City University of New York

School of Professional Studies

Predicting Hospital Readmission for Diabetes

Vijaya Cherukuri

Data698 – Research Project

December 7, 2020

Contents

[Abstract 3](#_Toc58237154)

[Introduction 3](#_Toc58237155)

[Background & Motive 4](#_Toc58237156)

[Data Source 5](#_Toc58237157)

[Prior Researches 5](#_Toc58237158)

[Methodologies 6](#_Toc58237159)

[Data Collection : 6](#_Toc58237160)

[Data Wrangling : 7](#_Toc58237161)

[Data Analysis : 7](#_Toc58237162)

[Feature Engineering 13](#_Toc58237163)

[Building Model and Evaluating: 13](#_Toc58237164)

[Data visualization and Dashboard 18](#_Toc58237165)

[Future Scope 20](#_Toc58237166)

[Technical References: 21](#_Toc58237167)

[Bibliography 21](#_Toc58237168)

# Abstract

Hospital Readmission for diabetes is one of the major healthcare quality concerns in recent days. Day by day the number of patient readmitting due to diabetes is increasing rapidly. There are multiple reasons behind the readmission, sometimes it might happen due to lack of proper medication or not going for follow up visits etc. In order to avoid hospital penalization and healthcare costs health insurance companies use the prediction model to identify the patients with diabetes to adjust for healthcare performance. In Healthcare insurance Care Coordination, Utilization management and Admission, Discharge and Transfer (ADT) teams needs better visualizations for diabetes patients for effective communication and frequent follow-up with patients. We will use Dash Plotly to publish a cloud interactive visualization dashboard. Earlier studies focus on key aspects like Length of stay, Acuity of admission and Emergency visits. We will use these aspects along with visualization. In order to deal with readmission prediction we will use UCI dataset collected from 130 US hospitals between 1999 - 2008. In this study we will use both simple linear (LR and LDA), nonlinear (KNN, CART, NB and SVM) algorithms for Patient Readmission for diabetes model. Generally Regression models are most commonly used models for risk standardization. Here we got the best results for SVM with 0.622 followed by Logistic regression with 0.614 and Linear Discriminant Analysis (LDA) with 0.609. Additionally we will use Random Forest, Extra Trees, AdaBoost and Gradient Boosting Classifier to check if there is any improvement. We got 0.622 for Gradient Boosting Classifier. SVM and Gradient Boosting are the best models in our study. There are certain limitations to this dataset as the data is collected only from 130 hospitals and readmission rows are less. Dataset doesn’t contain any hereditary health condition fields for patients as this is another important factor to consider. If we can consider latest data with more number of hospitals we can improve the accuracy of the prediction and there will be more future scope for the research.

# Introduction

Readmissions are one of the main reasons for increase in health care costs. The hospital readmission of patients within 30-days of discharge is crucial to healthcare quality and has involved with high costs in Medicare expenditures [1] [2]. Every year Health care insurance companies spend a lot to analyze the health insurance plans but it keeps on growing year after year. There are many incentives healthcare insurance companies’ uses to eliminate readmissions. Hospitals need to pay a lot if there are readmissions of patients with same reason. Diabetes is one of the main common readmission types. Despite the broad interest in readmission, relatively little research has focused specifically on readmission of patients with diabetes [3] [4] [5]. The burden of diabetes among hospitalized patients, however, is substantial, growing, and costly, and readmissions contribute a significant portion of this burden. Reducing readmission rates among patients with diabetes has the potential to greatly reduce health care costs while simultaneously improving care. Many patients with diabetic background who keep returning to acute care hospitals as unplanned readmissions within 30 days as these are the main reason for increased health insurance costs. Healthcare companies use a separate Care Coordination department to handle the diabetes patients by making frequent calls for checkup when they are due. Similarly there are others department like Utilization management and Fraud Waste and Abuse to determine that proper diagnosis is given to the patients and also to check if there is any extra charges or different ICD-10 codes claims are coming for diabetes patient [6] [7].

Patients with diabetes may be at higher risk of readmission than those without diabetes. In a study of 4769 medical patients, diabetes was associated with a statistically significant 40 % increased risk of readmission within 90 days. Diabetes is also associated with an increased risk of readmission in patients hospitalized for cardiac surgery [8] [9], heart failure [10], acute myocardial infarction, stroke, or liver disease. In contrast, recent studies have not found that diabetes is independently associated with an increased risk of 30-day readmission in medical patients or veterans. Considered together, these data suggest that the effect of diabetes on readmission risk may vary by length of follow-up and by primary reason for hospitalization.

There are earlier studies on readmissions that primarily focused on Cancer readmission and chronic conditions [11]. Even though some research concentrated on diabetes they concentrated on only inpatient visits ignoring outpatient visits. In their research they haven’t included visualization which is one of the key factors for analyzing data and provide better insight [12]

# Background & Motive

Diabetes is a chronic (long-lasting) health condition that affects how your body turns food into energy. Most of the food you eat is broken down into sugar (also called glucose) and released into your bloodstream. When your blood sugar goes up, it signals your pancreas to release insulin. Insulin acts like a key to let the blood sugar into your body’s cells for use as energy. Diabetes is a medical condition where patients have to undergo frequent checkups and medication. Insulin hormone is used to control blood sugar in the body. Main cause for diabetes is due to insufficient production of insulin from the pancreas.

Types of diabetes:

There are 2 types of diabetes that can be found in most of the patients.

* Type 1 : Body does not produce Insulin
* Type 2 : Body produces but unable to control blood sugars

Type 1 diabetes can be found in some children’s and young adults.

Type 2 diabetes is the form of disease that was found in majority of diabetes patients.

Diabetic patients with severe dysglycemia (uncontrolled high blood sugar or low blood sugar), are at increased risk for hospital readmission, despite their initial reason for hospitalization reports a new study in the Journal of General Internal Medicine.

This project will help me and my organization to analyze and do reporting and advanced analytics on ADT project (Admission, Discharge and Transfers). Health Insurance companies and Hospitals are penalized when patients are discharged and readmitted within 30 days.  Readmissions are associated with negative patient and financial outcomes. A readmission can be defined in multiple ways, including:

* Patients who are readmitted to the same hospital, or another applicable acute care hospital with same reason.
* Patients who are readmitted to the same hospital, or another applicable acute care hospital with different reason.

Health Insurance companies use this data to categorize how they can avoid the readmissions by analyzing the data. Both Health Insurance companies and Hospitals were very keen on analyzing the readmissions data. Our study is focused on patients with diabetic background who keep returning to acute care hospitals as unplanned readmissions within 30 days as these are the main reason for increased health insurance costs. Here using UCI dataset, we will attempt to create an interactive visualization that could help a hospital (business) to better understand their patient population, and what factors might determine if a patient will be readmitted.

# Data Source

The UCI Machine Learning Repository: Diabetes 130-US Hospitals for years 1999-2008 of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Data Set contains 100000 instances of data, with 55 attributes. The data contains attributes such as patient’s race, gender, age, time in hospital, number of lab tests performed, number of in-patient, outpatient, and emergency visits for prior year, etc.

# Prior Researches

In 2014 Nationwide Readmissions Database (NRD) developed by the Agency for Healthcare Research and Quality (AHRQ) Healthcare Cost and Utilization Project (HCUP), which includes data on nearly 15 million admissions from 2,048 hospitals. The NRD has the advantage of including all payers, including government and commercial insurers.

Their outcome was 30-day unplanned readmission created using the NRD Planned Readmission Algorithm. The NRD also includes patient-level information on demographics and up to 30 ICD-9 diagnosis codes and 15 procedure codes from each hospitalization. Among the diagnosis codes, the principal diagnosis code at discharge represents the primary reason for the hospitalization while the rest represent comorbidities for the patient. To improve computational efficiency they only included codes that appeared at least 10 times in the whole NRD database, reducing the number of ICD-9 diagnosis and ICD-9 procedure codes for inclusion in their analyses from 12,233 to 9,778 diagnosis codes and from 3,722 to 3,183 procedure codes, respectively.

They used four statistical models: 1) a hierarchical logistic regression model; 2) gradient boosting (using the extreme Gradient Boosting [XGBoost] approach, a widely-used, decision tree-based machine learning algorithm) using ICD-9 diagnosis and procedure codes represented as dummy variables (1 if present, 0 if absent); 3) an ANN model using a feed-forward neural network with ICD-9 codes represented as dummy variables; and 4) an ANN model in which ICD-9 codes were represented as latent variables learned through a word embedding algorithm. In their study they used hierarchical logistic regression as a baseline comparator given its ubiquitous use in health services and outcomes research. XGBoost is based on gradient boosted decision trees and it is designed for speed and performance.

Another study was conducted using health record data (without patient names) taken from 20 hospitals across Sutter Health, a large nonprofit hospital network serving Northern California. The Institutional Review Board (IRB) of Sutter Health (SH IRB # 2015.084EXP RDD) approved the study. The study used data from in-patient only visits and excludes outpatient visits. The researches also mapped the patient data using Google’s Geocoding API to determine coordinates of each patient home, using 2010 census data.

In their research they trained and tested a neural network model to predict the risk of patients’ re-hospitalization within 30 days of their discharge. The Model has several advantages over LACE, the current industry standard, and other proposed models in the literature including (1) significantly better performance in predicting the readmission risk, (2) being based on real-time data from EHR, and thus applicable at the time discharge from hospital, and (3) being compact and immune to model drift. Furthermore, to determine the classifier’s labeling threshold, we suggested a simple cost-saving optimization analysis.

The model was initially trained on all features from the dataset (1667 features), eventually being tuned to use a subset of the features (top 100), which showed a reasonable amount of near-optimal performance.

# Methodologies

In this research project we will follow these methodologies:

## Data Collection :

Data is collected from UCI web link (https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008). This data is published to Github and can be read from there and used in this project. Based on the need and usage this data can be loaded in to any traditional databases and to get a clean data which is effectively used for better Data Visualization.

## Data Wrangling :

We can observe the characteristics of the dataset from the UCI web link. It has multivariate characteristics which contains both Numerical and Nominal data types. Our data set has some missing values. These missing values are effectively handled in the data wrangling phase. Here we might impute the missing data a value or drop the field if is it doesn’t have any significance. For effective model first we have to concentrate on the data analysis part. This Data Analysis is the key to explore the data and have clear understanding about the attributes present in the dataset.

## Data Analysis :

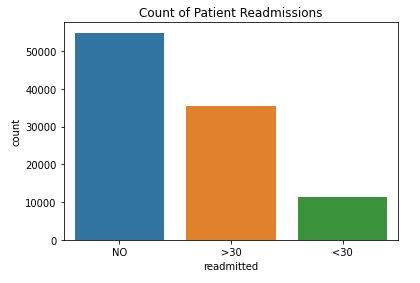
In this dataset we have data for ICD9. ICD9 is the diagnosis code which is used for our analysis. There are different diagnosis codes listed for diabetes. We can find the details about diagnosis codes and other diagnosis codes from below websites [13] [14].

Now we have loaded the data and we have good amount of data to deal with Readmissions. For our research the most important field is “readmitted”. We found more than 10K patients that are readmitted in less than 30 days and 35K patients readmitted in more than 30 days. We have 54K patients with no readmissions records in the dataset.

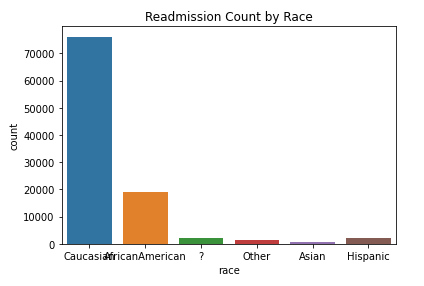
|  |  |
| --- | --- |
| **Readmitted** |  |
| <30 | 11357 |
| >30 | 35545 |
| No | 54864 |

We need to explore more on the data. Below are some of the analysis we found on the dataset. All the plots are visually available in the .ipynb file attached in this document.

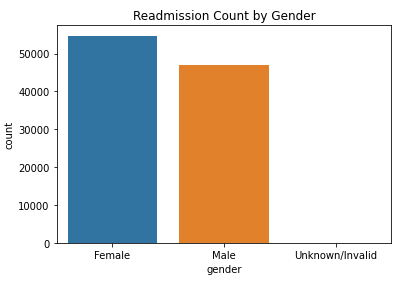
* Plot to analyze readmissions: This is to analyze how many readmissions are there in the dataset. It also tells us how many are readmitted before 30 days and after 30 days. In this dataset we can see that majority are not readmitted. When we check the data between readmissions happened before 30 and after 30 days we got more data for people readmitting after 30 days.



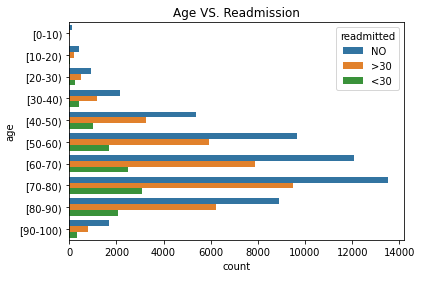
* Plot to analyze race field: This plot is used to get better understanding of how are the records on the race field. Here we found more number of patients in ‘Caucasian’ when compared to the other race.



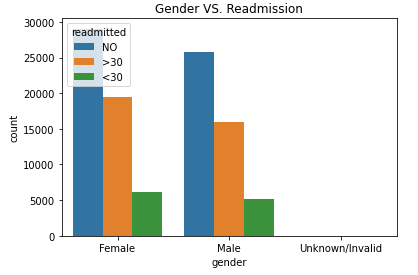
* Plot to analyze gender field: This is used to identify in which gender we have more diabetic patients. Apparently stas show that there are more number of Female readmissions happened than Male.



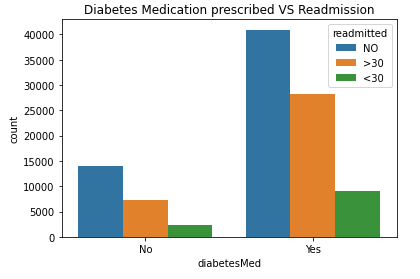
* Plot to analyze Age vs. Readmission: To identify the distribution of Age Vs Readmission. This helps us to categorize which age people are more affected to diabetic readmission. Plots suggest that there are more number of patients from 40 to 90 years which is expected.



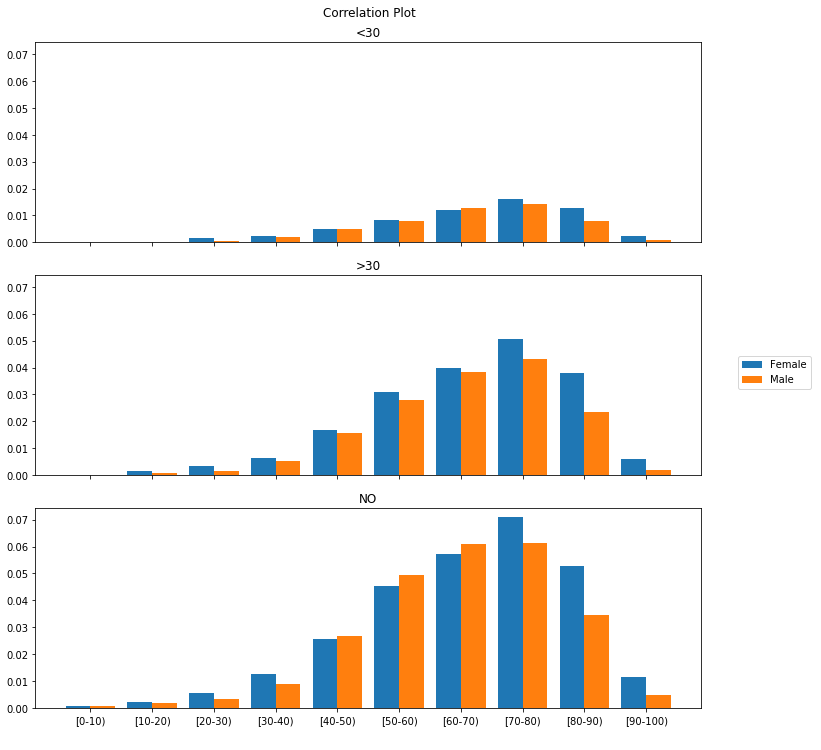
* Plot to analyze Gender vs. Readmission: To identify the distribution of Gender Vs Readmission. This helps us to categorize which age people are more affected to diabetic readmissions. The plot shows more number of Female people are readmitted before 30 days and after 30 days when compared to Male gender.



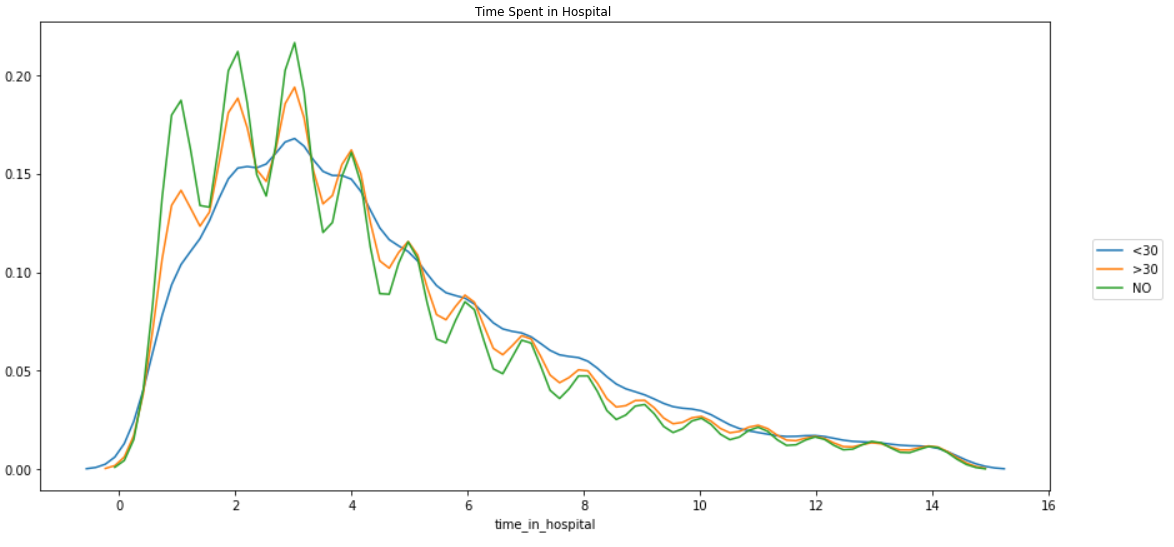
* Plot to analyze Diabetes Medication Prescribed vs Readmission: To identify the distribution of Diabetes Medication Prescribed vs Readmission. This helps us to categorize if Diabetes Medication was prescribed to the patients and still they got readmitted.



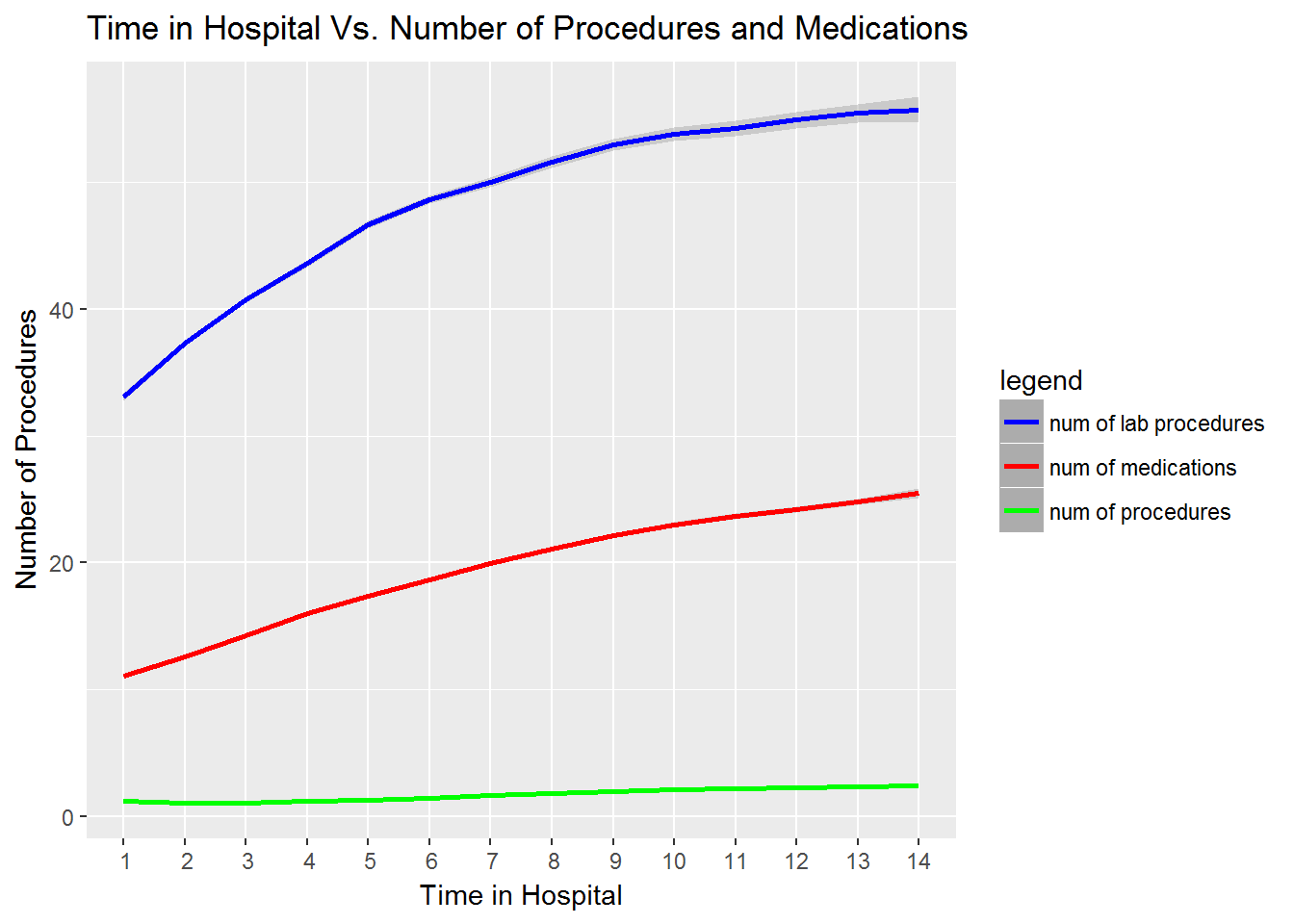
* Plot for Correlation: We tried to find the correlation between race, gender, age and how these features relate to the response variable ‘readmitted’. It seems like the number of patients are lowest who got readmitted in less than 30 days. Whereas Readmission rate appears to be highest for both male and female for more than 30 days period. There are significant number of patients who do not return to the hospital. The readmission data for the age group between 40 to 90 appears to have normal distribution and peaking for the age group 70-80 for both more than 30 days and less than 30 days period.



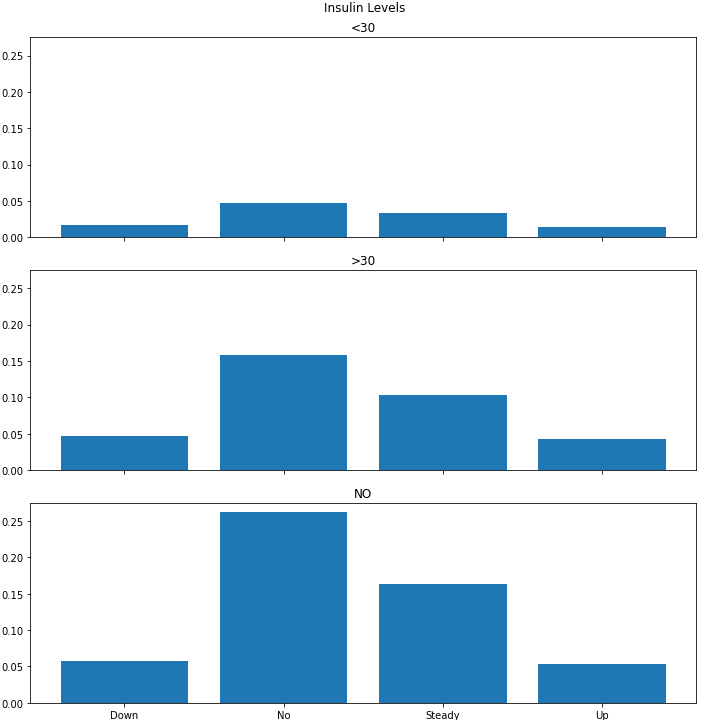
* Patient Discharge Graph/Time Spent in Hospital : As plot suggests that patients who are discharged between 2 to 5 days are having high chances of being readmitted to hospitals. The plot seems to be right skewed and number of patients who was not readmitted to hospitals also appears to be discharged within 2 to 5 days. So, we can assume that the length of stay appears to be one of the key predictor for the readmission but there are definitely other predictors available in the dataset which have larger impact on the readmission outcome.



* Time in Hospital Vs Number of Procedures: The number of lab procedures and number of medications seems to be having upward trend with respect to the length of stay in the hospital. We can also observe that the number of procedures remain low as patients stays longer in hospital. With this we can find number of lab procedures and number of medications might play significant role for readmission cases.



* Plot for Insulin Levels : The Plot shows the insulin levels for the patients that were admitted before 30 days, after 30days, and not readmitted. We can observe that in the patients readmitted before 30 days the data was almost flat line for different insulin levels, where “No” as being slightly higher than the others. Whereas for patients readmitted after 30 days the data show “No” being the highest recorded and followed by steady. For patients not readmitted or “No”, “No” appears to be the highest recorded insulin, followed by steady.



## Feature Engineering

Feature engineering plays a key role in this research project. We will add new variables necessary to add to the data frame. These new variables are used while creating our predictive model. In Feature engineering we will take count of both Numerical and Categorical fields.

Numerical Features: It is very easy for us to concentrate more on Numerical fields as these fields need less modification. If there are any missing values then we will impute them with appropriate values.

Categorical Features: Categorical fields are non-numeric or text fields. We will check for any missing values and impute with appropriate values.

## Building Model and Evaluating:

In this Research project we will create a model to predict that patients that are at high risk of being readmitted to hospitals after their discharge. We will use different modeling techniques for our readmission prediction. To better understand the effectiveness of the model we build the predictive model for readmission and run the model on different criteria and tools to predict if a patient is likely to be readmitted to a hospital within 30 days.

For better model building we need to do data pre-processing. Here are some of the steps we followed for data preprocessing.

* Drop features like payer code, weight, medical speciality
* Some diag codes like Diag1, Diag2 and Diag3 have very less missing values. We dropped records where all 3 diag codes are missing.
* Since we are counting only on readmission, we are dropping patients who have expired which can be identified in discharge\_disposition\_id.
* Mapping Diag codes according to ICD9 specifications as our dataset has ICD9 doag codes.
* Our primary focus is for patients readmitted before 30 days for risk assessment. If there are patients readmitted after 30 days we can mark them as “No” for identification.
* There are lot of features available in the dataset for consideration. We can leverage these features for our model building. We can start with simple feature and can create composite features and other enhanced techniques for better model building.
* Time\_in\_hospital
* Number\_lab\_procudures
* Number\_procedures
* Number\_medication
* Number\_inpatient
* Number\_outpatient
* Number\_diagnosis
* Number\_emergency

We will use multiple machine learning methodologies like

* Logistic Regression (LR)
* Linear Discriminant Analysis (LDA)
* K-Nearest Neighbors (KNN).
* Classification and Regression Trees (CART).
* Gaussian Naive Bayes (NB).
* Support Vector Machines (SVM).

We will use libraries like numpy/pandas etc. in building the model. We will have to consider our machines processing power, and the amount of data points available (especially for use of a neural network).

This is a mixture of both simple linear (LR and LDA) and nonlinear (KNN, CART, NB and SVM) algorithms. We reset the random number seed before each run to ensure that the evaluation of each algorithm is performed using exactly the same data splits. It ensures the results are directly comparable [15].

**Logistic regression:** Logistic regression is used to conduct when the dependent variable is dichotomous. It is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables. It is one of the classification technique used for discrete cases [16].

We got the accuracy value of 0.614. Here are the Confusion matrix and classification report

0.6149159870295765

[[9119 1850]

[5988 3397]]

precision recall f1-score support

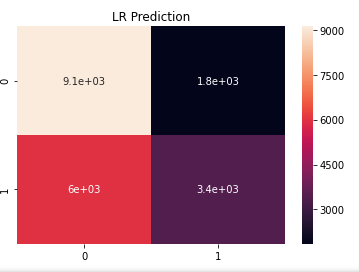
0 0.60 0.83 0.70 10969

1 0.65 0.36 0.46 9385

accuracy 0.61 20354

macro avg 0.63 0.60 0.58 20354

weighted avg 0.62 0.61 0.59 20354



**Linear Discriminant Analysis:** LDA is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications. LDA makes predictions by estimating the probability that a new set of inputs belongs to each class. The class that gets the highest probability is the output class and a prediction is made.

We got the accuracy value of 0.609. Here are the Confusion matrix and classification report

0.6095607742949789

[[9255 1714]

[6233 3152]]

precision recall f1-score support

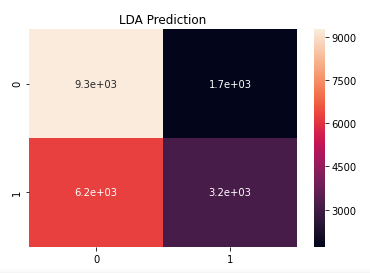
0 0.60 0.84 0.70 10969

1 0.65 0.34 0.44 9385

accuracy 0.61 20354

macro avg 0.62 0.59 0.57 20354

weighted avg 0.62 0.61 0.58 20354



**Support vector machine:** It is highly preferred by many as it produces significant accuracy with less computation power. SVM can be used for both regression and classification tasks [17] [18].

We got the accuracy of 0.622

0.6225803281910189

[[8357 2612]

[5070 4315]]

precision recall f1-score support

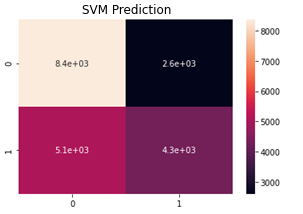
0 0.62 0.76 0.69 10969

1 0.62 0.46 0.53 9385

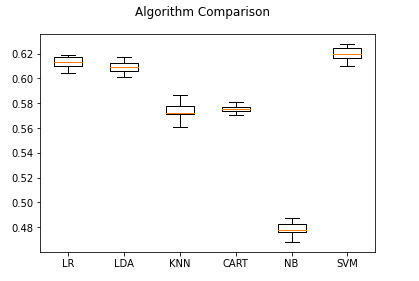
accuracy 0.62 20354

macro avg 0.62 0.61 0.61 20354

weighted avg 0.62 0.62 0.61 20354



Other models KNN, CART and GNB did not gave promising results. If we compare the results of all the 5 methods Support Vector Machine gave the best accuracy followed by Logistic Regression and Linear Discriminant Analysis.



Now lets check if bagging models and boosting algorithms improve the performance of the model [19].

We will use additional models like [20]

* Random Forest
* Extra Trees
* AdaBoost
* Stochastic Gradient Boosting

**Random Forest:** Random Forest is an extension of decision trees. Here large number of decision trees are created and each observation is fed into decision tree. The new observations are fed into all trees and taking majority vote for each classification model [21].

Here we got an accuracy of 0.581

RF: 0.581462 (0.004017)

**Extra Trees :** Extra Trees is an ensemble machine learning algorithm that combines the predictions from many decision trees. It is related to the widely used random forest algorithm. It can often achieve as-good or better performance than the random forest algorithm, although it uses a simpler algorithm to construct the decision trees used as members of the ensemble.

We got an accuracy of 0.580

ET: 0.580553 (0.005800)

**AdaBoost:** AdaBoost is an ensemble method that trains and deploys trees in series. AdaBoost implements boosting, wherein a set of weak classifiers is connected in series such that each weak classifier tries to improve the classification of samples that were misclassified by the previous weak classifier.

We got an accuracy of 0.620

AB: 0.620621 (0.005522)

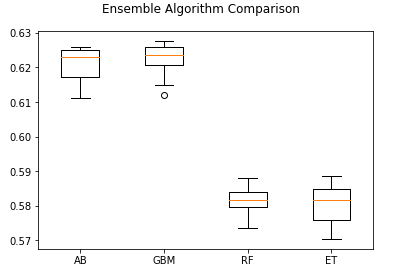
**Stochastic Gradient Boosting:** Stochastic Gradient Boosting is also called as Gradient Boosting Classifier which is an additive ensemble of a base model whose error is corrected in successive iterations (or stages) by the addition of Regression Trees which correct the residual. GB builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions.

We got an accuracy of 0.622

GBM: 0.622157 (0.004820)

If we compare these 4 models we got better results for GBM followed by AdaBoost, Random Forest and Extra Trees.

Here is the comparison of these 4 models.



## Data visualization and Dashboard

Data visualization is the graphical representation of information and data. It plays an important role in analyzing the data. Better visualization can really help users to analyze data quickly and take decisive actions. In our project one of the key goals is to provide better visualization to the business users.

Here we are trying to use Plotly to create a dashboard which serves the business needs. We will create an app using Plotly Dash to create interactive visualizations of our dataset. The visualizations will include various forms including:

* Patient demographics (age, sex, etc.)
* Medication history and how medications impact a patient being readmitted to the hospital

This visual dashboard will help business departments like Utilization Management, Claims Administrators, Fraud Waste and Abuse department to identify the number of readmissions happening and what is the root cause for the readmission. With diabetes how many patients are readmitting to hospitals, what is their length of stay etc. This dashboard also helps in taking the decision of how to deal with readmissions particularly for diabetes. Best example is frequent checkups and incentives that can offer to the patients to maintain better health.

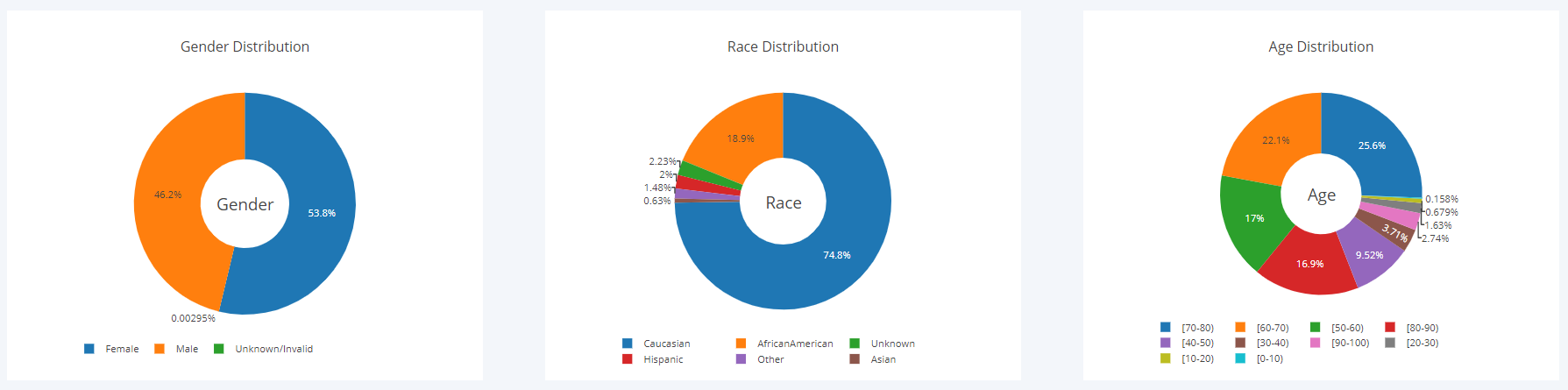
For visualization we will use Heroku for deploying our dashboard. Heroku as a well-loved platform for deploying hobbyist web applications. The Heroku Dashboard allows us to create multiple dashboard pages where each dashboard page contains multiple reports which are helpful for better decision taking.

Heroku Patient Readmission Dashboard link [22]: [https://dash-698.herokuapp.com/](about:blank)

Readmission dashboards contains following reports:

Demographics Page :

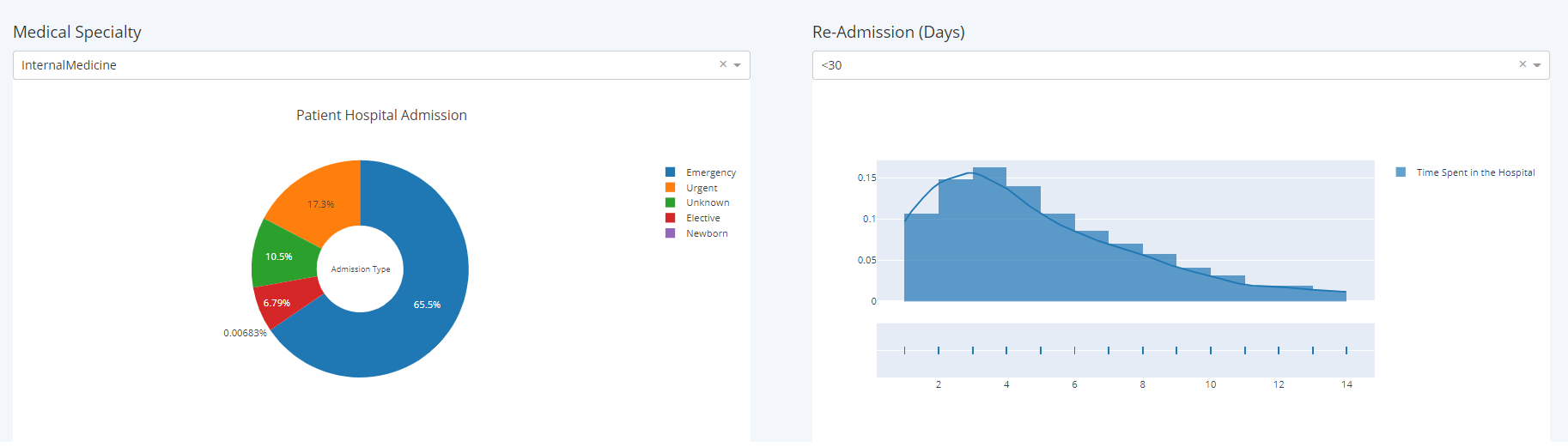
* Gender Distribution
* Race Distribution
* Age Distribution
* Race-Gender Distribution
* Age-Gender Distribution
* Race-Age Distribution





Medical Specialty And Readmission Page:

* Medical Specialty
* Re-Admission(Days)



# Future Scope

There is always a scope of improvement. Here are some of the aspects for improvement.

* More Data
* Invent Data
* Transform Data
* Feature Selection

**More Data:** The data we have is collected from 130-US Hospitals for years 1999-2008 for specific delivery network. Data Set contains only 100000 instances of data. If we can get the latest data from 2010 to 2020 with more data instances and good amount of hospitals we can improve the performance.

**Invent Data:** Here we have invented only few columns like readmitted\_30 to check the number of patients that are readmitted within 30 days and also fields indication only readmitted patients. If we can invent more data like genetic behavior then there is room for inventing new fields and better prediction.

**Transform Data:** Spot-check lots of different transforms of your data or of specific attributes and see what works and what doesn’t.

**Feature Selection:** We have dropped few fields and took only the necessary one. May be if we include or exclude some more features then it will have an impact on the outcome.

**Reframe Problem:** We have taken the readmission data and checked from diabetes. Diabetes can also occur through hereditary genes. So if we try to reframe the problem then there might be space for improvement.

# Bibliography

|  |  |
| --- | --- |
| [1] | M. M. A. M. Epstein. [Online]. Available: https://www.nejm.org/doi/full/10.1056/NEJMe0901006. |
| [2] | M. M. R. N. A. a. M. M. V. Williams. [Online]. Available: https://jamanetwork.com/journals/jama/article-abstract/645364. |
| [3] | K. S. C. Cook. [Online]. Available: https://journals.aace.com/doi/abs/10.4158/EP.15.3.263. |
| [4] | M. P. B. Draznin. [Online]. Available: https://care.diabetesjournals.org/content/36/7/1807. |
| [5] | M. k. M. Dungan. [Online]. Available: https://journals.sagepub.com/doi/10.1177/193229681200600508. |
| [6] | C. Mathews. [Online]. Available: https://via.library.depaul.edu/cgi/viewcontent.cgi?article=1316&context=jhcl. |
| [7] | T. M. W. a. D. Lessler. [Online]. Available: https://www.annualreviews.org/doi/full/10.1146/annurev.publhealth.23.100901.140529. |
| [8] | M. M. S. A. W. M. S. M. Dunlay. [Online]. Available: https://www.acpjournals.org/doi/10.7326/0003-4819-157-1-201207030-00004. |
| [9] | E. C. L.-L. J. H. Lichtman. [Online]. Available: https://www.ahajournals.org/doi/10.1161/STROKEAHA.110.599159. |
| [10] | A. l. o. o. p. H. T.AbrahamMD. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0002870307004140?via%3Dihub. |
| [11] | N. E. R. Solomon. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30925070/. |
| [12] | A. Long. [Online]. Available: https://towardsdatascience.com/predicting-hospital-readmission-for-patients-with-diabetes-using-scikit-learn-a2e359b15f0. |
| [13] | [Online]. Available: https://dexur.com/icd9/2500/. |
| [14] | [Online]. Available: https://greatplainsqin.org/wp-content/uploads/2016/01/ICD9-Code-Crosswalk-for-Hospital-Readmissions.pdf. |
| [15] | [Online]. Available: https://machinelearningmastery.com/compare-machine-learning-algorithms-python-scikit-learn/. |
| [16] | [Online]. Available: https://www.statisticssolutions.com/what-is-logistic-regression/. |
| [17] | [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47. |
| [18] | [Online]. Available: https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/. |
| [19] | [Online]. Available: https://machinelearningmastery.com/improve-deep-learning-performance/. |
| [20] | [Online]. Available: https://machinelearningmastery.com/ensemble-machine-learning-algorithms-python-scikit-learn/. |
| [21] | [Online]. Available: https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd. |
| [22] | [Online]. Available: https://dash-698.herokuapp.com/. |