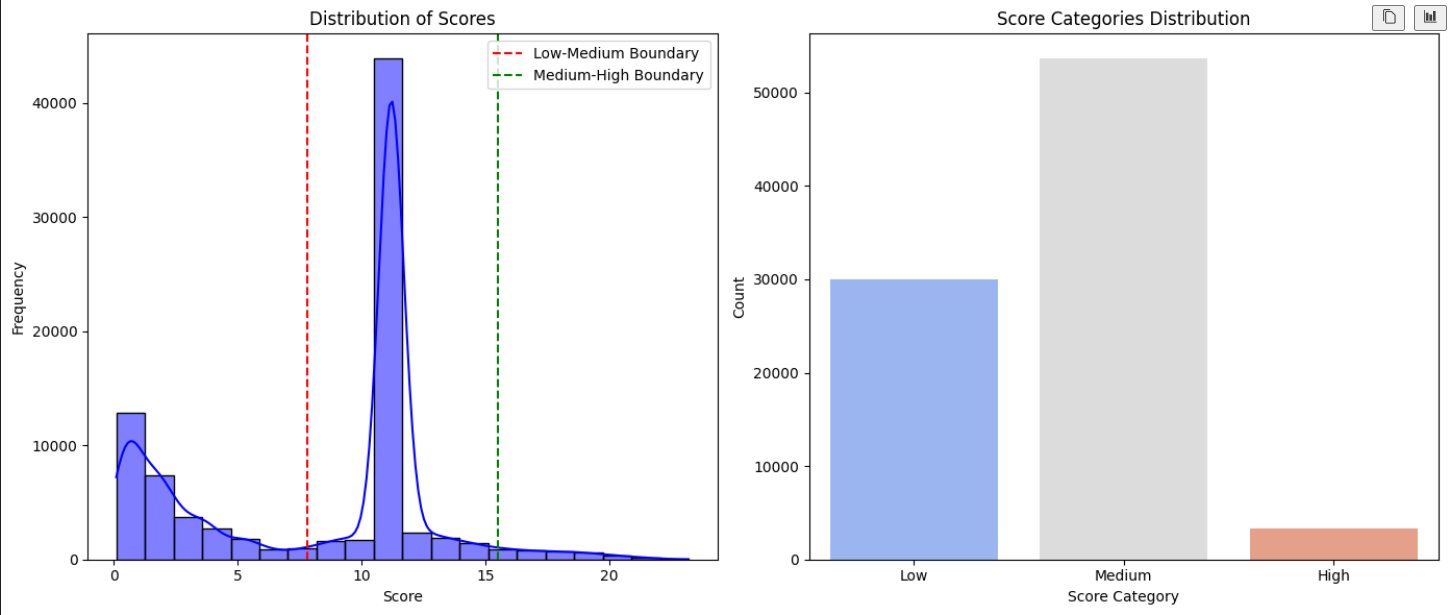
****

**Figure 1.** Top10 states with highest number of complications and deaths. (According to our data set – its Texas, California and Florida)

**Table 1.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Name** | **Accuracy** | **Precision** | **AUC/ROC** |
| Random Forest Classifier |  |  |  |
| Logistic Regression |  |  |  |

As you can see above out of all the models RFC has the highest accuracy.

The above figures – on the left hand side which represents the correlation heat map shows the following  
 1. Measure ID and Measure Duration are positively corelated  
 2. Lower Estimate and Higher Estimate are negatively corelated

The figure on the right side shows the distribution of scores as following  
 1. Three score categories – low, medium and high   
 2. Medium has the most number of records and few number of high indicating the complication rate

**4. Discussion**

Highlights the potential of Machine learning in predicting post operative risks. While Random Forest classifier demonstrated highest predictive power. Integrating these models into clinical workflow could enhance patient stratification. Future work should explore larger datasets and external validation to improve generalizability, integrating real time patient monitoring data could further enhance predictive accuracy.

5. Conclusions

This proposed machine learning framework can be useful in predicting the in-hospital mortality, as we can identify the death complication associated with it which helps improving clinical decision-making and postoperative care strategies. According to this data we can identify the hospitals with highest and lowest mortality rates.

**Supplementary Materials:** The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: title; Table S1: title; Video S1: title.

**Author Contributions:** Conceptualization, J.J. and J.U.; methodology, J.U.; software, J.J.; validation, J.J., J.U. and N.V.; formal analysis, J.J.; investigation, J.U.; resources, J.U.; data curation, N.V.; writing—original draft preparation, J.U., and J.J.; writing—review and editing, J.U and J.J.; visualization, J.U.; supervision, N.V.; project administration, J.J. All authors have read and agreed to the published version of the manuscript.”

**Funding:** The authors received no external funding for the research, authorship of this article.

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board (or Ethics Committee) of Saint Peters University (protocol code #### and date of approval is ).

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** The dataset used for this study is publicly available at   
<https://data.cms.gov/provider-data/dataset/ynj2-r877>

**Acknowledgments:** The authors acknowledge the administrative and technical support provided by Saint Peters University and our professor Mastouri Reda

**Conflicts of Interest:** There are no funders for this study. The authors declare no conflicts of interest

References

References must be numbered in order of appearance in the text (including citations in tables and legends) and listed individually at the end of the manuscript. We recommend preparing the references with a bibliography software package, such as EndNote, ReferenceManager or Zotero to avoid typing mistakes and duplicated references. Include the digital object identifier (DOI) for all references where available.

Citations and references in the Supplementary Materials are permitted provided that they also appear in the reference list here.

In the text, reference numbers should be placed in square brackets [ ] and placed before the punctuation; for example [1], [1–3] or [1,3]. For embedded citations in the text with pagination, use both parentheses and brackets to indicate the reference number and page numbers; for example [5] (p. 10), or [6] (pp. 101–105).

1. Smilowitz, N. R., Banco, D., Katz, S. D., Beckman, J. A., & Berger, J. S. (2021). Association between heart failure and perioperative outcomes in patients undergoing non-cardiac surgery. European Heart Journal - Quality of Care and Clinical Outcomes, 7(1), 68–75
2. Miyake, K., Iwagami, M., Ohtake, T., Moriya, H., Kume, N., Murata, T., Nishida, T., Mochida, Y., Isogai, N., Ishioka, K., Shimoyama, R., Hidaka. (2020). Association of pre-operative chronic kidney disease and acute kidney injury with in-hospital outcomes of emergency colorectal surgery: A cohort study. World Journal of Emergency Surgery, 15(1)
3. Castro-Dominguez, Y. S., Wang, Y., Minges, K. E., (2021). Predicting In-Hospital Mortality in Patients Undergoing Percutaneous Coronary Intervention. Journal of the American College of Cardiology, 78(3).
4. Author 1, A.B. (University, City, State, Country); Author 2, C. (Institute, City, State, Country). Personal communication, 2012.
5. Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference (Day Month Year).
6. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.
7. Title of Site. Available online: URL (accessed on Day Month Year).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.