Instagram Likes Analysis

Vishnu J G
Computer Science and Engineering
PES University
Bangalore, India
jgvishnu2001@gmail.com

Ajay Kumar K
Computer Science and Engineering
PES University
Bangalore, India
ajayk.sm2001@gmail.com

Praneeth A L
Computer Science and Engineering
PES University
Bangalore, India
praneeth.arasada@gmail.com

***Abstract*—Instagram likes analysis is finding the influence of date and time of post on the number of likes the post gets. This analysis gives us an idea of the most suitable time to make a post, so that maximum likes can be obtained on that post.**

***Index Terms*—Instagram, likes, post, influence, time, date**

I. Introduction and Background

Conventional method of promoting a product such as TV advertisements, radio broadcasts, billboards and banners, pamphlets and events are fairs are nowadays being dominated by social media platforms. Influence of social media on product promotion and sales has increased tremendously in the past two years(2019 and 2020) and is expected to increase further this year as well.

Even if a company or business sector does not have a social media presence, it can take the help of social media influencers to promote their products. These influencers are the key elements in improving the popularity of many budding companies and industries. Many small scale industries and home based businesses' only mode of marketing is social media. As we have seen many home based businesses emerged during the period of pandemic. These businesses were able to make a good sale of their products from their social media presence.

Even though social media influence can impact sales, it is important that some experience in the sales field necessary. Being an influencer or a stand-alone home-based business, it is important to understand the type of customers they are dealing with and how the product concerns them. One of the key things to take into consideration while advertising a product on a social media platform is the time when the post for advertising the product is made.

The number of likes on a post majorly depends on the number of people who view the post, so finding the right time to post in general translates to finding the right time when more people are online on Instagram. This way the post may be viewed by as many people as possible. Understanding this based on historic data is important for an influencer for deciding the right time to make a post in order to get maximum likes. He/she has to understand when his/her followers are more likely to view their content.

One of the social media platforms, Facebook market, helps in building an online presence for many businesses. This way it is easier to connect both the seller and someone who is in need of a particular product(supply on demand), improving the supply chain. Post may also be categorised based on hastags, here the most popular posts(posts with maximum views/lines) are found at the top. People searching for a product are directed to such a space where posts are categorised based on hashtags, where the user is presented with the most popular posts first.

It is clear that increasing the number of likes is the goal of influencers. And to do so, they need their post to be viewed by as many people as possible. This can be done if we knew when most people will be online. Not only the time but also the date, day and month also matters for their post to get more likes.

The main objective of this research is to understand the influence of date, time, day and month on the likes obtained on the post. Based on a prediction we can say what would be the right time to make a particular post so that maximum people are able to view the post and this in turn helps in reaching the target customers.

II. Previous work

Social Media is one of the major platforms that facilitates sharing of ideas, thoughts, information and promoting brands through posts [1]. A post on a social media platform such as Facebook can receive many type of reactions such as likes, cheers, claps, angers and hearts. The authors [1] in their work have considered all these as likes or in general reaction to a post. This may not be exactly correct as an anger reaction on a post is not the same a cheer or like on a post. Contrary to this in our work we have considered social media platform Instagram which only allows likes on a post. This way there can only be a like and dislikes on a post.

Considering all the reactions as likes, the authors [1] built a SUR model(generalised OLS model) is built. But the consideration of each day is not take into account. The authors [1] have generalized days as weekdays or weekends. This is a more generalized approach as the heteroscedastic property in an assumption in a SUR model. In contrary in this research paper we have considered each day seperately, each month and hour of the day is taken into account while making predictions.

A. Assumptions

Having considered all the users separately, we are not taking into account the influence of hashtags, comments and tags on the number of likes on a post. The reason for this assumption is, the above mentioned attributes matter when making a comparative study, say we are doing a competition between 2 user as to who can get more likes. But in our case each user is isolated and not comparison is done with other users. The only factor that the likes depends now is the number of followers and time of post.

This leads to our second assumption that the followers count for an user is constant over time. This is done as the influence of changes in the followers count for an user on the number of likes is minuscule. To accommodate this minuscule change we also have decided to change using number of likes as a metrics to a like score. This like score takes into account changes in likes due to change in followers count as well.

III. Proposed Solution

The Models that we have built is used for predicting Like score.Based on the results we can conclude that which time is good for posting to get max number of likes.

A. For a particular user

Results for a particular user (say user 2) is as follows. From

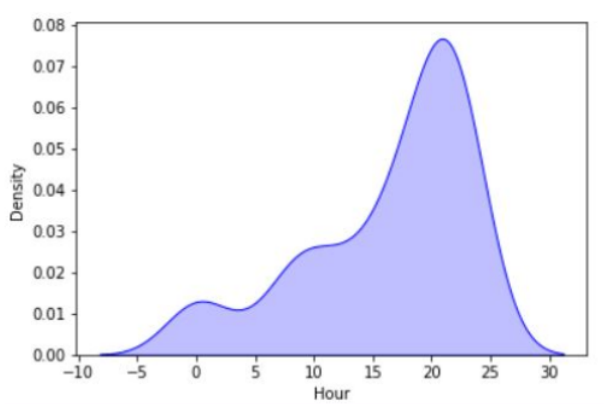

Fig. 1. Single user

the fig.1 we can say that the correct time for posting is from 17:00 hour. By posting at the right time users may get most number of likes.

B. For all users

Similarly computing a general result taking into consideration all users at a time.

From the above image we can say that by posting in at mid-night and later in the evening would be a good time to

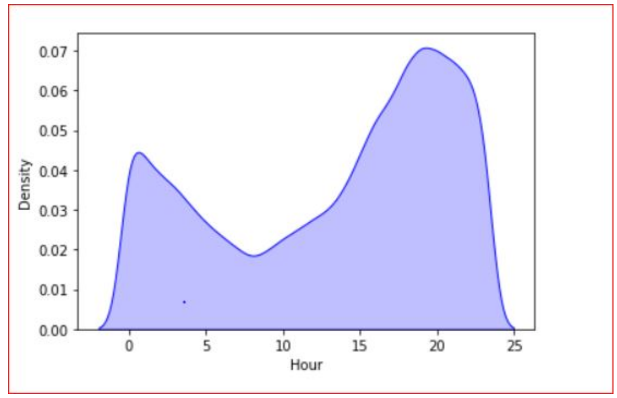

Fig. 2. All Users

post in order to get maximum number of likes.

The above graphs(fig.1 and fig.2) shows hour as the only attribute to predict likes. We can extend the same to the other attributes under consideration such as day, month and year.

IV. Models

There are 4 models that are used for predicting the results
1.Linear Regression
2.Ridge
3.Theilsen Regressor
4.Lasso

A. Linear Regression

Linear Regression is perhaps one of the most well known and well understood algorithm. This model assumes linear relationship between explanatory variables and is used to predict the value based on the explanatory variable . Linear Regression is the process of finding a line that best fits the data points available on the plots ,so that we can use it to predict the output values.

B. Ridge

Ride regression or Ride analysis is a way to create a parsimonious when the number of predictor variable is set exceeds the number of observation

C. TheilSen Regressor

Theil-sen estimator is robust multivariate regression model. this a method for robustly fitting a line into sample points.

D. Lasso

Lasso regression is a type of regression that uses shrinkage. shrinkage is where data points are shrunk towards the central point like mean.this encourages simple and spares models

The attributes under consideration for training the model are shown in the fig.5. The list included hour of the day, Month of the year and day of the month.
These play a significant role in influencing the likes score value for each user, but these attributes are not the only ones

```
models['Linear'] = LinearRegression()
models['TheilSen'] = TheilSenRegressor()
models['Ridge'] = Ridge()
models['Lasso'] = Lasso()
```

Fig. 3. Models List

that influence the Likes Score, but since our paper is focused on influence of time and day on likes, we only consider these attributes that relate to date and time.

V. Metrics

Metrics tells how good the model is. The lower the metric value, the better the model results are. The metrics used are as follows :
1. Mean squared error
2. Root mean squared error
3. Mean absolute error

```
Linear / 0.0008713314832295958 / 0.029518324532899827 / 0.13897878579327055
TheilSen / 0.0009030769530949632 / 0.03005123879468138 / 0.14038969278573002
Ridge / 0.0008713237355091567 / 0.029518193296832324 / 0.13897865136772666
Lasso / 0.0008444118112890501 / 0.029058764792899407 / 0.1384023668639053
```

Fig. 4. Metrics

From the above Fig.4 we can see that the MSE value for TheilSen regression model is the least as compared to the MSE value of the other models. So for this particular user TheilSen model gives more accurate results.

```
model.fit(train[['Month','Day','Hour']],train['Likes Score'])
predres=model.predict(test[['Month','Day','Hour']])
```

Fig. 5. Attributes used to train the models

The reason for not using the year attribute is that, the social media platforms grew to such an extent that it is today only after 2019. This revolution is like a cyclical change and can not be accounted for as our data is limited to just 8 years.

Other time related attribute such as minutes are ignored as the influence of them is captured by the hour attribute. Number of tags and number of comments is not the focus of our study, even though they play a significant role on the likes score.

Experimental Result

From fig.3 Metrics, we can clearly see that a better results with accurate output is seen in case of ThielSen model. This may or may not be true. To check this theory we compute the same for another user(randomly chosen) say user 512.

In the above fig.7 we see the hour vs Density for user 512 and comparing this with user 2's, shown in fig.1 we

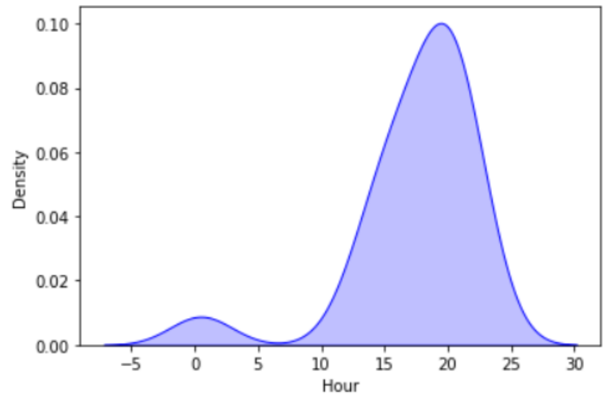

Fig. 6. User 512 Hour vs Density

can see that the graph shows similarity. Can we assume the that ThielSen works best for this user? No, testing the hypothesis we get the following results(fig.8) in the order(model name/MSE/RMSE/MAE).

```
Linear / 0.0010772150881463618 / 0.03282095501575726 / 0.1460651207101641
TheilSen / 0.0013499903316336l1 / 0.03674221457171044 / 0.1462192851579217 6
Ridge / 0.0010771831356466098 / 0.03282046824234246 / 0.146051486894994 3
Lasso / 0.0010604628117326098 / 0.032564747991234475 / 0.13897297297297298
```

Fig. 7. metrics for user 512

In this we can see that Lasso regression model gave a more accurate result than any other model. So a more robust method of finding the right time would be to go ahead a build all the models and choose the one with the least error. But this is something close to a ensemble model where we use many weak learners to get one good model.

Since we are run the set of models for individual users and not to the whole dataset this is a feasible solution as the model training set is minuscule.

Some other information that we derived while working on the data include distribution of post over the month, like on which day the user made more posts, also which month of the year he made more posts. These metrics help in deciding on the model to be chosen for a given user.

As specified before we are not taking into account the changes in the followers count, as this is assumed to be minuscule. But when the change in the followers count is not negligible then this attribute directly affects the likes on a post. This models also don't take into account the use of hashtags on the post. This increases global users which are not even followers of the user in consideration. This affects the likes count tremendously.

This attribute has been skipped as the number of hashtags are not available in the dataset we have chosen. The other attributes such as number of tags and number of comments do not actually show significant influence on the user. The reason for this being, the count of number of tags is given in the dataset which does not help, as tagging some one popular user has more affect then tagging 100s of not popular users. And also the count is really some and in most cases 0.

It might have come to the notice that we are also not using Year attribute. The reason for this is most social media platforms blew up in the year of 2019 and 2020. Before these years, the number of posts made were comparatively low. The ratio is 1/129 for a particular day in 2018 to 2020 i.e., there are 129 posts made in the year 2020 for a single post made in 2018.

CONCLUSION

In conclusion, we can say that even though time, day, month and year bear a huge influence on the number of likes obtained by a post, these are the only attributes. So, coming to a conclusion based on only these may not be completely right. But these attributes give a fairly better predictions as compared to other attributes taken individually.

Time plays a significant role in predicting the number of likes on a post. As seen above we have evidence that each user gets relatively more like score when the post is made between 20:00 hour and 01:00 hour.The reason for this can be numerous, free time for user, after office times, after school and college time, leisure time for a few users to name a few reasons.

Day of the month also play a significant role. In the fig.9 we can see that the beginning of the month shows a significantly more posts than the end of a month.

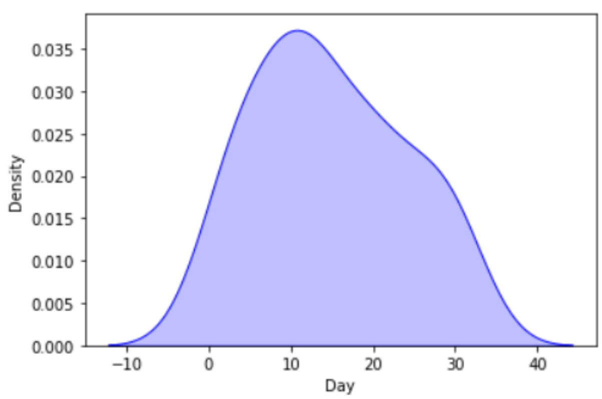

Fig. 8. Day vs Density

Month of the year plays a significant role based on festivals, holidays, seasons, etc,. We see that the graph is skewed to the right as all the major holidays, namely Christmas, Halloween and Thanksgiving coming in the end of the calendar year.

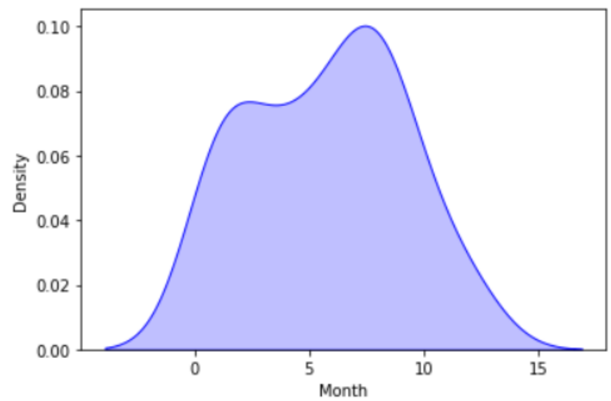

Fig. 9. Year vs Density

REFERENCES

[1] Tri Hanifawati, Utan Sahiro Ritonga, and Euis Evi Puspitasari, "MANAGING BRANDS' POPULARITY ON FACEBOOK: POST TIME, CONTENT AND BRAND COMMUNICATION STRATEGIES" Journal of INDONESIAN ECONOMY and BUSINESS. Volume 34, Number 2, 2019.