Cedric Kim
Softdes Spring 2016

League of Legends Tribunal

Project Overview:

As a League of Legends enthusiast, I wanted to something related to League of Legends. After a while, I stumbled upon a particular website. This website had the chat files of all cases where the user became banned due to different reasons. I wished to analyze the different user's verbal language in order to determine which user was banned. I emailed the person in charge of the person, and was able to receive 10,000 files to analyze.

Implementation:

In order to analyze these files, I had to look at the text from different angles. Each case consisted of different characteristics. Each one had multiple messages the different users sent that I had to sort. The majority of the functions used were to sort the jumbled text files into neat readable and analyzable lists according to the user.

The one major algorithm I used was to create a value for each negative word (noob, idiot, bad, dumb) as well as profane language. The more profane the language used, the higher 'toxicity' value the user got. The the total value was taken as a percentage of toxicity that occurred that game. The algorithm takes the person who was most toxic, and brands them as the offender (person who was banned). Then it checks if the person was actually banned.

One decision I had to make was how to value the different words used in the person's chat. I generally had to use my intuition in order to rate these words. We can see how this affected the results in the end.

Results:

My program takes any case number you present, and returns the highest toxic percentage, as well as the champion the user used. Then the program shows if the prediction was correct.

The program can also show the toxicity levels of the other champions, as well as display the chat log via website.

Another part of the program was showing how many predictions were correct. The program had an average accuracy of around 36% This accuracy number will change depending on your word/toxicity value. If we were to look at more chat logs, there could be words and values that we can add to gather a more accurate representation.

Reflection:

One thing I did well was the implementation of the program. I was able to achieve a 36% accuracy (depending on how you look at this number, it could be good or bad). Remember that verbal language is not the only way to get banned (you could be banned for being afk, intentionally dying, etc. which could not be analyzed with the team's chat log). One thing I could have done was also process the user's stats (kills, deaths, and assists) which would be interesting to look at if I had additional time. I also wanted to use the league tribunal API to get the data, but it was giving some trouble, so I used the raw data that was sent to me. I believe that this was a good project to push me forwards. I learned a lot about text analysis (cleaning up text) and in addition created a lot of comments so I wouldnt get confused looking back at the code.