BUS 243 Name:

Quiz 3.

Problem. In our previous class, we explored the effectiveness of LSA topics in predicting spam emails. Specifically, we encoded the spam email data using the TF-IDF technique and applied PCA to obtain the topic vectors. Then, we utilized linear discriminant analysis to predict whether a message is a spam or not based on these topic vectors. What’s the advantage/disadvantage of this process compared to predictions based on TF-IDF document word vectors?