Data preprocessing

1. Remove URL
2. Remove timedelta – nothing to do with the article (assuming shares have stabilized)
3. Replace concepts with one single column is\_life\_style\_social\_media
4. Remove columns of the form: Kw\_{min,max}\_{min,max,avg}
5. Remove columns of the form: self\_reference\_{min,max}\_shares
6. Remove all weekdays columns – weekend column captures the required information
7. Replace LDA\_[0-9]{2} with max\_closeness containing the maximum of the 5 values
8. Remove {min, max}\_positive\_polarity
9. Remove {min, max}\_negative\_polarity
10. Remove title\_subjectivity and title\_sentiment\_polarity – keep the absolute values which is 0.5 subtracted
11. Remove n\_non\_stop\_words – all values seem 1

