**Diabetes Prediction**

**Data Exploration and Cleaning:**

* Few feature data types are incorrect.
* The features are
* The size of the data is 100000 \* 8.
* The data set has 4 numerical features and 4 categorical features after.
* The dataset doesn’t contain any missing values.
* Features “hypertension”, “heart\_disease” types were improper. So, I have changed their type from int to object.
* We didn’t detect missing values in the data set.
* This data set contains 3854 duplicate values.