Risk Test of People with Type II Diabetes

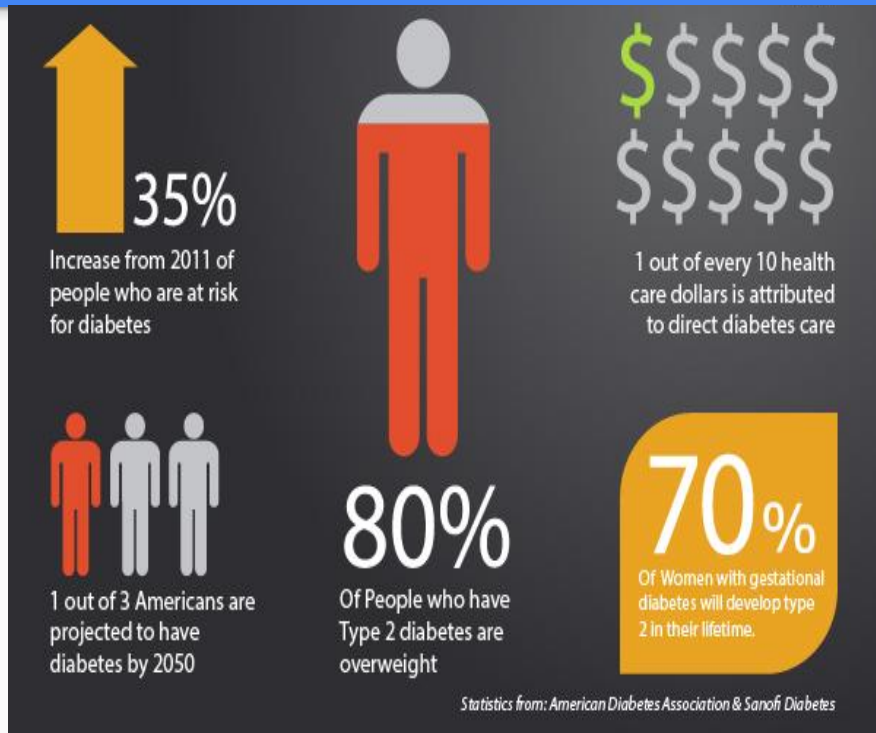
Are you at risk?

Yijing Li
Sonia Mahop

Diabetes in the US

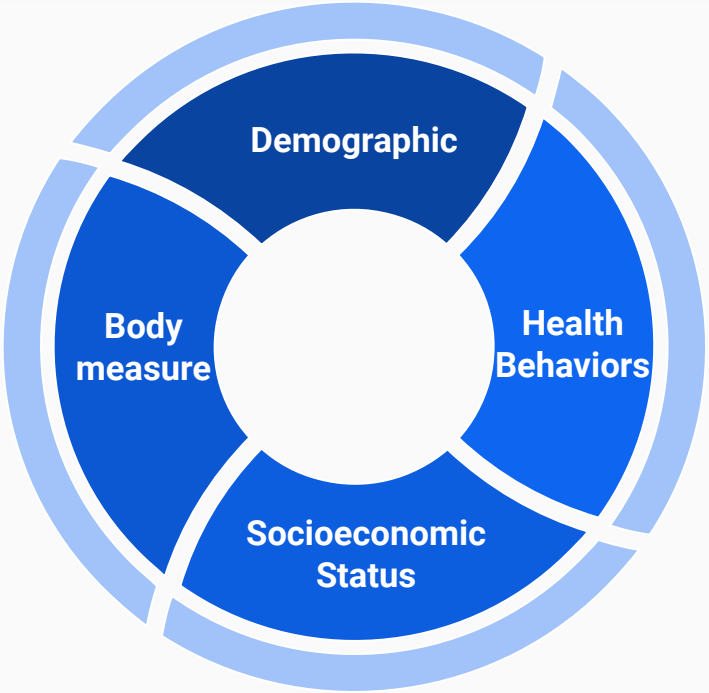
- An estimated 23.1 million people are diagnosed with diabetes at a cost of more than $245 billion per year.

- Individuals with diabetes are at greater risk for many common problems, including coronary heart disease, stroke, hypertension, and depression.



Are you at risk for type II diabetes?

- National Health and Nutrition Examination Survey (NHANES)
  - Designed to study health and nutritional status of adults and children in the United State.
  - Note that we only take the type II diabetes into account.

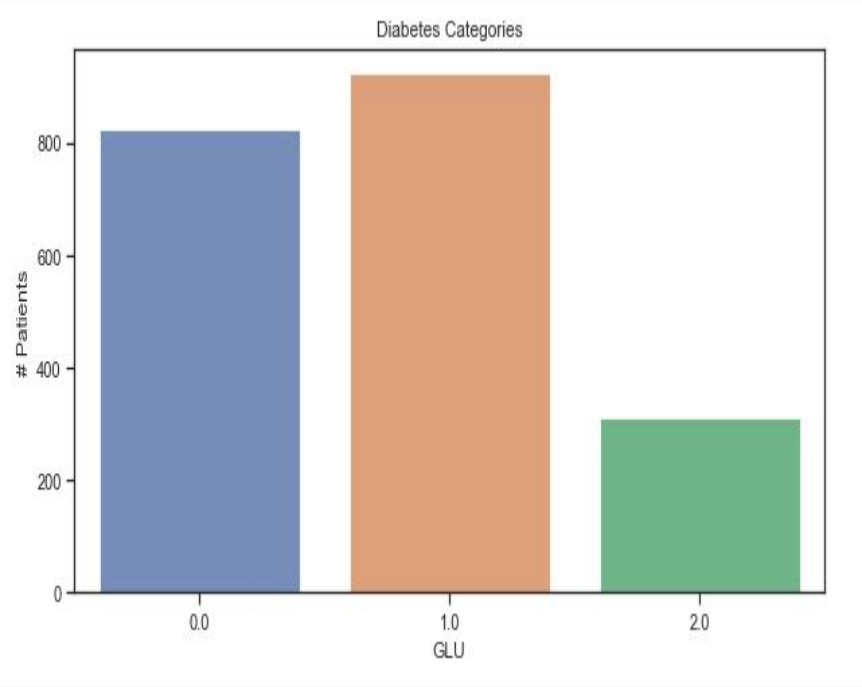


National Health and Nutrition Examination Survey

Methodology

- Dummy
- Logistic Regression
- K-Nearest Neighbors
- Support Vector Machine
- Decision Tree
- Random Forest

Methodology



Target

0 Normal — Fasting blood glucose level less than 100 mg/dL (5.6 mmol/L)

1 Prediabetes — Fasting blood glucose level between 100 to 125 mg/dL (5.6 - 6.9 mmol/L)

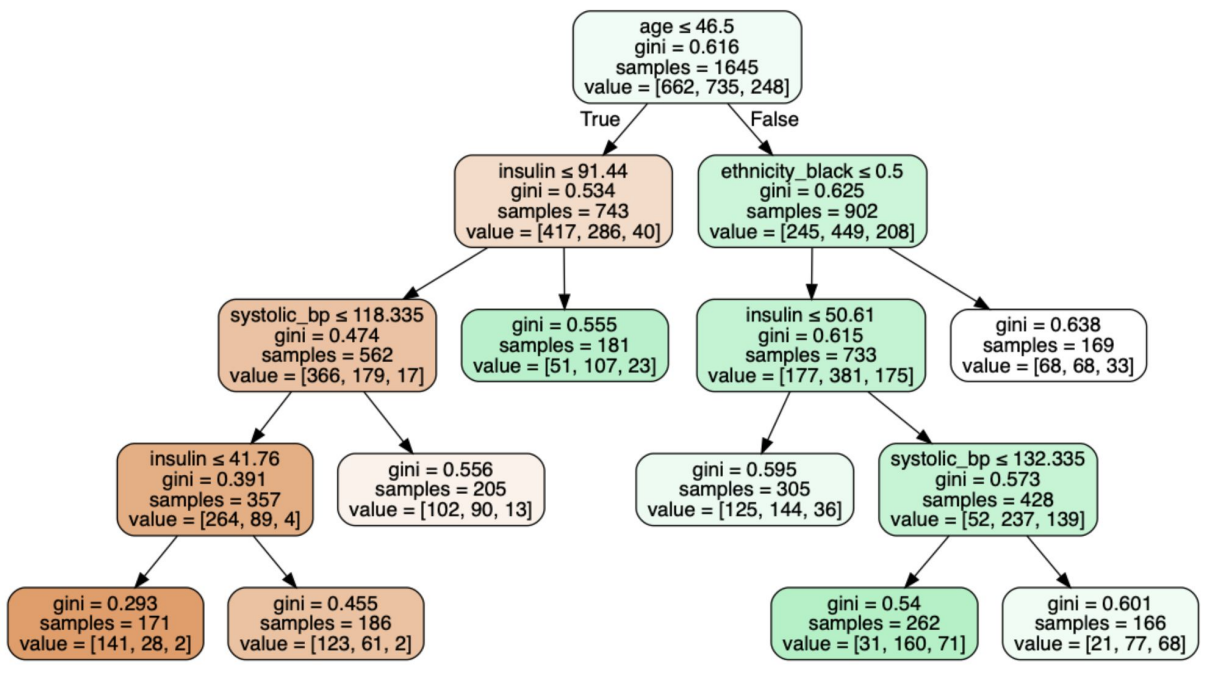
2 Diabetes — Fasting blood glucose level is 126 mg/dL (7 mmol) or above

Results Comparison

Evaluation	Baseline Model	Logistic Regression	K-Nearest Neighbors	Support Vector Machine	Decision Tree	Random Forest
F-1 Score	0.283	0.529	0.529	0.528	**0.562**	**0.566**
Accuracy Score	0.453	0.558	0.541	0.568	**0.570**	**0.594**
Precision Score	0.206	0.586	0.537	0.615	**0.564**	**0.603**
Recall Score	0.454	0.558	0.541	0.568	**0.570**	**0.594**

*** In the case of multiclass classification, evaluation scores are adjusted*

Decision Tree

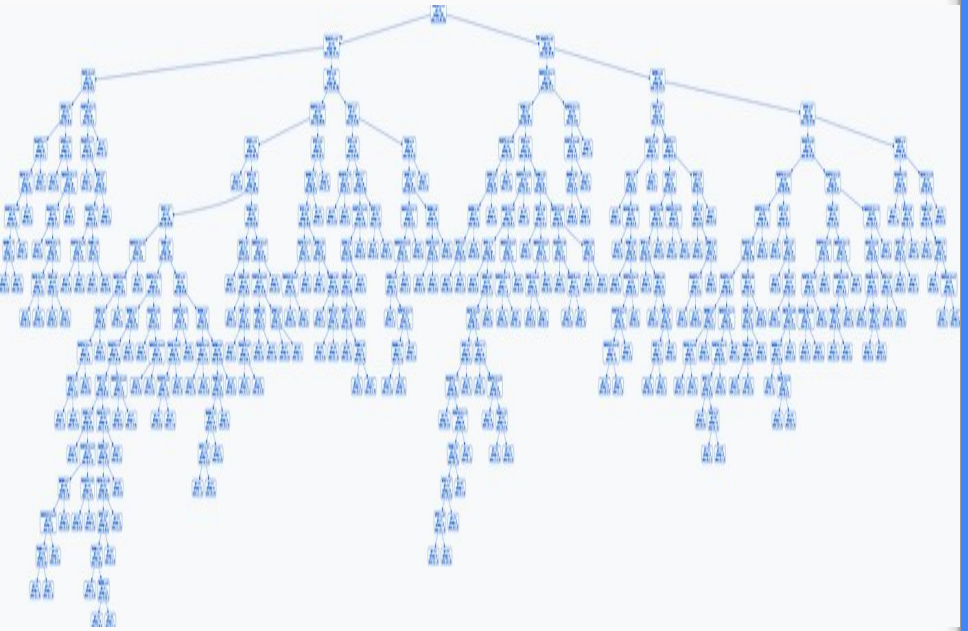


Most important features:

- Age
- Insulin level
- Systolic blood pressure
- Ethnicity

Best Model

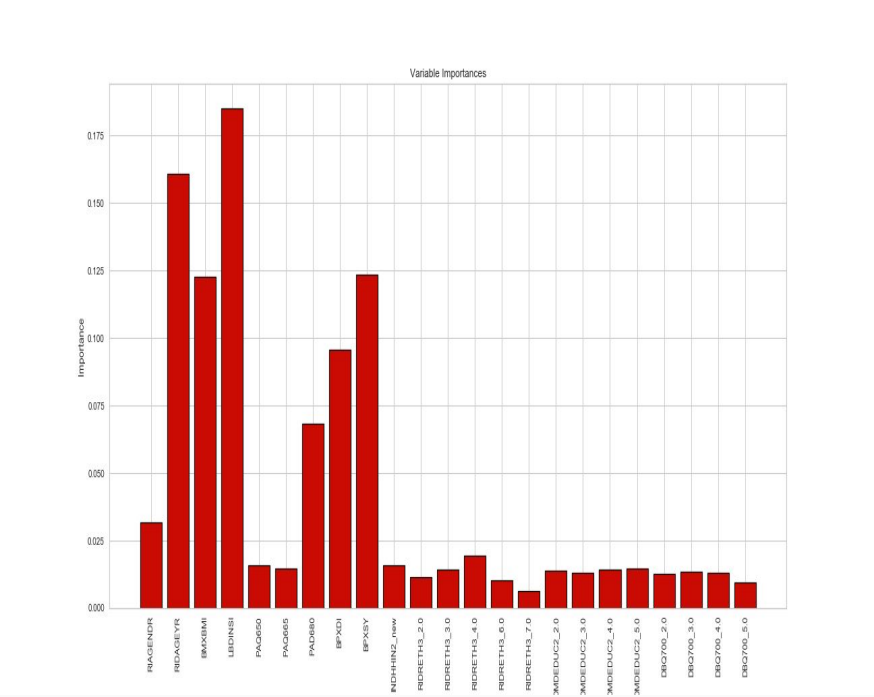
Random Forest Classification



- F-1 Score: **0.566**
- Recall/Sensitivity Score: **0.594**

Best Model

Random Forest Classification



Most important features Ascending:

- Insulin Level
- Age
- Systolic blood pressure
- BMI
- Diastolic Blood Pressure
- Sedentary Time

Recommendation

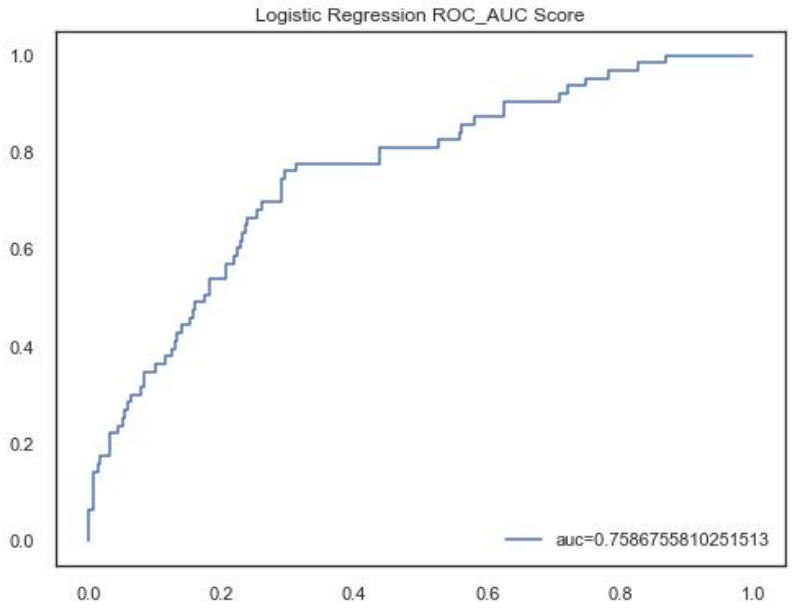
- Are you at risk?
  - Age over 46
  - Insulin level
  - Blood pressure
  - BMI
  - Physically inactive

Limitation

- Number of features included in the analysis
- Better understanding of grid search, randomize search, pipeline

Next Step

- Include more variables
- Reduce to two classes:
  - ‘Normal’
  - Diabetes
- Tune the hyperparameters of the algorithm
- Try different algorithms



Thanks!