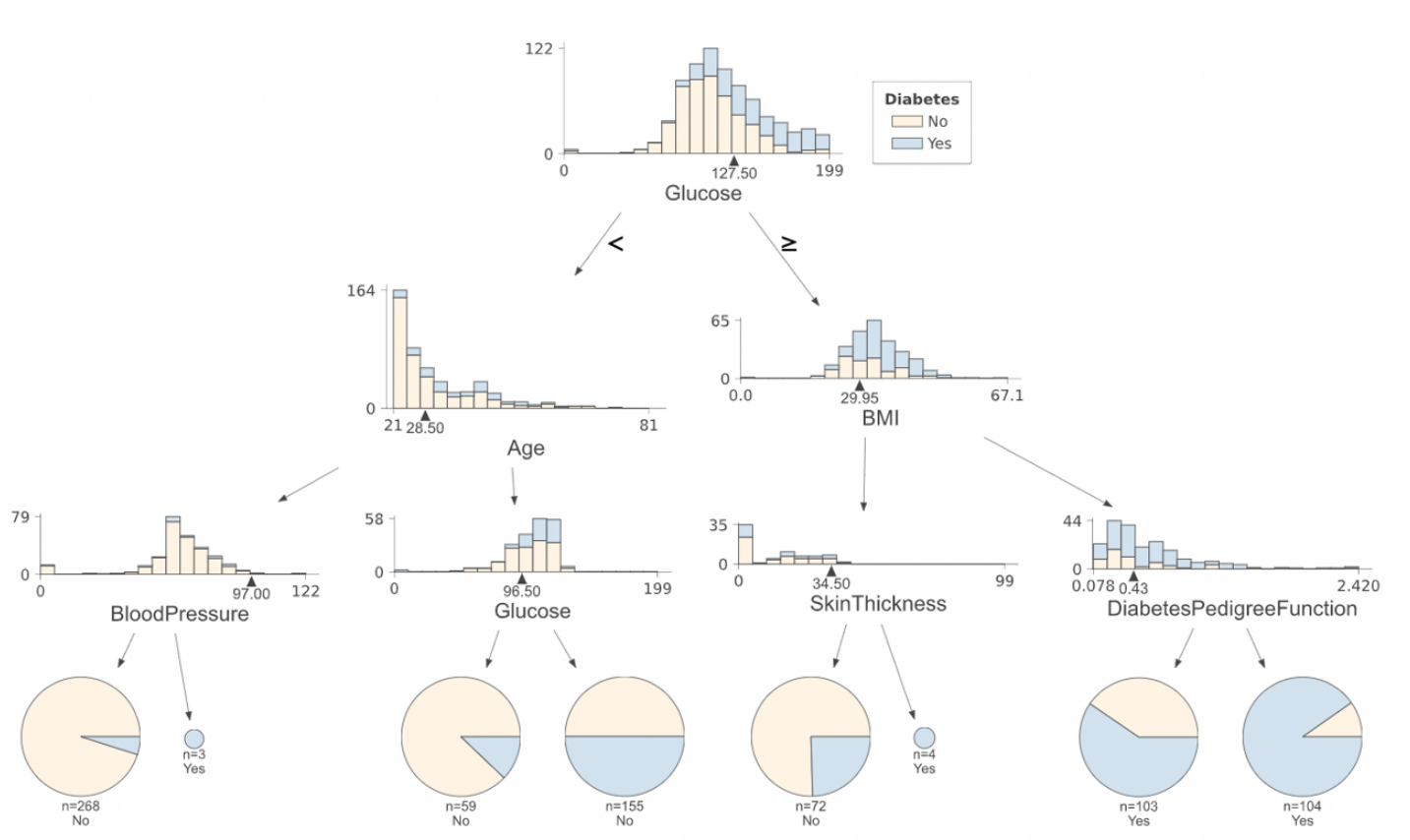
Exercise 4.1: Bagging vs Random Forest

□□□ Bagging vs Random Forest

How does Random Forest improve on Bagging?

The goal of this exercise is to investigate the correlation between randomly selected trees from *Bagging* and *Random Forest.*



Instructions:

- Read the dataset `diabetes.csv` as a pandas data-frame, and take a quick look at the data.
- Split the data into *train* and *validation* sets.
- Define a `BaggingClassifier` model that uses `DecisionTreClassifier` as its base estimator.
- Specify the number of bootstraps as 1000 and a maximum depth of 20.
- Fit the `BaggingClassifier` model on the *train* data.
- Use the helper code to predict using the mean model and individual estimators. The plot will look similar to the one given below.
- Predict on the test data using the first estimator and the mean model.

- Compute and display the *validation* accuracy
- Repeat the modeling and classification process above, this time using a `RandomForestClassifier.`

Hints:

```
pandas.DataFrame.Drop()
```

Drop specified labels from rows or columns.

```
sklearn.train_test_split()
```

Split arrays or matrices into random train and test subsets.

```
sklearn.ensemble.BaggingClassifier()
```

Returns a Bagging classifier instance.

```
sklearn.tree.DecisionTreeClassifier()
```

A Tree classifier can be used as the base model for the Bagging classifier.

```
sklearn.ensemble.andomForestClassifier()
```

Defines a Random forest classifier.

```
sklearn.metrics.accuracy_score(y_true, y_pred)
```

Accuracy classification score.