News Articles Share Prediction Machine Learning Project

Tasks:

- Exploratory Data Analysis (EDA)
- Feature Engineering
- Data Preprocessing
- Build and Evaluate Regression Models to predict shares for the news article
- Find the best model to predict the share per article.
    - ✓ I completed all of the required stages in a single code file.
    - ✓ To estimate the share per article, I created and assessed Linear Regression, Lasso Regression (L1 Regularization), Ridge Regression (L2 Regulazisation), and Ensemble (Assuming a combination of models) in my project implementation. These are all examples of Regression.
    - ✓ Calculate the R2_Score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) of the predictions to assess the model's performance. These measure accuracy by comparing predicted values to actual values in the test set.

Dataset:

1. *article_id:* Unique identifier for article
2. *title:* Title of the news article
3. *text:* Content of the article
4. *published_date:* date on which the article was published
5. *unique_tokens_rate:* Rate of unique words in the content
6. *num_hrefs:* number of links in the article
7. *num_imgs:* number of images in the article
8. *num_videos:* number of videos in the article
9. *average_token_length:* Average length of the words in the content
10. *num_keywords:* number of keywords in metadata
11. *data_channel:* data channel of the article
12. *min_avg_key:* minimum shares for average keywords
13. *max_avg_key:* maximum shares for average keywords
14. *avg_avg_key:* average shares for average keywords
15. *href_avg_shares:* average shares for links in the article

16. *global_subjectivity:* text subjectivity
17. *global_sentiment_polarity:* text sentiment polarity
18. *global_rate_positive_words:* rate of positive words in the text
19. *global_rate_negative_words:* rate of negative words in the text
20. *title_subjectivity:* title subjectivity
21. *title_sentiment_polarity:* title sentiment polarity
22. *shares:* number of shares for the article (target variable)