My philosophy is as follows:

1. a new variable as 'polarity' which will tell whether the post was -ve or +ve.

2. cluster, I used as a variable, you can see it in 'HeadlineWords$cluster'

3. I started with glm. using summary, I extracted the significant variables.

4. Then I tried both 'rf' and 'nnet', to check which one will do better.

5. NNet had better performance, so one of my submission was NNet.

6. Finally, as per the suggestion of a forum post, I took avg. of both 'rf' and 'nnet' outputs. Which did improve my performance by 0.004. and that was my 2nd final submission.