



INF1771 – T3

INTELIGÊNCIA ARTIFICIAL

MARCELO PAULON, RODRIGO SILVA, GABRIEL MEDEIROS, RENAN DA FONTE

MODELO 1 - WEKA

- Pré-processamento (Weka):
 - Unificação da part1 e part2 do dataset
 - TextDirectoryLoader para importar .txt's
 - StringToWordVector
 - Lista de 630 stop-words em inglês
 - TF-IDF
 - Snowball Stemmers Library
 - AttributeSelection – 59 atributos
 - Remoção manual de atributos – 52 atributos



Weka

ANNOYING **SCRIPT** WORSE FUNNIEST EMOTIONS BUNCH
BAD WORST EXCELLENT
RIDICULOUS FANTASTIC CRAP HORRIBLE
MINUTES AMAZING PAINFUL UNUSUAL BOTHER AMATEURISH
POINTLESS PATHETIC WASTED
TERRIBLE UNDERRATED ENJOYED BORING AWFUL PREDICTABLE LAME
OUTSTANDING POORLY WASTE **GREAT** DULL UNFUNNY FAILS STUPID LAUGHABLE MESS
LOVED TODAY BEAUTIFULLY
RECENTLY **LOVE**
SUPPOSED BEAUTIFUL REASON
POOR MONEY

MODELO 1

- Classificação:
- IBk (k-nearest neighbor) – cross validation:10 folds
- Correctly Classified Instances 38462 76.924 %
- Incorrectly Classified Instances 11538 23.076 %
- Kappa statistic 0.5385
- Mean absolute error 0.3014
- Root mean squared error 0.4122
- Relative absolute error 60.2731 %
- Root relative squared error 82.439 %
- Total Number of Instances 50000

MODELO 1

- Classificação:
- J48 (C45) – cross validation: 10 folds
- Correctly Classified Instances 38500 77 %
- Incorrectly Classified Instances 11500 23 %
- Kappa statistic 0.54
- Mean absolute error 0.3187
- Root mean squared error 0.4059
- Relative absolute error 63.7386 %
- Root relative squared error 81.1795 %
- Total Number of Instances 50000

MODELO 1

- Classificação:
- LibSVM - cross validation:10 folds
- Correctly Classified Instances 39458 78.916 %
- Incorrectly Classified Instances 10542 21.084 %
- Kappa statistic 0.5783
- Mean absolute error 0.2108
- Root mean squared error 0.4592
- Relative absolute error 42.168 %
- Root relative squared error 91.8346 %
- Total Number of Instances 50000

MODELO 1

- Classificação:
- SMO - cross validation:10 folds
- Correctly Classified Instances 39480 78.96 %
- Incorrectly Classified Instances 10520 21.04 %
- Kappa statistic 0.5792
- Mean absolute error 0.2104
- Root mean squared error 0.4587
- Relative absolute error 42.08 %
- Root relative squared error 91.7388 %
- Total Number of Instances 50000

MODELO 1

- Classificação:
- MultiLayerPerceptron (Rede Neural) - cross validation:10 folds
- Correctly Classified Instances 38804 77.608 %
- Incorrectly Classified Instances 11196 22.392 %
- Kappa statistic 0.5522
- Mean absolute error 0.2732
- Root mean squared error 0.3996
- Relative absolute error 54.6316 %
- Root relative squared error 79.9156 %
- Total Number of Instances 50000

MODELO 2 – WEKA + JAVA

- Pré-processamento (Java):
 - Unificação da part1 e part2 do dataset
 - Importação dos dados – Apache Commons IO
 - StrikeAMatch para corrigir erros de grafia



MODELO 2

- Classificação:

OBRIGADO