# News Articles Share Prediction Machine Learning Project

1. Problem Statement

   Predicting the popularity of news articles is an important task in the field of journalism and online media. By accurately predicting the number of shares an article is likely to receive, media companies can optimize their content distribution strategies and reach a wider audience.

2. EDA: For better analysis of data, the numerical and textual features are separated. To keep the interior and exterior of numerical features and categorical features are analyzed. Popularity of news based on the share is also shown. Also count of popular or unpopular news. Over different days of weeks as well as different data channels are shown with bar graph using the approach of mean and median of shares where median has clearly shown better distributions. Shares were better during weekdays or weekends are also shown.

3. Data preprocessing: Using data preprocessing techniques, textual data like title and text are thoroughly cleaned by using techniques like removing punctuation. removing links and converting them to lower cases etc. Number of tokens in title as well as texts are determined.

4. Feature Extraction: Categorical features like data channel are further divided into their respective categories such as. Technology, business, entertainment., social media, etc. Similarly, from the column published date, days of share were extracted. Outliers were detected using box plot and were tried to be eliminated.

5. Models used: Linear Regression, Ridge Regression and Ensemble.

6. Results acquired:

   LR (Test R2 score: 0.128, Test MAE: 0.409, Test RMSE: 0.460)
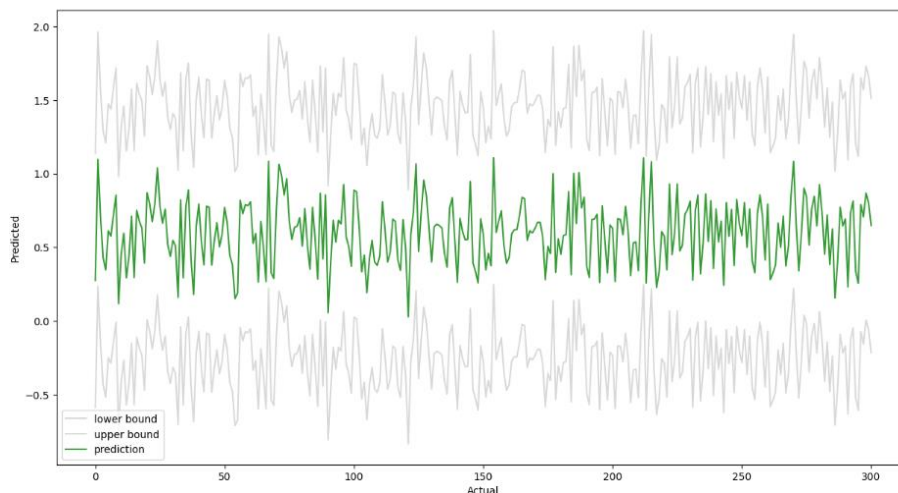
RR (Test R2 score: 0.0751, Test MAE: 0.432, Test RMSE: 0.474)
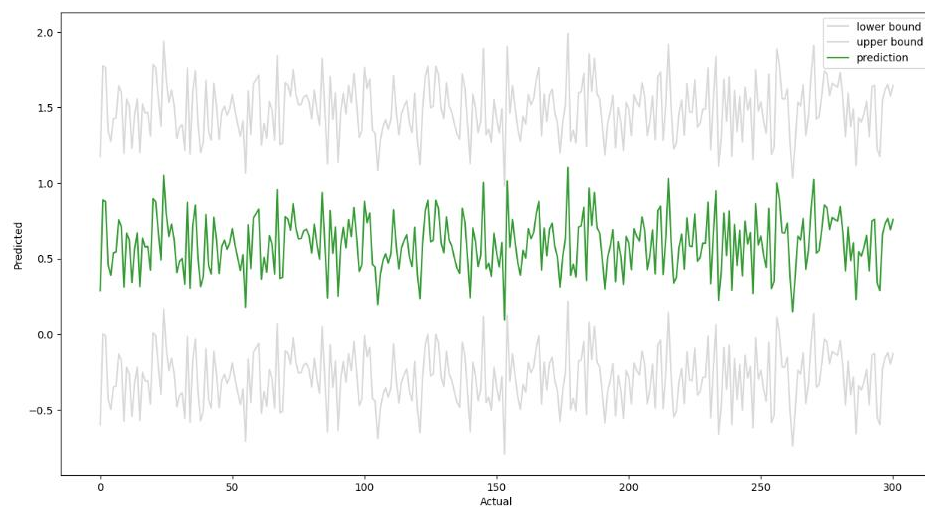
Ensemble (Test R2 score: 0.110, Test MAE: 0.440, Test RMSE: 0.465)