**Survey of student’s favourite sport**

**Summary:**

We conducted a survey to know what is the favourite sport of most of the students in the campus.

So, by collecting some sample data we finalized 3 games and a sport activity. We decided to collect more

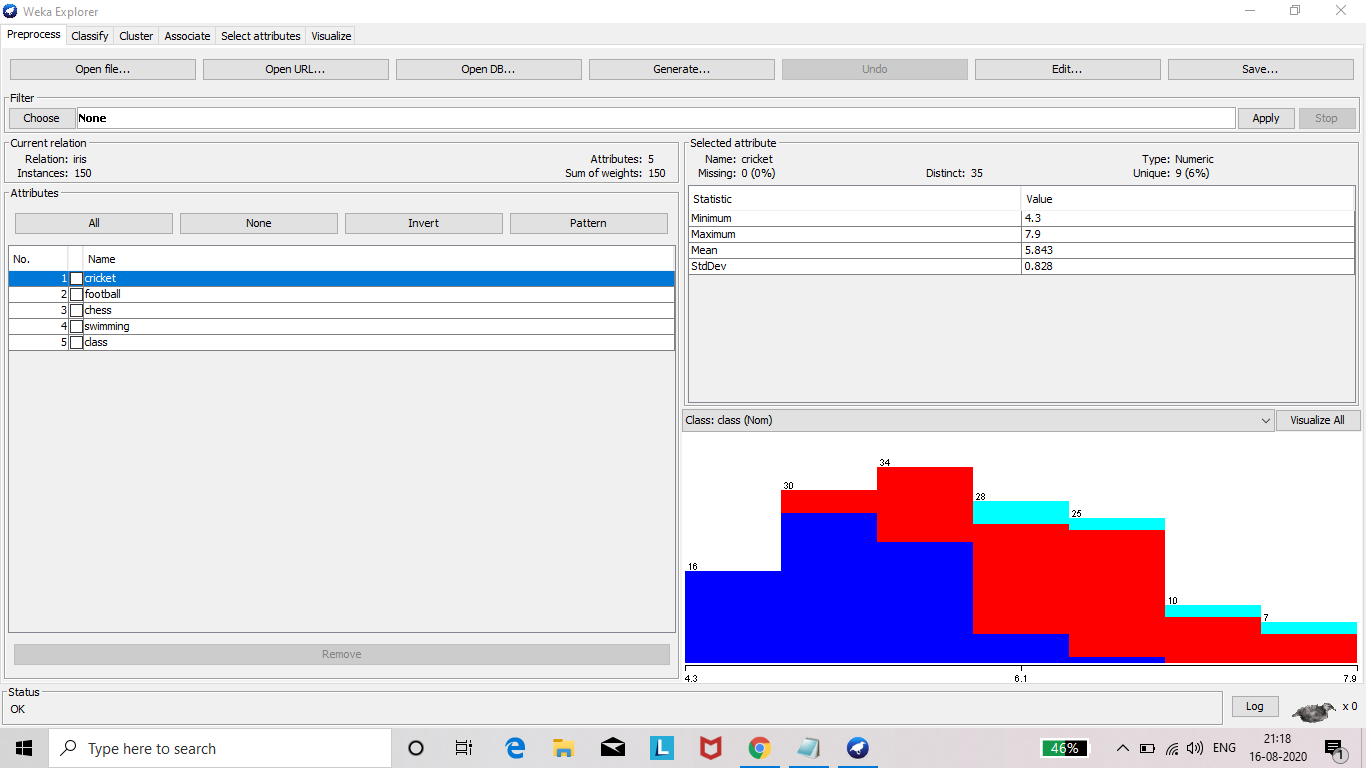
Data on these games that are cricket, football, chess and swimming.

A google form was circulated so that the dataset would be more diverse. Around 150 students registered their opinions on each game by giving them a star rating.

**Testing:**

**Weka** tool was used to test the dataset. The acquired dataset was a CSV file. So, it was converted into **”.arff ”** file using Weka. Then two classification functions were used to get the results.

Those two functions are **ZeroR** and **Randomforest** classification methods. Randomforest gave the highest accuracy among all.



**ZeroR and Randomforest:**

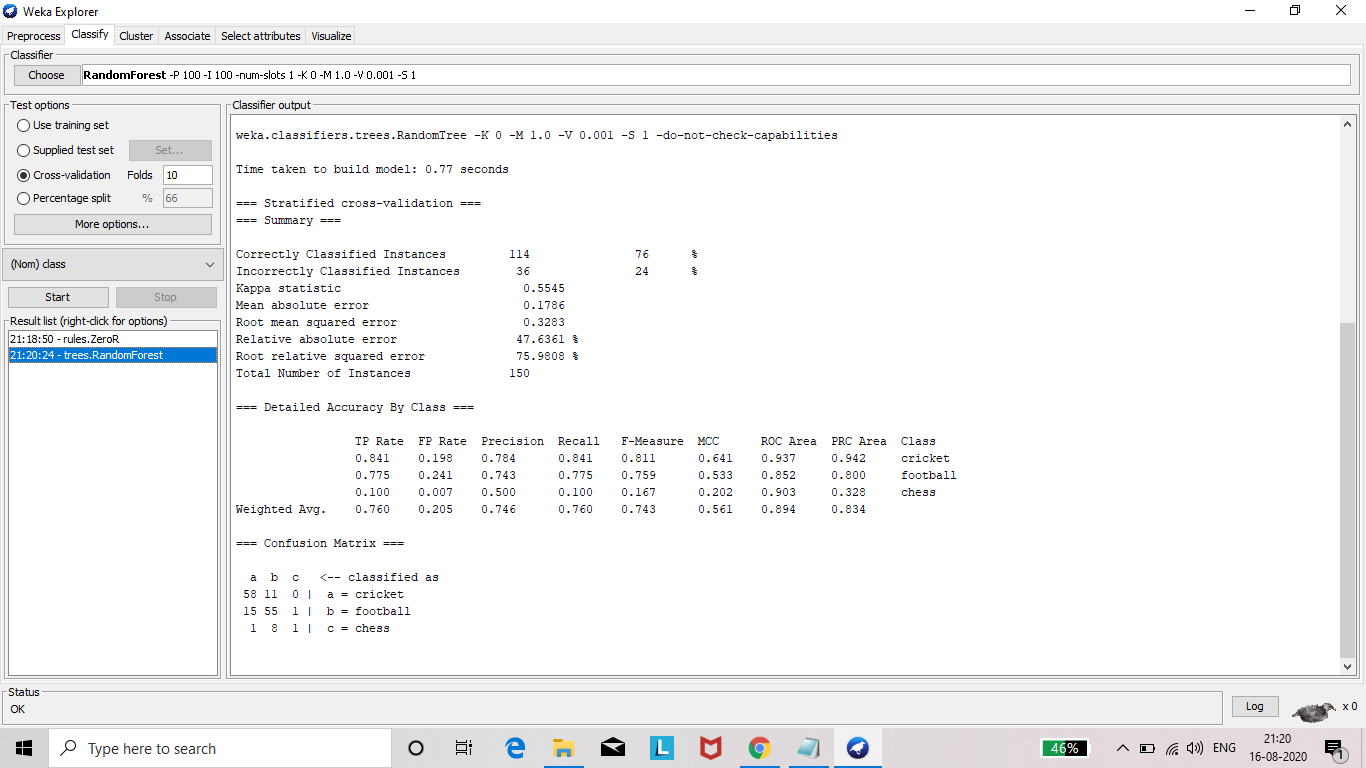
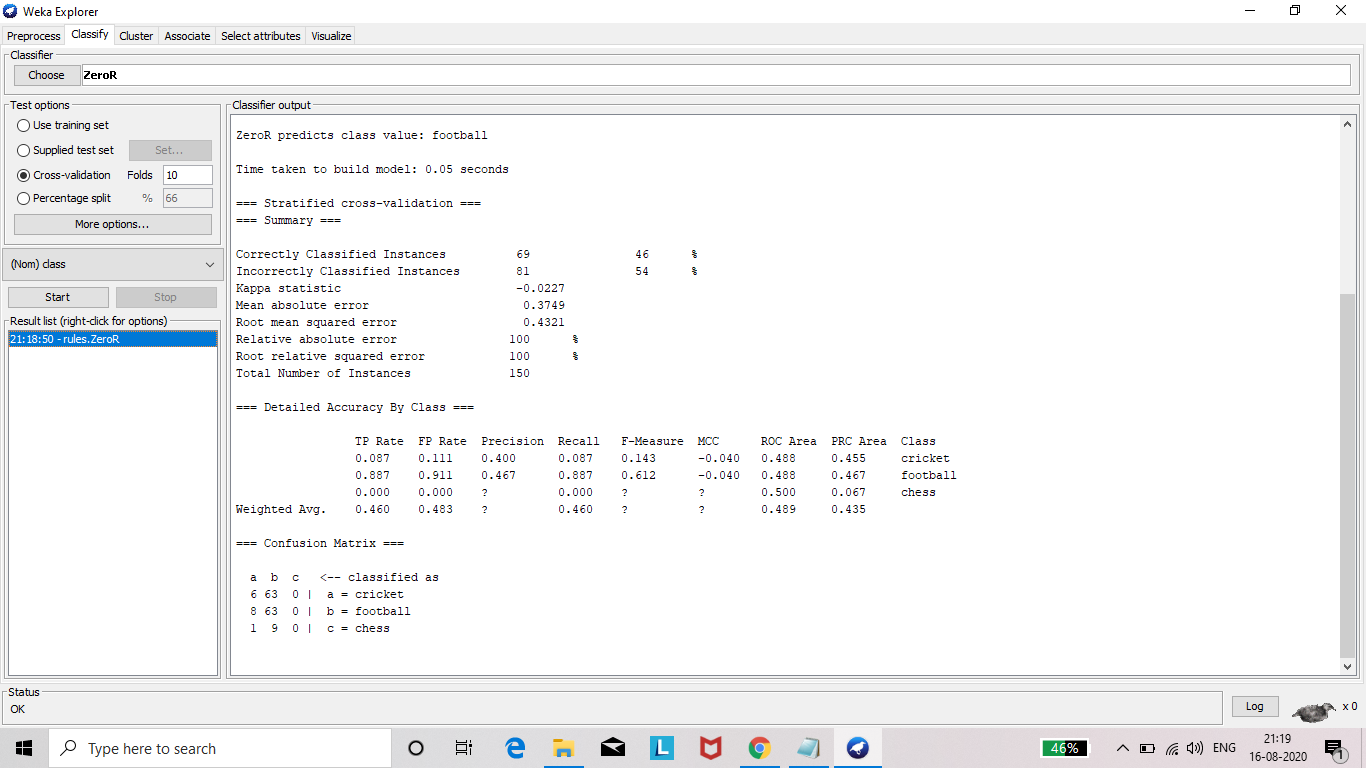
**ZeroR:**

**ZeroR** is the simplest classification method which relies on the target and ignores all predictors. **ZeroR classifier** simply predicts the majority category (class). Although there is no predictability

power in **ZeroR**, it is useful for determining a baseline performance as a benchmark for other classification methods.

**Randomforest:**

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.



**Result:**

In these two classification functions Randomforest’s accuracy is 74% and ZeroR’s accuracy is 66%.

Submitted by,

Gowthamraj A – 2017103014

Sanjeev K - 2017103036