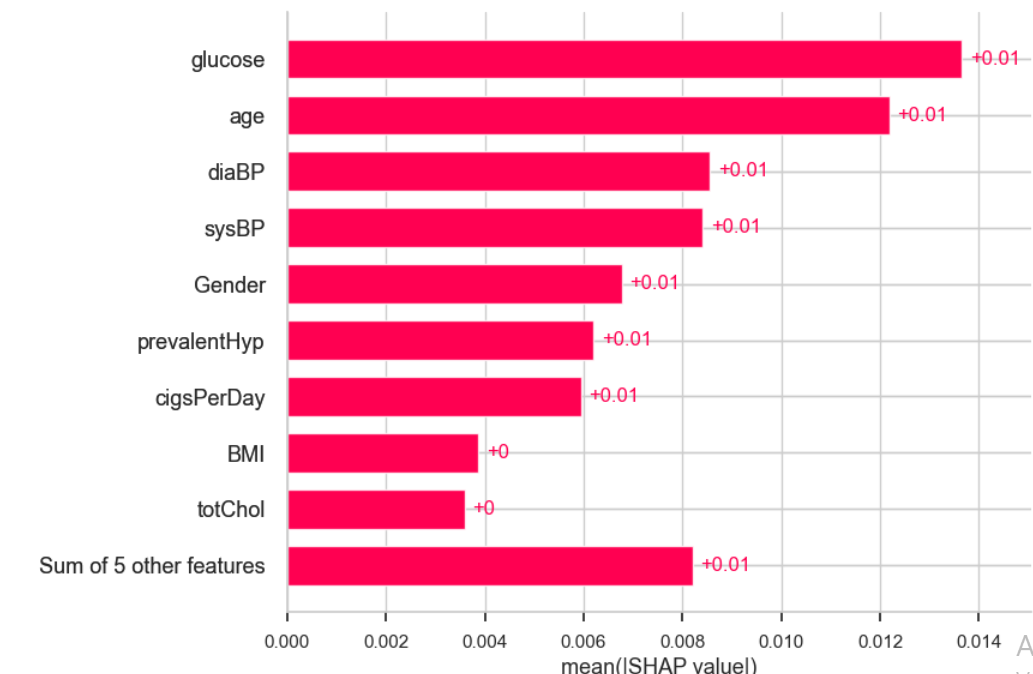
In order to use SHAP, we need a tree-based model so we decide to apply the random forest classification. A random forest is a supervised algorithm that uses an ensemble learning method consisting of a multitude of decision trees, the output of which is the consensus of the best answer to the problem. A random forest is nothing more than a series of decision trees with their findings combined into a single final result. They are so powerful because of their capability to reduce overfitting without massively increasing error due to bias.

SHAP (SHapley Additive exPlanations) values are a way to explain the output of any machine learning model. It uses a game theoretic approach that measures each player's contribution to the final outcome. In machine learning, each feature is assigned an importance value representing its contribution to the model's output.

SHAP values show how each feature affects each final prediction, the significance of each feature compared to others, and the model's reliance on the interaction between features.

The reason we decide to apply SHAP system here is mainly to see and compare if the best features that feature selection gives us is the same as what SHAP throws at us. The SHAP most important figures are the following ones:



19 – Shap Value

Here the features are ordered from the highest to the lowest effect on the prediction. It takes in account the absolute SHAP value, so it does not matter if the feature affects the prediction in a positive or negative way. We can see that the features order changed, for example or top 1 feature now is glucose while top 1 feature from feature selection is age. In machine learning, each feature is assigned an importance value representing its contribution to the model's output. SHAP values show how each feature affects each final prediction, the significance of each feature compared to others, and the model's reliance on the interaction between features.