Predicting Hospital Readmissions

PRESENTED BY: ALLY CHRYSTAL

PRESENTED TO: ABLETO ANALYTICS TEAM

Goal

To build a model to predict which encounters will be followed by the patient being readmitted to the hospital within 30 days for diabetic patients

- How accurately can we predict hospital readmission?
- Which features during a hospital visit are the main predictors of readmission?

Why is this important?

- For the patient: readmission rates indicate the quality of patient care
- For insurance providers and hospitals: avoidable readmissions are costly ($ millions/ year)
- For the doctors: Hospitals and doctors can be penalized for high readmission rates

Data Used

- Claims from 130 US Hospitals (1999 -2008)
- 101,766 hospitalizations for diabetic patients
- 50 Features including patient demographics, procedures, tests and medications

Data Cleaning & Feature Engineering

Null Values

- Columns Dropped:
  (~ half or more null values is bad data)
  - Weight
  - Medical Specialty
  - Payer Code
- Nulls replaced with own category (low # of nulls that will not significantly affect model)
  - Race
  - Diag_1
  - Diag_2
  - Diag_3

Categorical Columns Converted to Binary or Numerical

- Medication columns
  - 0=No
  - 1=Steady/Up/Down
- Diagnoses:
  - Converted to categories 0-8
- Max_glu_serum
- A1Cresult
- diabetesMed
- Change
- Gender
- Readmitted
- Age

Features Created

- Total_Meds
  - Total unique medications patient is prescribed (sum of medication columns after binary conversion)
- Total_Visits
  - Sum of total encounters in the preceding year
  - Sum of Outpatient, emergency, inpatient visits

Eliminated Columns

- Columns dominated by a single value (>=95%)
  - Many medications
  - Weight (nulls)
- Eliminated due to noise:
  - Diag_2
  - Diag_3

Admission, Discharge & Admission Source IDs

- Converted from numerical to categorical
- Many ID codes could be consolidated

Target: Readmission

Goal is to predict readmission less than 30 days:

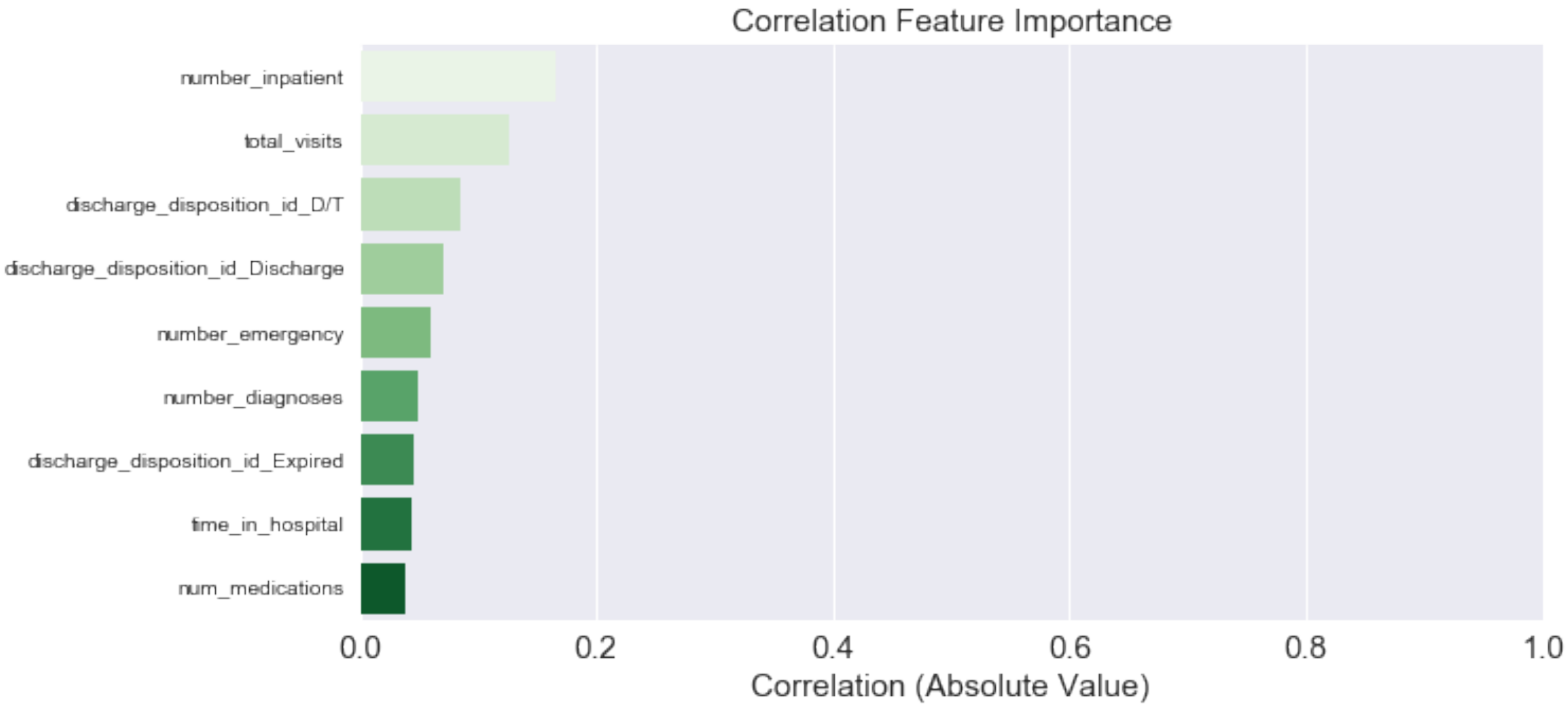
- 1 = < 30   (11% of data)
- 0 = NO and > 30 (89% of data)

Imbalanced target can lead to overfitting

Baseline accuracy = 89%

- Meaning if we predict 0 (negative) for every case we will be 89% accurate – highly accurate, but does that teach us anything?

Target: Readmission



Top 9 Features Correlated to Readmission

- All features are under 20% (abs value) correlated to our target
- Weak correlations could potentially mean weak predictors of readmission

A weak model can be made stronger by ensemble modelling

Modelling

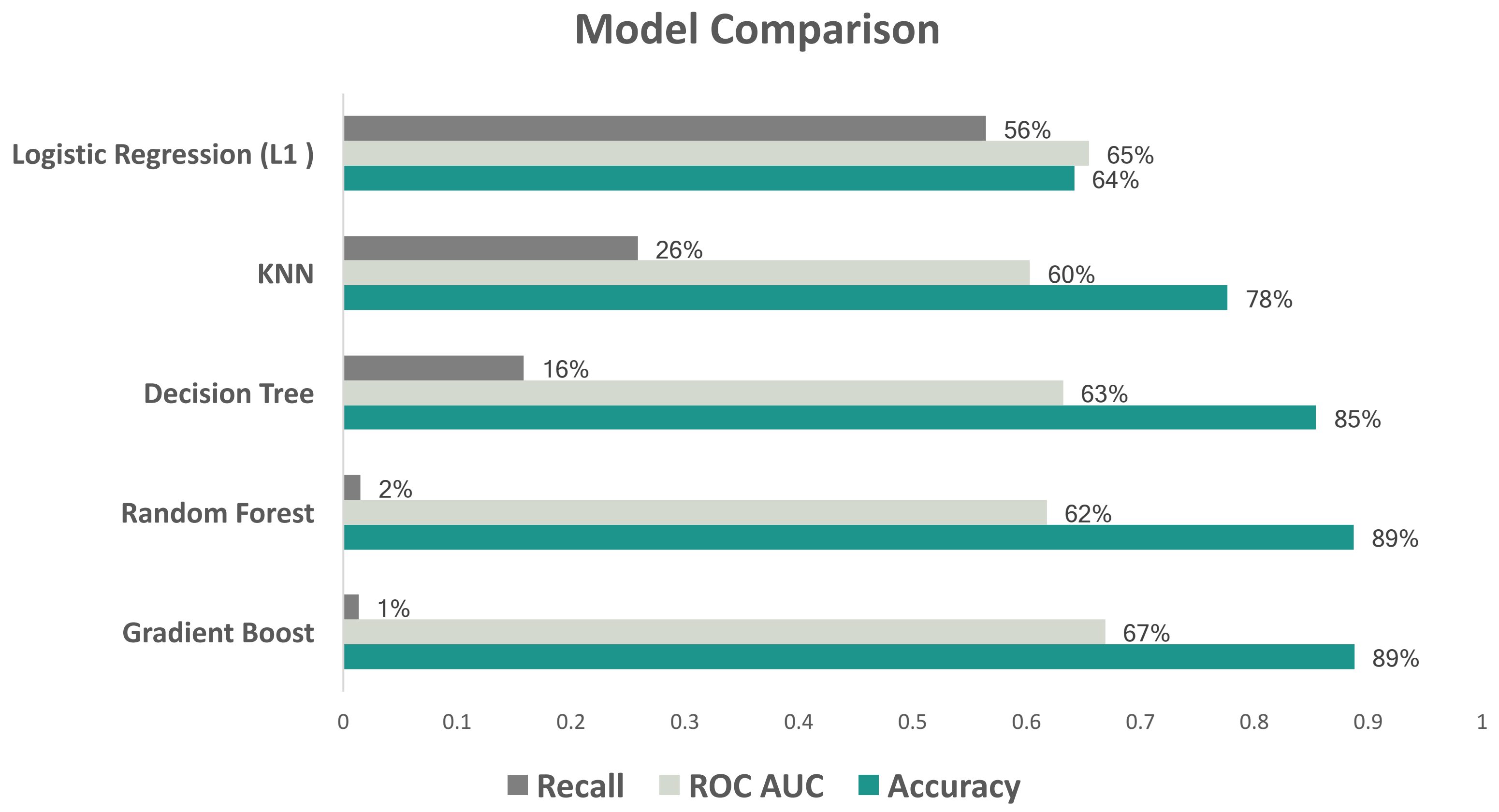
Preprocessing

- SMOTE (Synthetic Minority Oversampling Technique) to balance classes
- Standard Scaler to standardize the variable ranges to reduce bias
- Train 80% / Test 20% Split

Model Selection

- Tested on 5 classification models
- Accuracy is good, but we're most considered with catching positive cases and learning what predictors leads to readmission in order to corrective measures
- Based model selection off of:
  - **Accuracy**: What percentage of our predictions are correct?
  - **Recall**: What percentage of positive cases (readmission within 30 days) did we catch?
    - Important for hospitals as ability to predict positive readmissions is costly (quality and financially)
  - **AUC-ROC**: The area under the ROC (Receiver Operating Characteristic) curve representing a relation between recall and specificity
    - Specificity or True Negative Rate: What % of negative cases were correct?

Model Performance Comparison



Decision Tree Classification

Hyper-parameters:

- max_depth = 5
- max_features = None
- min_samples_split = 3

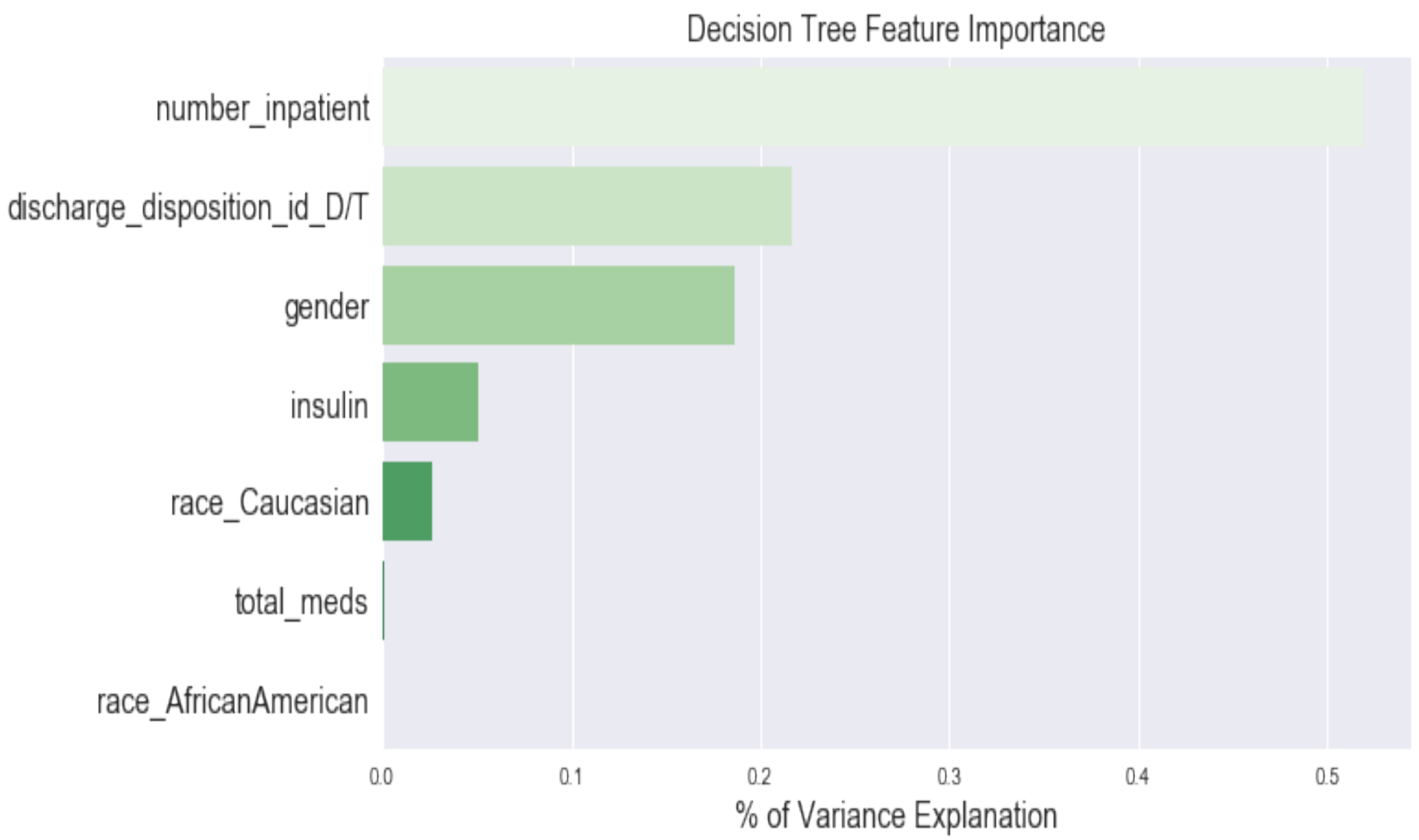
Performance:

- 89% accuracy – equal to baseline

Decision Tree Confusion Matrix

	Predicted No	Predicted Yes
Actual No	17,023	1,041
Actual Yes	1,927	363

No = 0 = Not Readmitted   Yes = 1 = Readmitted



Next Steps

Lower threshold to improve recall

- Default: If probability of readmission is 50% or higher, classified as readmission
- Example: Decrease threshold to 30% or higher, classified as readmission
- Will decrease accuracy but increase recall and Type II errors

More feature engineering

Add more features