Online News Popularity Data Set

This dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. The goal is to predict the number of shares in social networks, i.e. how popular any given article is.

Mashable Inc. is a digital media website founded in 2005. It has been described as a "one stop shop" for social media. It has over 6,000,000 Twitter followers and over 3,200,000 fans on Facebook.

The Data

Attributes:

1. **url:** URL of the article (non-predictive)
2. **timedelta:** Days between the article publication and the dataset acquisition (non-predictive)
3. **n_tokens_title:** Number of words in the title
4. **n_tokens_content** Number of words in the content
5. **n_unique_tokens:** Rate of unique words in the content
6. **n_non_stop_unique_tokens:** Rate of unique non-stop words in the content
7. **num_hrefs:** Number of links
8. **num_self_hrefs:** Number of links to other articles published by Mashable
9. **num_imgs:** Number of images
10. **num_videos:** Number of videos
11. **average_token_length:** Average length of the words in the content
12. **num_keywords:** Number of keywords in the metadata
13. **self_reference_min_shares:** Min. shares of referenced articles in Mashable
14. **self_reference_max_shares:** Max. shares of referenced articles in Mashable
15. **self_reference_avg_sharess:** Avg. shares of referenced articles in Mashable
16. **global_subjectivity:** Text subjectivity
17. **global_sentiment_polarity:** Text sentiment polarity
18. **global_rate_positive_words:** Rate of positive words in the content
19. **global_rate_negative_words:** Rate of negative words in the content
20. **rate_positive_words:** Rate of positive words among non-neutral tokens
21. **rate_negative_words:** Rate of negative words among non-neutral tokens
22. **title_subjectivity:** Title subjectivity
23. **title_sentiment_polarity:** Title polarity
24. **abs_title_subjectivity:** Absolute subjectivity level
25. **abs_title_sentiment_polarity:** Absolute polarity level
26. **shares:** Number of shares (target)