

Project Synopsis

Predicting Instagram Influencers using ML

Prepared by

Aishwarya More

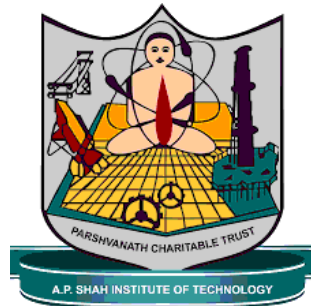
Shweta Patil

Preksha Vora

Moksha Shah

Under the guidance of

Prof. Shafaque Ma'am



Bachelor of Engineering (BE)

Department of Computer Engineering

A.P Shah Institute of Technology, Thane

University of Mumbai

2021-2022

Predicting Instagram Influencers using ML

Abstract:

Predicting the popularity of posts on social networks has taken on significant importance in recent years, and several social media management tools now offer solutions to improve and optimize the quality of published content and to enhance the attractiveness of companies and organizations. Scientific research has recently moved in this direction, with the aim of exploiting advanced techniques such as machine learning, deep learning, natural language processing, etc., to support such tools. In light of the above, in this work we aim to address the challenge of predicting the popularity of a future post on Instagram, by defining the problem as a classification task and by proposing an original approach based on Gradient Boosting and feature engineering, which led us to promising experimental results.

- **Introduction**

Benefit of Machine Learning is allows the user to feed a computer algorithm an immense amount of data and have the computer analyze and make data-driven recommendations and decisions based on only the input data. The basic objective of AI (also called heuristic programming, machine intelligence, or the simulation of cognitive behavior) is to enable computers to perform such intellectual tasks as decision making, problem solving, perception, understanding human communication (in any language, and translate among them).

Basis on this , we are trying to make system that will be helpful for brands which are trying to collaborate with Instagram influences based in their engagement rate and growing and declining phases in basis of their recent activities.

Instagram influencers are regular Instagram users but with a unique ability to influence others, well-established credibility, and a considerably large audience. From a marketing perspective, an influencer is defined as a person with the power to influence potential buyers of a product or service.

Objective is to make system through which brands will be able to find their influencers which will increase their potential buyers with help of Machine learning Algorithms and Sentimental Analysis.

- **Literature Review**

With the advent of technology and digital transformation, the Internet has evolved from being an information exchange source from limited sources to an open platform where everyone on this planet can access and create usergenerated content. Some user generated content gave rise to Influencers who assist certain marketing activities by promoting a product or services and influencing its audience (viewers) to buy/use the product or services (Duncan Brown, 2008). Influencer Marketing is flourishing and gearing up by giving out effective results and more marketing managers are opting Influencer Marketing as an alternative to reach a wider audience on social media with a very short span of time (Evans, 2017). However, it is important for managers to select

right influencers for their products and services to broadcast a right image in the minds of the audience.

- **Project Concept**

We will be using machine learning Algorithm to calculate engagement rates of Instagram Influencers using python libraries and ML .

- 1.Retrieving data from Instagram influencers using Selenium and BeautifulSoup.
- 2.Preprocessing data starts from data cleansing, feature engineering, feature selection and etc until the data is ready to be consumed by the Machine Learning model.
- 3.Modeling uses Machine Learning Algorithm (Random Forest)
- 4.Interpreting of results from prediction output of Machine Learning.

We will be using basic python programming and various libraries like -

1. Pandas and Numpy libraries for data analysis tools.
2. Matplotlib and Seaborn libraries for data visualization.
- 3.. Scikit-Learn Library for Machine Learning.
- 4.. Selenium and BeautifulSoup libraries for acquiring instagram data
5. Jupyter Notebook.

After getting dataset online, we will be taking post links from every influencer on Instagram using Selenium.

Next, Retrieve information from each post such as number of likes, number of comments, captions etc. using BeautifulSoup.

Next, Preprocessing data starts from data cleansing, feature engineering, feature selection until the data is ready to be consumed by the Machine Learning model.

Next, Interpreting the results

Making Front-end using Web development tools, for Brands to select which category they want influencers and List of all Influencers will be displayed, when brand select on any influence it will be shown whether the influencer is in growing phase of declining.. On that basis the brand can decide whether to collaborate with that Influencer or not.

- **Problem Definition**

The challenge is to recommend influencer(s) for a particular marketing campaign. Companies might be interested in micro influencers for a certain industry/region and they would also like to understand which content works for the audience of a particular influencer.

After finding the right influencers they might want to start a marketing campaign with them and monitor the efficiency of such a campaign.

we built a smart search algorithm that would take into account these attributes as well as various metrics. Finally, we designed a machine learning based recommender that links content with the influencers based on their audience. This content would serve as an inspiration to the campaign manager.

- **Objectives**

1. Study Influencer Marketing and its process as one of the emerging field of marketing in digital space.
2. Overview of Sentiment Analysis and its effects and consequence
3. Exploring the practice for deciphering the parameters of influence within the application of sentiment analysis to evaluate the effectiveness of Influencer
4. Explore the opportunities and challenges of using sentiment analysis to evaluate Influencer Marketing

- **Scope**

1. The main idea behind this project is to develop a system which can recommend influencer(s) for a particular marketing campaign.
2. Analyze the sentiment of the comment section of the post and identify the number of positive and negative comments.
3. Classify by % of female and male followers in profile. I.e. e.g. It would be easy for a cosmetic brand to analyze and influence them with maximum female followers because profiles maximum male followers will not give them so much commitment.

- **Benefits For Environment and society**

- You can develop a more insightful, data-based marketing strategy. Nothing beats data-based strategy. ...
- Understand your customers. ...
- Measure your marketing campaign. ...
- Take a look at brand perception. ...
- Find industry leaders and influencers. ...
- Give extra boost to your customer service.

- **Technology Stack**

- Python is a general-purpose interpreted, interactive, object-oriented, and high-level programming language. Python is designed to be highly readable.
- Scikit-learn is a free software machine learning library for the Python programming

language.

- Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python.
- NumPy stands for Numerical Python. NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices. NumPy aims to provide an array object that is up to 50x faster than traditional Python lists.
- Matplotlib is a low level graph plotting library in python that serves as a visualization utility. Matplotlib is open source and we can use it freely.
- Algorithm:
- Random forest
- Random forest is a flexible, easy to use machine learning algorithm that produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity (it can be used for both classification and regression tasks). In this post we'll learn how the random forest algorithm works, how it differs from other algorithms and how to use it
- It is a supervised learning algorithm. ... The general idea of the bagging method is that a combination of learning models increases the overall result. Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.
- Another great quality of the random forest algorithm is that it is very easy to measure the relative importance of each feature on the prediction.
- NumPy is a Python package. It stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of arrays.
- Numeric, the ancestor of NumPy, was developed by Jim Hugunin. Another package
- Numarray was also developed, having some additional functionalities. In 2005, Travis Oliphant created the NumPy package by incorporating the features of Numarray into the
- Numeric package. There are many contributors to this open source project

