

Article Popularity Prediction

Motivation

With the increased access of handheld devices and availability of the internet across the world. People are rapidly making their way from printed sources of news and media to online sources such as The Daily Mail and The Mashable. So to cater to these increasing new audiences, these websites track their data based on the number of likes, shares etc. In turn, it's easier for these companies to make use of Machine learning to make sure that every single viewer gets exactly what they want and what type of articles will have the maximum retention for these articles. With this project we aim to help the various stakeholders to have an easier way to tell how good a new article would do based on our training data, content of the article and the length of the article.

Related work

[Predicting and evaluating the popularity of online news by He Ren and Quan Yang](#)

[Predicting the popularity of online news from content metadata](#)

Timeline

Week	Work
1	Literature review
2	Data cleaning and preprocessing
3	Feature finalization/ selection
4	Training the models
5	Hyperparameter tuning
6	Analyzing the various types of model

Week	Work
7	Selection of best models
8	Evaluation of the project
9	Reporting the results

Individual Tasks

Rupin Oberoi:

Project topic selection, Data collection, Literature Review, Data cleaning, preprocessing, Data feeding.

Aditya Bugalya:

Data preprocessing, training data on a lot of models like linear regression, logistic regression. Data testing using K-fold cross validation.

Sarthak Dixit:

Project topic selection, Model selection, Normalization, Logistic regression training , Calculation of accuracy ,Parameter selection.

Ashwin Tomer:

Further processing of the data models, evaluation of the results obtained from the various models.

Final Outcome

At the end of this project, we hope to be able to identify the features that lead to success of a news article and also to help the newspaper and the various stakeholders to decide whether to include the particular article in their article/newspaper or not. We, the viewers, too would get only those articles that interest us, saving a lot of time.