



MEASUREMENT AND PREDICTION OF INFLUENCE ON SOCIAL MEDIA

GAURAV DUBEY (11919) | SHIVAM GOYAL (11941) | YASHU GARG (11949)
MENTORED BY: DR. SACHIN KUMAR

Cluster Innovation Centre
University of Delhi

ABSTRACT

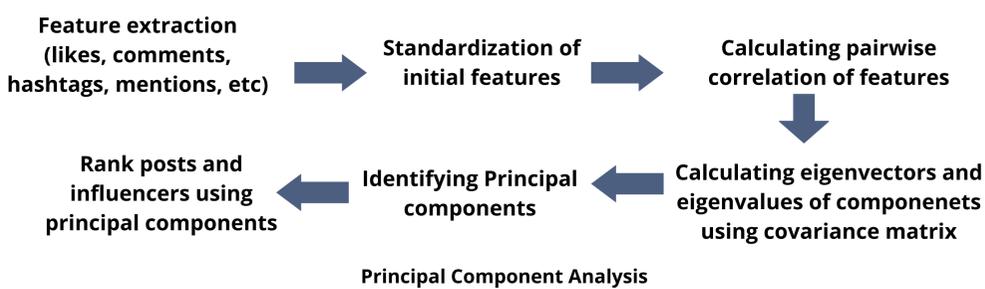
Instagram is one of the most popular and widely used social network platforms. It is used as a tool to connect with users and also to influence them for marketing purposes. Influence of popular users is broadly determined by the posts, engagement rate in terms of likes, comments and shares and the number of followers.

Having a measure of popularity is necessary in order to understand the factors that will help make an influencer marketing campaign successful. This research attempts to take various features of an influencer and posts dataset and come up with a model that accurately measures and predicts the influence of a user on Instagram.

The research is based on datasets of topmost regional influencers and posts based on categories which are signified through hashtags and captions. Studies have been done to develop a predictive model by applying supervised learning on a labelled training data that contains engagement rate as a target variable and derived feature lags. Our research attempts to achieve a model using Principal Component Analysis to quantify posts and influencers to rank them solely based on their post activity, as engagement rate might be biased because of certain features.

METHODOLOGY

To build and train our model, posts and users related to certain categories/hashtags were collected with their Instagram handles and post links for further processing. Python libraries like BeautifulSoup and Selenium were used to extract user and post information (date, caption, likes, comment counts, etc.). To extend the scope of our analysis we used publicly available dataset of top regional influencers having post activity of around two months and carried out feature extraction to get relevant features for PCA analysis.



Using PCA analysis relative feature importance was calculated. Components of eigenvalues > 1 were considered for analysis as these variables were responsible for about 70 percent of variance in the dataset. A comparison of users was done on the basis of the number of followers and on the basis of average engagement rate calculated for two months. The resultant ordering of users based on principal components was then compared with the aforementioned ranks. Parity in ranks of users in the three comparisons was calculated using root mean squared error (RMSE).

CONCLUSION

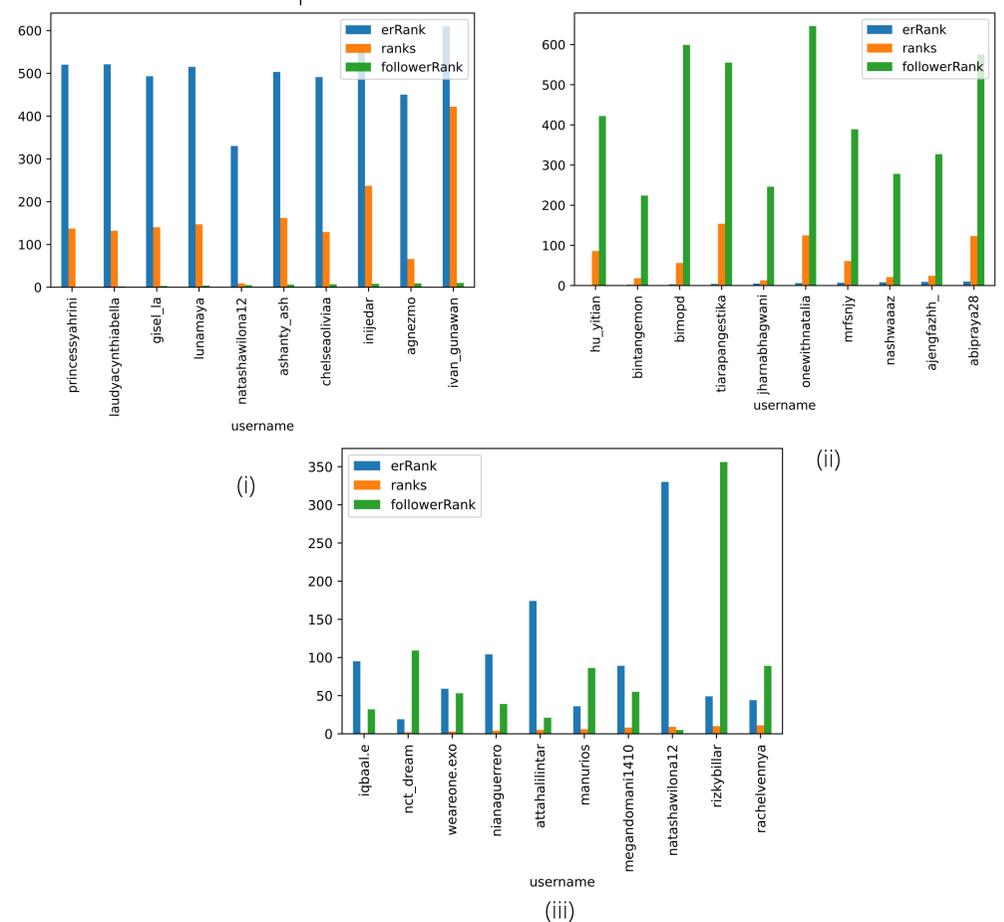
- An unsupervised learning model was developed using Principal Component Analysis to quantify posts and rank influencers. The results obtained were compared with the results obtained from supervised predictive models that ranks users based on average engagement rate.
- It can be concluded that number of followers and engagement rate are not the most accurate parameters to predict the future influence of social media users. The accuracy of our analysis lies in between the aforementioned metrics as it is not biased towards users who have higher followers.

INTRODUCTION

Existing research results show that both attitudinal and behavior engagements with the brand community have significant effects on purchase intention. The findings indicate that the more the customers actively engage with the online community, the more likely they will purchase products or services through the community. Hence, a central influencer index is essential to standardize the process of identifying key influencers. This would involve extracting relevant features out of a data frame that would only contain dates, likes, comments, captions to get more useful information like time elapsed, caption length, type of media, number of hashtags used, number of users mentioned etc., and aggregating them to train a model and test its accuracy on the given data, using number of followers and engagement rate.

RESULTS AND DISCUSSION

- A detailed analysis of posts based on features like comments count, likes count, caption, media, etc., and users based on posting activity was carried out on data based on categories as well as top regional influencers.
- Ranks based on engagement rate, followers count and principal components selection were compared.



Comparison of ranks of top 10 users based on (i) followers (ii) avg. engagement rate (iii) principal components

REFERENCES

1. Arora, A., Bansal, S., Kandpal, C., Aswani, R., & Dwivedi, Y. (2019). Measuring social media influencer index- insights from Facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, 49, 86–101. <https://doi.org/10.1016/j.jretconser.2019.03.012>
2. Kiss, C., & Bichler, M. (2008). Identification of influencers — measuring influence in Customer Networks. *Decision Support Systems*, 46(1), 233–253. <https://doi.org/10.1016/j.dss.2008.06.007>
3. Muzamhindo., S. (2017). Principal component analysis as a ranking tool - A case of world universities. *International Journal of Advanced Research*, 5(6), 2114–2135. <https://doi.org/10.21474/ijar01/4650>