

Lie detection

Scenario:

Some people claimed that machine learning algorithms can figure out whether a person is lying or not. Do you believe that? To test this claim, we have collected a collection of customer reviews, some are true some are fake, and you are going to test how good multinomial NB and SVMs can be for fake review detection.

This data set also has sentiment label for each review. You will also MNB and SVMs performance in sentiment classification.

For both tasks, try tune parameters and report results in a table like

Parameter setting	Overall accuracy	Precision in category I	Recall in category I	Precision in category II	Recall in category II

Explain your initial parameter tuning strategy (which parameter to tune, to what option, and theoretical foundation for your choice). Does your strategy help you get better results?

Compare performance difference in sentiment classification and lie detection, and tell us which task is harder, and try to explain why.

For each task, use GainRatio and Chi2 to rank the features and list top20 features from each method. Based on these top features, can you understand what patterns the classifiers have learned from the data?