PREDICTING HOSPITAL READMISSION FOR DIABETIC PATIENTS

Module 4

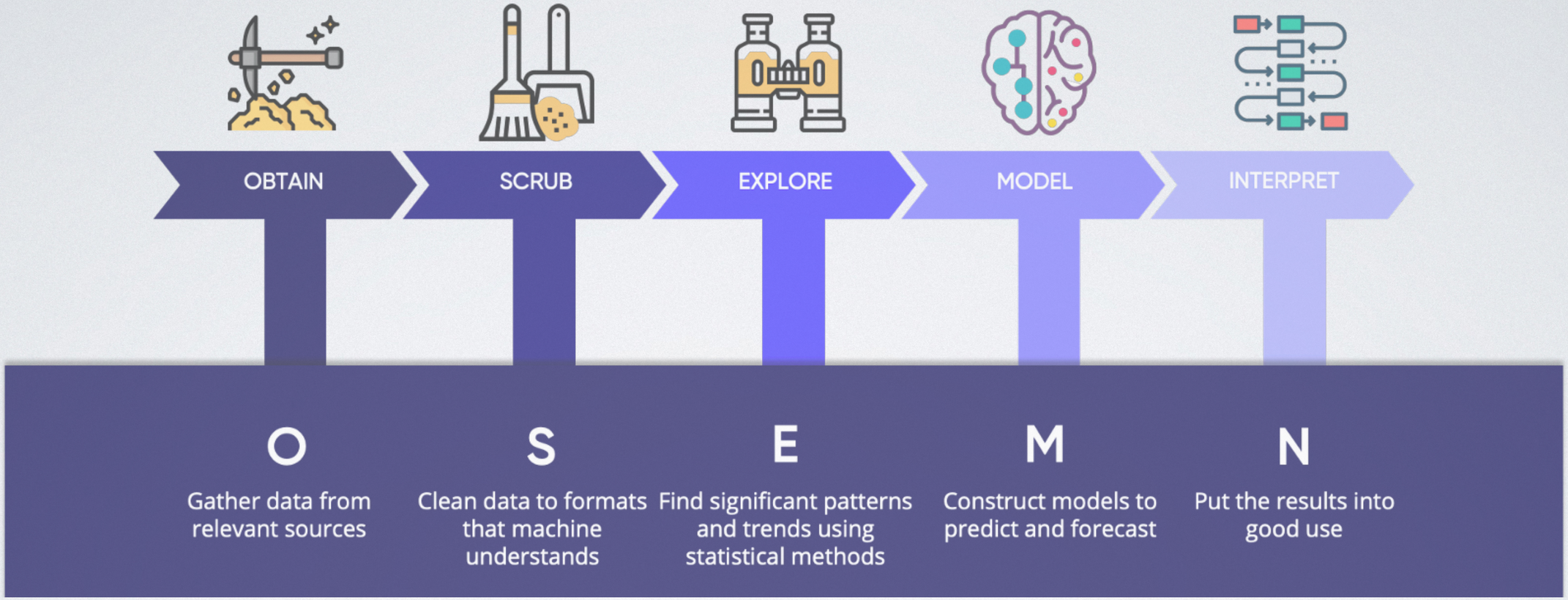
ds-ft-041519

Paul Woody

PREDICTING HOSPITAL READMISSION IN DIABETIC PATIENTS

Background | **EDA** | **Models** | **Conclusion**

METHODOLOGY



Perform exploratory data analysis to identify meaningful trends.
Use models and hypothesis tests, when appropriate, to evaluate significance.

BACKGROUND

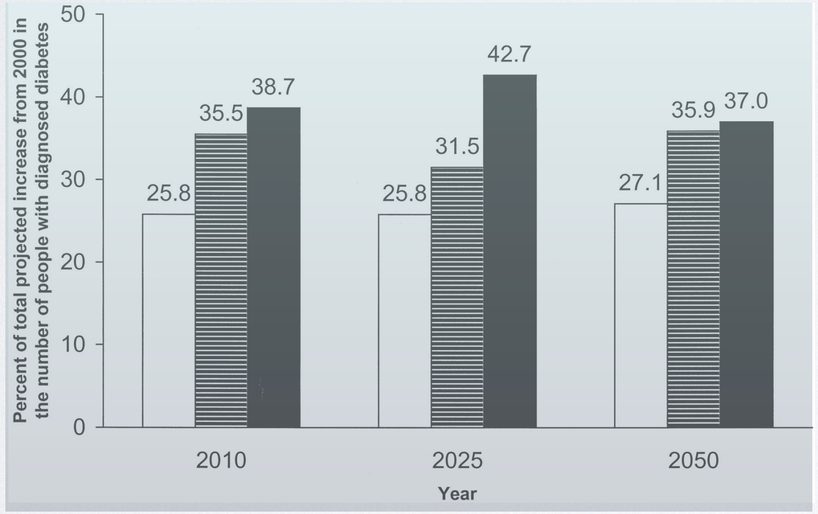


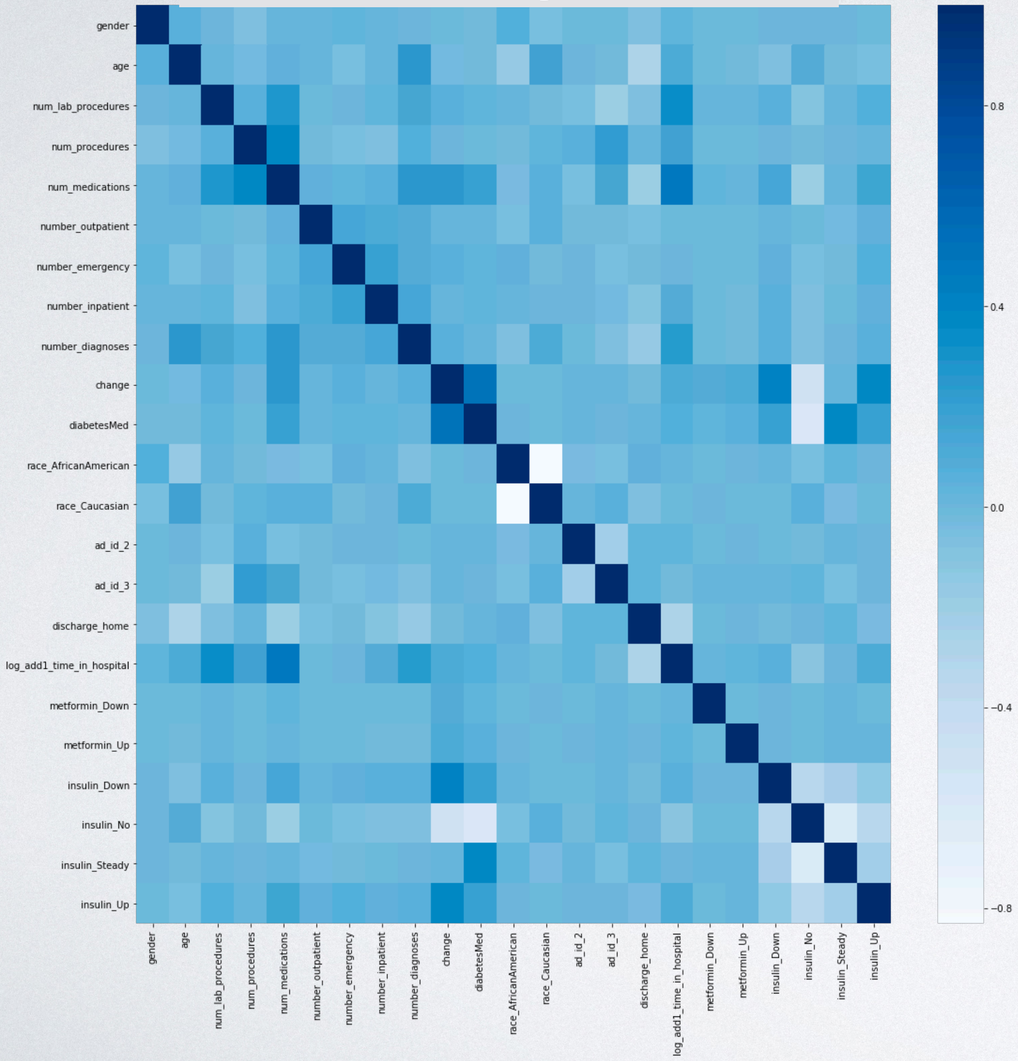
Figure: Percentage breakdowns of the projected increase in the total number of people with diagnosed diabetes due to population growth.

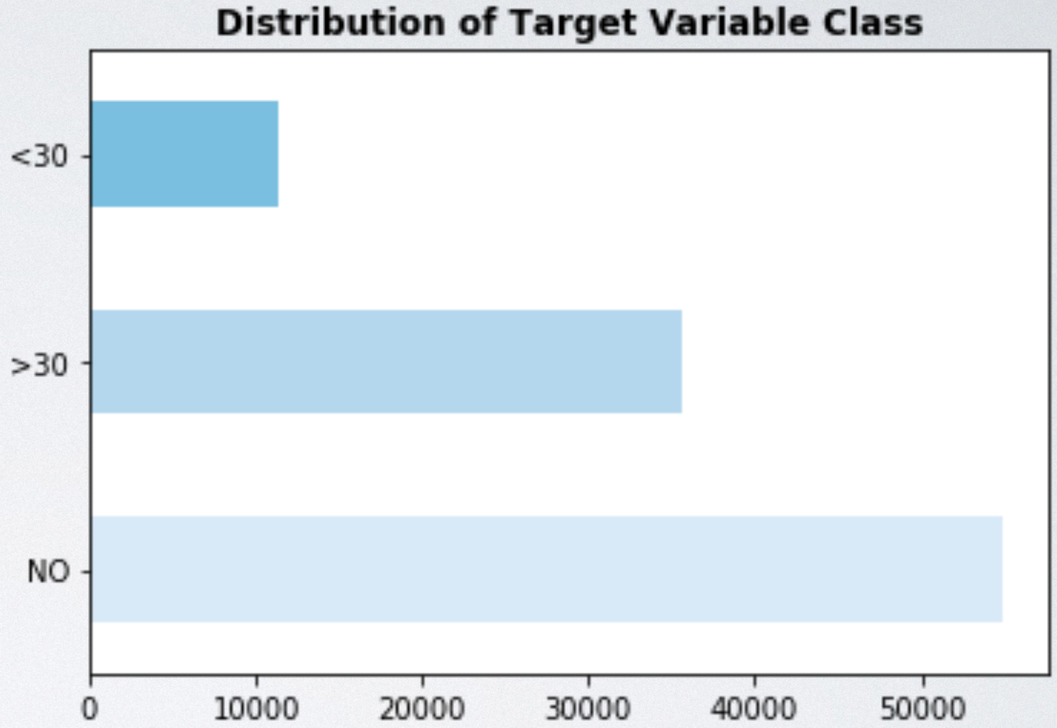
According to a study published by the American Diabetes Association in 2001, the number of Americans diagnosed with diabetes is projected to increase by 165% by 2050.

In this investigation, a dataset containing ~100,000 patient encounters was used in order to better understand, identify, and (ultimately) improve efficacy of medical treatment plans for diabetic patients.

EXPLORATORY DATA ANALYSIS

Multicollinearity Evaluation

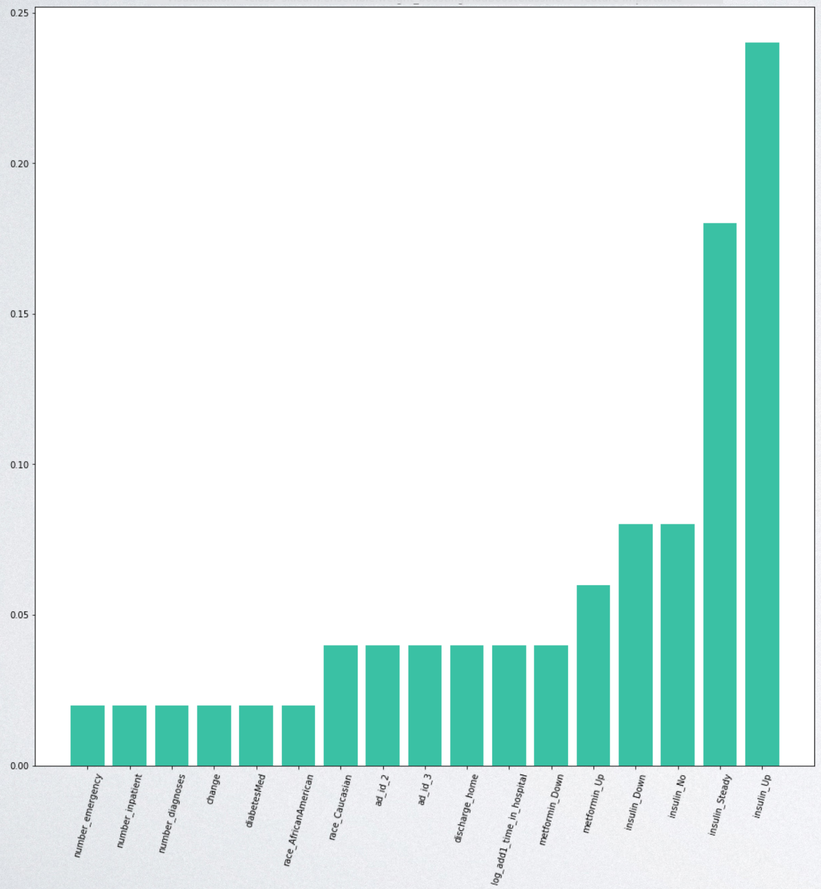




Note Classes represent readmission time in days.*

MODELING PT 1:
DECISION TREE, RANDOM FOREST CLASSIFIERS

Feature Importance:



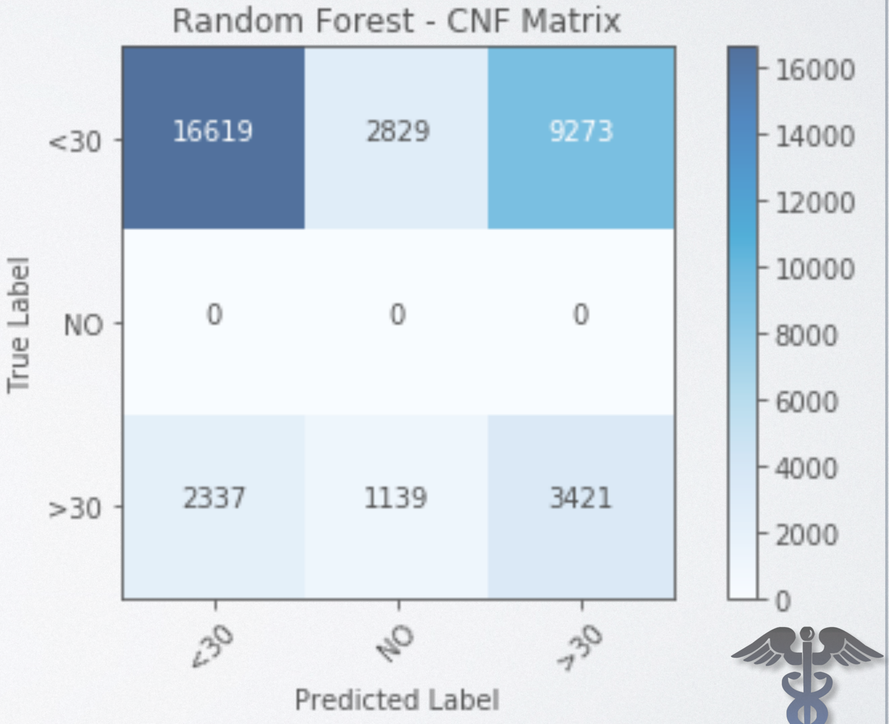
As seen in this figure, the top 4 predictors (by weight) are all related to changes in insulin dosage.

Best Performing Model:

Decision Tree pipeline test accuracy: 0.506
Random Forest pipeline test accuracy: 0.509

As seen by the predictive accuracy in the confusion matrix, the model did not predict the minority class.

Poor recall ability precludes this model use in healthcare, as it completely disregards an important target class.



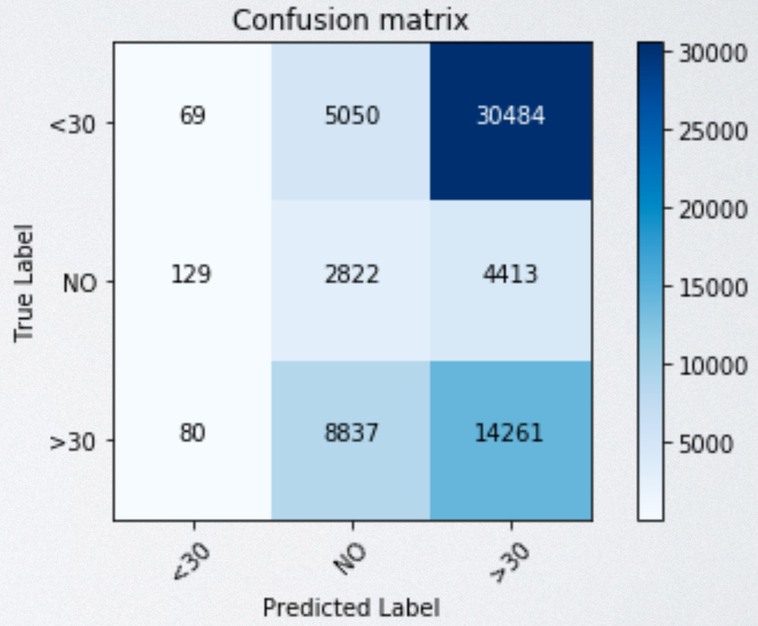
MODELING PT 2: DEEP LEARNING

To begin, a simple Deep Learning model with no hyper parameter adjustment was fit and evaluated, in order to determine whether the minority class was being predicted.

```
model_base = models.Sequential()
model_base.add(layers.Dense(64, input_shape=(27,), activation='relu'))
model_base.add(layers.Dense(32, activation='relu'))
model_base.add(layers.Dense(4, activation='relu'))
model_base.add(layers.Dense(3, activation='softmax'))

model_base.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

history = model_base.fit(X_train, y_train, epochs=128, batch_size=256)
```



As seen by the confusion matrix, the initial deep learning model did predict the minority class, albeit poorly. Next, SMOTE oversampling was performed in order to artificially increase population size of the minority class in the dataset.

To improve results, this analysis should be repeated as the sample population of the dataset increases.

MODELING PT 3: OVERSAMPLING USING SMOTE

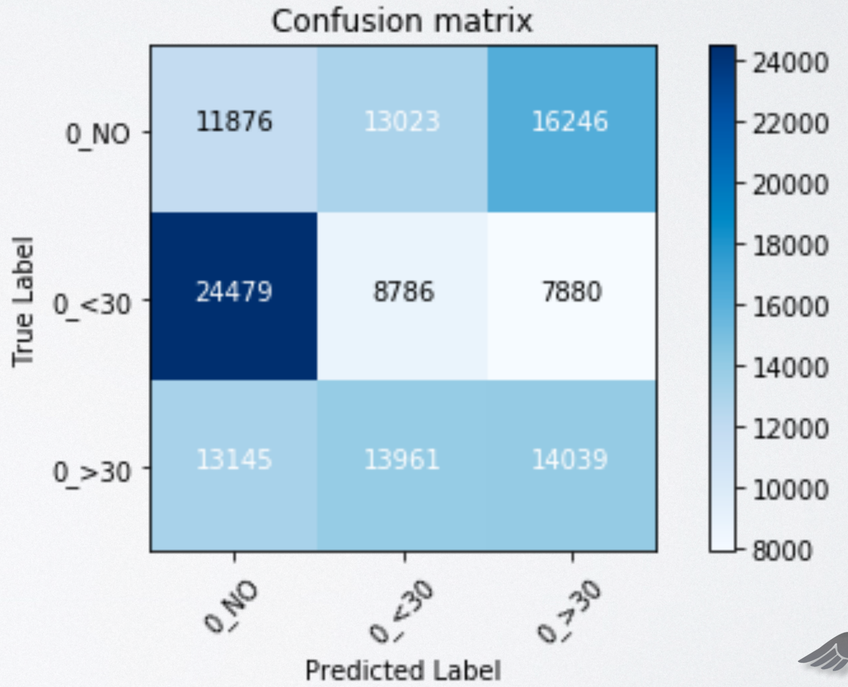
Feature Importance:

Weight	Feature
0.0558 ± 0.0022	insulin_No
0.0536 ± 0.0026	number_inpatient
0.0478 ± 0.0008	insulin_Steady
0.0377 ± 0.0010	num_procedures
0.0365 ± 0.0011	ad_id_1
0.0346 ± 0.0019	discharge_home
0.0304 ± 0.0016	num_medications
0.0299 ± 0.0009	metformin_Steady
0.0289 ± 0.0015	log_add1_time_in_hospital
0.0260 ± 0.0017	number_diagnoses
0.0243 ± 0.0008	race_Caucasian
0.0239 ± 0.0009	metformin_No
0.0231 ± 0.0003	age
0.0226 ± 0.0008	ad_id_3
0.0219 ± 0.0006	num_lab_procedures
0.0218 ± 0.0011	insulin_Up
0.0212 ± 0.0007	insulin_Down

Best Performing Model:

Training accuracy: 51.2%
Testing accuracy: 46.2%

Despite a reduction in accuracy (especially in test performance), the results of of modeling after performing SMOTE improve on the predictive ability of the Random Forest Classifier et al used previously in this investigation.



INSIGHTS

While not achieving the predictive ability hoped prior to the investigation, there are several useful conclusions we can draw.

In all of the models fitted in this investigation, changes in insulin dosage was among the most significant features in determining whether patient was readmitted.

Because insulin can be expensive, as much as $700 per month for some patients, according to a WebMD report published in 2019, patients that cannot afford the medication may fail to fill prescriptions given by doctors because they can't afford them.

In a future study, income and self-reported insulin dosing regiment should be included in order to more accurately evaluate diabetic patient prognosis.

THANK YOU