**Association rules exploitation**

The dataset used for this part is the same used in the data preparation and understanding part, also used for clustering. After applying the algorithm for all these attributes, the one that came out to be the most interesting is the *emotion*.

**“sex”**



Figure 1: classification report of "sex" attribute by exploiting association rules

Using all rules with a degree of confidence greater than 80%, i.e. five for F and four for M, we obtain a prediction model with 77% accuracy, which is very close to the minimum confidence threshold.

In general we have a worsening of the values compared to the forecast with Naive Bayes, except for the precision of the M class, whiche increases from 0.91 to 0.94. The value that decreased the most is the recall of M, from 0.96 to 0.68. Although the overall accuracy has dropped from 0.93 to 0.77, this model remains the most reliable of the four that we calculated by exploiting association rules.

**“emotional\_intensity”**

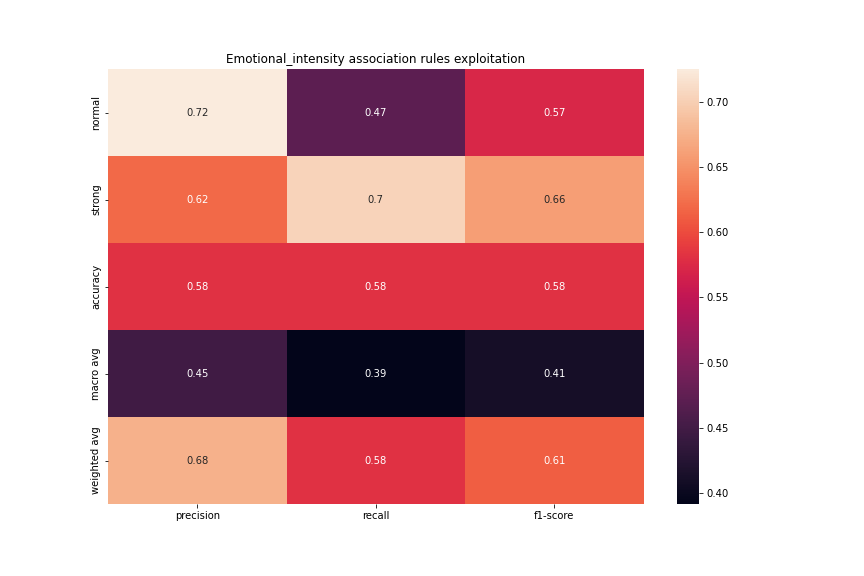


Figure 2: classification report of "emotional\_intensity" attribute by exploiting association rules

Taking the four most recurring rules for each value of emotional\_intensity we obtain a prediction model with 58% accuracy. This value is very close to the confidence threshold, which is 60%, but it's not very high, considering the number of rules selected (four each).

As seen for the Sex attribute, here too we have a general decrease in values, except for the recall of the strong class, which increases from 0.61 to 0.70, and for the precision of the normal class, which increases slightly from 0.71 to 0.72.

**“emotion\_positivity”**



Figure 3: classification report of "emotion\_positivity" attribute by exploiting association rules

Using two rules for each possible value of emotion\_positivity, for a total of six rules, we obtain a model with an accuracy of 47%, a very low value but influenced by the minimum confidence threshold of 30% of 1emp\_ (positive emotions). Indeed, the other two classes have a higher confidence threshold, 40% and 60%.

This is the least reliable model among those calculated with association rules: all values decrease with respect to the classification made with Naive Bayes.