IST736 Text Mining

HW4

Multinomial Naïve Bayes

**Use Weka’s Multinomial Naïve Bayes algorithm to build models to classify the customer reviews by (1) sentiment (positive or negative) (2) authenticity (true or fake, lie detection). (6 points)**

For each of the two classification tasks, use MNB to build the models, and evaluate them using 10-fold cross validation methods. You can choose your own options for tokenization, and make sure to report if you change any default setting.

For each model (lie detection and sentiment classification), report the 20 most indicative words that the models have learned. Based on these words, do you think the models have learned the concepts (lie or sentiment) that they are expected to learn?

Also, compare the difficulty level of sentiment classification vs. lie detection. Discuss whether you believe computers can detect fake reviews by the words.

**Submission:**

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

**Grading rubrics:**

Same as HW1.