



# **Testing Document**

Name: Cian Manley

Student Number: 15563637

Supervisor: Mark Roantree

Project title: Classifier Analysis Application

# 1 USER TESTING

---

User testing was a big part of this project. Throughout the year I got continuous feedback from other students in the course on what they would be looking for out of an application like this. Most of them gave feedback at the beginning of the year, which helped out my research. Once I had my first desktop shell working, I asked a few people to try and use it to see how they did. Unfortunately, this didn't work out well as the layout of the desktop application was not user friendly and people were poorly interacting with it.

The screenshot shows a desktop application titled "Classifications". It features a file browser with "Iris.csv" selected. Four classifiers are listed with checkboxes: Naive Bayes, Support Vector Machine, Decision Tree, and K-Nearest Neighbour. Below these is a "- Value" section with a dropdown menu showing "Categorical" and "Numerical". A "Run" button is at the bottom left. On the right, there are two columns: "Attributes to be used" and "Selected Attribute". The "Attributes to be used" column lists "Id", "SepalLengthCm", "SepalWidthCm", "PetalLengthCm", "PetalWidthCm", and "Species". The "Selected Attribute" column lists the same attributes, with "Species" highlighted. A checkbox "Give column headings" is at the top right. At the bottom, a text box displays the accuracy of each classifier: Naive Bayes (0.93), SVM (0.98), Decision Tree (0.96), and K-Nearest Neighbours (0.98).

| Attributes to be used | Selected Attribute |
|-----------------------|--------------------|
| Id                    | Id                 |
| SepalLengthCm         | SepalLengthCm      |
| SepalWidthCm          | SepalWidthCm       |
| PetalLengthCm         | PetalLengthCm      |
| PetalWidthCm          | PetalWidthCm       |
| Species               | Species            |

Accuracy of Naive Bayes classifier: 0.93  
Accuracy of SVM classifier: 0.98  
Accuracy of Decision Tree classifier: 0.96  
Accuracy of K-Nearest Neighbours classifier: 0.98

Once I had collected the feedback from this initial phase, I attempted to make the web application more user friendly. I also reached out to a fellow colleague in work, to get his advice on it and how it should be done. He said the same thing as everyone else when I showed the desktop application. It seemed like it was going in the wrong direction after everyone had the same thing to say. We had a discussion then on what I should be looking for and going for.

## 2 AD HOC TESTING

---

Throughout the development of this application, it was critical that I acted like an end user. I had to do this in order to see if I was able to break my application. At the beginning it was very easy as there was a number of defects and bugs present in the program. When development went on, it became difficult to break the program as I began to cover up most of the form fields with validators to ensure they were being filled out correctly.

## 3 ALGORITHM TESTING

---

In order to test the different algorithms, I used Jupyter Notebooks as both a learning suite and to make sure that I was receiving accurate results. At the beginning it was more conventional to program into these notebooks as it updates each code block one at a time and you are able to write an explanation in markdown to what is occurring.

## 4 STRESS TESTING

---

In order to get a good idea of the limitations that come with the project. I attempted to load large CSV files into the application. I first tried a 774mb CSV file. This wasn't going to work in the first place as it is too large to convert to a BLOB type and inputted into the SQLite database. So, I brought it down to a simpler 5mb. The file instantly loaded into the database, when I kept going through the program it gave me an error that I didn't think to even look for. The application was unable to turn a large string to a float.

```
File "c:\users\cman1\documents\college\2019-ca400-manleyc4\src\backend\env\lib\site-packages\sklearn\utils\validation.py", line 756, in check_X_y
    estimator=estimator)

File "c:\users\cman1\documents\college\2019-ca400-manleyc4\src\backend\env\lib\site-packages\sklearn\utils\validation.py", line 567, in check_array
    array = array.astype(np.float64)

ValueError: could not convert string to float: 'San Diego Padres Manager Bud Black ejected by HP umpire Jordan Baker. '
```

This led me to believe that this an end-user issue as once I looked at the dataset, it was not pre-processed and was missing multiple amounts of data.