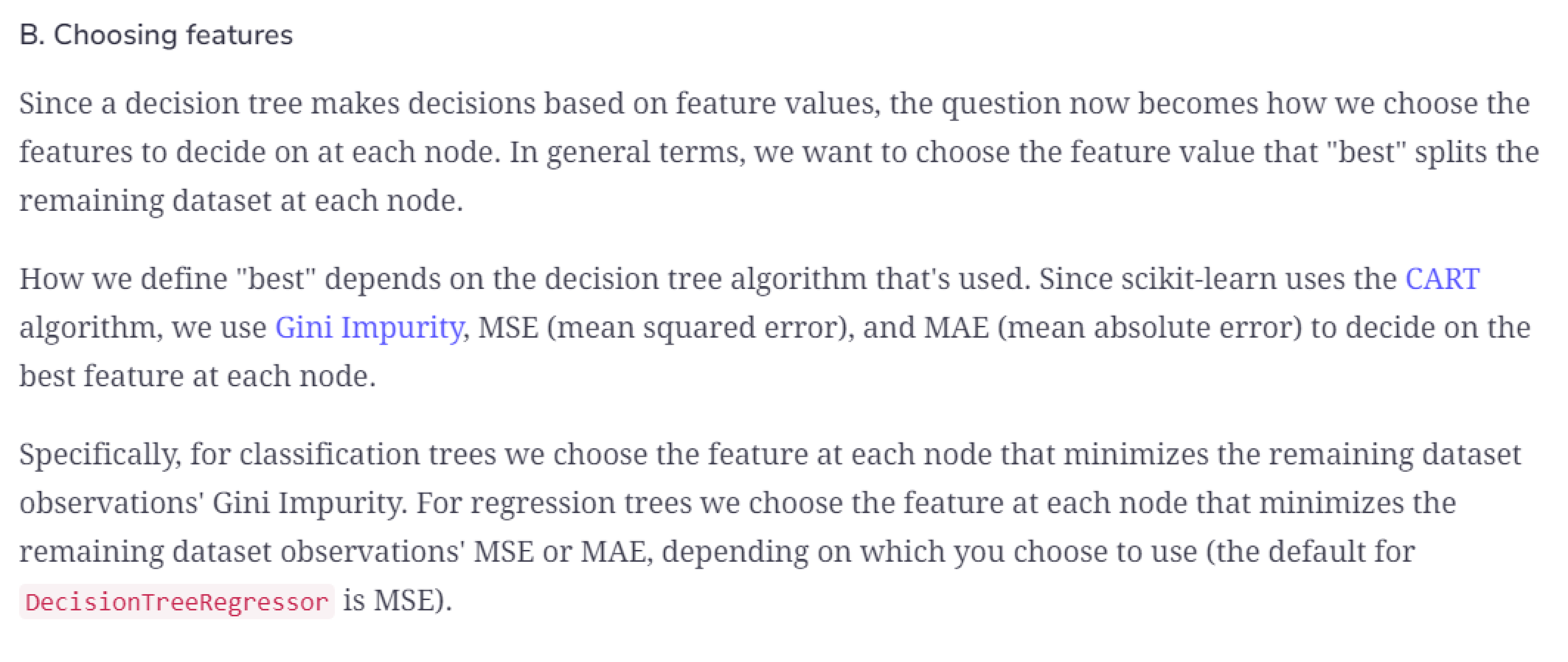
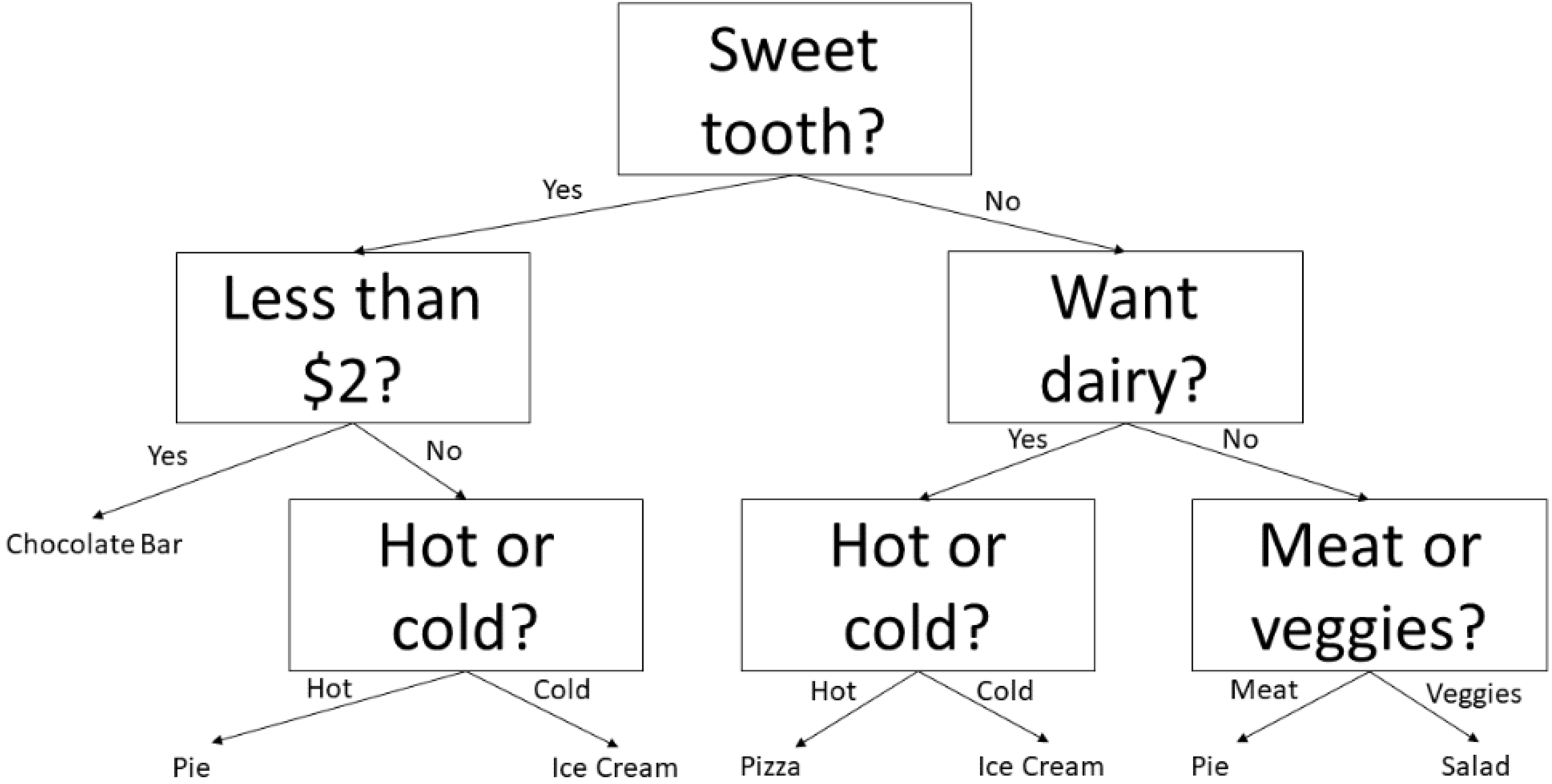


B. Choosing features

Since a decision tree makes decisions based on feature values, the question now becomes how we choose the features to decide on at each node. In general terms, we want to choose the feature value that "best" splits the remaining dataset at each node.

How we define "best" depends on the decision tree algorithm that's used. Since scikit-learn uses the CART algorithm, we use Gini Impurity, MSE (mean squared error), and MAE (mean absolute error) to decide on the best feature at each node.

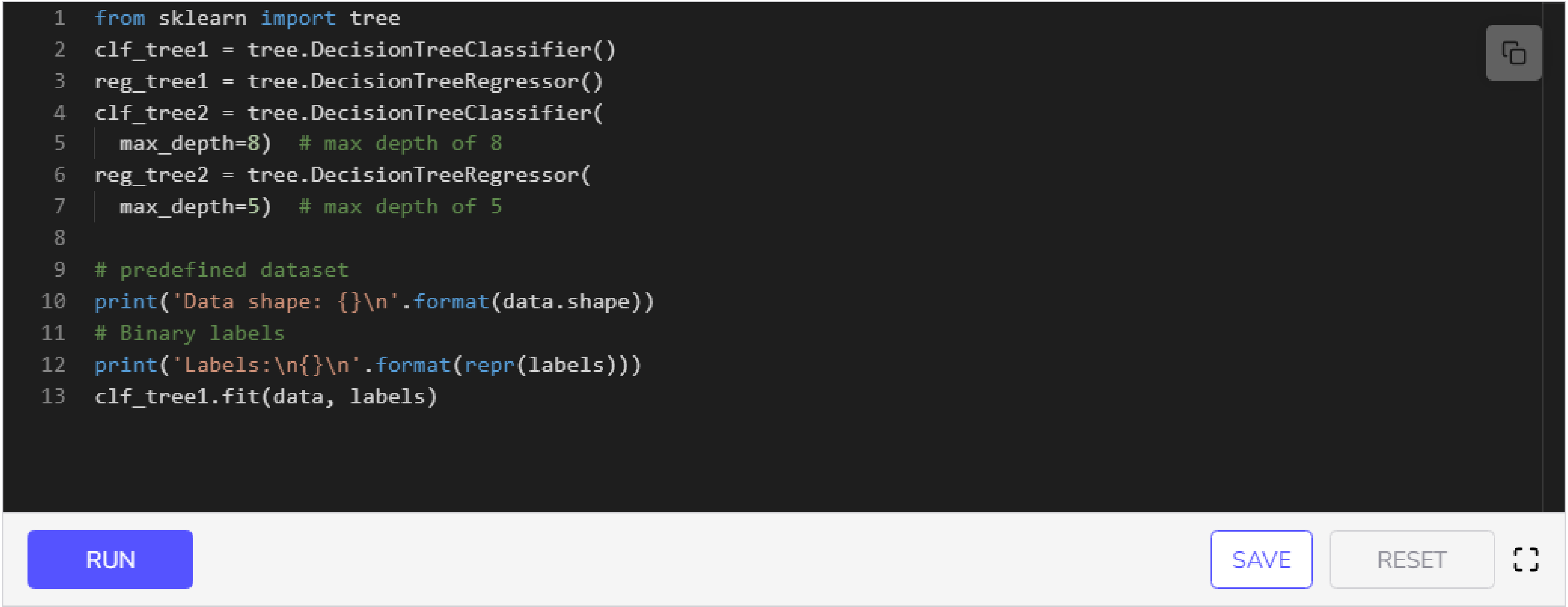
Specifically, for classification trees we choose the feature at each node that minimizes the remaining dataset observations' Gini Impurity. For regression trees we choose the feature at each node that minimizes the remaining dataset observations' MSE or MAE, depending on which you choose to use (the default for `DecisionTreeRegressor` is MSE).



A decision tree for deciding what to eat. This is an example of multiclass classification.

In scikit-learn, we implement classification decision trees with the `DecisionTreeClassifier` object, and regression trees with the `DecisionTreeRegressor` object. Both objects are part of the `tree` module in scikit-learn.

The code below demonstrates how to create decision trees for classification and regression. Each decision tree uses the `fit` function for fitting on data and labels.



```python
from sklearn import tree
clf_tree1 = tree.DecisionTreeClassifier()
reg_tree1 = tree.DecisionTreeRegressor()
clf_tree2 = tree.DecisionTreeClassifier(
  max_depth=8)  # max depth of 8
reg_tree2 = tree.DecisionTreeRegressor(
  max_depth=5)  # max depth of 5

# predefined dataset
print('Data shape: {}\n'.format(data.shape))
# Binary labels
print('Labels:\n{}\n'.format(repr(labels)))
clf_tree1.fit(data, labels)
```

The `max_depth` keyword argument lets us manually set the maximum number of layers allowed in the decision tree (i.e. the tree's maximum depth). The default value is `None`, meaning that the decision tree will continue to be constructed until no nodes can have anymore children. Since large decision trees are prone to overfit data, it can be beneficial to manually set a maximum depth for the tree.