

decision_trees_real_data_practice_workbook

September 22, 2020

1 DECISION TREES REAL DATA PRACTICE WORKBOOK

2 IMPORT THE DATA

```
In [ ]: # import packages going to use
In [ ]: # import the data to colab

        # read the data into google colab

In [ ]: # print out the headder
In [ ]: # look at the information for the data
In [ ]: # do you have a problem with missing data?
In [ ]: # select the data you are going to use
        # survived, pclass, mf, age, sibsp, parch

In [ ]: # create a new data set from this data selection and drop the nans
In [ ]: # look at data set
In [ ]: # import decision tree classifier if you haven't done so already
In [ ]: # set up your data to test on as sibsp and parch
        # set up your answer as survived

In [ ]: # run the model on the data selection

In [ ]: # look at the graph
!apt-get install python-pydot
!pip install pydotplus
from sklearn.tree import export_graphviz
from sklearn.externals.six import StringIO
from IPython.display import Image
import pydotplus
tree_data = StringIO()
export_graphviz(titanic_tree, out_file=tree_data)
graph = pydotplus.graph_from_dot_data(tree_data.getvalue())
Image(graph.create_png())
```

```
In [ ]: # evaluate how good your decision tree is at predicting the correct result

In [ ]: # repeat with pclass and age

In [ ]: # is this any better?

In [ ]: # try with pclass and sibsp

In [ ]: # is this any better?

In [ ]: # what do you think are the best categories to use to create a decision tree?
        # pick 2 and create a decision tree an calculate csv scores

In [ ]: # is that better

In [ ]: # use this last decision tree to estimate the survival of
        # a person of your choosing

In [ ]: # what were their survival chances?
```