



Predicting ER Readmissions for patients with diabetes

[Fatima Loumaini](#)- Metis Summer 2019



Project at a glance



More than 30 Million americans have diabetes

Hospitals with high ER readmission rates pay penalties to Medicare & Medicaid programs



ER Readmissions within 30 days of discharge cost \$41B in 2017(comparable to \$68B for the US Education budget)

10% of diabetes patients are readmitted to the ER within 30 days for an equivalent cost of \$452M annually.

Dataset



- Data from [UCI](#) Irvine Machine Learning Repository.
- Dataset represents 10 years of clinical care at 130 US hospitals between 1999-2008. It includes 50 features representing patient characteristics, conditions, tests and medications

Methodology



Pre processing

- Data Cleaning
- Medication and Diagnoses feature removed
- Feature Encoding
- Interaction Terms
- Scaling features

Feature engineering

- Feature creation to understand medication dosage change
- Messy features, for ex: discharge_disposition_id had 34 categories.
- Balanced data by SMOTE
- Regularization

Modelling

Dataset shape: (97874, 57)

- Logistic Regression
- Random Forest

Models



- **Logistic Regression: Train/Test score: 91%, 91%**

	precision	recall	f1-score	support
0	0.85	1.00	0.92	43385
1	1.00	0.83	0.91	43320
accuracy			0.91	86705
macro avg	0.92	0.91	0.91	86705
weighted avg	0.92	0.91	0.91	86705

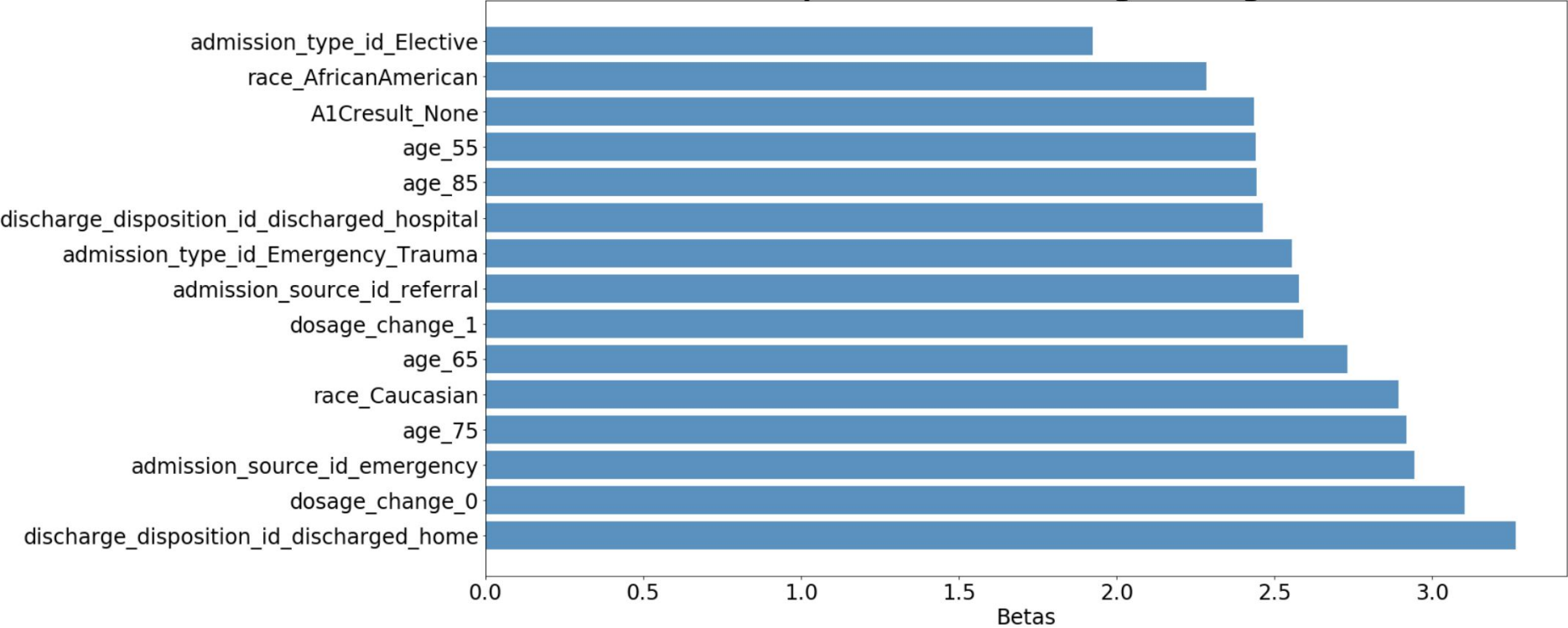
- **Random Forest: Train/Test Scores: 96%, 92%**

	precision	recall	f1-score	support
0	0.87	0.99	0.93	43385
1	0.99	0.85	0.91	43320
accuracy			0.92	86705
macro avg	0.93	0.92	0.92	86705
weighted avg	0.93	0.92	0.92	86705

Feature Importance



Most important features - Logistic Regression



Future Work



- Include the medications and diagnoses relationships in my model (Academic publications available)
- Do more interaction terms and feature engineering



Thank you!