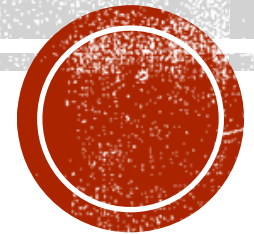


# INFLUENCER POPULARITY PREDICTOR



By

Joyna Joy Perinchery

# ABSTRACT

This project, Influencer Popularity Predictor, aims to analyse and predict the popularity of social media influencers, with a focus on YouTube, while also considering their presence on platforms such as Twitter and Instagram. The goal is to classify influencers such as Viral, Trending, Emerging, or Niche, based on metrics like subscriber count, video views, likes, engagement rate, and sentiment from user comments.

I began by manually collecting a list of influencer YouTube handles from Socialblade and saved them in a CSV file. Using the YouTube Data API and Python, to automatically fetch additional details for each influencer

Using this dataset, I am performing a classification task to predict an influencer's popularity label. Additionally, I'm analyzing whether a YouTube influencer is also popular on Twitter or Instagram by comparing their follower counts cross-platform popularity.

The final part of this project involves building a Flask web application where users can enter influencer metrics and get a real-time prediction of their popularity level. This app can also suggest if the influencer is gaining more traction on YouTube, Twitter, or Instagram.



# INTRODUCTION TO DATASETS

The dataset contains metrics for influencer profiles across YouTube, Twitter, and Instagram, used to predict popularity trends and classify influencers' popularity on social platforms. The data includes 1514 profiles of influencers and 24 columns

## Column Explanations

**youtube\_handle**: unique youtube channel identifier (@mrbeast).

**subscriber\_count**: number of youtube subscribers ( 402m).

**total\_videos**: total videos uploaded on youtube ( 875).

**Country**: influencer's country of origin ( USA).

**influencer\_name**: name of the influencer or channel (MrBeast).

**channel\_id**: unique youtube channel id (UCX6OQ3DkcsbYNE6H8uQQQuVA).

**Most\_viewed\_video\_url**: url of the channel's most viewed video.

**Content\_Type**: type of content (Entertainment, Music..).

**upload\_frequency\_per\_Month**: average videos uploaded per month (5).

**channel\_created\_year**: year the youtube channel was created (2012).



**avg\_views\_per\_video**: average views per video (113717113.3).  
**avg\_likes\_per\_video**: average likes per video ( 2586568.9).  
**avg\_comments\_per\_video**: average comments per video (33567.3).  
**sentiment\_label**: sentiment of content (neutral, positive, negative).  
**toxicity\_score**: measure of content toxicity (0.08).  
**popular\_on\_twitter**: twitter popularity status (highly popular, less popular, no account).  
**twitter\_follower\_count**: number of twitter followers ( 45.9m).  
**popular\_on\_instagram**: instagram popularity status (highly popular, less popular).  
**instagram\_follower\_count**: number of instagram followers (115m).  
**popularity\_label**: overall popularity category (viral, trending, emerging, niche).  
**Virality Score**: metric indicating content virality (2.27).  
**GrowthRate\_over\_time**: rate of channel growth over time 129814.06).



# FUTURE FOCUS

## Data Preprocessing + EDA (Exploratory Data Analysis)

1. cleaning data, preparing it for the model
2. handling missing values and duplicate data.
3. normalized numbers
  - scaled big numbers (like subscriber count or views) so everything is on a similar range
4. encoded platform data
5. created new features to make the model smarter:
  1. engagement rate =  $(\text{likes} + \text{comments}) / \text{followers}$  — tells how active the audience is.
  2. post frequency =  $\text{total posts} / \text{account age}$  — shows how often the influencer posts content.
6. explored the data (eda)
  - i looked at the overall distribution of numbers (who's getting how many likes, views, etc.).
  - made correlation heatmaps to see which features are related to each other.



**THANK YOU**

