Instagram Popularity Forecast for Optimum Digital Marketing Using Machine Learning

Shashank Kotian, Mayuresh Jadhav, Samiksha Jadhav,
Shrutika Khetam and Priya Mule

Instagram Popularity Forecast for Optimum Digital Marketing using Machine Learning

Mr.Shashank Kotian, *Mr.Mayuresh* Jadhav, *Ms.Samiksha* Jadhav, *Ms.Shrutika* Khetam, and *Mrs.Priya* Mule

[1]Ramrao Adik institute of technology,NERUL: Department of EXTC

Abstract. Ever since the platform moved away from reverse-chronological and to a curated content feed. For digital Marketeers they are like little data puzzles just waiting to be solved. Recent data suggest that 72 percent of executives surveyed, social data is already the top data source used to inform business decisions, even more so than market research. 85 percent agree that social data will be primary source of business intelligence and 90 percent of marketeers agree that social data enables them to stay ahead of their competitors, resulting in help of social media listening tools

1 Introduction and Explanation

Ever since Covid-19 hit us harshly there was a sudden drop in business profits and revenue's. In order to get back in routine many small scale businessman,businesswoman and entrepreneurs had to shift their entire line of business on numerous online social media platform. Instagram being popular was a hub for local businessman and big game tycoons to market and sell their commodities and services amongst the masses. Our work and aim for this project is to help local businessman and small time upcoming influencers to upscale their game and our model will forecast likes on their reels and pictures.In order for the successful implementation of this project we had to employ and get well versed with many soft-wares,tools and programming languages. We have elaborated and expounded more on it in the below stanzas. Now, Ever since the platform moved away from reverse-chronological and to a curated content feed. For digital Marketeers they are like little data puzzles just waiting to be solved. Recent data suggest that 72 percent of executives surveyed, social data is already the top data source used to inform business decisions, even more so than market research. 85 percent agree that social data will be primary source of business intelligence and 90 percent of marketeers agree that social data enables them to stay ahead of their competitors, resulting in help of social media listening tools.

2 Languages and Tools

2.1 Python

Python or python programming knowledge is a Sophisticated,catholic and an all round programming language.Its uses are wide ranged but is mainly used in Machine learning applications and Web designing. Since its codes and programs are comparatively timid to those of C++ and

Table 1. Box of Languages and Tools

Python	Pandas	Numpy
NLP	Ensemble larning	Matplotlib
Wordcloud		

java- Its been widely used and hence for time saving purpose we have employed python for our project. Python is being employed by IT giants all across the globe. Companies such as Facebook,Google and Amazon have employed python for various projects. Since our project deals with Instagram and Machine learning. Python is the most feasible option.

2.2 Pandas

Pandas is basically used for supporting Multi dimensional arrays and hence vastly used in the field of Data Science, Data Analysis and Machine learning. Since our project deals with a data set that needs to be imported and analysed for the further steps, Pandas needs to be employed. Pandas help in exploring a correlation between columns and rows and tells us about the minimum and maximum value and also the average of the data set

2.3 Numpy

Numpy or Numerical Python acts as a library which wholly consists of arrays and collection of objects which will be beneficial for the processing of those arrays. Due to to easy accessability and usage (i.e a person with any background can be versed with Numpy) we are using Numpy The data structure is Numpy is called as ndarray or N dimensional array. Array are basically the set of values that we will be collecting for our project.

2.4 NLP

NLP or Natural Language processing helps in making the Natural Human language that can be understood and used further for computer programs. NLP or Natural Language processing with python pre-process the un-arranged, un-structured Human readable data before analysing it in the program. Since we will be using a huge unstructured data, organising it on manual basis will be a tedious and laborious task. Hence, NLP is used to organise the unstructured data and is surely beneficial for our project.

2.5 Ensemble Learning

In Ensemble learning many base models are employed to produce one single optimal predictive tools. Since our project will help predict likes of a post, Ensemble technique is the key tool. It also helps to increase the generalisation of the data. it is the best technique to increase the performance of a predictive model. Predictive model are the essence of machine learning.Better the accuracy,Better the model.Its a sought of a direct relationship.

2.6 Matplotlib

Matplotlib helps with the visualisation in python. As humans, we bette that things that we can visualiseor thast we actually see. The rate of inference is better if we visualise an object or a concept. It helps in creating 2D plot of arrays that we discuseed above. It consists of several plots such as lines, bars and histogram.This willbe used by us to track the popularity of posts on a Daily/Monthly/Yearly basis based on the user. It will help in increasing our spectrum on the understanding of correlations and patterns.

2.7 Word Cloud

Word cloud is employed to determine the size and frequency of each word along with its importance in the dataset. It is also used to analyse data from social networking sites and hence will help us in th prediction model. One distinct feature of the word cloud is that the bigger and bolder the word appears,more distinct and frequently used it will be in the data set.

3 Working

Firstly we will be needing a data set, for which we will be creating our own data set starting with 200 entries and more entries can be added later on for better persistence and accuracy.

3.1 Data Import

We will start with the job of utilizing python to develop a social media algorithm using Machine Learning. We will we creating our own data set staring with 200 entries. It will consist of various categories such as Hashtags,followers,caption,time since posted and last but not the least-Likes.
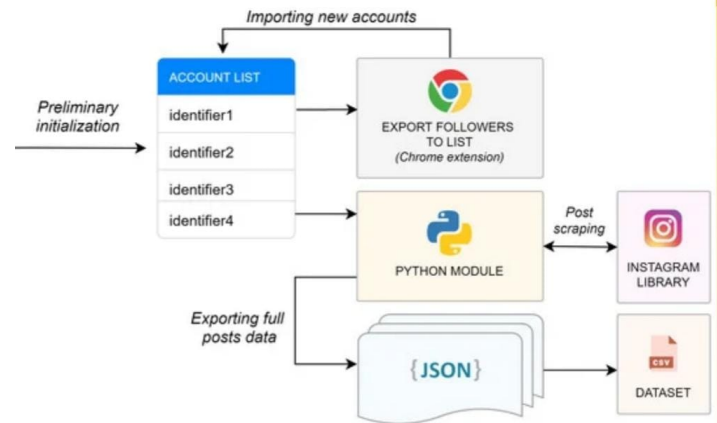

Figure 1. Flow diagram of working

3.2 Understanding frequency of words

We have imported some more Python libraries and write a function to analyse the word cloud in order to figure out what the most common types of words are used in Instagram.We have also used Word cloud plotter to plot the frequency of words used in Instagram.There are several words used in Instagram and all of it has been plotted on the Word cloud plotter.

3.3 Plotting of Regplot

Here based on the number of followers we have since the time the picture or reel is posted on Instagram.We can forecast the number of likes on a particular post/reel. Expounding on the point made prior, we have plotted 2 reg plots. First is followers to likes graph and second is followers to time since posted graph.In the first graph the relationship is directly proportional i.e less the followers-lesser the number of likes and more the followers- more are the number of likes. Whereas in the second regression plot does not really establish any relationship. The number of likes that is predicted is scattered throughout the x-y plane.

3.4 Training of Model

For our model to forecast value, that is number of likes in our case we firstly created our own data set. A gradient regressor plot is used that is basically a x-y plane graph that depicts the true value and predicted values in the former and the latter. The output concentration us found to be more towards the lower end values in the x-y plane. Less concentration towards the higher values. This was mainly because we were using a data set of only 200 entries, that is only 200 followers of our Instagram account. Secondly, a graph was plotted in the x-y plane depicting hours since

posted and the number of likes. This was a bit tricky as the engagement of our followers came in picture while depicting this output. This particular graph was plotted with 100 constant followers.Than we entered 100 more followers that makes it 200 followers and repeated the same step. We had accurate results and engagement also increased which increased the number of likes. Last but not the least we added 800 more followers to the data-set making it a set of 1000 followers and the results were highly forecasting the predicted number of likes in real time.

4 Features

There as certain key features of our model as discussed and expounded below.

4.1 Discovering new sales leads easily

Our model is trained in such a way that any user who are mainly businessman can discover new sales leads easily on Instagram by keeping themselves updated with the changing trends. Instagram changes its algorithm and trends almost once every year. So our model will adapt to these changing trends first and hence show users the most accurate number of likes they will receive.

4.2 Reaching balanced accuracy

The more data and categories are fed to our model, better will be its accuracy. At no state we are saying its 100 percent accurate, it needs to be constantly fed with data. While we started we 200 entries our forecasting was not up to the mark. But now since we have reached 1000 entries we receive a much better prediction in terms of likes.

4.3 User Friendly

We have built our model in such a way that any indivisual across the globe who owns an electronic device and has access to the internet can easily use of model wherever he/she resides.

5 Motivation and Scope

The main motivation behind the idea of implementing this project is to implement our knowledge of Machine learning and Digital marketing so as to create a model which will be very beneficial for small businessman/businesswoman who had taken a huge blow ever since the covid era in terms of sales and profits. We wanted to create something for the betterment of small business minded people with big dreams. The total cost for the preparation,building and training of the model is zero and we are quite certain this will change the phase of online business's on social media particularly Instagram. The scope is simple- our model will be very crucial for small scale business people to improve their quality of posts on Instagram which in return will be helpful for more of their sales.A small time influencers who wants to make a career on Instagram as an "Instagram Influencer"will find our model to be very helpful for his/her up-scalement and betterment. For better accuracy more entries can be added and our model is very user friendly so in lehman's english anyone and everyone who has the access to an electronic device and internet can use our model.

6 Conclusion

Based on the prediction above, we can observe one thing: if you have a higher number of subscribers, your post is more likely to get more likes in attendance for life. This was explained in the 'Training of Model' (Refer section 3.4). But the maximum number of likes will not increase more than 20 times the number of followers so the number of followers only contributes to an increase of 20 percent in likes.

References

[1] Stockholm, Sweden 2018 deep text mining of Instagram data Journal **1**, page numbers (2018)

[2] Salvatore Carta, *Pouplarity Prediction Of Instagram posts* (Diego Reforgiato,2021)

[3] Pradeep Kumar and Abdul Wahid, Social media analysis for sentiment classification using gradient boosting 2017