User Manual

Name: Cian Manley

Student Number: 15563637

Supervisor: Mark Roantree

Project title: Classifier Analysis Application

1 TABLE OF CONTENTS

2 DESCRIPTION

This project is based around the field of data science and machine learning. The ideal scenario for the user is to understand what they are looking for when using this program. There are 4 different classifiers used in the program, Naïve Bayes, Decision Tree, K-Nearest Neighbours and Support Vector Machines. Each of these will produce a different result based on the inputted csv file.

3 INSTALLATION

To install the application, you will need to work through the command line terminal. It is important that you also have Python installed as this is the main language throughout the project.

First, you are going to have to download the project manually or through the terminal with the command below.

```
$ git clone https://gitlab.computing.dcu.ie/manleyc4/2019-ca400-manleyc4
```

Next you will have to navigate to the src folder in order to run the code.

```
$ cd 2019-ca400-manleyc4/src
```

Once in this directory you will have to make a virtual environment for the project in order for it to run all of its dependencies.

```
$ virtualenv flaskpython
```

Once the virtual environment has been setup, it should remain active. Now you install all the dependencies listed out in the requirements folder. This is a simple step as python uses pip in order to install these packages.
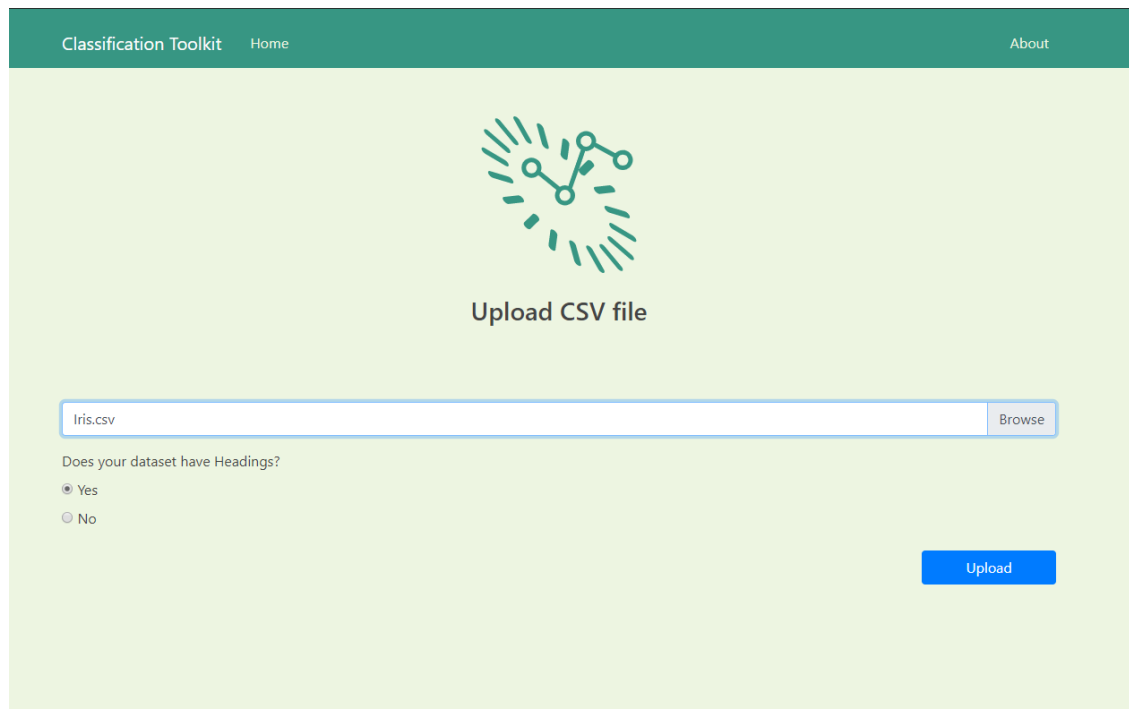
```
$ pip install -r requirements.txt
```

Once everything is installed last thing is to start up the database in order to be able to store the files. This is also straight forward as the commands are already built in.

Then we just use flask run and everything should spin up and start working.

4 USER GUIDE

4.1 UPLOAD CSV FILE

You will be greeted on the first page with an upload file page. All you have to do is click into the select file bar and you will get a pop up allow you to browse for your file. If you do not upload a file it will raise an error and ask you to put in a file.

You will also have the option to insert column names in the case that your data does not contain any headers. If you do not, headers with the labels "Column 1", "Column 2" and so on will be added to your data. This is to aid you in the selection of your attributes in the next stage of the processing.

4.2 SELECT WHAT PARAMETERS TO USE

This is where you are going to input all the different parameters you want to use are. First you will see two dropdown boxes containing the different column names in them. Use these to choose which columns you are looking to use and to select your target class.

There are different parameters for each of the classifiers. You can choose whether you want to use them or not. For decision trees you have the option of choosing to use either K-Nearest neighbours you will have to choose your K-Value, which distance method you want to use between Euclidean and Manhattan.

4.3 RESULTS

The results page shows of different outputs from the use of the program. After running it, you will see outputs specifying the:

- Classifier Used
- Chosen Columns
- Target Class
- Different Parameters
- Accuracy Rating
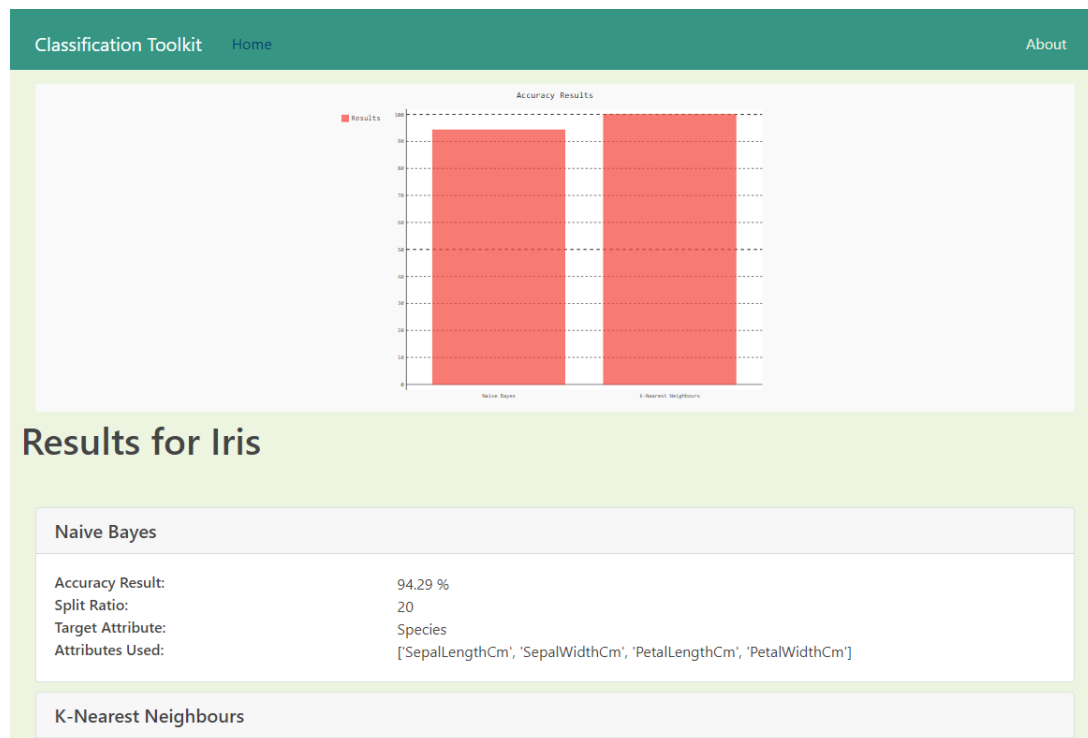
4.4   ABOUT PAGE

The about page is just something on the side giving some detail about the different algorithms used and how to use the overall program. It tells about the different parameters used for each of the algorithms and such.