greatlearning

Prediction of Early Hospital Readmission of Diabetic patients

Batch : Group 2 Online MARCH - 2022

Team Details

1. H H Arvind
2. Ram Krishna Namdev
3. Yerrabachu Keerthy Rao
4. Prangi Sharma
5. Tarun Tadikonda

Mentor

Mr. Jayveer Nanda

Introduction and Importance of the problem

Hospitals engaging in any model are likely to face penalties if their providers cannot improve hospital readmission rates. In recent years, government agencies and healthcare systems are increasingly focused on 30-day readmission rates as a way to improve quality Health care.

To avoid value-based penalties readmission rates Hospitals should reduce early readmission by identifying the Diabetic patients who are having high probability of compared to other patients who do not have Diabetes.

As cost of inpatient care & readmission rates are higher in patients with Diabetes Mellitus (DM) compared to other diagnosis focusing on reducing early readmissions, cost of readmissions to avoid value-based penalties from the government is primary research goal.

greatlearning

Business objective

The main objective of our work is to come up with the predictive model which helps Hospital Management systems to predict the risk of early readmission of patients who are having Diabetes Mellitus which can further address

- Enhanced patient care, patient Engagement, glucose monitoring, Transitional care & Post discharge follow up

- Reducing cost of early readmission there by reducing penalties to the Hospital which increases its commercial value in terms of reputation in health care & escape value based penalty.

greatlearning

Data set information

'Diabetes 130-US hospitals for years 1999-2008 Data Set'

To achieve our business objective we have considered the Analysis of a large clinical database which was undertaken to examine historical patterns of diabetes care in patients with diabetes admitted to a US hospital during the period 10 year period (1999 – 2008) . The dataset has been retrieved from UCI Machine learning Repository which is related to Diabetic inpatient encounters in 130 Hospital across United States.

- ✓ Dataset contains 101766 unique encounters corresponding to more than 70000 patients
- ✓ Dataset has 50 features including target label related to diabetes of which
  - ❑ Numerical features – 8
  - ❑ Categorical features – 41 (23 medicines + 18 patient information)
  - ❑ Target

Analysis of Target Variable

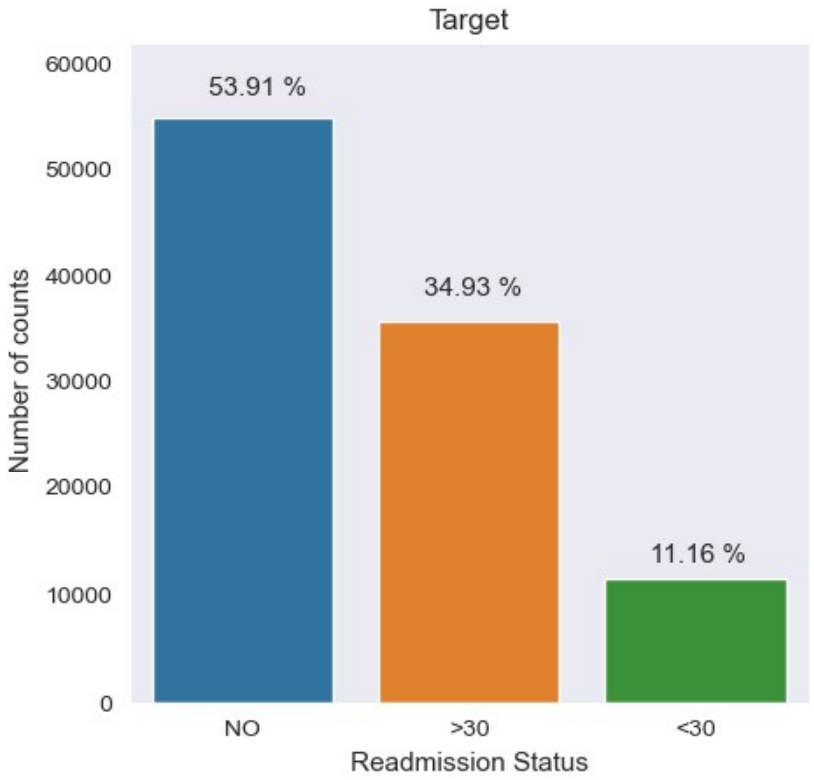
- ✓ 'Readmitted' is the target variable and it is of categorical (Nominal) data type. Hence our problem is a classification problem

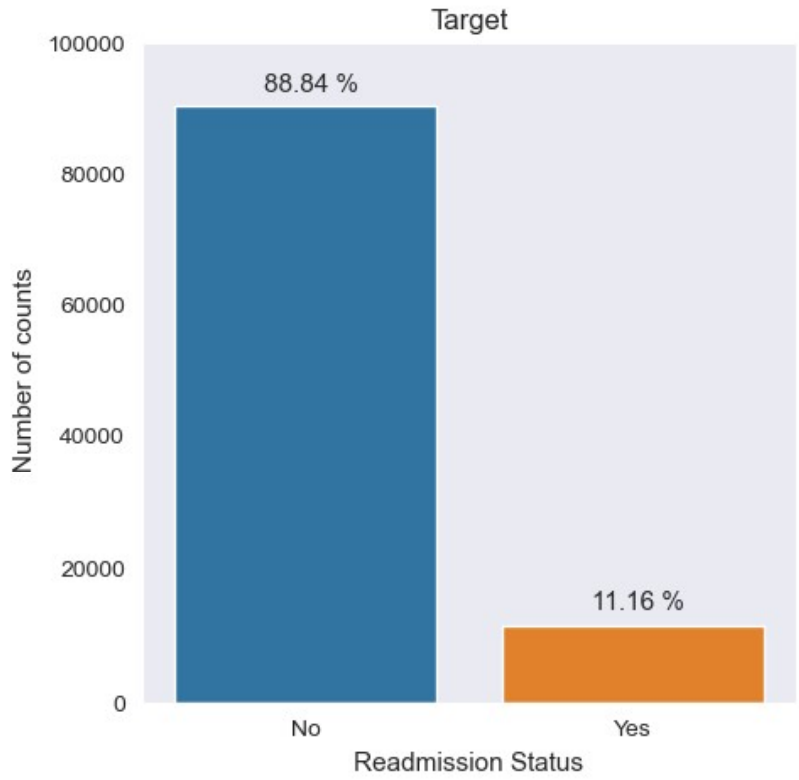
- ✓ There are 3 levels in the Target variable **:**
  - ❑ No : No Readmission
  - ❑ < 30 Days : Readmission within a month
  - ❑ > 30 Days : Readmission after a month

- ✓ A data imbalance is observed for the target variable which might effect the accuracy of the model.
- ✓ Since our business objective is to predict 'Early Hospital Readmissions' preferably (<30 days). we will redefine Multi – class Target into Binary Target.

greatlearning

Redefining Target label (Binary classification)





✓Categorized the target variable into two levels due to class imbalance:

- ❑ 0 : Not Readmitted (NO or > 30 days) : 'No'
- ❑ 1 : Readmitted (< 30 days) : 'Yes'

Missing value treatment

• **Gender :** As 'Unknown/Invalid' are only 3 observations, so we have imputed using mode.

• **Diagnosis :** In diagnosis columns,first three digits are coded from ICD9, so from research study we categorized the data into description using ICD9 codes.

Variable	%
weight	96.85%
Medical specialty	49.08%
Payer code	39.55%
Race	2.23 %
Diagnosis 1	0.02 %
Diagnosis 2	0.35 %
Diagnosis 3	1.40 %

- **Medical specialty** is an important information related to Readmissions, we have considered missing values as it as 'Not mentioned' category
- **Race, Diagnosis1, Diagnosis2 & Diagnosis3** are being imputed using Iterative imputer (KNN Algorithm) using only numeric values as features which found to significant with Target (Features which contain missing values)
- We dropped weight which had more than 90% missing values & Payer code which does not give any information regarding readmissions.

greatlearning

Key business findings (EDA)

- ✓ Most number of patients early readmissions are in the age range of 50-80 With the highest patient readmission between 70-80. As the age increases the Early readmission cases increase.
- ✓ patients who have not undergone HBA1c test / glucose serum test, they have high chance of early readmission.
- ✓ patients who are taking diabetic medication have higher chance of early readmission which has relevance in our business objective.
- ✓ risk of readmission decreased over the time with frequent visits.
- ✓ on an average every patient is under 15 medications, average lab procedures around 40 and average time spent in hospital is 4 days.
- ✓Insulin, Metformin are the major medication prescribed to patients which are taken by more than 70 % of patients

Considerations –

- ✓ There are 23 medicines in total out of them we are considering 16 features.
- ✓ we have also dropped patients who are expired, patients who are sent to hospice facilities.

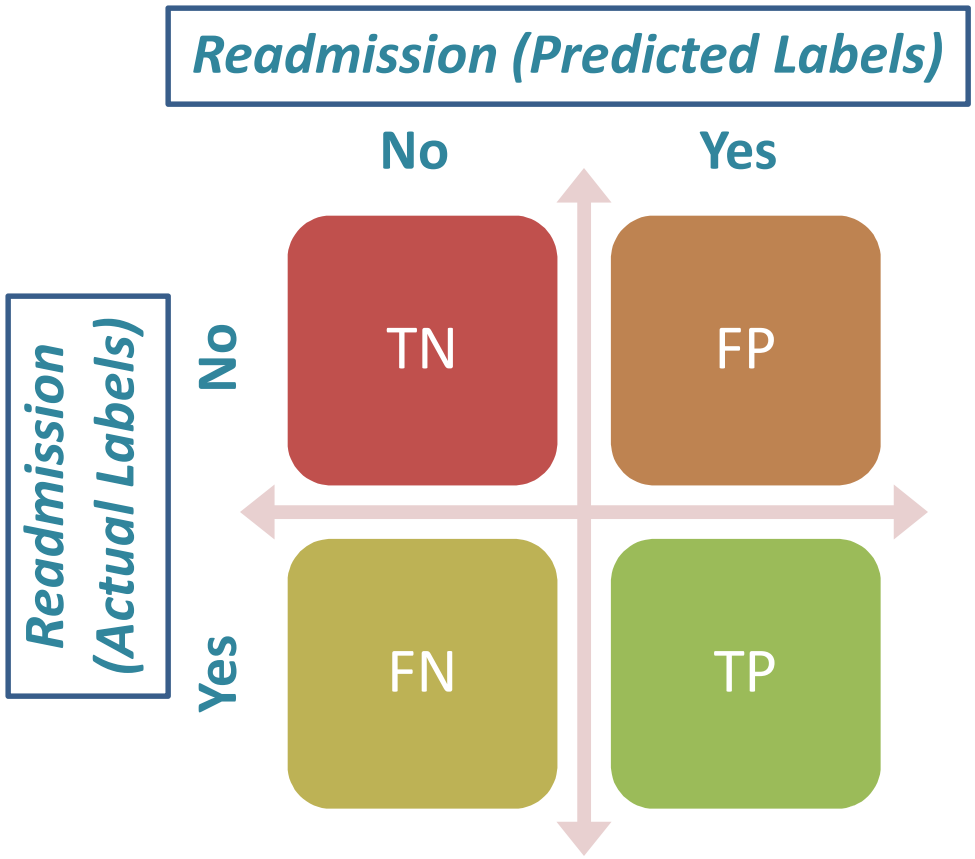
Evaluation Metrics

An evaluation metric which quantifies the performance of a predictive model related to our business objective.

In our case, if readmitted patient is predicted as non-admitted then there is a risk of patient life.

It will leads to hospital law suits, loss of reputation, huge business loss so our main target is to decrease the false negatives which can be done by using

- ✓ **RECALL**
- ✓ **F1 SCORE**



$$\text{Recall} = \frac{TP}{(TP + FN)}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

Feature Engineering

Health index :

Based on frequency of patient's visit to the hospital is high then we can say that patient is less healthier and less healthier patient tends to readmit quickly. Higher the health index lesser the chance that person will readmit (inversely proportional)

Health Index = (number emergency +
number inpatient +
number outpatient)

Severity of Disease :

Severity of disease is the feature created based on time spent in hospital, number of procedures, medications taken by patient. For probabilistic interpretation we divided it by total values.

Severity of disease = 1 / (time_in_hospital (in days) +
number of procedures +
number of medications +
number of lab procedures +
number of diagnoses)

greatlearning

Results (after Resampling, feature engineering)

	Accuracy	Precision	Recall	F-1 score
Logistic Regression	0.90	0.90	0.89	0.90
XG-Boost	0.93	0.99	0.86	0.92
Decision Tree	0.82	0.75	0.87	0.81
Random Forest	0.95	0.99	0.86	0.92
KNN	0.79	0.74	0.89	0.82
Naïve Bayes	0.66	0.57	0.88	0.73
Adaptive Boosting	0.89	0.92	0.86	0.89
Gradient Boosting machine	0.89	0.92	0.87	0.89

Resampling technique
K-means SMOTE

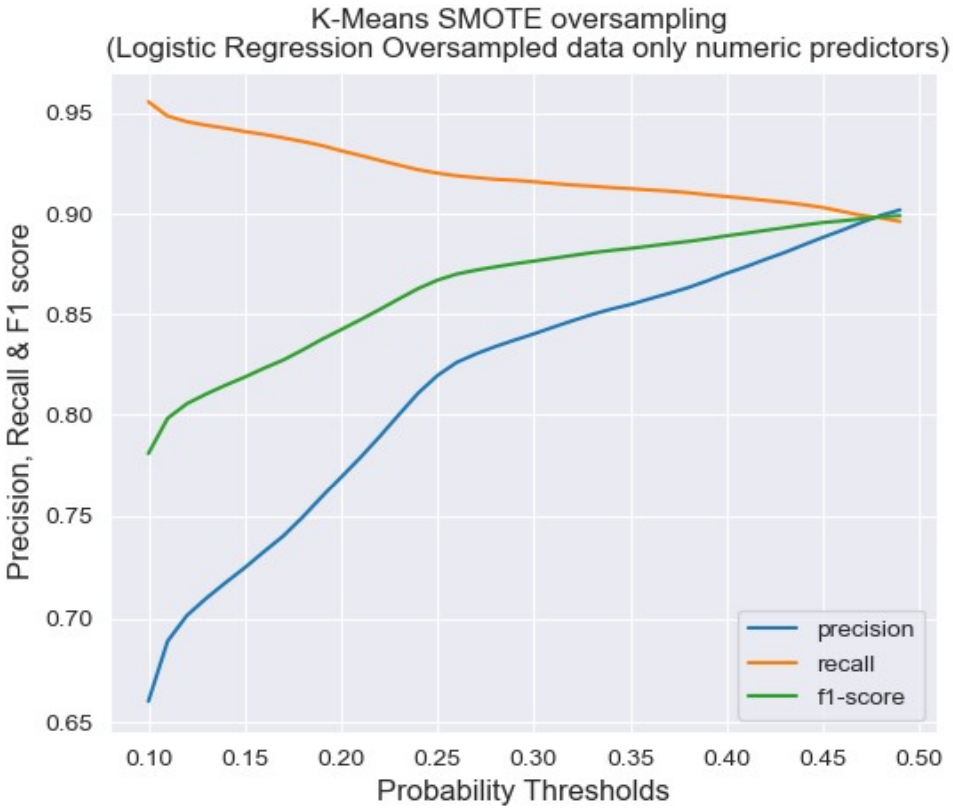
Number of features
10 features

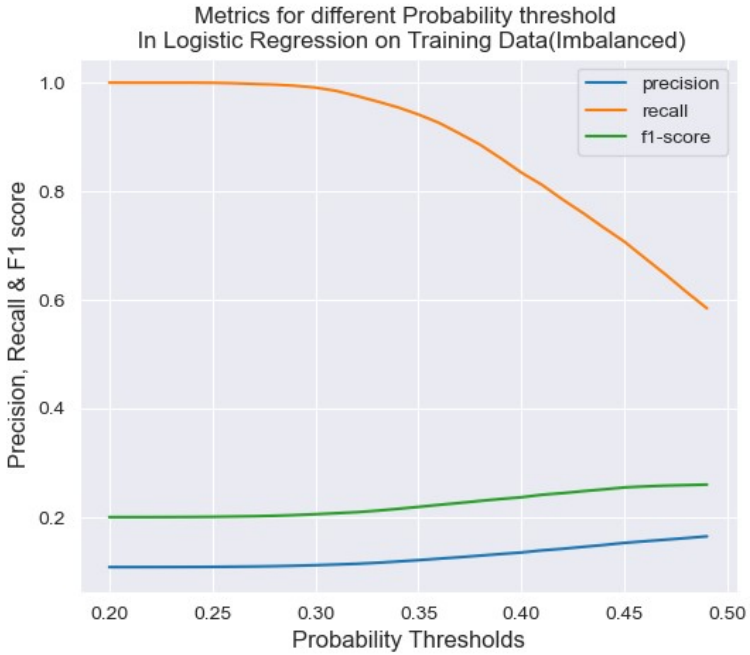
Final model
Logistic regression (10 features)

K-Means SMOTE hyperparameters
cluster balance threshold = 0.134
n neighbors = 12
sampling strategy = 0.99

Feature selection
fit on numerical and categorical predictors separately
(Adhoc approach)

Logistic regression (Probability threshold)





- At probability threshold between 0.45 and 0.5 yield all metrics converging to 90%

greatlearning

Additional work :

- ✓ Explored different types of encoding techniques.
- ✓ Tested various feature selection techinques.
- ✓ Applied different scaling techniques.
- ✓ Worked on Grid Search CV and Randomized CV for tuning hyperparameters.

Challenges :

- ✓ Due to insufficient domain knowledge, couldn't able to impute missing values in better way.
- ✓ Couldn't able to build multi-calssification model as data is highly imbalanced.
- ✓ could not interpret the coefficients/feature importance due to insufficient domain knowledge

Recommendations

- ✓ Increase glucose monitering capabilities (HBA1c / serum test)
- ✓ Support patient medication adherence to prevent rehospitalization
- ✓ Identify risk factors for readmission
- ✓ Follow up with patients after discharge (Post discharge care)

Deployment

The Machine learning model which will meet the business requirements which will enhance the early hospital readmission in case of diabetes.

Web application – deployed using streamlit cloud

Link - Readmission app

greatlearning