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 have done all of the required steps in one code file.
    - In My Project Implementation, built and evaluated ***Linear Regression***, ***Lasso Regression (L1 Regularization), Ridge Regression (L2 Regularization), and Ensemble (Assuming a combination of models)*** to predict the share per article. All of these are types of Regression.
    - Evaluate the performance of the model by calculating the R2_Score, Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), of the predictions. These indicate the accuracy by comparing the predicted values to the actual values in the test set.

Dataset

article_id: Unique identifier for article

title: Title of the news article

text: Content of the article

published_date: date on which the article was published

unique_tokens_rate: Rate of unique words in the content

num_hrefs: number of links in the article

num_imgs: number of images in the article

num_videos: number of videos in the article

average_token_length: Average length of the words in the content

num_keywords: number of keywords in metadata

data_channel: data channel of the article

min_avg_key: minimum shares for average keywords

max_avg_key: maximum shares for average keywords

avg_avg_key: average shares for average keywords

href_avg_shares: average shares for links in the article

global_subjectivity: text subjectivity

global_sentiment_polarity: text sentiment polarity

global_rate_positive_words: rate of positive words in the text

global_rate_negative_words: rate of negative words in the text

title_subjectivity: title subjectivity

title_sentiment_polarity: title sentiment polarity

shares: number of shares for the article (target variable)