



Tecnológico
de Monterrey

Luis Daniel Medina Cazarez

A01651070

Homework 4

09/26/19

Intelligent Systems
TC-2011

PhD Octavio Loyola

Process

For this homework I used 3 different algorithms based on pattern to predict comments' rating. The algorithms used in this homework were:

- J48
- REPTree
- Random Forest

For training, I used a dataset of 2000 comments extracted from Kaggle and for testing I used a dataset of 500 different comments, extracted from the same page.

For each comment I extracted 10 different features (score_tag, agreement, subjectivity, confidence, irony, anger, joy, fear, sadness and surprise) to get those features I used Indico's Emotion library and meaningcloud's sentiment analysis library.

Result

- **J48**
 - With J48 I got an accuracy of 19.6%. This one is the lowest of the 3 algorithms. I don't know why is this so low and why it's the lowest one, but if I get more features, this might improv.

resultJ48Cleaned.csv

7 hours ago by Daniel

J48

0.19600

Weka Explorer

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose

J48 -C 0.25 -M 2

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

Set...

Folds 10

% 66

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

16:53:29 - misc.InputMappedClassifier

16:54:14 - misc.InputMappedClassifier

16:54:30 - misc.InputMappedClassifier

Classifier output

Relation: training - Copy

Instances: 1994

Attributes: 11

score_tag

agreement

subjectivity

confidence

irony

anger

joy

fear

sadness

surprise

class

Test mode: user supplied test set: size unknown (reading incrementally)

=== Predictions on test set ===

inst#,actual,predicted,error,prediction

1,1:?,1:1,,0.5

2,1:?,3:3,,1

3,1:?,4:4,,0.833

4,1:?,4:4,,1

5,1:?,5:5,,1

6,1:?,5:5,,0.833

7,1:?,1:1,,0.933

8,1:?,2:2,,1

Status

OK

Log

x 0

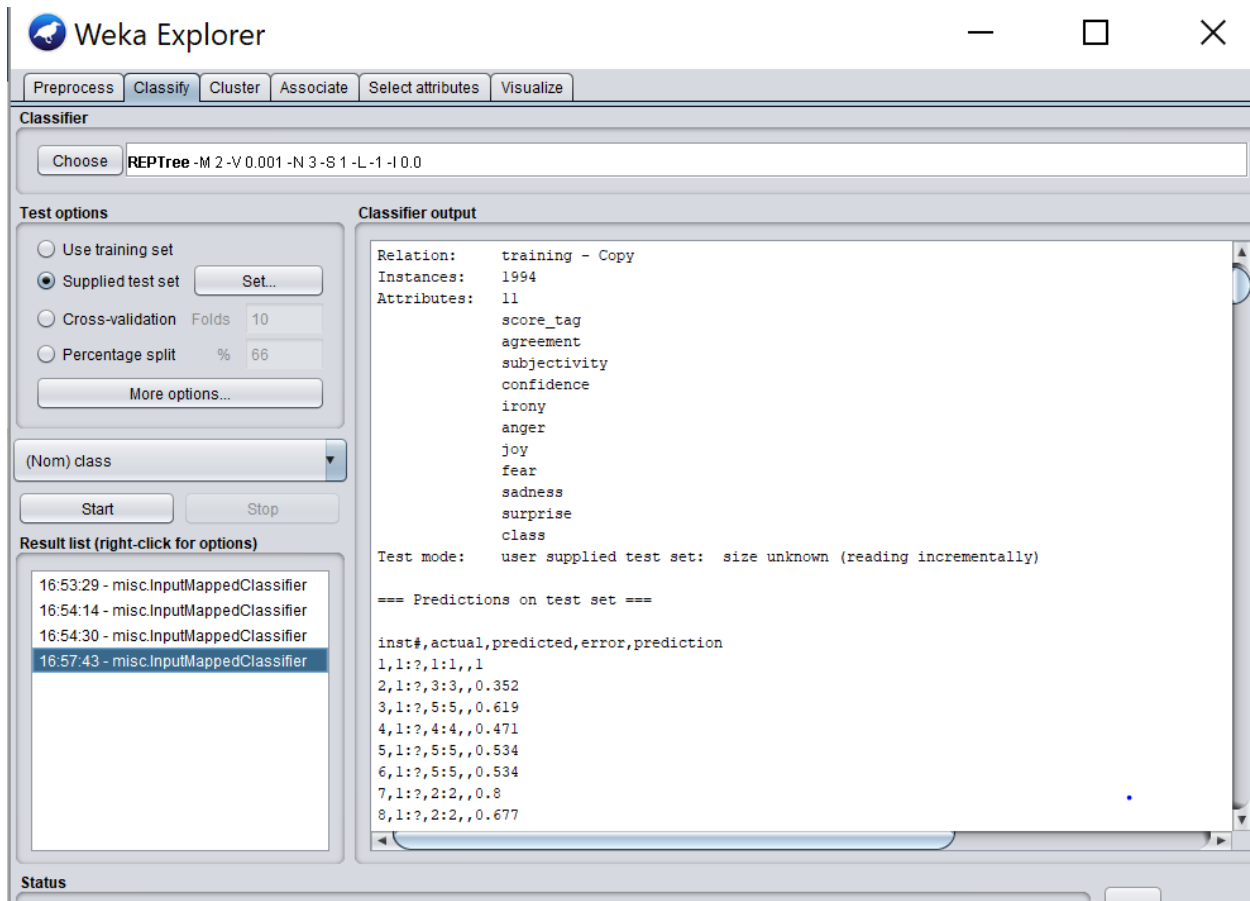
- **REPTree**

- With REPTree I got the same accuracy than Random forest although the predictions weren't the same.

resultREPTreeCleaned.csv
7 hours ago by Daniel
REPTree

0.22400





- **Radom forest**

- With Random I got the same accuracy than REPTree forest although the predictions weren't the same.

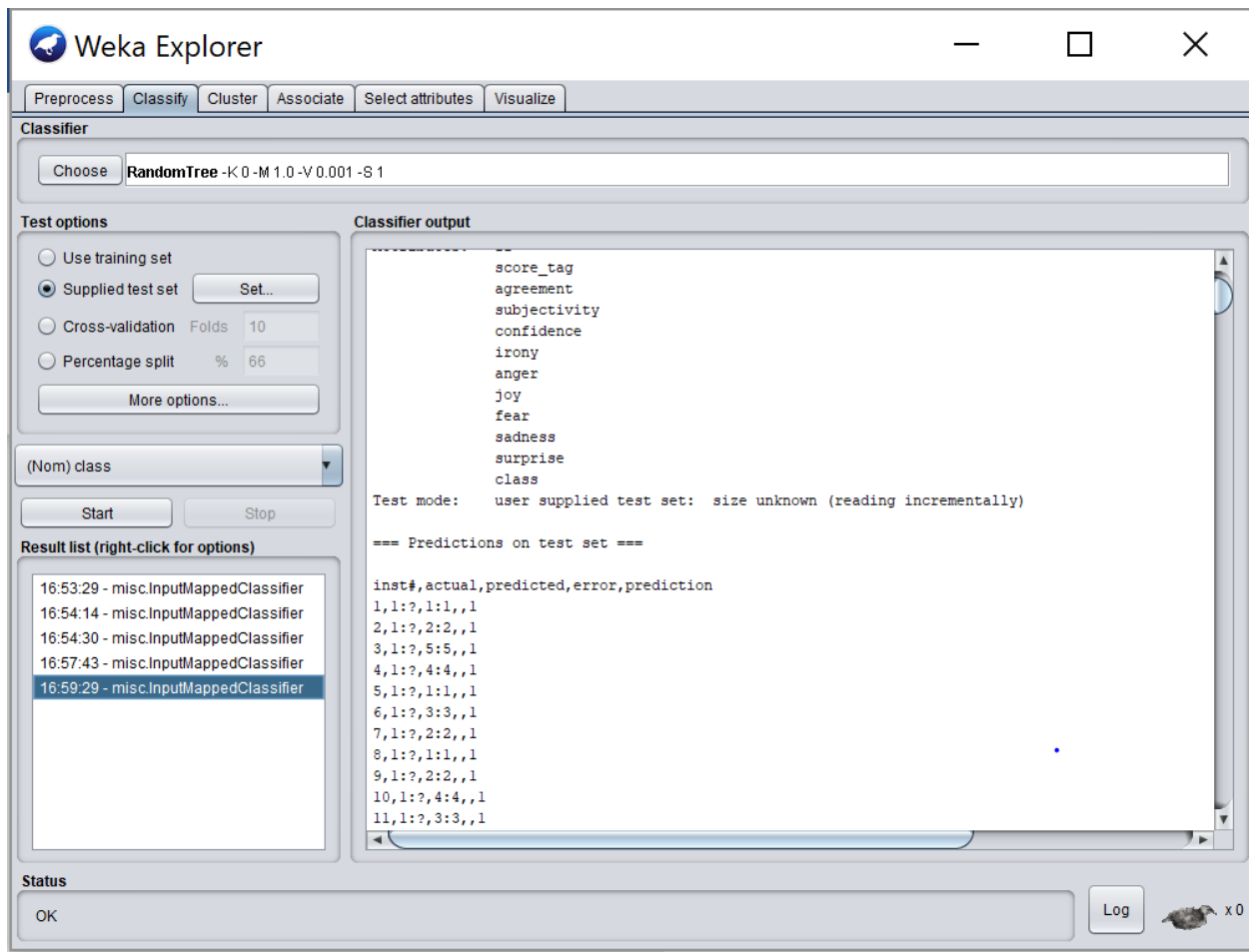
[resultRandomForestCleaned.csv](#)

7 hours ago by [Daniel](#)

Random Forest

0.22400





Conclusion

I got different results, but, none of them was so much better than making random prediction, it is necessary get different features or extract more. I need to analyze the output of the different algorithms, and the features in the training dataset to see if I can find a significant features or if there are any feature that is being noisy.