IST736 Text Mining
HW6 (10 points)

Benoulli and Multinomial Naïve Bayes in scikit-learn

For this assignment you will use the Benoulli and multinomial naïve Bayes algorithms in sklearn to repeat the customer review classification by sentiment and authenticity (same data from HW4 Task 2).

Remember that the Benoulli model takes Boolean vectors as input, and the multinomial model takes frequency vectors as input. Use several vectorization options, such as removing stopwords, n-grams, etc., to create three Benoulli models and three multinomial models. Explain the rationale for your vectorization choices, and compare the performance of these models using appropriate evaluation methods and metrics.

Submission:

Submit a Word document in research-paper format, similar to previous homework.

Grading rubrics:

Same as HW1.