**Introduction:**

The growing number of diabetes type 1/ type 2 diagnoses across the United States has sparked major concerns for the overall health its residents. As workers in healthcare/pharmaceuticals, we found this topic quite interesting and impactful to our careers. Simply put, diabetes is a disease in which the body is unable to properly produce or respond to hormone insulin, resulting in deviant metabolism of carbs and higher concentrations of glucose in blood and urine. Type 1 diabetes is a hereditary disorder, often apparent at a young age. Meanwhile, type 2 diabetes usually develops over time and is mainly related to diet. We wanted to build an effective model to accomplish two goals: predicting the risk of diabetes based upon measurable, physical characteristics and determining the probability of hospital readmission for those already affected by the disease.

Shape

Description automatically generated with medium confidence

Source: https://commons.wikimedia.org/w/index.php?search=Diabetes+mellitus&title=Special:MediaSearch&type=image

**System Requirements:**

This project required many languages and programs utilized throughout this course including Python 3, Pandas, Matplotlib, Jupyter Lab, flask app, pickle, and html.

**Installation Packages:**

Packages include the following:

Sklearn.model\_selection import GridSearchCV, train\_test\_split

Sklearn.preprocessing import StandardScalar

Sklearn.ensemble import Random Forest Classifier

Sklearn.linear\_model

Pip install upgrade category\_encoders

**Data Engineering:**

We pulled two csv files of diabetes data from Kaggle- both of which are in the Notebooks folder of our GitHub. These were uploaded to a Jupyter Lab interface for data analysis.

**Data Analysis:**

1. Assigned each data set to separate ipynb files and converted to pandas’ data frames that were cleaned.
2. Data frames were separated into test/training buckets according to each variable (X/y). These were standardized and scales.
3. Base and edited Random Forest Classifier models were created and tested for accuracy.
4. Employed further model improvement methods to improve accuracy results.

**Conclusion:**

We realized that the Random Forest Classifier is an effective model for predicting the risk of acquiring diabetes and the probability of hospital readmission for those affected among small and large sets of data, respectively.

A crowd of people at a concert

Description automatically generated with medium confidence

Source:https://commons.wikimedia.org/w/index.php?search=insulin&title=Special:MediaSearch&go=Go&type=image