DIABETES PATIENTS' READMISSION

ADEDOYIN OYEBANJI

November 28, 2022

Introduction

- Machine learning has proven to give insight on a lot of healthcare related issues.
- Diabetes patients' readmission being one of them has not only proven costly but also risks the patients' medical condition.
- Major studies propose that if there is unplanned readmission within 30days, it indicates treatment or diagnosis. However, if readmission is after 30days, it depends on the patients' lifestyle or several factors. So an early prediction of readmitting the patients becomes an important task.
- Diabetes Mellitus (DM), a chronic disease where the blood has high sugar level, occurs when the pancreas does not produce insulin, or when the body cannot effectively use the insulin it produces (WHO).
- In recent years, government agencies and healthcare systems have increasingly focused on a 30-day readmission rates to determine the complexity of their patient populations and to improve quality.

Problem Definition

To identify patients who are at high risk of being diabetic and correspondingly to predict those who are most likely to get readmitted within 30days.

Data Exploration

- Data Source: The data was gotten from Kaggle, available in CSV format, is composed of electronic records spanning ten years (1999 through 2008) with various demographic and clinical variables per patient.
- Contains 10, 1766 rows and 50 columns
- Consist of numerical and categorical attributes
- Demographic information of the patient is stored as categorical variables, including gender and race as well as age, which was intervals measured in years (e.g., [0-10), [10-20), [20-30),up until (80-90].
- About 25 features associated to different medications given to diabetic patients, each one indicating if the drug, or a change in its dosage, was prescribed. The possible values for all these 24 columns are ‘NO’ (not prescribed), ‘Steady’ (no change in dosage), ‘Up’ (increased dosage) and ‘Down’ (decreased dosage). The class attribute indicates if the patient was readmitted after the visit and its possible values are ‘NO’ (i.e., no readmission), ‘<30’ (i.e. readmission occurred within 30 days) ‘>30’ (i.e., readmission occurred after 30 days).
- Some columns were removed because most of the values in them were unknown or missing or because they do not give insights on the readmission of the patients (namely encounter id, patient nbr, weight, admission type id, race etc).
- Categorical features were changed to numerical representation (replacing them with dummy variables), while the numerical features were kept intact.

VECTOR INSTITUTE

Model Description

<u>Supervised Learning</u>

- Logistic Regression
- K Nearest Neighbour,

These models were chosen since our aim was prediction of readmission of diabetic patients. The following were used to analyse the performance of the model using Classification Report (Accuracy, Recall, Precision, F1 support)

VECTOR INSTITUTE

Results and Findings

The first steps involve exploring the dataset and evaluating its features follows:

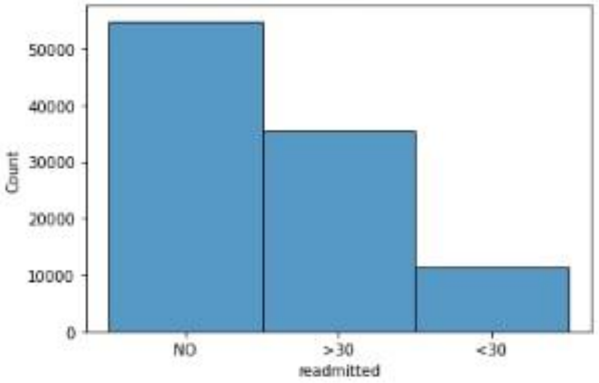
Feature name	Type	Description and values
Encounter ID	Numeric	Unique identifier of an encounter
Patient number	Numeric	Unique identifier of a patient
Race	Nominal	Values: Caucasian, Asian, African American, Hispanic, and other
Gender	Nominal	Values: male, female, and unknown/invalid
Age	Nominal	Grouped in 10-year intervals: [0, 10), [10, 20), . . ., [90, 100)
Weight	Numeric	Weight in pounds.
Admission type	Nominal	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available
Discharge disposition	Nominal	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available
Admission source	Nominal	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital
Time in hospital	Numeric	Integer number of days between admission and discharge
Payer code	Nominal	Integer identifier corresponding to 23 distinct values, for example, Blue Cross\Blue Shield, Medicare, and self-pay
Medical specialty	Nominal	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family\general practice, and surgeon
Number of lab procedures	Numeric	Number of lab tests performed during the encounter
Number of procedures	Numeric	Number of procedures (other than lab tests) performed during the encounter
Number of medications	Numeric	Number of distinct generic names administered during the encounter
Number of outpatient visits	Numeric	Number of outpatient visits of the patient in the year preceding the encounter
Number of emergency visits	Numeric	Number of emergency visits of the patient in the year preceding the encounter
Number of inpatient visits	Numeric	Number of inpatient visits of the patient in the year preceding the encounter
Diagnosis 1	Nominal	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values
Diagnosis 2	Nominal	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values
Diagnosis 3	Nominal	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values
Number of diagnoses	Numeric	Number of diagnoses entered to the system
Glucose serum test result	Nominal	Indicates the range of the result or if the test was not taken. Values: ">200," ">300," "normal," and "none" if not measured
A1c test result	Nominal	Indicates the range of the result or if the test was not taken. Values: ">8" if the result was greater than 8%, ">7" if the result was greater than 7% but less than 8%, "normal" if the result was less than 7%, and "none" if not measured.
Change of medications	Nominal	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: "change" and "no change"
Diabetes medications	Nominal	Indicates if there was any diabetic medication prescribed. Values: "yes" and "no"
24 features for medications	Nominal	For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: "up" if the dosage was increased during the encounter, "down" if the dosage was decreased, "steady" if the dosage did not change, and "no" if the drug was not prescribed
Readmitted	Nominal	Days to inpatient readmission. Values: "<30" if the patient was readmitted in less than 30 days, ">30" if the patient was readmitted in more than 30 days, and "No" for no record of readmission.

VECTOR INSTITUTE

Results and Findings

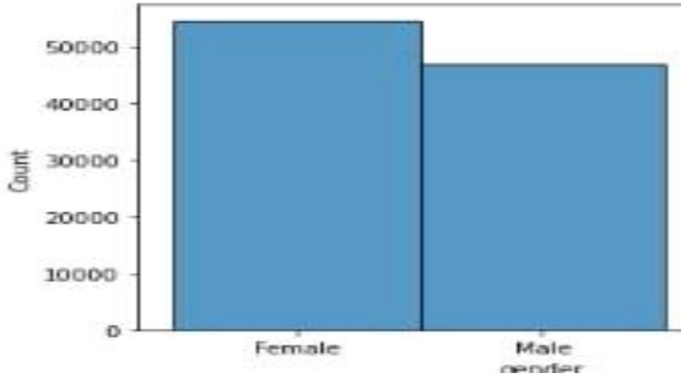
```
sns.histplot(x ="readmitted", data =df)
```

```
<AxesSubplot:xlabel='readmitted', ylabel='Count'>
```



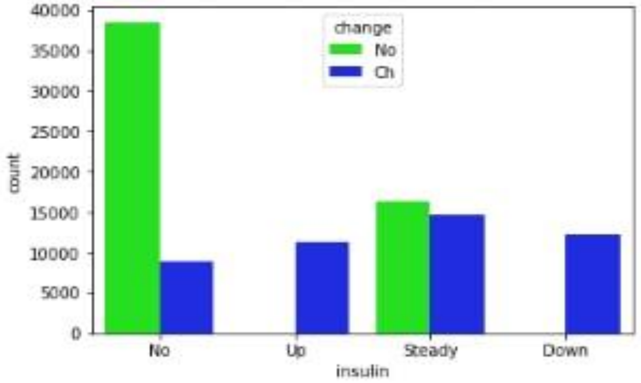

```
sns.histplot(x ="gender", data =df)
```

```
<AxesSubplot:xlabel='gender', ylabel=
```




```
sns.countplot(x= "insulin", hue ="change", data = df, color ="blue", palette ="hsv")
```

```
<AxesSubplot:xlabel='insulin', ylabel='count'>
```

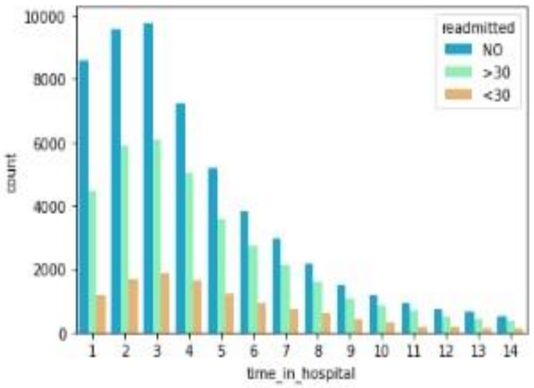


VECTOR
INSTITUTE

Results and Findings

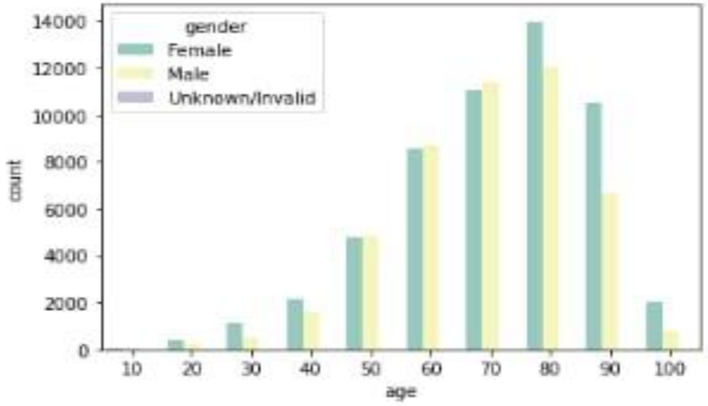
```
sns.countplot(x= "time_in_hospital", hue ="readmitted", data = df, color ="black", palette="rainbow")
```

```
<AxesSubplot:xlabel='time_in_hospital', ylabel='count'>
```




```
sns.countplot(x="age", hue="gender",
              data=df, palette="Set3", color ="blue")
```

```
<AxesSubplot:xlabel='age', ylabel='count'>
```

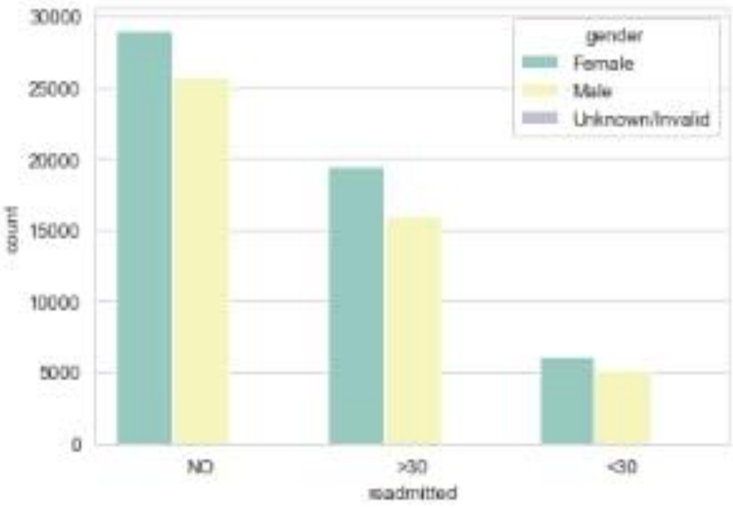


VECTOR INSTITUTE

Results and Findings

```
sns.set_style('whitegrid')
sns.countplot(x="readmitted", hue="gender",
                    data=df, palette="Set3", color ="blue")
```

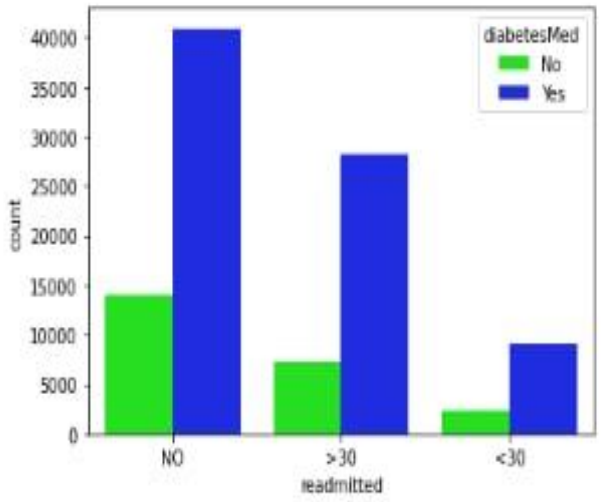
```
<AxesSubplot:xlabel='readmitted', ylabel='count'>
```



Results and Findings

```
sns.countplot(x= "readmitted", hue ="diabetesMed", data = df, palette ="hsv", color ="lightblue")
```

```
<AxesSubplot:xlabel='readmitted', ylabel='count'>
```



VECTOR INSTITUTE

Conclusions

Logistic Regression has been shown to be one of the most efficient algorithms in building prediction and models, combining KNN with Logistic Regression serves as a factor in improving the performance of the model. The data shows we had more people who were not readmitted, the persons that were not readmitted had the highest number of Diabetes Med that was not given, but not all of them were readmitted, this could be for many reasons, the diagnoses of diabetics is on the rise. We also had more female compared to male, hence more female were readmitted after 30 days than the male. The persons that were not readmitted, readmitted after >30days, and readmitted after <30days spent about 3-4hours in the hospital. The top 4 age range that had diabetes more were the 50-60, 60-70, 70-80, 80-90).

REFERENCES

1) B. Strack, J. P. DeShazo, C. Gennings, J. L. Olmo, S. Ventura, K. J. Cios, and J. N. Clore, “Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records,” BioMed Research International, vol. 2014, 2014.

2) International Diabetes Federation IDF Diabetes Atlas, 10th edn.Brussels, Belgium:2021. Available at Diabetes: https://idiabetesatlas.org

3) W. Boulding, S. W. Glickman, M. P. Manary, K. A. Schulman, and R. Staelin, “Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days,” The American Journal of Managed Care, vol. 17, no. 1, pp. 41–48, 2011.

4)R. N. Axon and M. V. Williams, “Hospital readmission as an accountability measure,” Jama, vol. 305, no. 5, pp. 504–505, 2011

5) M. T. Kassin, R. M. Owen, S. D. Perez, I. Leeds, J. C. Cox, K. Schnier,V. Sadiraj, and J. F. Sweeney, “Risk factors for 30-day hospital readmission among general surgery patients,” Journal of the American College of Surgeons, vol. 215, no. 3, pp. 322–330, 2012

6. Dungan KM. The effect of diabetes on hospital readmissions. J
Diab Sci Technol. 2012;6(5):1045–52.

7. Rubin, D. J. (2015). Hospital Readmission of Patients with Diabetes. Current Diabetes Reports, 15(4). doi:10.1007/s11892-015-0584-7