Private: 19th (could not believe it!)

New to kaggle, new to R, but I have a background in data processing & text analytics.

I tried a zillion things that didn't work, in the end settled on a pretty simple RF model.

Features:

- newsdesk, section, subsection

- wordcount, log(wordcount) [using both was helpful, weirdly]

- cluster [used it as a feature, helped a little]

- did headline: contain a question? match a regex of boring and interest words? [ devious hack ]

- how long was the headline?

- time: day, hour, primetime [ hacked variable of daylight hours + sunday evening ]

I tried tuning my RF, but the OOB settings seemed to work the best.