Sports articles for objectivity analysis Data Set

Download: Data Folder, Data Set Description

Abstract: 1000 sports articles were labeled using Amazon Mechanical Turk as objective or subjective. The raw texts, extracted features, and the URLs from which the articles were retrieved are provided.

Data Set Characteristics:	Multivariate, Text	**Number of Instances:**	1000	**Area:**	Social
Attribute Characteristics:	Integer	**Number of Attributes:**	59	**Date Donated**	2018-04-09
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	3518

Source:

Yara Rizk, American University of Beirut (yar01 '@' aub.edu.lb)
Mariette Awad, American University of Beirut (mariette.awad '@' aub.edu.lb)

Data Set Information:

Some of the features are retrieved using the Stanford POS tagger and the tags are as defined in Penn Treebank Project: [Web Link]

Attribute Information:

TextID text file name
URL link to article
Label objective vs. subjective
totalWordsCount total number of words in the article
semanticobjscore Frequency of words with an objective SENTIWORDNET score
semanticsubjscore Frequency of words with a subjective SENTIWORDNET score
CC Frequency of coordinating conjunctions
CD Frequency of numerals and cardinals
DT Frequency of determiners
EX Frequency of existential there
FW Frequency of foreign words
INs Frequency of subordinating preposition or conjunction
JJ Frequency of ordinal adjectives or numerals
JJR Frequency of comparative adjectives
JJS Frequency of superlative adjectives
LS Frequency of list item markers
MD Frequency of modal auxiliaries

NN Frequency of singular common nouns
NNP Frequency of singular proper nouns
NNPS Frequency of plural proper nouns
NNS Frequency of plural common nouns
PDT Frequency of pre-determiners
POS Frequency of genitive markers
PRP Frequency of personal pronouns
PRP$ Frequency of possessive pronouns
RB Frequency of adverbs
RBR Frequency of comparative adverbs
RBS Frequency of superlative adverbs
RP Frequency of particles
SYM Frequency of symbols
TOs Frequency of 'to' as preposition or infinitive marker
UH Frequency of interjections
VB Frequency of base form verbs
VBD Frequency of past tense verbs
VBG Frequency of present participle or gerund verbs
VBN Frequency of past participle verbs
VBP Frequency of present tense verbs with plural 3rd person subjects
VBZ Frequency of present tense verbs with singular 3rd person subjects
WDT Frequency of WH-determiners
WP Frequency of WH-pronouns
WP$ Frequency of possessive WH-pronouns
WRB Frequency of WH-adverbs
baseform Frequency of infinitive verbs (base form verbs preceded by â€œtoâ€)
Quotes Frequency of quotation pairs in the entire article
questionmarks Frequency of questions marks in the entire article
exclamationmarks Frequency of exclamation marks in the entire article
fullstops Frequency of full stops
commas Frequency of commas
semicolon Frequency of semicolons
colon Frequency of colons
ellipsis Frequency of ellipsis
pronouns1st Frequency of first person pronouns (personal and possessive)
pronouns2nd Frequency of second person pronouns (personal and possessive)
pronouns3rd Frequency of third person pronouns (personal and possessive)
compsupadjadv Frequency of comparative and superlative adjectives and adverbs
past Frequency of past tense verbs with 1st and 2nd person pronouns
imperative Frequency of imperative verbs
present3rd Frequency of present tense verbs with 3rd person pronouns
present1st2nd Frequency of present tense verbs with 1st and 2nd person pronouns
sentence1st First sentence class
sentencelast Last sentence class
txtcomplexity Text complexity score

Relevant Papers:

Nadine Hajj, Yara Rizk, and Mariette Awad, 'A Subjectivity Classification Framework for Sports Articles using Cortical Algorithms for Feature Selection,' Springer Neural Computing and Applications, 2018.
Yara Rizk, and Mariette Awad, 'Syntactic Genetic Algorithm for a Subjectivity Analysis of Sports Articles,' International Conference on Cybernetic Intelligent Systems, Limerick, Ireland, 2012.

Citation Request:

Nadine Hajj, Yara Rizk, and Mariette Awad, 'A Subjectivity Classification Framework for Sports Articles using Cortical Algorithms for Feature Selection,' Springer Neural Computing and Applications, 2018.