Hospital Care for Diabetes in Relation to Patient Demographics

By: Viswa Kasireddy

Business Objective 1

Try and find out if a patient will need to be readmitted to the hospital after a diabetes related visit

Business Objective 2

Determine if a patient's demographic information significantly affects their chances of readmission

Data

Race - Caucasian, African American, Asian, Hispanic, Other

Gender - Female, Male, Unknown

Age - [0-10), [10-20), [20-30), [30-40), [40-50), [50-60), [60-70), [70-80), [80-90), [90-100)

Weight - pounds/lbs

Number of Lab Procedures - Number of lab tests performed during the encounter

Number of Procedures - Number of procedures (other than lab tests) performed during the encounter

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

Models

K Nearest Neighbors - Assigns classifications based on surrounding data points

Decision Trees - Makes decisions based on attributes (represented as nodes).

K Means Clustering - Assign initial k centroids, adds to the cluster based on distance of to the centroid, recalculate the centroid of the cluster

Random Forest - Ensemble model comprised of decision trees, results of which are aggregated together

Results

No replacement data records

Decision Tree - 48.9% accuracy

K Nearest Neighbors - 52.7% accuracy

K Means -

47371 vs 54864 no readmission

30862 vs 35545 readmissions after 30 days

23533 vs 11357 readmissions within 30 days

Random Forest - 51.2%

With replacement data records

Random Forest - 50.3%

No replacement gender split

Marginal difference

50.2% male vs 49.7% female

Conclusions

Given the data, the conclusion can be made that the provided features do not strongly relate to patient readmission.

Why?

One possible reason - Data is not drawn from one particular hospital.

Future Use

For future work on the project, Ichoose a different dataset. Here there was no clear relation to readmission. But there are other outcomes that could have more of a relation, such as proper diagnosis of diabetes, that other datasets would be centered around.