

Introduction to Machine Learning and Text Mining for Trade Policy

Problem: Lecture 3.

Ben Shepherd, Principal.

1. Read the same UNCTAD TRAINS data used for the demonstration.
2. Retain the descriptive text, and the HS field. That field lists the HS codes of products affected by a measure.
3. Isolate the first two characters of each entry in the HS field, i.e. the first two characters of the first product listed.
4. Transform those two characters to numerical values. Verify that you have a list of numbers 1-99, corresponding to the HS 2-digit code for the first listed product affected by each measure.
5. Remove NA entries from the data.
6. Using the demonstration code as a guide, use an artificial neural network to try and classify text descriptions by HS 2-digit code.
7. To the extent possible, experiment with different hyperparameter sets to obtain a model that performs well.
8. When you have a preferred model, evaluate it on the testing sample.