

```

1  # weka_implementation.py
2
3  from weka.classifiers import Classifier
4  from weka.core.converters import Loader
5  import weka.core.jvm as jvm
6  from weka.core.dataset import create_instances_from_lists
7  from weka.filters import Filter
8  from weka.core.classes import Random
9  from weka.classifiers import Classifier, Evaluation, PredictionOutput
10 import weka.plot.graph as graph
11 import graphviz
12 import time
13 import pandas as pd
14 import numpy as np
15 from graphviz import Digraph
16
17 # Louise Kilheeney - 16100463
18 # Taking in the data file location, and the train/test split proportion
19 # Build a weka C4.5 implementation using the Python Weka Wrapper API
20 def build_weka_tree(split):
21     jvm.start()
22
23     accuracy, time_to_build = build_weka(split)
24
25     jvm.stop()
26     return accuracy, time_to_build
27
28 # Louise Kilheeney - 16100463
29 def build_weka(split):
30     # Load the data file
31     loader = Loader(classname="weka.core.converters.CSVLoader")
32     train = loader.load_file('train_data_generated.csv')
33     test = loader.load_file('test_data_generated.csv')
34
35     # Store the target values for the test data
36     # so that the accuracy of the formula can be checked
37     test_target = pd.read_csv('test_data_generated.csv')['style'].values
38     # Get the dataset used to train the model,
39     # so that we can identify what the key values for the class are.
40     # As data is split randomly, we cannot assume it is in [ale, lager, stout] order
41     train_classes = pd.read_csv('train_data_generated.csv')['style'].values
42
43     # Set the class to be column 3 - the style column
44     train.class_index = 3
45
46     # Set the class to be column 3 - the style column
47     test.class_index = 3
48
49     # Store the time before starting to build the tree
50     starttime = time.time()
51
52     # initialise the time_to_run and accuracy to 0
53     time_to_run = 0
54     accuracy = 0
55
56     # Build and Train the weka tree
57     cls = Classifier(classname="weka.classifiers.trees.J48")
58
59     # Check that the data is valid
60     # If so, Build and Train the weka tree
61     if len(list(np.unique(train_classes))) != 1:
62         cls.build_classifier(train)
63         graph = cls.graph
64         # Store the time once the tree has been built
65         endtime = time.time()
66
67         # Create a list to store the predicted values in

```

```

68     pred = []
69     accurate = []
70     # Get the class labels in the order that they were allocated when training the
    model
71     classes = pd.Series(train_classes).drop_duplicates().tolist()
72
73     correct = 0
74     total = 0
75     # loop through test dataset, incrementing total every time
76     # and incrementing count if the predicted value was correct
77     for index, inst in enumerate(test):
78         total = total + 1
79         predicted = classes[int(cls.classify_instance(inst))]
80         pred.append(predicted)
81         act = test_target[index]
82         if predicted == test_target[index]:
83             correct = correct + 1
84             accurate.append(1)
85         else:
86             accurate.append(0)
87
88     # Get the accuracy of the weka implementation
89     accuracy = (correct/total)
90
91     # store the results in a csv file - the predicted class and the actual class
92     df = pd.DataFrame()
93     df['Actual'] = test_target
94     df['Predicted'] = pred
95     df['Accuracy'] = accurate
96     filename =
    "results/weka-results-"+str(round(split,2))+ "-" + str(round(accuracy,2)) + ".csv"
97     df.to_csv(filename, index=False, header=True)
98
99     time_to_run = endtime-starttime
100    # If the data is invalid, create a node to indicate failure
101    elif len(train) == 0:
102        graph = Digraph('python_tree_implementation')
103        graph.node(name='A', label="Fail", shape='box', style='filled')
104    else:
105        graph = Digraph('python_tree_implementation')
106        graph.node(name='A', label=train_classes[0], shape='box', style='filled')
107
108    # Render a png image of the weka tree to display in the PySimpleGUI popup
109    g = graphviz.Source(graph)
110    g.format = "png"
111    g.render('weka-test.gv', view=False)
112    return round(accuracy*100, 2), time_to_run

```