Project1 Naive Bayes

#Cornell/CS5740

1. vectorizer(input='filename')
   then a whole file is a row, but expected prediction is on line-level.
   use default input='file'
2. different Vectorizer (feature extraction method)
   CountVectorizer works best on validation set
3. different naive bayes models
   GaussianNB works best on validation set
4. ngram
   accuracy on validation set increases with the value of n,
   but when n = 5, MemoryError: Unable to allocate array with shape (1600, 695555) and data type float64
5. test method accuracy on validation set
6. predict on (train+validation)