Group - 7

THE UNIVERSITY
OF TEXAS AT DALLAS

Programming for Data Science

DIABETIC PATIENT READMISSION

Objective

To predict the hospital re-admission probability of a DIABETIC patient by using appropriate Data Science techniques.

WHY THIS DATA?

- People Affected by Diabetes :
  WORLD: 425m, USA: 26m (8.3% of the population)
- Expenditure on Diabetes:
  WORLD: $727 billion, USA: $327 billion
- People will be affected by Diabetes : 629 million
- Penalties paid by US Hospitals due to readmission of patients: $528 million
- Readmission Rates of diabetes patients are readmitted with 30-days of discharge : 20.3%

KNOW THE DATA

DIABETIC PATIENT ENCOUNTER

- 130 HOSPITALS ACROSS USA
- RECORDED BETWEEN 1999-2008
- 50 FEATURES, INCLUDING 23 MEDICINES
- GENDER : 53% FEMALE AND 46% MALE
- AGE : MOST FREQUENT (70-80) YEARS

	count	mean	std	min	25%	50%	75%	max
gender	57678.0	0.466001	0.498847	0.0	0.0	0.0	1.0	1.0
age	57678.0	71.089670	15.529343	10.0	60.0	70.0	80.0	100.0
admission_type_id	57678.0	2.092080	1.506279	1.0	1.0	1.0	3.0	8.0
discharge_disposition_id	57678.0	3.636846	5.278340	1.0	1.0	1.0	3.0	28.0
admission_source_id	57678.0	5.685963	4.152320	1.0	1.0	7.0	7.0	25.0
time_in_hospital	57678.0	4.337702	2.963873	1.0	2.0	4.0	6.0	14.0
num_lab_procedures	57678.0	43.193245	19.952895	1.0	31.0	44.0	57.0	132.0
num_procedures	57678.0	1.431291	1.757343	0.0	0.0	1.0	2.0	6.0
num_medications	57678.0	15.858768	8.261490	1.0	10.0	14.0	20.0	81.0

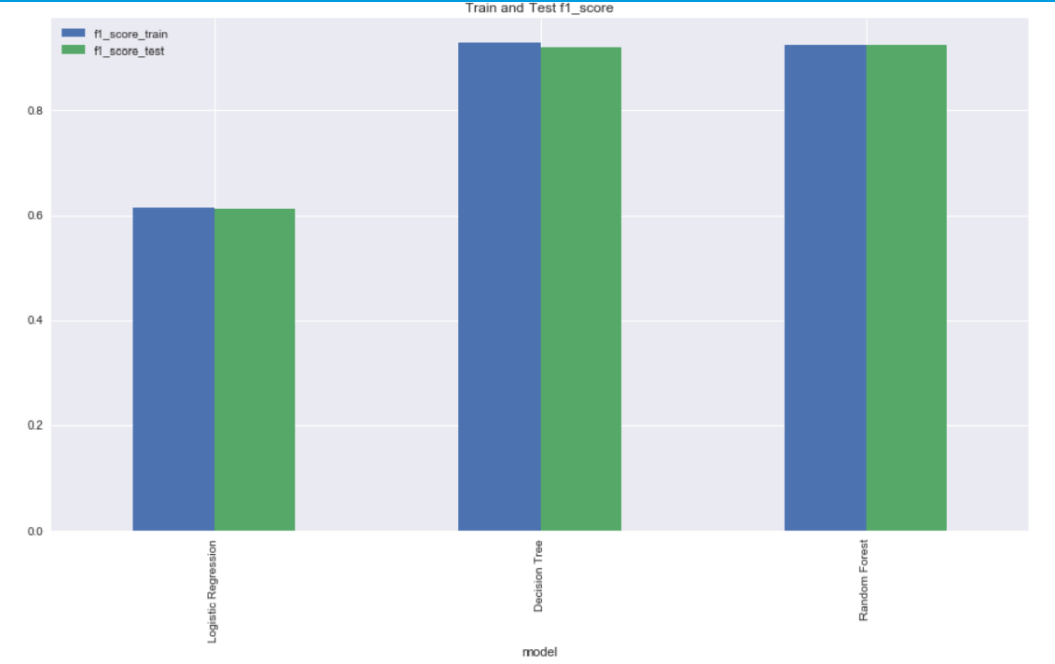
HEAT- MAP

- NO STRONG CORRELATION
- AGE HAS NO CORRELATION SIGNIFICANCE WITH TIME IN HOSPITAL & LAB PROCEDURES

PROJECT OUTLINE

STEPS WE TOOK TO SOLVE THE PROBLEM AND MAKE AN AWESOME MODEL

model	f1_score_train	f1_score_test
Logistic Regression	0.614356	0.612979
Decision Tree	0.929377	0.920059
Random Forest	0.925341	0.923664



FEATURE CREATION

Number of medication change

Number of medicine prescribed

Service utilization

FEATURE ENGG.

Feature Encoding

Collapsing Multiple Encounters

Standardization & Log Transform

Oversampling by SMOTE

MODELLING

Logistic Regression

Decision Tree

Random Forest

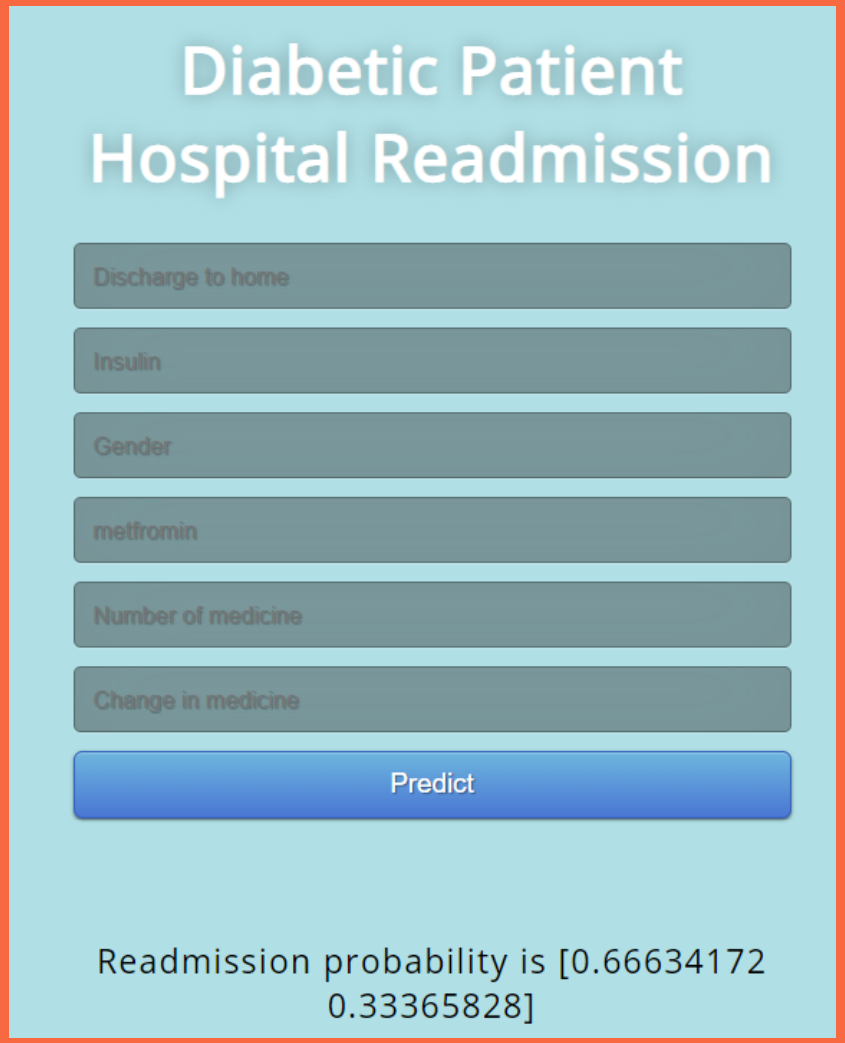
10175

18

1422

8712

PRODUCT INTERFACE

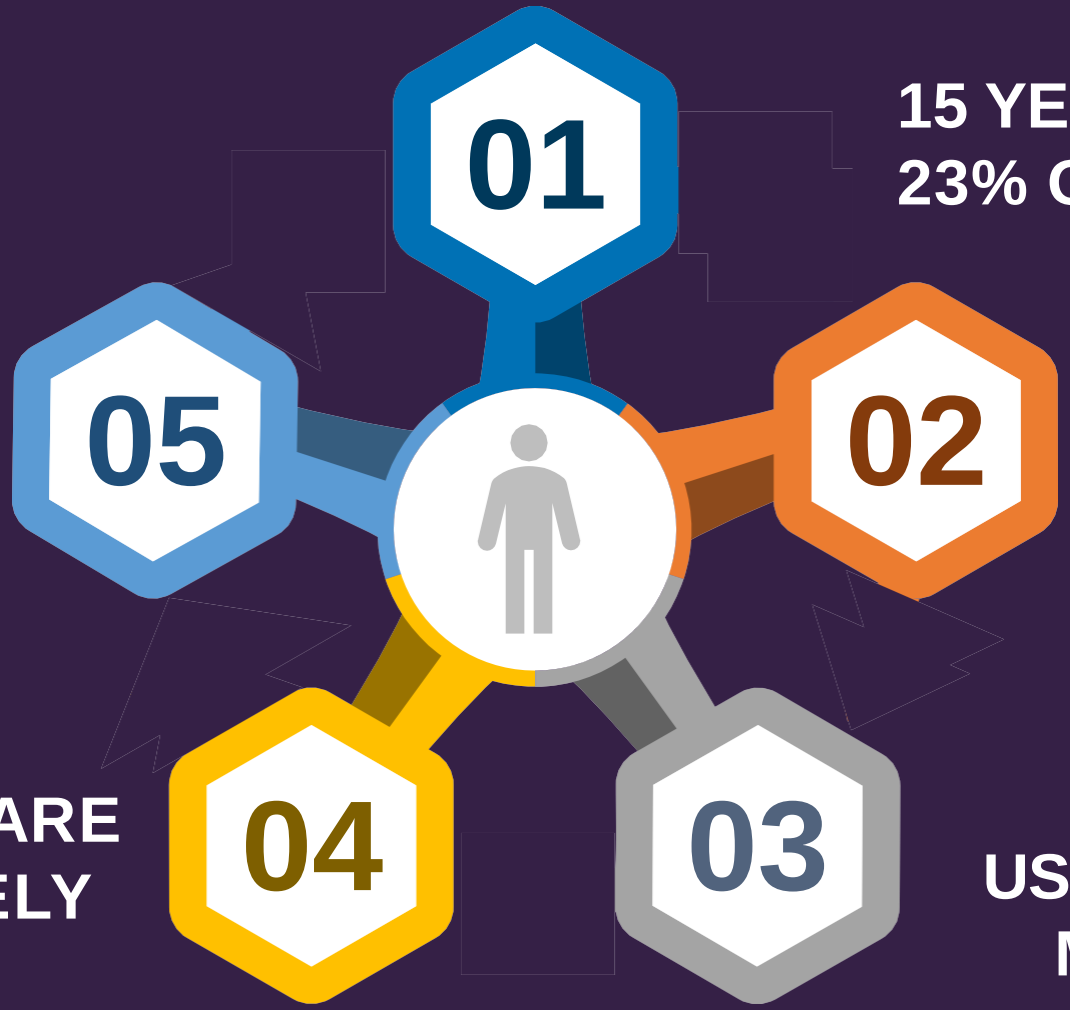


GROUP - 7

PRODUCT DEMO

PREDICTING
DIABETIC PATIENT READMISSION

INSIGHTS FOR READMISSION



15 YEARS AGE ↑
23% ODDS OF RE-ADM.

PROMINENT FACTOR DISCHARGE TO HOME

USE OF INSULIN & METAFROMIN INCREASES CHANCE

CAUCASIAN ARE HIGHLY-LIKELY

MALE HAS MORE ODDS COMPARED TO FEMALE

IMPROVEMENTS

01 USING MORE FEATURES IN MODEL

02 ADDING MORE STAT IMP FIELDS ON INTERFACE

03 ATTAINING CURRENT DATA

04 UTILIZING UNDER-SAMPLING TECHNIQUE

05 EMPLOYING OTHER MACHINE LEARNING METHODS

THANK YOU