

Niels Louis

40077435

Quang Jimmy

Nguyen

40066773

COMP
472 ASSIGNMENT
3

Dataset Analysis

Dataset is small => yields better (code)performance for BOW, but less accurate

Lower-case Semantic Trimming

Yes-class: 61.75%; No-class: 38.25%

Classification is non-linear; features are not easily discriminable

Vocabulary Analysis

Original Vocabulary length is 3575

Filtered Vocabulary length is 1030

OV accuracy is 54.55%; FV accuracy is 69.09%

OV perClassPrecision: Yes-1.00 No-1.00; FV perClassPrecision: Yes-1.00, No-1.00

OV perClassRecall: Yes-0.788 No-0.181; FV perClassRecall: Yes-1.00, No-0.227

OV perClassF1: Yes-0.788 No-0.181; FV perClassF1: Yes-1.00, No-0.227

*remember class distribution(Yes-class: 61.75%; No-class: 38.25%)

Model Analysis

OV:

Accuracy: 54.55%

Per Class Precision: Yes-100%;No-100%

Per Class Recall: Yes-78.79%;No-18.18%

FV:

Accuracy: 69.09%

Per Class Precision: Yes-100%;No-100%

Per Class Recall: Yes-100%;No-22.7%

LSTM:

Accuracy: 67.27%

Per Class Precision: Yes-65.31%;No-83.33%

Per Class Recall: Yes-96.97%;No-22.73%