
WISDOM ML
105
Early stage
Diabetes
Detection
using
Machine
Learning

The Problem

Diabetes is a common condition. The CDC found that 30.3 million adults in the U.S. have diabetes. They also estimate that 80.4 million adults have prediabetes. Diabetes can be difficult to live with and be expensive to treat. Therefore, diagnosing diabetes early can be helpful to a person's way of life and wallet.

The Goal

Develop a machine learning algorithm that can read common symptoms for diabetes and predict whether a person will get diabetes.

By knowing ahead of time, the person can work on minimizing or eliminating some symptoms to minimize the chance that they will get diabetes.

Data

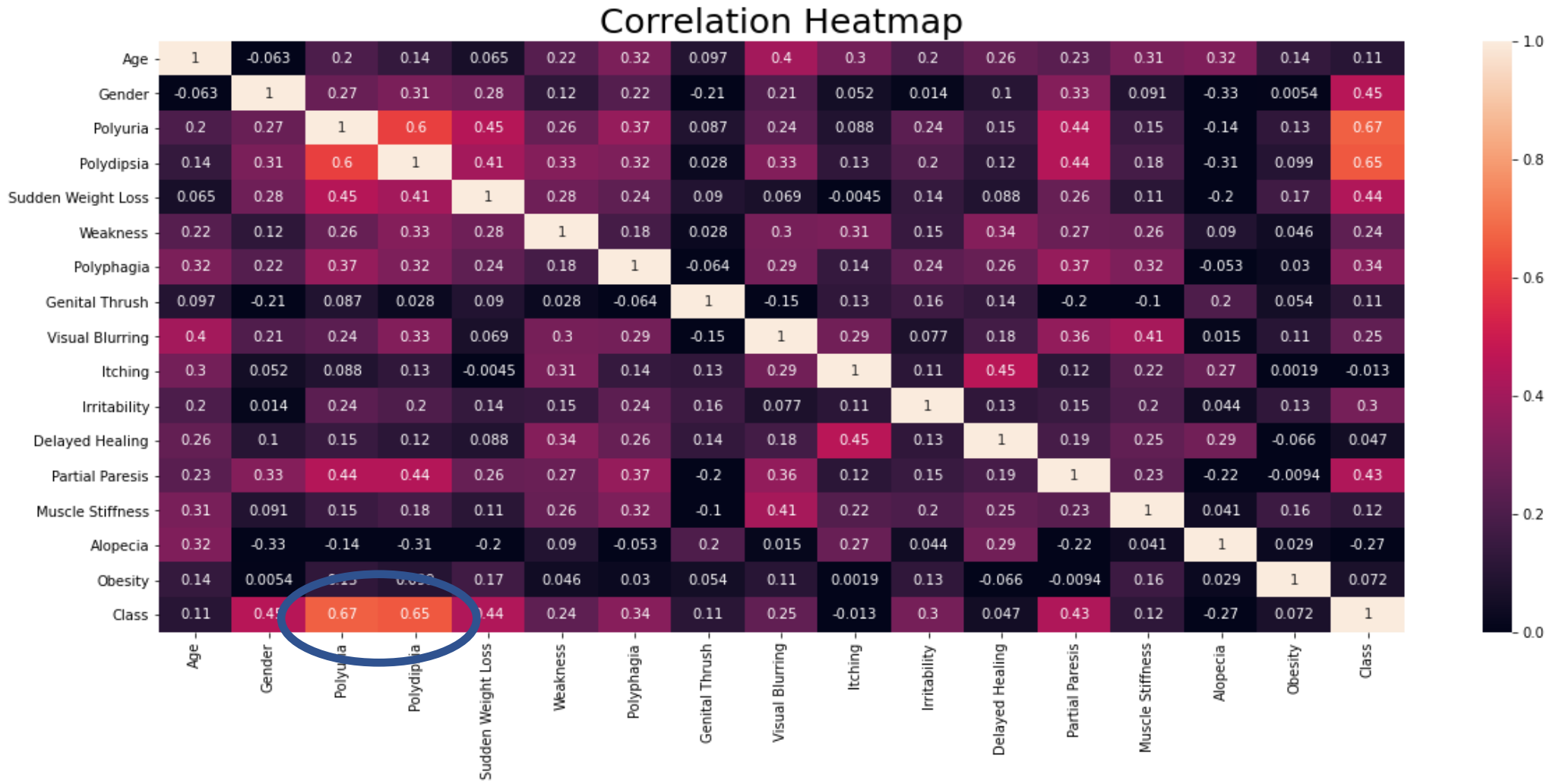
kaggle


UCI
Machine Learning Repository
Center for Machine Learning and Intelligent Systems

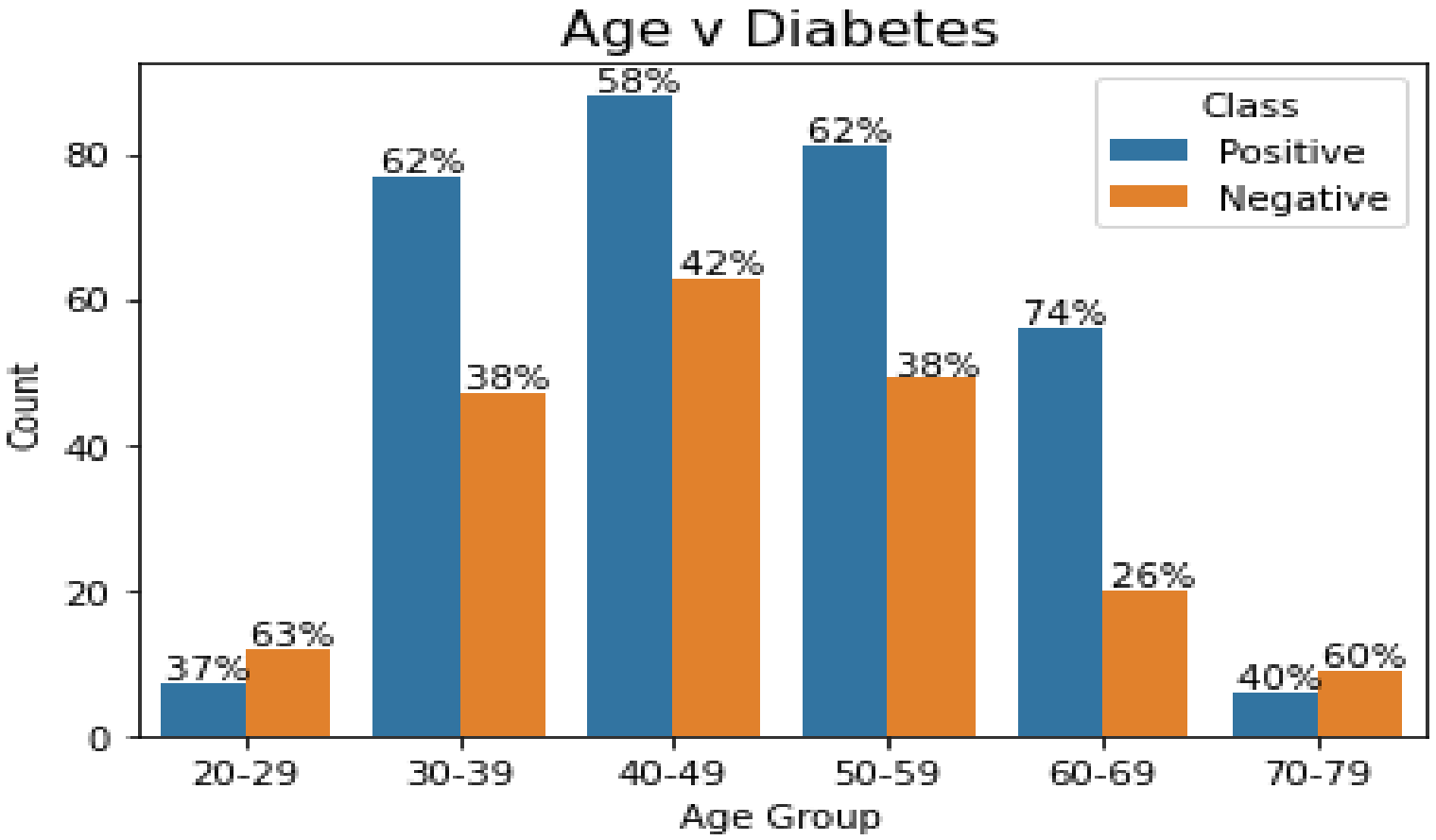
Data Wrangling

The dataset did not have any missing values.

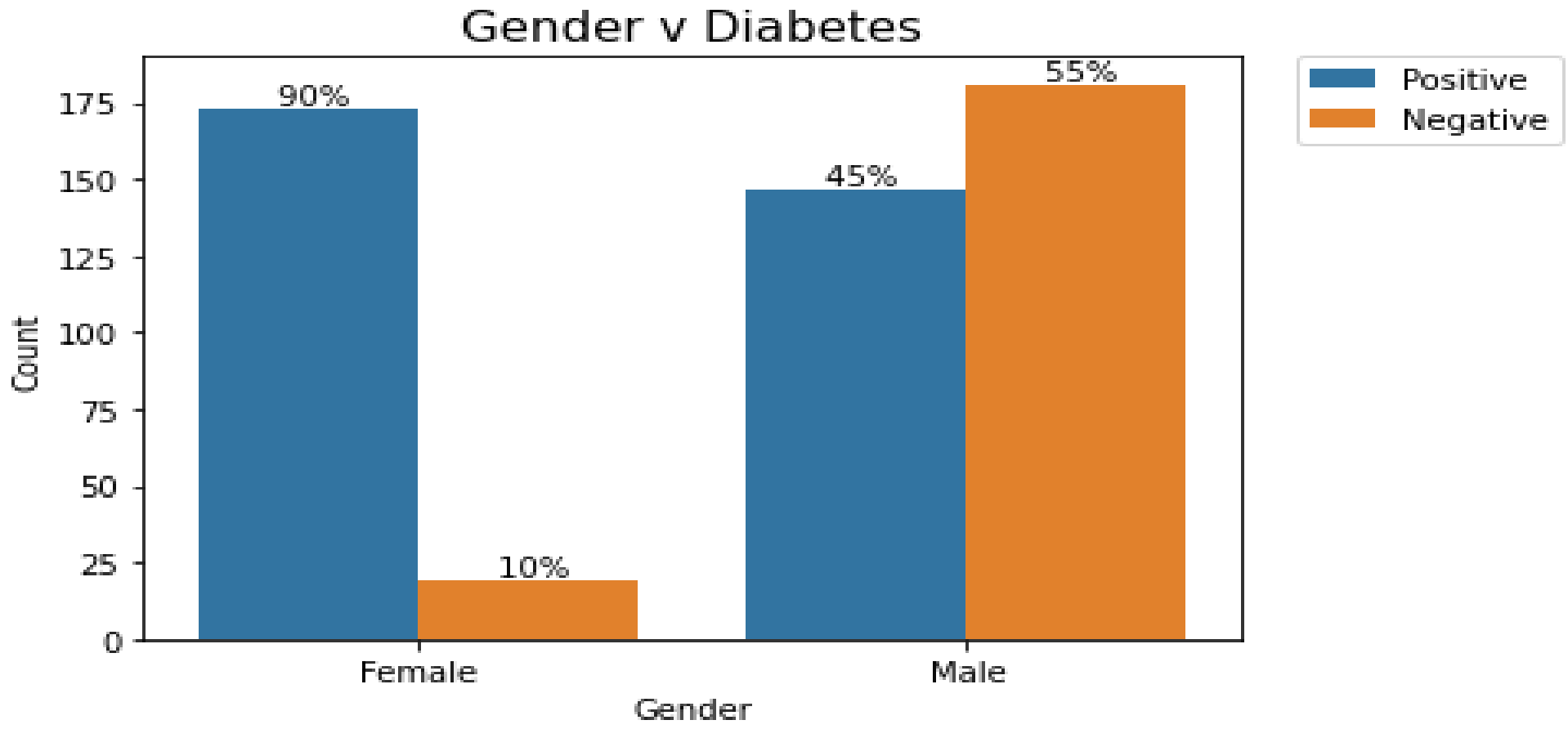
Exploratory Data Analysis



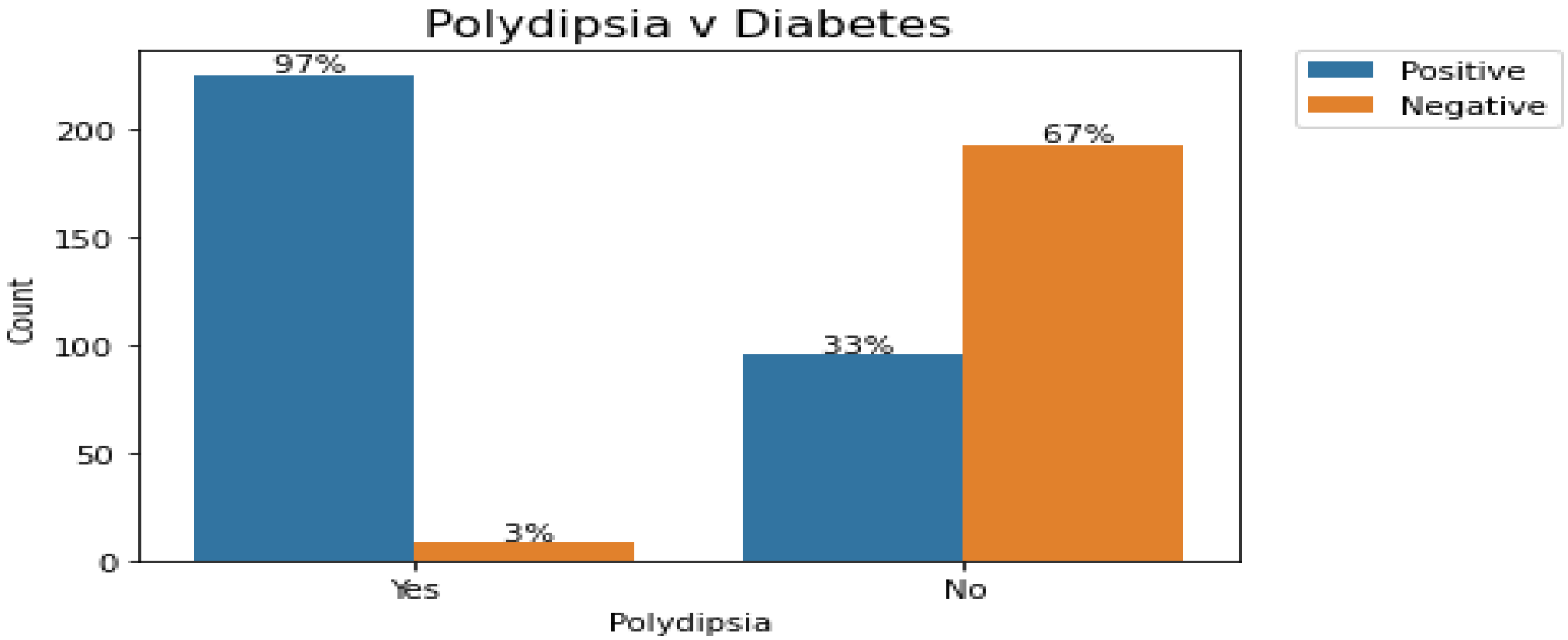
Feature Analysis



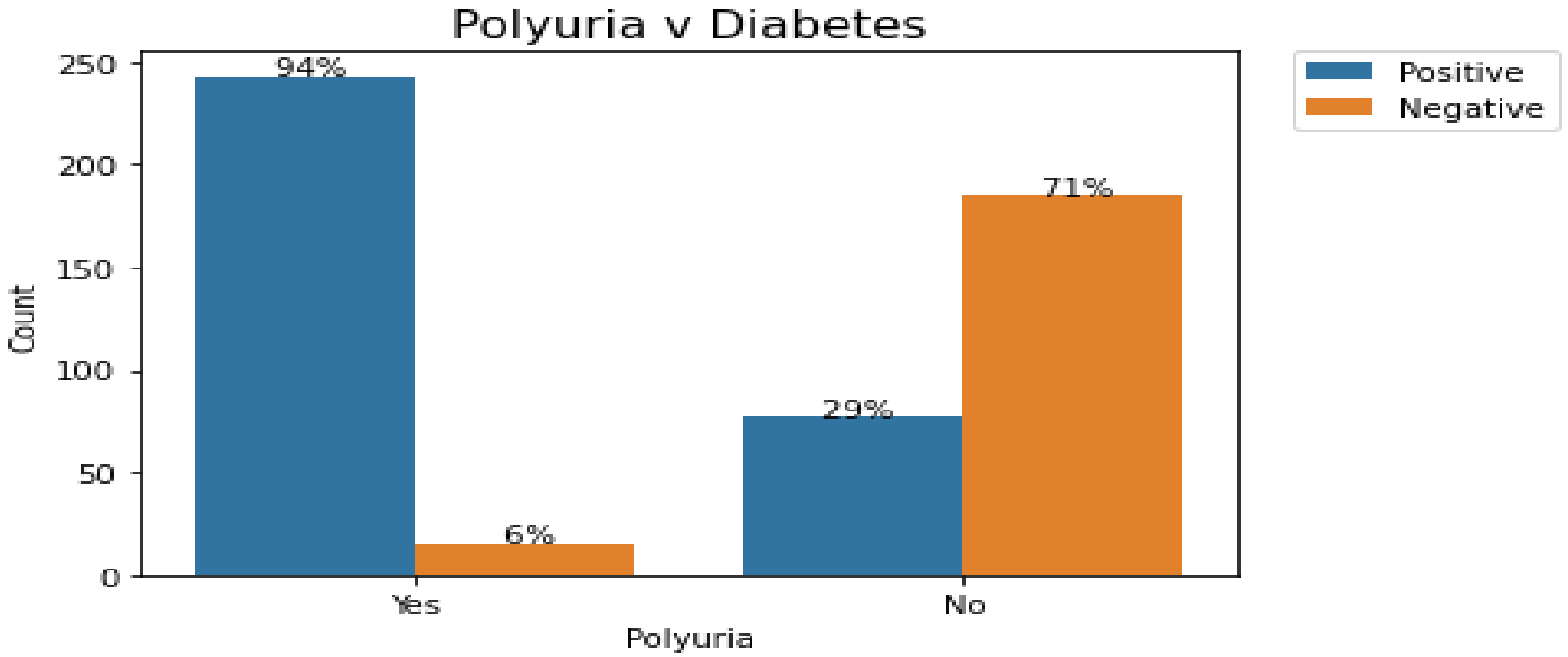
Feature Analysis



Feature Analysis

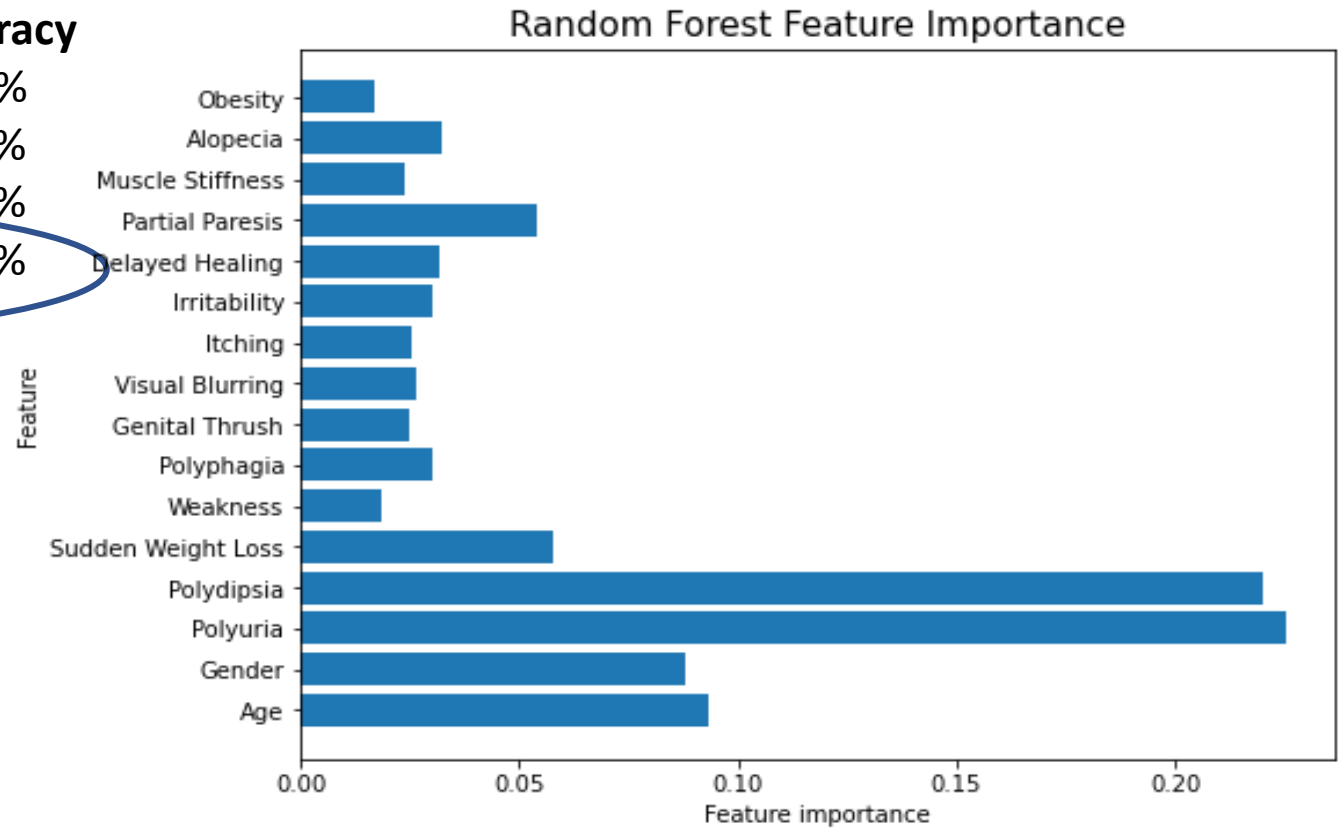


Feature Analysis



Model Analysis

Modelling Algorithm	Accuracy
Logistic Regression with Scaling	93%
Decision Tree	94%
Gradient Boosting	95%
Random Forest	98%



Summary

Correlation Heatmap	Polyuria	67%
Correlation Heatmap	Polydipsia	65%
Feature Analysis	Polydipsia	97%
Feature Analysis	Polyuria	94%
Feature Analysis	Gender – Female	90%
Feature Analysis	Age – 60’s	74%
Model	Logistic Regression with Scaling	93%
Model	Decision Tree	94%
Model	Gradient Boosting	95%
Model	Random Forest	98%

For Future Consideration

The data had only 520 instances. Collecting more data may help to make the model prediction more accurate.