

Predict

Shreya Sajal,Vineet Agarwal.,Siddharth S. P.,Saman G.,Rishika P.

Aim:

Often important linkedin posts by organizations don't get the required reach due to poor post designing or use of inaccurate community hashtags.Our project, PredictIn aims at solving this problem by predicting a popularity score for a post beforehand and help the organizations in designing the posts in an effective way that maximizes their popularity.

Data:

The data has been obtained by scraping the linkedin pages of companies and influencers.Information like Number of followers,company size,post content,number of likes,number of comments,comments' content,etc are being used as the features.

We have used selenium and beautifulsoup for scraping

Working Approach:

- After the data was collected ,a popularity metric using the features(Number of followers,company size,post content,number of likes,number of comments) was made.These were made after proper research on LinkedIn metrics .
- We added additional Features like Post type by classification into 5 different categories using unsupervised classification technique that made use of Flair Embeddings and calculated Cosine Similarity Scores on keywords used in a particular type of post. Another Feature that we added was the Relevance score of the post's content with the company/influencer's about information we had scraped.This was done using Smooth Inverse Frequency over Glove embeddings and then calculating the final Cosine similarity scores.
- The most suitable Machine learning model with proper hyperparameter tuning was then fitted on our training set and used to predict the popularity score of any test post.
- Further some Exploratory Data Analysis was done to analyze the extent to which a particular feature impacts a particular post's popularity(Mainly, post-content based)
- The model was then deployed.