

H2 – Serializing data using WEKA

Introduction

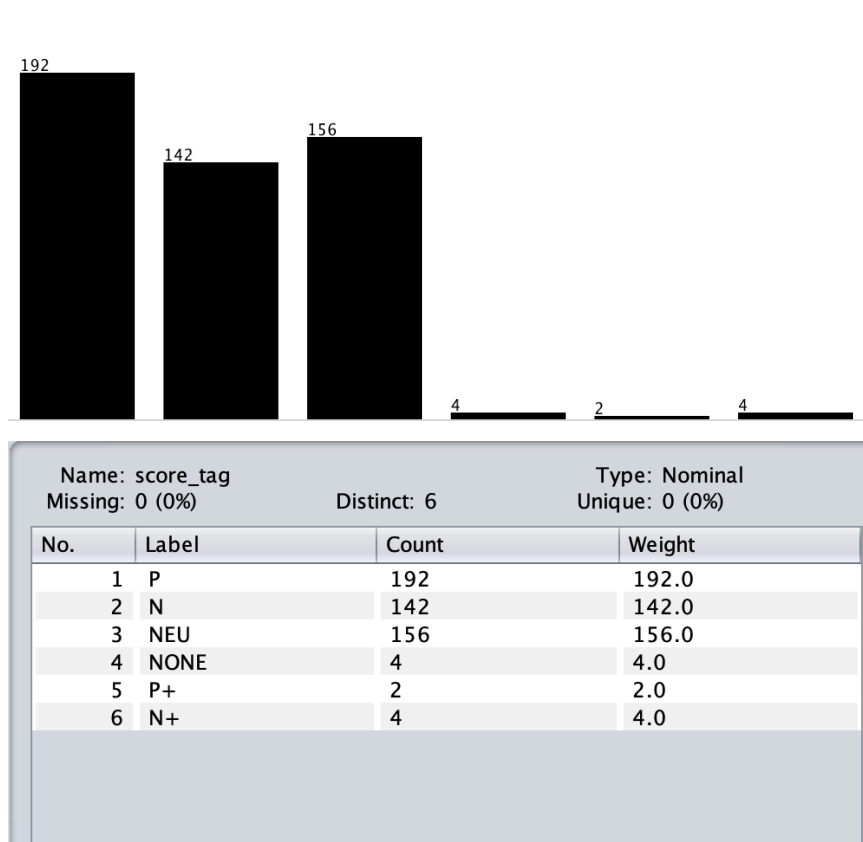
Weka is an open source framework created by the University of Waikato that is used for Data Mining and Data Analysis.

Development

First of all, the Training.csv file obtained from Homework 1 Had to be converted to a Weka File (.artff). Then Data was analyzed involving the different twelve features of the Data Set.

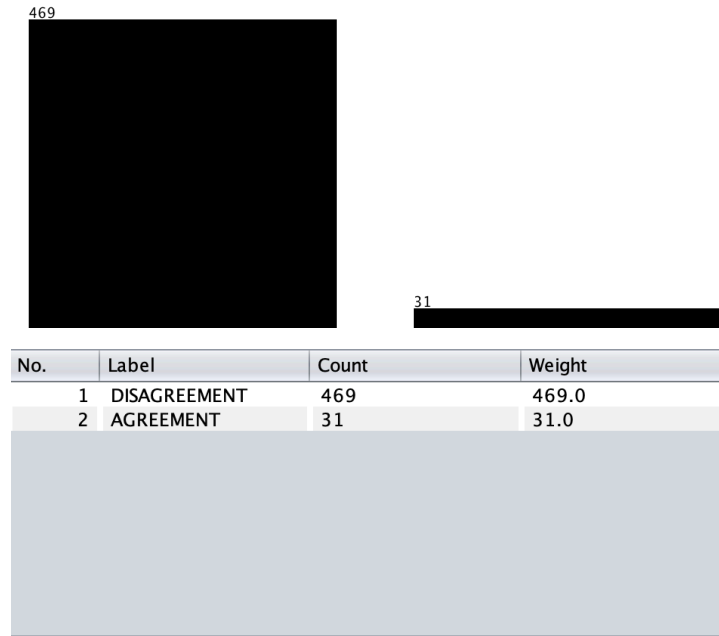
Score_tag

In score_tag, it is appreciated that P has the higher score among all the possible options, but, the gap between P and N & NEU is not as big compared with NONE, P+ and N+, which have extremely low values.



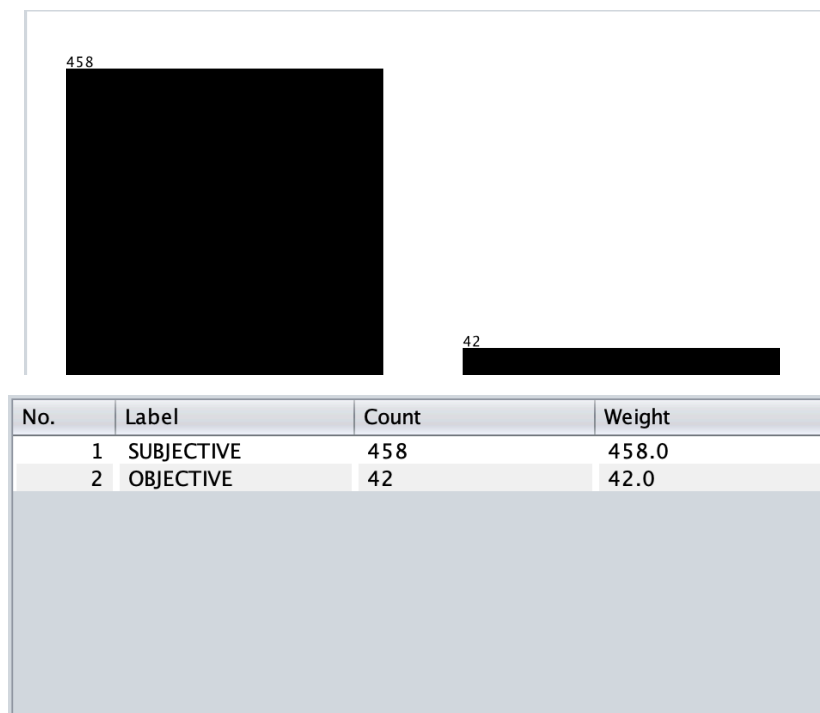
Agreement

According to the following histogram, it is appreciated that disagreement outnumbered agreement, therefore it can be concluded that the majority of the fake news disagree.



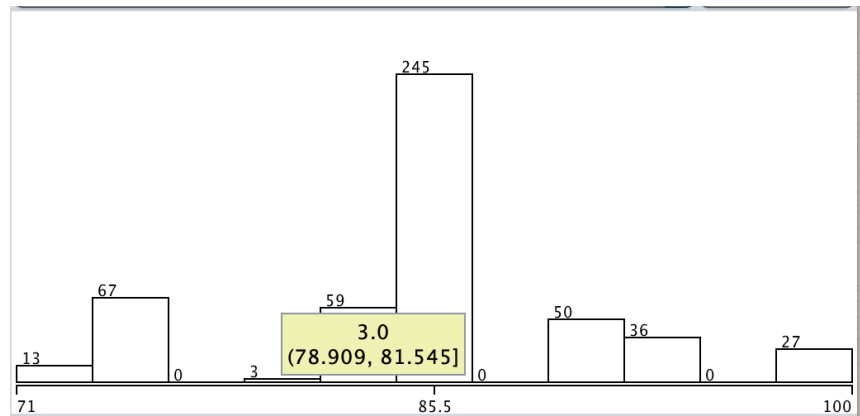
Subjectivity

In this case, subjective is greater than objective, hence, it can be concluded that fake news is more likely to be subjective.



Confidence

Confidence looks to be neutral in this chart, but in fact it is a little bit closer to the right, so it is concluded that fake news tend to look confident.



Statistic	Value
Minimum	71
Maximum	100
Mean	85.81
StdDev	6.168

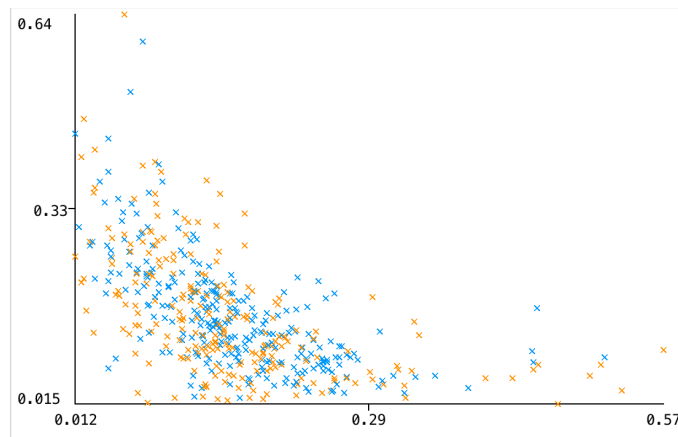
Irony

The NONIRONIC column is almost 5x bigger than the IRONIC column, so it is concluded that fake news are not usually ironic.



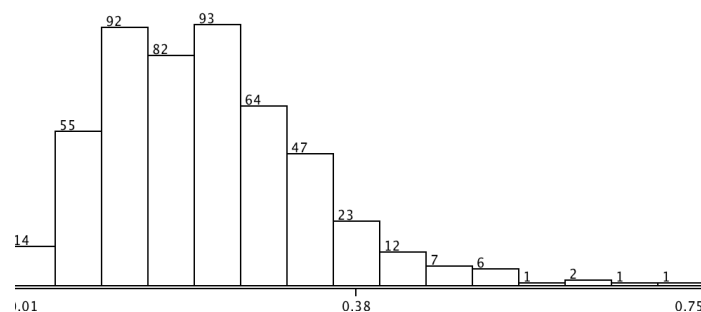
Happy

In this plot, happy is plotted against angry. It can be observed that there exists a certain negative correlation between both features.



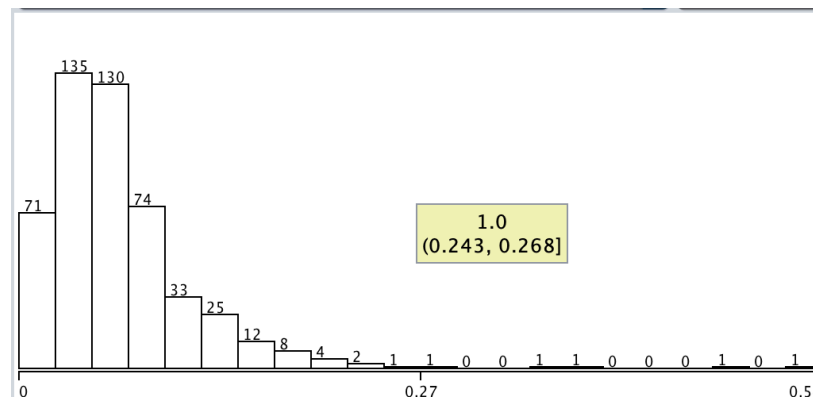
Angry

In this plot, it can be observed that the plot has left skewness, indicating that in general, most of the fake news do not present angeriness.



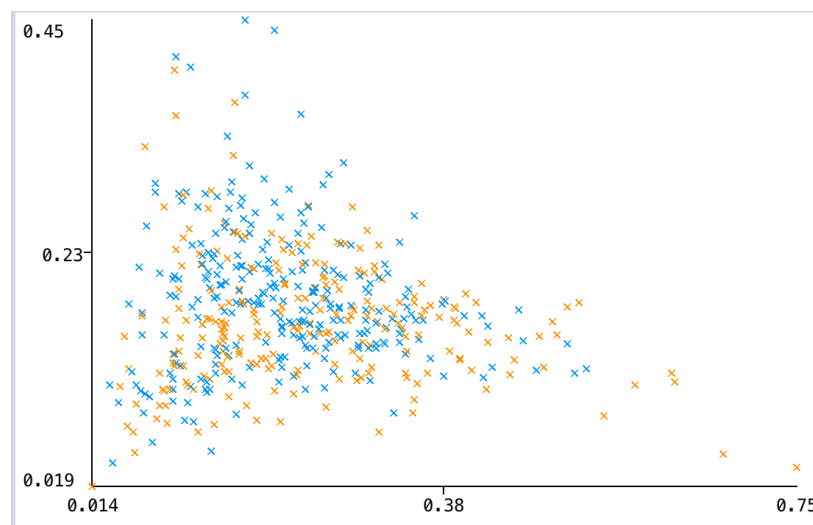
Bored

In this plot, it can be observed that the plot has a lot of skewness to the left, so there are not many bored emotions in fake news.



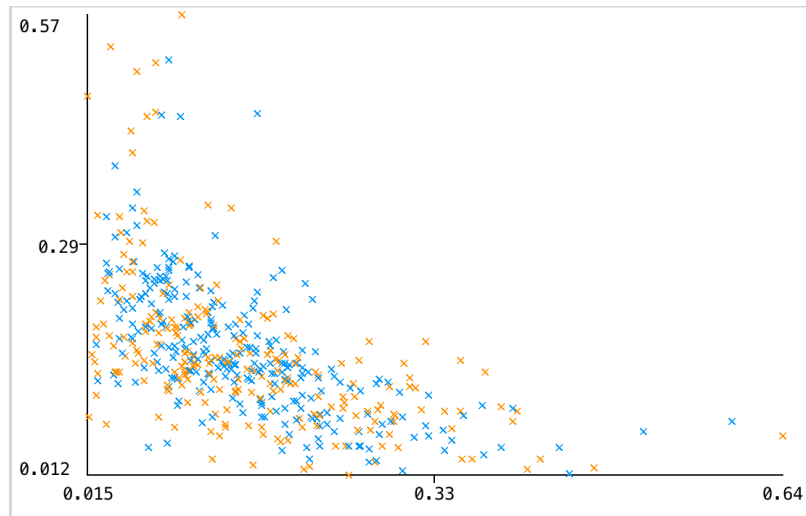
Fear

In this plot, it can be observed that there is no correlation between fear and angry.



Sad

In this plot, it can be observed that it exists a certain negative correlation between sad and happy in fake news.



Conclusion

Data Analysis and Statistics are powerful tools for interpreting data and help us to make better decisions based on numbers and not on intuition. Weka resulted to be an excellent tool for making these tasks due to all the tools that it contains.