



# Readmission of Diabetes Patients

A review of the US hospital system from 1999 - 2008

**Center for Clinical and Translational Research** (Virginia, USA) provides a de-identified diabetes patient dataset from 1999-2008 across **130 US hospitals** and integrated delivery networks. It includes over **101k data points** across **50 features** representing patient and hospital outcomes. The information extracted from the database focuses on set criteria:

1. *It is an inpatient encounter (hospital admission)*
2. *It is a diabetic encounter, that is, one during which any kind of diabetes was entered to the system as a diagnosis*
3. *The length of stay was at least 1 day and at most 14 days*
4. *Laboratory tests were performed during the encounter*
5. *Medications were administered during the encounter*





# Objective

**Use machine learning to predict a diabetes patient's likeliness of readmission to hospital within the proceeding 30 days following discharge.**

This is based on a number of factors such as age, race, gender, time spent in hospital, lab procedures, medications etc.

Outcomes may support the hospital system in determining risk & the level of support required post discharge of a patient.

# Process

## Data cleaning & pre-processing

- Data is available in CSV format.
- Read via python script
- Dataframe created, N/As & invalid data accounted for
- Unnecessary columns dropped
- For some categories, binning or grouping of features required

## Exploratory analysis of key variables

Age, gender, race, admission type, medications, number of lab tests, discharge disposition, admission source, time in hospital, number of procedures, glucose serum testing, AC1 testing

Utilised **matplotlib**, **pandas**, **numpy**, **seaborn**

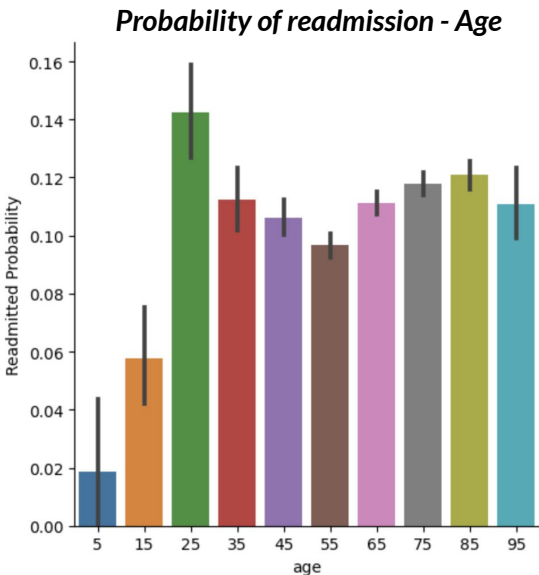
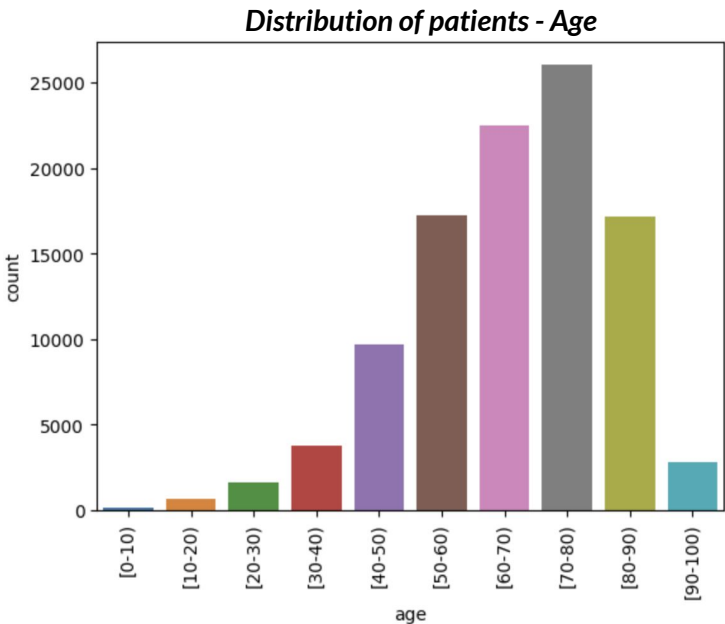
## Machine Learning

Finding the best model to predict diabetes pt readmission to hospital post discharge.

Utilised **sklearn** (Standard Scalar, Logistic Regression), **Random Forest**, **Imblearn** (oversampling)

# Exploratory Highlights - Age

We understand that we are dealing with an older population, however the highest risk of readmission sits with the 20-30 age group

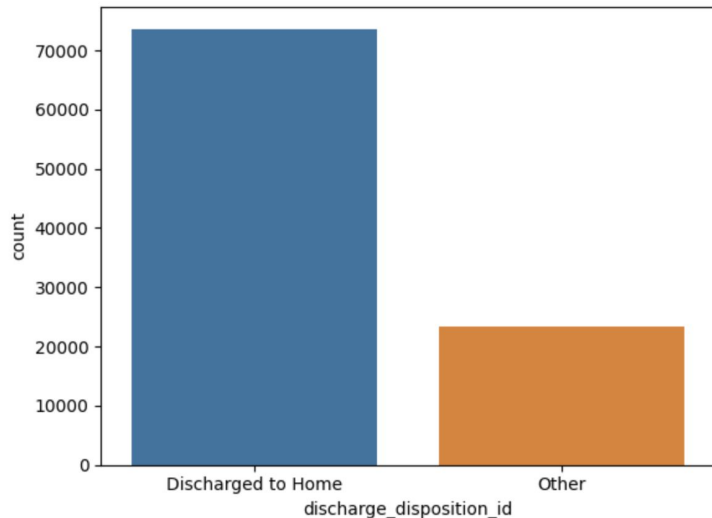


# Exploratory Highlights - Discharge location

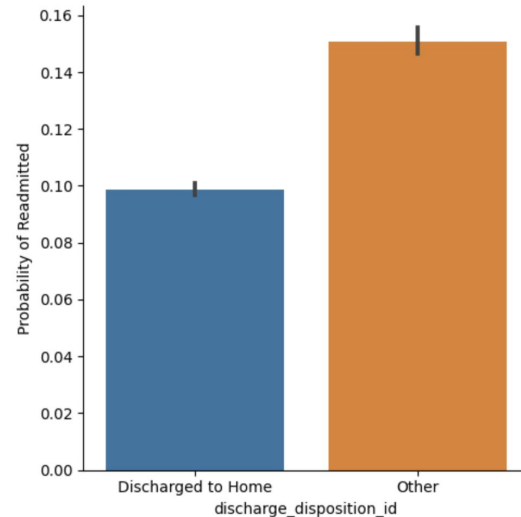


Those who were discharged to a place that was not their home (e.g. ongoing care home, nursing home) are recognised as a higher risk group for re-admission.

*Distribution of patients - Discharge*



*Probability of readmission - Discharge*



# Machine Learning

Using Random Forest, we are able to understand the importance of each feature.

1. Number of medications
2. Number of lab procedures
3. Time in hospital

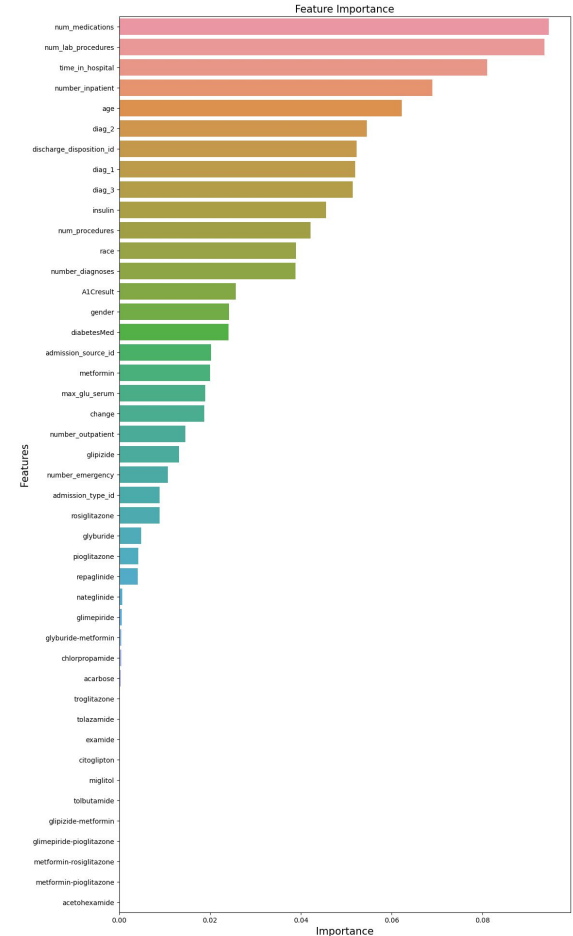
## Key Statistics

RF score (using oversampling): 0.9623

Accuracy score: 0.90

## Confusion Matrix

		Actual values	
		Positive	Negative
Predicted values	Positive	27	0
	Negative	3	0



# Conclusions



Having the ability to predict a patient's readmission rate is critical in managing hospital resources and patient wellbeing post discharge.



We understand that there is a clear link between the number of medications, number of lab procedures & time spent in hospital with readmission in less than 30 days. Other features such as age, discharge disposition & race will also play a role.



Ideal progression: Development of a dashboard for hospitals to input specific patient details and be advised on risk of pt readmission and steps to mitigate



## *Data Limitations*

Data is now dated (1999-2008) and represents a pre-covid hospital setting. Data is only representative of the USA and not other countries.

