**Risk Analysis Explanation**

1. **Prior Probability:** The prevalence of metabolic syndrome in the dataset. It is calculated by counting up every instance of metabolic syndrome and dividing it by the total number of rows or data points. In this case it was 0.34 or 34%
2. **Likelihood:** Likelihood is defined as he probability of observing a specific feature given that an individual has metabolic syndrome. It tells you how likely it is to see a particular feature (e.g., high BMI) if the person has the disease.

So, for example if we want to calculate the likelihood of a Male being above the age of 45 given that he has Metabolic Syndrome, we calculate it by counting up every instance in which the sex is male, the age is >= 45 and has metabolic syndrome, and we divide it by the total cases of individuals who have metabolic syndrome. In this specific case, the likelihood of this scenario was 34.3%

In the results, the column with the value age\_male \_risk will represent this value.

Same is done for female, but the age that is considered is >=55.

Here is a table that summarizes the likelihood results

|  |  |  |
| --- | --- | --- |
| Condition 1 | Condition 2 | Likelihood |
| Waist >= 88 | Sex = Female | 0.933 |
| Blood Glucose >= 100 | - | 0.855 |
| Waist >= 102 | Sex = Male | 0.768 |
| BMI >= 30 | - | 0.624 |
| HDL < 50 | Sex = Female | 0.610 |
| Age >= 45 | Sex = Male | 0.343 |
| Age >= 55 | Sex = Female | 0.270 |
| Uric Acid > 6.0 | Sex = Female | 0.320 |
| Uric Acid > 7.0 | Sex = Male | 0.291 |
| Triglycerides > 150 | - | 0.547 |
| HDL < 40 | Sex = Male | 0.468 |
| Race = White | - | 0.410 |
| Race = Black | - | 0.220 |
| Race = MexAmerican | - | 0.125 |
| Race = Hispanic | - | 0.125 |
| Race = Asian | - | 0.099 |
| Race = Other | - | 0.021 |
| UrAlbCr >=30 | - | 0.203 |
| Albuminaria > 1 | - | 0.159 |

1. **Posterior Probability:** This is the updated probability of having metabolic syndrome after considering a specific feature value. It combines the prior probability and the likelihood to give a revised belief about the probability of the disease given the feature.

This is calculated by first calculating the probability of certain factors.

For example, we want to calculate the posterior probability of Male above or equal to the age of 45 having metabolic syndrome.

The code first finds out the probability of instances of Sex = Male and age >= 45 and divides it by the entire length of the dataset. Now we have the probability of that prior knowledge.

We then calculate the likelihood of Sex = Male, and Age >=45 given Metabolic Syndrome is present.

We also have the prior probability of the disease.

Then the posterior probability of this case is calculated by

(likelihood \* prior)/probability of feature

The posterior probabilities are calculated for all the conditions for which the likelihoods were calculated.