

Esercizio NewsClassification

lunedì 22 maggio 2023 09:38

- Era presente un errore che non permetteva di fare l'inferenza corretta nel metodo `.Append()`.

Risultati Algoritmi:

- PREDIZIONI con Headline e Short_Description:
 - SdcaMaximumEntropy:
 - MicroAccuracy: 0,361
 - MacroAccuracy: 0,274
 - LogLoss: 4,677
 - LogLossReduction: -,426
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: "MONEY"
 - SdcaNonCalibrated:
 - MicroAccuracy: 0,361
 - MacroAccuracy: 0,274
 - LogLoss: 7,296
 - LogLossReduction: -1,224
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: "MONEY"
 - LbfgsMaximumEntropy:
(Ha richiesto troppo tempo)
 - MicroAccuracy: ???
 - MacroAccuracy: ???
 - LogLoss: ???
 - LogLossReduction: ???
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: ???
 - LightGbm:
(Non è possibile riuscire a trovare il metodo)
 - MicroAccuracy: ???
 - MacroAccuracy: ???
 - LogLoss: ???
 - LogLossReduction: ???
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: ???
 - NaiveBayes:
 - MicroAccuracy: 0,171
 - MacroAccuracy: 0,024
 - LogLoss: 34,536
 - LogLossReduction: -9,53
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: "POLITICS"
- Predizioni sono con il metodo SdcaMaximumEntropy modificando man mano dei parametri:
 - Esame iniziale precedente on Headline e Short_Description:
 - MicroAccuracy: 0,361
 - MacroAccuracy: 0,274
 - LogLoss: 4,677
 - LogLossReduction: -,426
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: "MONEY"
 - Utilizzando una porzione di train maggiore (0.9 Training, 0.1 Testing):
 - MicroAccuracy: 0,365
 - MacroAccuracy: 0,28
 - LogLoss: 4,694
 - LogLossReduction: -,432
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predicted: "MONEY"
 - Utilizzando sempre la proporzione 0.9/0.1 ma utilizzando come Feature anche "Authors" (trasformato con Categorical.OneHotEncoding):

- MicroAccuracy: 0,458
- MacroAccuracy: 0,371
- LogLoss: 3,682
- LogLossReduction: -,123
- Single Prediction:
 - Expected: "U.S. NEWS"
 - Predected: "MONEY"
- Utilizzando sempre la proporzione 0.9/0.1 ma utilizzando come Feature anche "Authors" (trasformato con Text.FeatuizeText):
 - MicroAccuracy: 0,458
 - MacroAccuracy: 0,374
 - LogLoss: 3,62
 - LogLossReduction: -,104
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predected: "MONEY"
- Utilizzando sempre come Feature anche "Authors" (trasformato con Text.FeatuizeText), ma questa volta il set di dati per il Training è tutto (1.0), ma lo è anche quello di Testing (che quindi sono uguali):
 - MicroAccuracy: 0,474
 - MacroAccuracy: 0,393
 - LogLoss: 3,614
 - LogLossReduction: -,099
 - Single Prediction:
 - Expected: "U.S. NEWS"
 - Predected: "MONEY"