1)

Predict ? ! or . based on polarity

Higher rating uses positive sentiment

select cast(rating as int) as r,sum(polmatch),count(\*) from data group by r;

1,6891,8514

2,2890,3655

3,3437,4414

4,3370,4452

5,4383,6094

6,5902,8268

7,9041,12591

8,11751,16390

9,10350,14432

10,19037,26613

select cast(rating as int) as r,sum(polmatch),count(\*) from data where Adj1Pol='POS' group by r;

1,2271,3248

2,1023,1509

3,1317,1911

4,1409,2109

5,2153,3291

6,3363,4885

7,5701,7943

8,7669,10543

9,7083,9628

10,13796,18582

Use distribution in ratings as a prior on positive polarity

Could be used to help identify clusters as positive or negative

2)

Emoticons

Punctuation (ie more or fewer exclamation points)

Words like ex

Grammatical correctness

General text quality

Use of pronouns (people talk about themselves more when being negative or positive)

Capitalization – IMMD would have more words in all caps and no caps (may be determined by site rules)

Cursing

Spelling (or OOV)