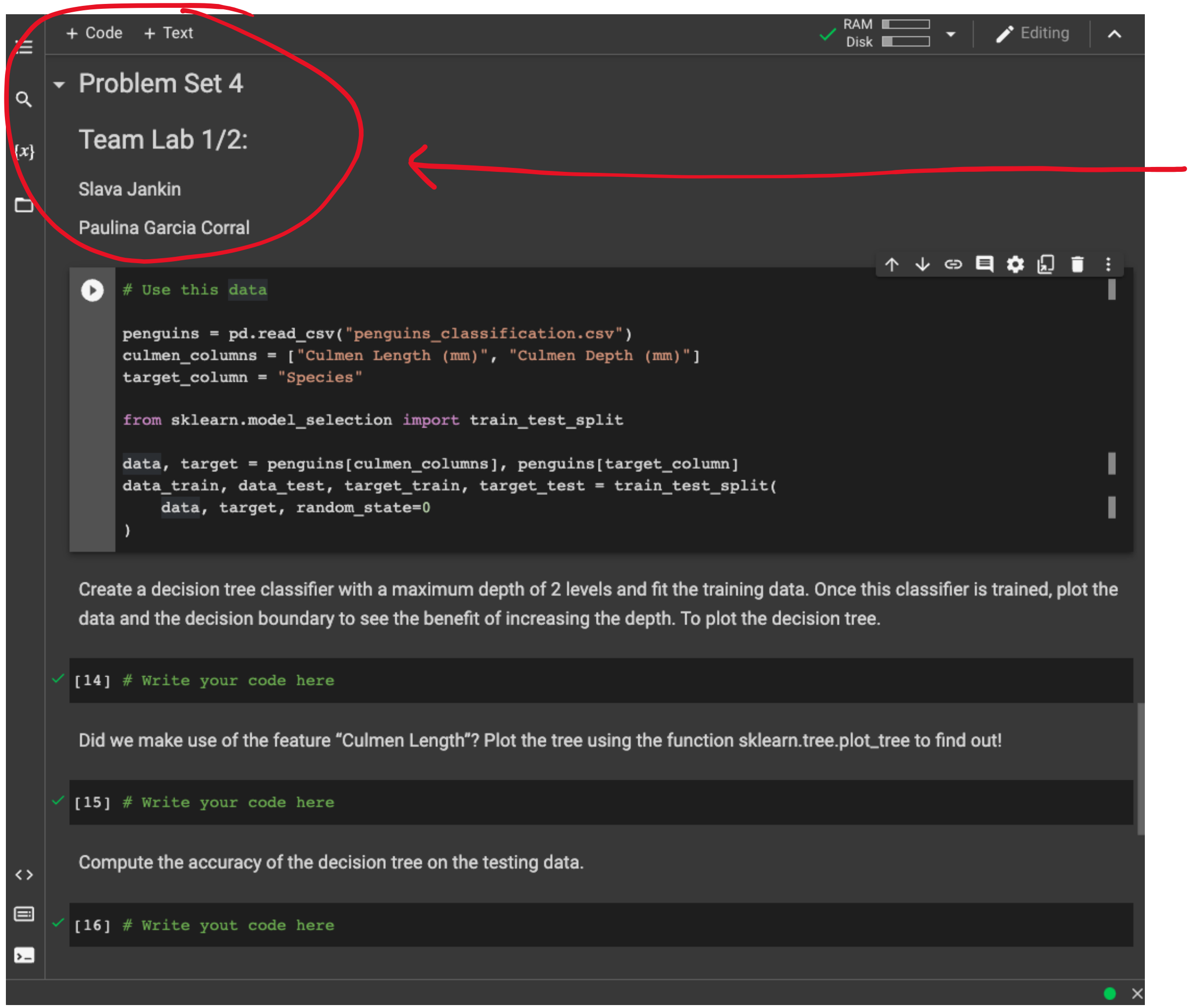


Problem Set 4

Team Lab 1/2:

Slava Jankin

Paulina Garcia Corral

```
# Use this data

penguins = pd.read_csv("penguins_classification.csv")
culmen_columns = ["Culmen Length (mm)", "Culmen Depth (mm)"]
target_column = "Species"

from sklearn.model_selection import train_test_split

data, target = penguins[culmen_columns], penguins[target_column]
data_train, data_test, target_train, target_test = train_test_split(
    data, target, random_state=0
)
```

Create a decision tree classifier with a maximum depth of 2 levels and fit the training data. Once this classifier is trained, plot the data and the decision boundary to see the benefit of increasing the depth. To plot the decision tree.

```
[14] # Write your code here
```

Did we make use of the feature “Culmen Length”? Plot the tree using the function sklearn.tree.plot_tree to find out!

```
[15] # Write your code here
```

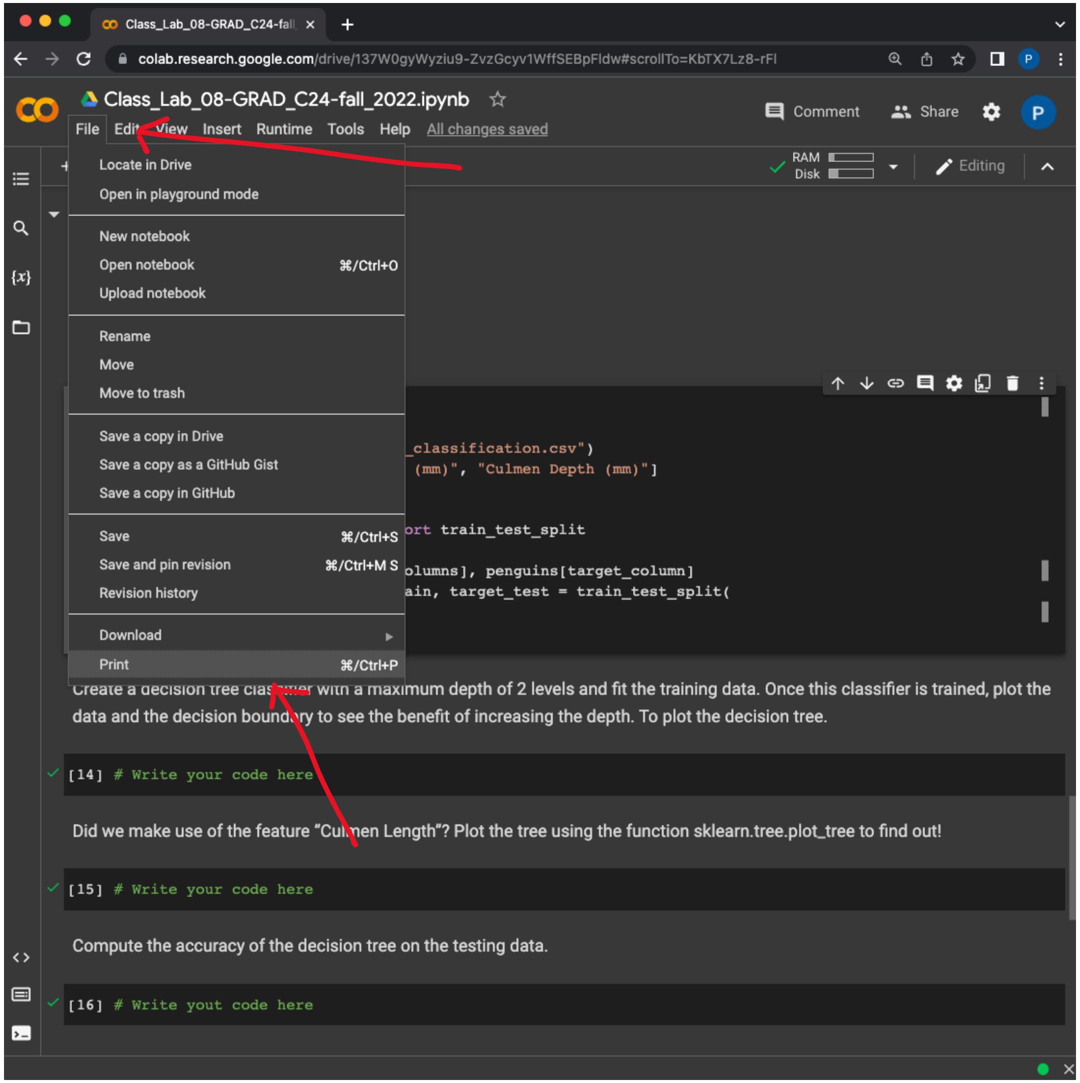
Compute the accuracy of the decision tree on the testing data.

```
[16] # Write yout code here
```

Create a new google collab file that only contains the code for the exercises.

Please add a text cell that includes:

- title “problem set 4”
- team lab 1 or 2
- name of the team members.


Class_Lab_08-GRAD_C24-fall
colab.research.google.com/drive/137W0gyWyziu9-ZvzGcyv1WffSEBpFldw#scrollTo=KbTX7Lz8-rFl
Class_Lab_08-GRAD_C24-fall_2022.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
P
RAM
Disk
Editing
Locate in Drive
Open in playground mode
New notebook
Open notebook ⌘/Ctrl+O
Upload notebook
Rename
Move
Move to trash
Save a copy in Drive
Save a copy as a GitHub Gist
Save a copy in GitHub
Save ⌘/Ctrl+S
Save and pin revision ⌘/Ctrl+M S
Revision history
Download
Print ⌘/Ctrl+P

```
_classification.csv")
(mm)", "Culmen Depth (mm)"]

ort train_test_split

olumns], penguins[target_column]
ain, target_test = train_test_split(
```

Create a decision tree classifier with a maximum depth of 2 levels and fit the training data. Once this classifier is trained, plot the data and the decision boundary to see the benefit of increasing the depth. To plot the decision tree.

```
[14] # Write your code here
```

Did we make use of the feature “Culmen Length”? Plot the tree using the function sklearn.tree.plot_tree to find out!

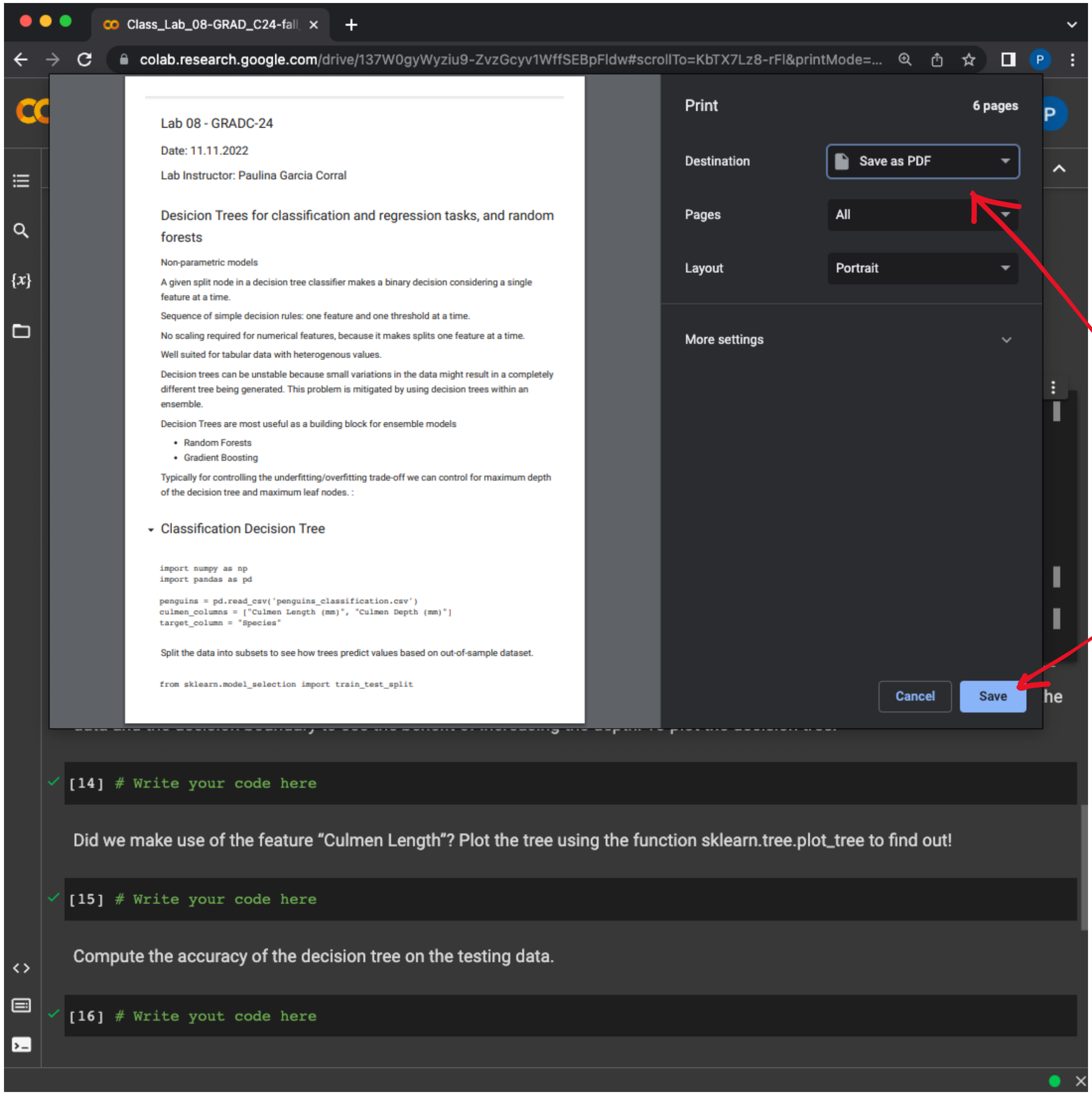
```
[15] # Write your code here
```

Compute the accuracy of the decision tree on the testing data.

```
[16] # Write yout code here
```

Submit the code as a PDF with all the plots visible.

Go to “file” and “print”



Class_Lab_08-GRAD_C24-fall
colab.research.google.com/drive/137W0gyWyziu9-ZvzGcyv1WffSEBpFldw#scrollTo=KbTX7Lz8-rFl&printMode=...

Lab 08 - GRADC-24

Date: 11.11.2022

Lab Instructor: Paulina Garcia Corral

Desicion Trees for classification and regression tasks, and random forests

Non-parametric models

A given split node in a decision tree classifier makes a binary decision considering a single feature at a time.

Sequence of simple decision rules: one feature and one threshold at a time.

No scaling required for numerical features, because it makes splits one feature at a time.

Well suited for tabular data with heterogenous values.

Decision trees can be unstable because small variations in the data might result in a completely different tree being generated. This problem is mitigated by using decision trees within an ensemble.

Decision Trees are most useful as a building block for ensemble models

- Random Forests
- Gradient Boosting

Typically for controlling the underfitting/overfitting trade-off we can control for maximum depth of the decision tree and maximum leaf nodes. :

Classification Decision Tree

```
import numpy as np
import pandas as pd

penguins = pd.read_csv('penguins_classification.csv')
culmen_columns = ["Culmen Length (mm)", "Culmen Depth (mm)"]
target_column = "Species"
```

Split the data into subsets to see how trees predict values based on out-of-sample dataset.

```
from sklearn.model_selection import train_test_split
```

Print 6 pages

Destination: Save as PDF

Pages: All

Layout: Portrait

```
[14] # Write your code here
```

Did we make use of the feature “Culmen Length”? Plot the tree using the function sklearn.tree.plot_tree to find out!

```
[15] # Write your code here
```

Compute the accuracy of the decision tree on the testing data.

```
[16] # Write yout code here
```

Select “Save as a PDF” for destination.

Then save.

Submit this file to Moodle, if and only if:

- only exercises code
- has Lab #
- all the team member names
- PDF file