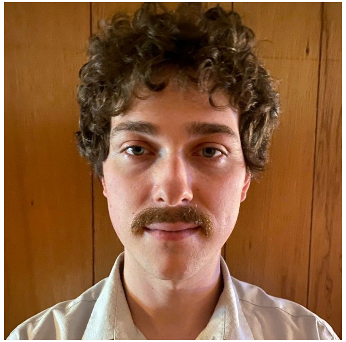
Analysis of Hospitalized Diabetes Patients between 1999-2008

Keith-Jordan Wilkinson, Eddy Doering, Cory Chitwood, Kai Gui

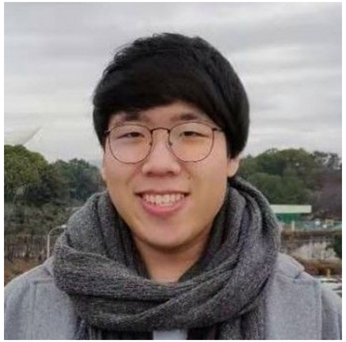
Introduction

Keith-Jordan Wilkinson



Eddy Doering

Cory Chitwood



Kai Gui

linkedin.com/in/keith-jordan-wilkinson
linkedin.com/in/efdoering
linkedin.com/in/cchitwood
linkedin.com/in/kaigui

Tools:

IDEs

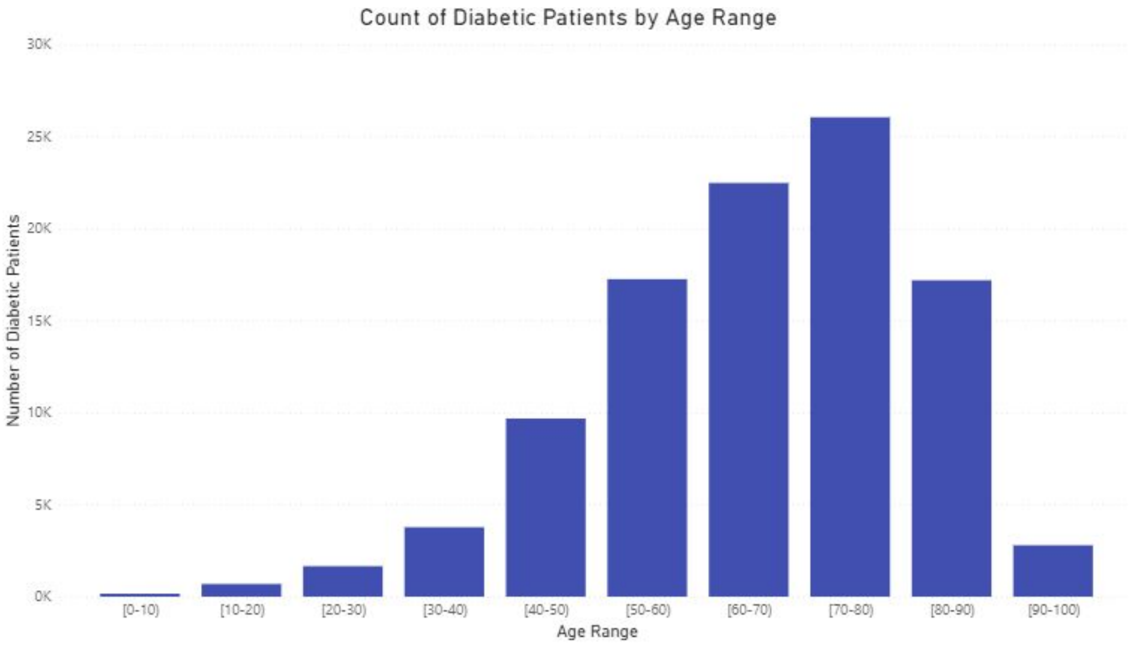
Introduction

- Diabetes affects 10% of the U.S. population, 8th leading cause of death.
- Hospital readmittance is associated with poor patient outcomes.
- Dataset: Diabetes 130-US hospitals for years 1999-2008 (UCI)
    - Can ML predict if hospitalized diabetes patients will be readmitted?
- This information may improve patient outcomes and diabetes care

Questions

1. What is the most common age range for diabetes patients?
2. What are common diagnoses with diabetes patients?
3. How does time spent hospitalized change with patients age?
4. Are there any racial disparities in hospital visits for diabetes patients vs. national demographics?
5. What features can predict readmittance rates?

Visuals (1/4)

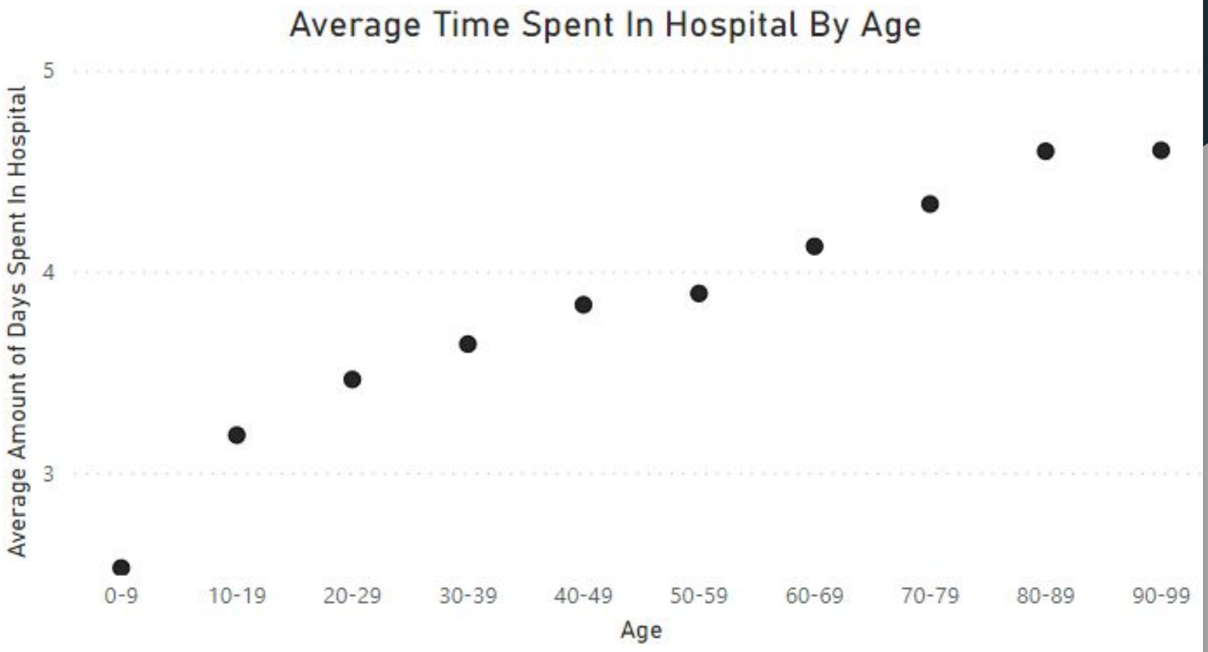


- Patient count increases from ages 40 to 80
- Patient count drops steeply at ages 80-100
- Between ages 0-30, there are very few hospitalized diabetes patients

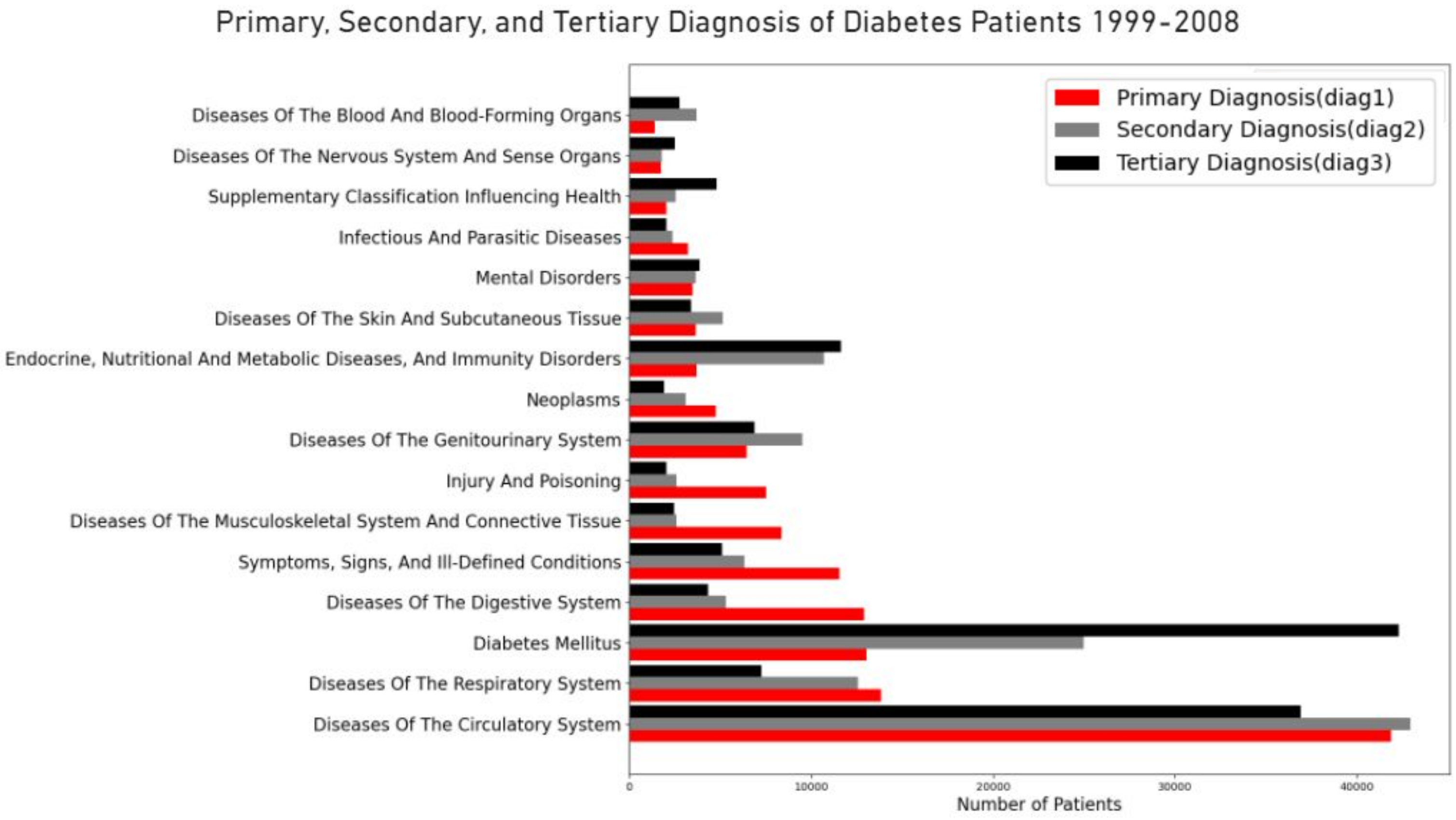
Visuals (2/4)

- Average hospitalization duration increases with age

- Complications associated with diabetes and surgical procedures may increase with age

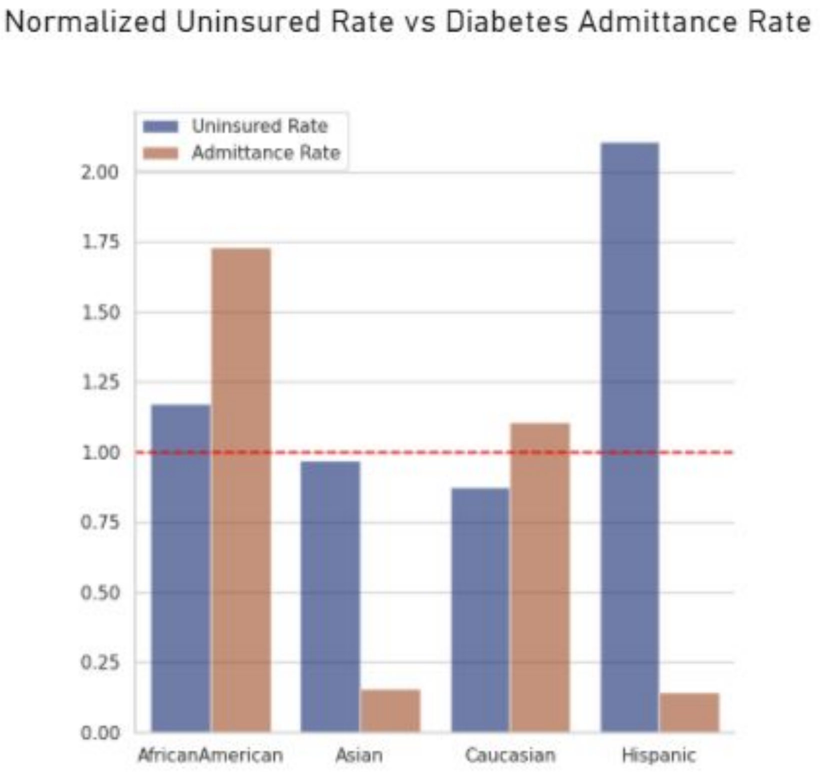


Visuals (3/4)



Primary, Secondary, and Tertiary Diagnosis of Diabetes Patients 1999-2008
Primary Diagnosis(diag1)
Secondary Diagnosis(diag2)
Tertiary Diagnosis(diag3)
Diseases Of The Blood And Blood-Forming Organs
Diseases Of The Nervous System And Sense Organs
Supplementary Classification Influencing Health
Infectious And Parasitic Diseases
Mental Disorders
Diseases Of The Skin And Subcutaneous Tissue
Endocrine, Nutritional And Metabolic Diseases, And Immunity Disorders
Neoplasms
Diseases Of The Genitourinary System
Injury And Poisoning
Diseases Of The Musculoskeletal System And Connective Tissue
Symptoms, Signs, And Ill-Defined Conditions
Diseases Of The Digestive System
Diabetes Mellitus
Diseases Of The Respiratory System
Diseases Of The Circulatory System
0
10000
20000
30000
40000
Number of Patients

Visuals (4/4)



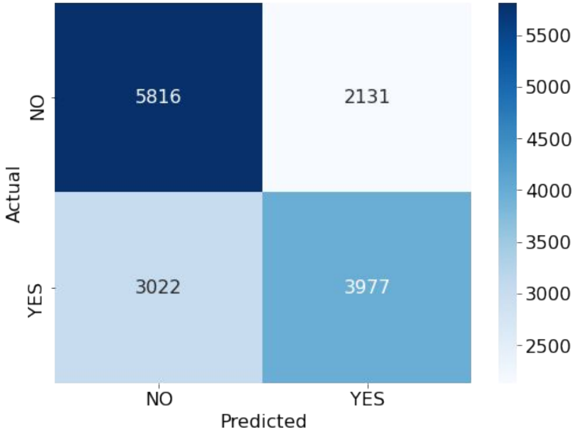
- Does higher uninsured percentage correlate with lower hospitalization rate?
- Hipanics and Caucasians follow hypothesis, African Americans and Asians do not.
- Other possible factors?
  - Language/cultural barriers?

Predicting Readmittance with Machine Learning

- Goal: Predict patient readmittance based on features of the initial stay
- Four different machine learning classification models were employed:
  - Random Forests
  - K-Nearest Neighbors
  - Ridge
  - SGD Classifier.
- Hyperparameters were tuned. Model performance was compared.

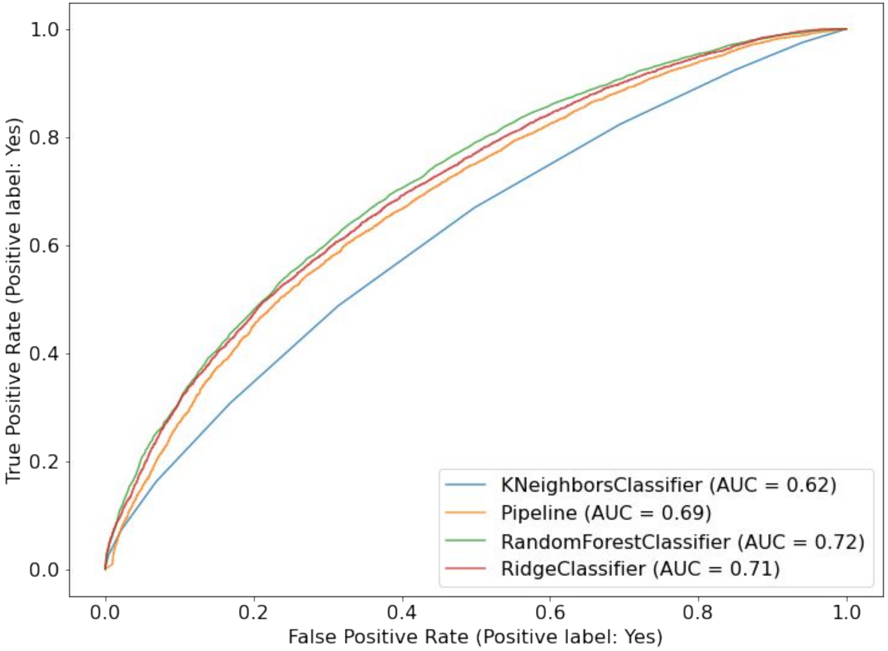
Summary of ML Models

Model	Base Accuracy	Optimized Accuracy	Improvement	Optimized AUC	Approx. Runtime (min.)
Random Forest Classifier	0.58421	0.65523	12.156%	0.72	5
Ridge Classifier	0.57886	0.65068	12.406%	0.71	2
SGD Classifier	0.54353	0.63233	16.339%	0.69	15
KNN Classifier	0.49331	0.63648	29.023%	0.62	25
Ensemble Voting Classifier	N/A	0.65175	N/A	N/A	60



- ML Classifiers Optimized: Random Forest, Ridge, SGD, K-Nearest Neighbors
- Random Forest and Ridge were the strongest performers in terms of accuracy, runtime, and AUC.

ROC Plot and Top Features



Feature	Importance
num_lab_procedures	0.0514
num_medications	0.0486
number_inpatient	0.0399
time_in_hospital	0.0365
number_diagnoses	0.0281
num_procedures	0.0248
number_outpatient	0.0173
number_emergency	0.0163
gender_Male	0.0137
insulin_Yes	0.0102

Conclusion

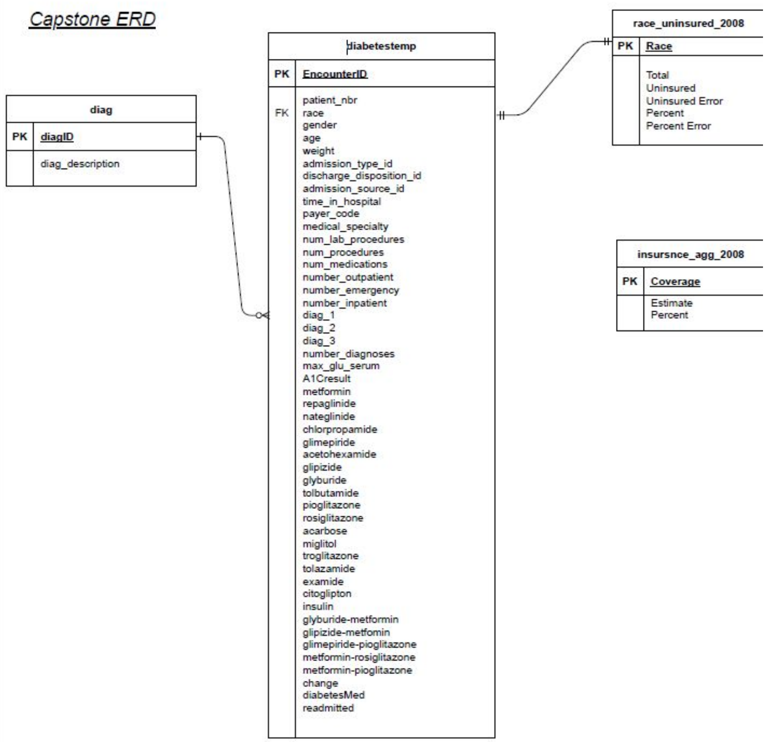
- Demographics of hospitalized diabetes patients were explored.
- Optimized ML Models predict hospital readmission better than chance.
- ML Model has room to improve. For future work, we recommend:
    - Develop models specific for select diagnoses/medical specialties.
    - Explore original database for additional features.

https://github.com/crchitwood/DiabetesCapstone

Sources

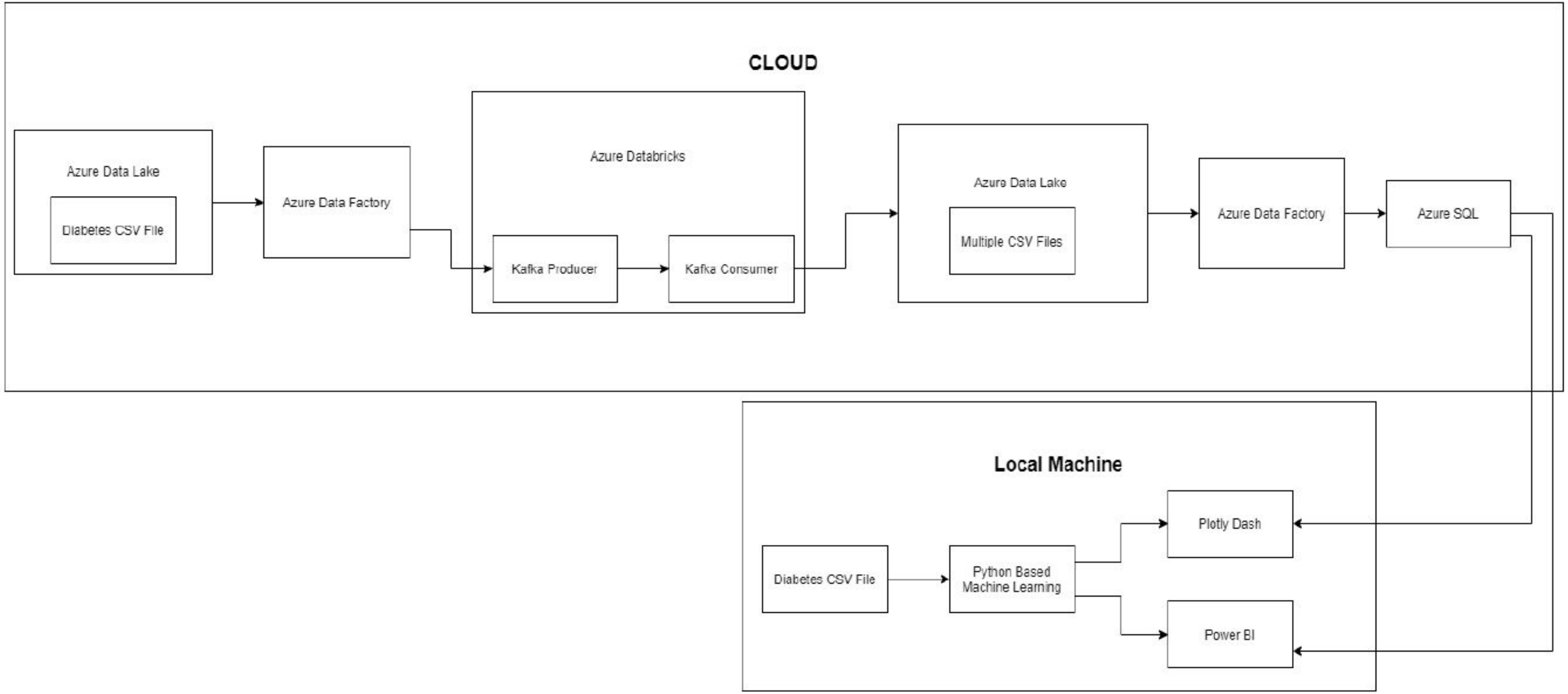
1. Ahmad FB, Anderson RN. The Leading Causes of Death in the US for 2020. JAMA. 2021;325(18):1829–1830. DOI: 10.1001/jama.2021.5469
2. Centers for Disease Control and Prevention. National Diabetes Statistics Report, 2020. Atlanta, GA: Centers for Disease Control and Prevention, U.S. Dept of Health and Human Services; 2020. Link
3. McIlvennan CK, Eapen ZJ , Allen LA. Hospital Readmissions Reduction Program. Circulation. 2015;131:1796–1803. DOI: 10.1161/CIRCULATIONAHA.114.010270
4. Strack B, DeShazo JP, Gennings C, Olmo JL, Ventura S, Cios KJ, Clore JN. Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records. Biomed Res Int. 2014:781670. DOI: 10.1155/2014/781670
5. U.S. Census Bureau. HIC-9_ACS. Population Without Health Insurance Coverage by Race and Hispanic Origin: 2008 to 2019. Health Insurance Historical Tables - HHI Series. Link
6. Cherney, K. (2018, July 6). Age of onset for type 2 diabetes: Risk factors and more. Healthline. Retrieved November 8, 2021, from https://www.healthline.com/health/type-2-diabetes-age-of-onset.

Questions?


Capstone ERD
diag
PK diagID
diag_description
diabetestemp
PK EncounterID
FK
patient_nbr
race
gender
age
weight
admission_type_id
discharge_disposition_id
admission_source_id
time_in_hospital
payer_code
medical_specialty
num_lab_procedures
num_procedures
num_medications
number_outpatient
number_emergency
number_inpatient
diag_1
diag_2
diag_3
number_diagnoses
max_glu_serum
A1Cresult
metformin
repaglinide
nateglinide
chlorpropamide
glimepiride
acetohexamide
glipizide
glyburide
tolbutamide
pioglitazone
rosiglitazone
acarbose
miglitol
troglitazone
tolazamide
examide
citoglipton
insulin
glyburide-metformin
glipizide-metfomin
glimepiride-pioglitazone
metformin-rosiglitazone
metformin-pioglitazone
change
diabetesMed
readmitted
race_uninsured_2008
PK Race
Total
Uninsured
Uninsured Error
Percent
Percent Error
insursnce_agg_2008
PK Coverage
Estimate
Percent

Data Platform

Outline of Data Platform



Model	Refactor "Readmitted	Encode MedList	Refactor MedList (Y/N)	Tuned	Drop Medical Specialty	Drop Emergency	Score	Percent Improvement
Ridge Classifier	N	N	N	N	Y	N	0.578860612	0
Ridge Classifier	N	N	N	Y	Y	N	0.583805733	0.854285317
Ridge Classifier	Y	N	N	N	Y	N	0.639636023	10.49914433
Ridge Classifier	Y	N	N	Y	Y	N	0.649939783	12.27915147
Ridge Classifier	Y	Y	N	N	Y	N	0.640773451	10.69563862
Ridge Classifier	**Y**	**Y**	**N**	**Y**	**Y**	**N**	**0.650675766**	**12.40629483**
Ridge Classifier	Y	Y	N	N	N	N	0.64893617	12.10577415
Ridge Classifier	Y	Y	N	Y	N	Y	0.647326507	11.82769981
Ridge Classifier	Y	N	Y	Y	Y	N	0.649404523	12.18668356
Random Forest Classifier	N	N	N	N	Y	N	0.584207775	0
Random Forest Classifier	Y	Y	N	Y	N	Y	0.645278726	10.45363531
Random Forest Classifier	Y	Y	N	Y	N	N	0.650006691	11.26293042
Random Forest Classifier	Y	Y	N	Y	Y	N	0.648735448	11.04532922
Random Forest Classifier	**Y**	**N**	**Y**	**Y**	**Y**	**N**	**0.655225478**	**12.15624062**
SGD Classifier	N	N	N	N	N	N	0.543526258	0
SGD Classifier	N	N	N	N	Y	N	0.546455996	0.539024331
SGD Classifier	N	N	N	N	Y	Y	0.550074378	1.204747712
SGD Classifier	Y	N	N	N	Y	Y	0.602219274	10.79856132
SGD Classifier	Y	N	Y	N	Y	Y	0.607928276	11.84892488
SGD Classifier	Y	N	Y	Y	Y	Y	0.632332248	16.33885923

Model	Refactor "Readmitted"	Encode MedList	Refactor MedList (Y/N)	Tuned	Drop Medical Specialty	Drop "Emergency/Trama" Medical Specalty	Drop "?" Medical Specalty	Drop BUT "Internal Medicine" Medical Specalty	Score	Percent Improvement
KNN Classifier	n	n	n	n	n	n	n	n	0.493309247	0
KNN Classifier	n	n	y	n	n	n	n	n	0.49484812	0.311949003
KNN Classifier	n	y	n	n	n	n	n	n	0.492573264	-0.149193001
KNN Classifier	y	y	n	n	n	n	n	n	0.568714037	15.28550115
KNN Classifier	y	y	n	y	n	y	n	n	0.59	19.60043402
KNN Classifier	y	y	n	y	y	n	n	n	0.586243811	18.83900719
KNN Classifier	y	y	n	y	n	y	y	y	0.636484687	29.0234658
KNN Classifier	y	y	n	y	n	y	y	n	0.614925373	24.65312121