Experiments:

1. Change order of HMM varying from 1 to 4.
2. Select relevant subset of features to build model based on some p-value.
3. Performed 5-fold cv to see performance with each subset of features.
4. Hold out validation (relevant to their experimental settings not to ours).