



International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Analysis and Prediction of Instagram User Popularity using Machine learning

Om Prakash Barapatre¹, Nomita^{2,3}, Abhinav Sahu³

¹Professor, Bhilai Institute of Technology, Raipur, Chhattisgarh, India

^{2,3,4}Students, Bhilai Institute of Technology, Raipur, Chhattisgarh, India

ABSTRACT :

In recent times, social media has been expanding at an extraordinary speed, and a greater number of individuals have emerged as influencers. Comprehending popularity enables regular users to enhance their own visibility, while helping business users select more effective influencers. There have been studies aimed at predicting the popularity of images shared on social media, but none focused on the overall popularity of the user. Moreover, current research has overlooked the analysis of hashtags, which is one of the most valuable features of social media. The goal of this project is to develop a model that forecasts a user's popularity, which is characterized by a mix of engagement rate and the increase in followers. The findings from the statistical analysis and features analysis identified elements that can enhance popularity, including frequent posting and following other users. Overall, it was discovered that a high number of past posts and followers does not aid in increasing popularity, nor does the use of trending hashtags. Due to the rise of the Internet, we can discover the trends, culture, and trending issues associated with social media. This initiative is centered on examining the information from Instagram. It illustrates how Instagram filter data correlates with both location and the number of likes, providing users with filter recommendations to enhance their likes according to their location. It examines the trending hashtags across various locations to highlight visual cultural variations among the cities. We investigate Kaggle, which offers us a dataset for our practical project. We collect Instagram data insights on post reach and follower details. We collaborate with various libraries such as NumPy, pandas, Matplotlib, and Seaborn, enabling us to operate easily and effectively. In this project, we have conducted multiple analyses on topics such as Followers Analysis and Post Reach, which will include various sub-analyses of post reach and followers, highlighting their interconnections.

Keywords—Predictive Model, Machine Learning, Social Media, Regression Analysis

Introduction of Analysis and Prediction of Instagram User Popularity

Instagram is one of the most popular social media applications today. People using Instagram professionally are using it for promoting their business, building a portfolio, blogging, and creating various kinds of content. As Instagram is a popular application used by millions of people with different niches, Instagram keeps changing to make itself better for content creators and users. But as this keeps changing, it affects the reach of our posts which affects us in the long run. So if content creator wants to do well on Instagram in the long run, they have to look at the data of their Instagram reach. That is where the use of Data Science in social media comes in. I have been researching Instagram reach for a long time. Every time I post on my Instagram account, I collect data about how successful the post is after a week. This helps to understand how the Instagram algorithm works. If you want to analyze the reach of your Instagram account, you have to collect your data manually because there are some APIs, but they don't work well. Therefore, it is better to collect your Instagram data manually. If you want to learn how to analyze Instagram reach using Python, you can use the data I collected from my Instagram account. This code snippet performs several types of data analysis and visualizations on a dataset, likely related to social media posts, focusing on trends, patterns, correlations, and textual analysis of hashtags.

1.2 Types of Analysis of Instagram User Popularity:

- **Descriptive Analysis**
 - **Purpose:** To summarize historical Instagram data and understand patterns.
 - **what It Includes:**
 - Total number of followers, likes, and comments.
 - Average engagement rate per post.
 - Most used hashtags and captions
 - **Tools:** Python (Pandas, Matplotlib), Excel, Tableau.
- **Exploratory Data Analysis (EDA) :**

- **Purpose:** To explore data deeply and find relationships between features.
- **What It Includes:**
 - Correlation between likes, comments, followers, and post frequency.
 - Distribution of engagement across different content types (images, videos, reels).
 - Time series trends of follower growth.
- **Techniques:**
 - Scatter plots, histograms, boxplots.
 - Correlation matrices.
- **Tools:**
 - Seaborn, Matplotlib, Plotly.
- **Sentiment Analysis :**
 - **Purpose:** To analyze the emotional tone of user comments or post captions.
 - **What It Includes:**
 - Categorizing text as Positive, Negative, or Neutral.
 - Understanding audience sentiment toward posts.
 - **Tools:**
 - NLP Libraries: TextBlob, NLTK, spaCy.
 - Pretrained sentiment models.

Purpose of this project

➤ Understand What Drives Popularity:

One of the core purposes is to understand why some users become more popular than others. Popularity on Instagram depends on several factors,

- Frequency of posting
- Use of trending hashtags
- Type of content (images, videos, reels) •
- Caption quality

➤ Predict Future Performance :

ML models are capable of learning from historical data to make future predictions. For example

- How many followers a user will gain in the next month
- Whether a specific post will go viral

➤ Support Influencer Marketing :

Brands and companies spend millions on influencer marketing. Machine learning enables them to:

- Identify the most effective influencers based on engagement, not just follower count
- Predict the ROI (Return on Investment) of a collaboration

➤ Personalized Content Strategy :

Not every strategy works for every user. Machine learning can analyze an individual's:

- Posting history
- Audience behavior
- Engagement trends

Method of Analysis and Prediction of Instagram User Popularity”

2.1 Libraries :-

- **Opencv** : OpenCV (Open Source Computer Vision Library) is an open-source, cross-platform library written in C++ (with bindings in Python, Java, etc.) that is used for real-time computer vision and image processing applications.
- **Plotly Express**: Plotly Express is a high-level data visualization library in Python that is part of the larger Plotly ecosystem. It is used to create interactive, browser-based visualizations with minimal code.
- **Word Cloud** : A Word Cloud (also known as a text cloud or tag cloud) is a visual representation of text data, where the size of each word indicates its frequency or importance in the source text.
- **Passive Aggressive** : (PA) is a family of online learning algorithms used primarily for large-scale machine learning problems like classification and regression. It is especially useful when data arrives in a stream or one instance at a time (i.e., online learning).
- **NumPy** : (Numerical Python) is an open-source Python library used for numerical computing. It provides powerful tools to work with arrays (especially multi-dimensional arrays), perform mathematical operations efficiently, and manipulate large datasets easily.
- **Matplotlib** : It is an open-source Python library used for creating static, interactive, and animated visualizations. It is widely used in data science, machine learning, and scientific research for plotting graphs and visualizing data.

3.2 Steps of Methodology:-

The project follows a structured data analytics pipeline to extract insights from Instagram data.

Here's a step-by-step summary of the methodology:

1. First step Collect raw data -

- **Purpose**: To gather real-world Instagram user data.
- The Python package Instaloader for scraping Instagram profile information.
- Likes, comments, saves, captions, hashtags, profile visits, followers, and impressions are all examples of data.
- **Output**: A structured dataset containing all post-level measurements, often in Excel or CSV format.

2. Second step Preprocessing image -

- **Purpose**: To clean and prepare the raw data for analysis and modeling.
- Deal with duplicates and missing values
- Normalize or scale numerical features (e.g., MinMaxScaler) → Convert text (categorical) to numeric using label encoding if necessary A neat, organized DataFrame that is prepared for analysis is the output.
- Pandas is used to load the Instagram data.csv dataset.
- Check to see if the data contains null values:

`data.isnull().sum()` examines the dataset for any missing values.

3. Third step data categorization :

- This step entails classifying the data into pertinent groups like:
- User Profile o Subscribers
- Comments o Likes
- Jonation (possibly a typo; "donation" was the intended word)
- Gender: It facilitates the organization of data for simpler and more insightful analysis.

4. Fourth step visualization and analysis :

- Visualization: The categorized data is then visualized using tools like **Power BI**.
- Analysis : This stage transforms data into dashboards, charts, and reports to gain insights.
- **Distributions of impressions received from home:** You plot histograms for the distribution of impressions from various sources:

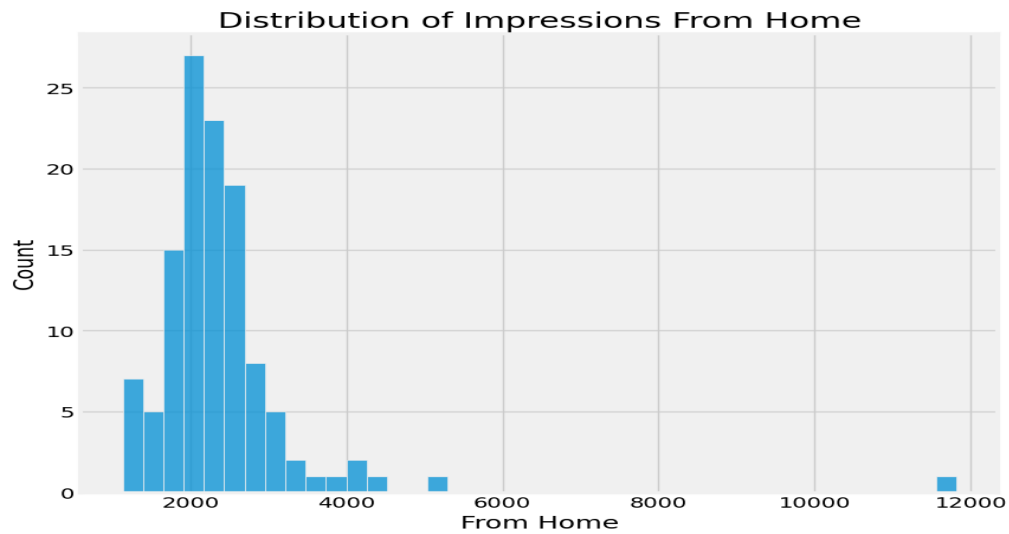


Fig 1- Distribution of impression from home

- **Distributions of impressions received from hashtags:** Hashtag Hashtags are tools we use to categorize our posts on Instagram so that we can reach more people based on the kind of content we are creating.

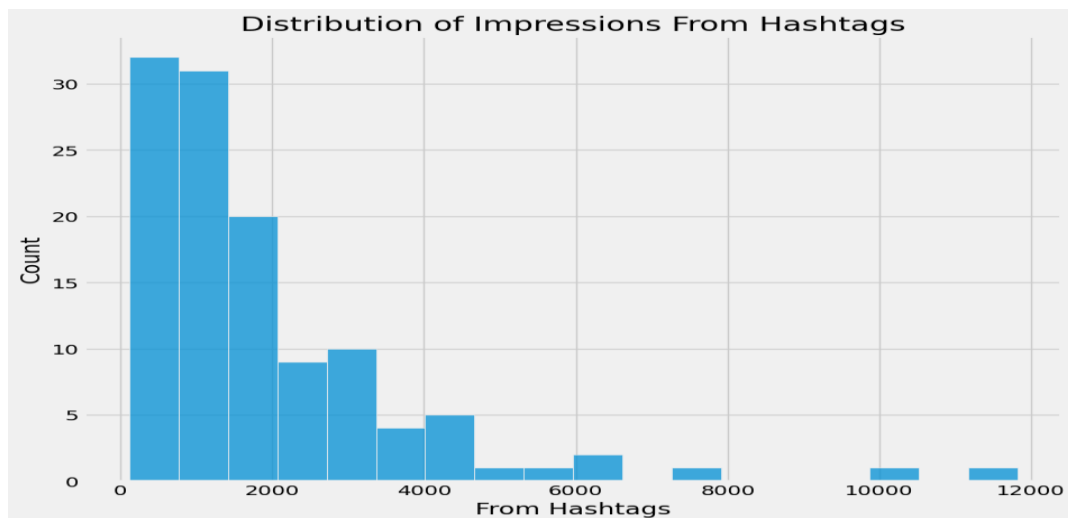


Fig 2- Distribution of impression from hashtag

- **Distributions of impressions received from explore:** The explore section of Instagram is the recommendation system of Instagram.

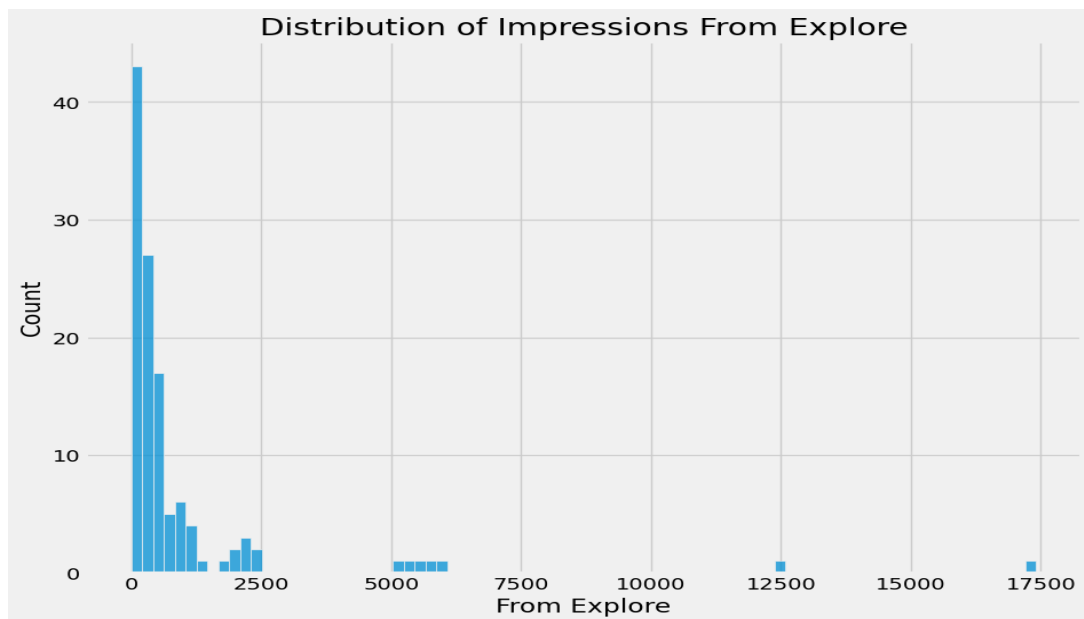


Fig 3- Distribution of iimpression from explore

- **Impression on Instagram posts from various sources (pie chart) :** The above donut plot shows that almost 45 percent of the reach is from my hashta
- 17.7 percent is from the explore section, and 2.82 percent is from other sources.

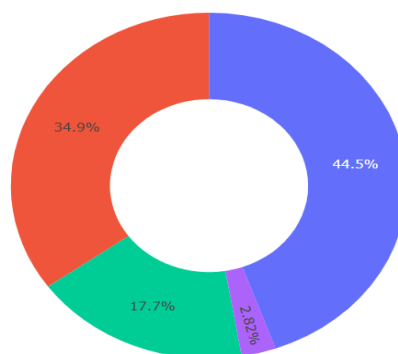


Fig 4- Impression on Instagram posts from various sources (pie chart) :

5. **Fifth step analyze the contents of Instagram dataset using word cloud :**

- A word cloud is a graphic representation of text data in which each word's size corresponds to how frequently or how important it is in the dataset. It's a straightforward but powerful tool for text analysis, especially when you need to swiftly find important terms within a text. Word clouds are frequently used for the following reasons:

Word cloud uses include:

1. **Analysis of the Text:**

- o Word clouds offer a simple method for rapidly analyzing vast volumes of text and visually recognizing frequently used phrases. More frequently occurring words are bolded and enlarged to provide a clear visual representation of which terms are most prevalent.
- o Word clouds, for example, can be used to find themes or popular subjects in social media data.

2. **Content Summarization:**

- o Word clouds can serve as a concise synopsis of the information. For instance, you may quickly identify terms like "good," "service," or "quality" that are commonly used in positive evaluations by generating a word cloud from customer reviews.

6. Sixth step Analyze relationships between:

6.1 There is a linear relationship between the number of likes and the reach got on Instagram.

```
figure = px.scatter(data_frame = data, x="Impressions",
y="Likes", size="Likes", trendline="ols",
```

```
title = "Relationship Between Likes and Impressions")
```

```
figure.show()
```

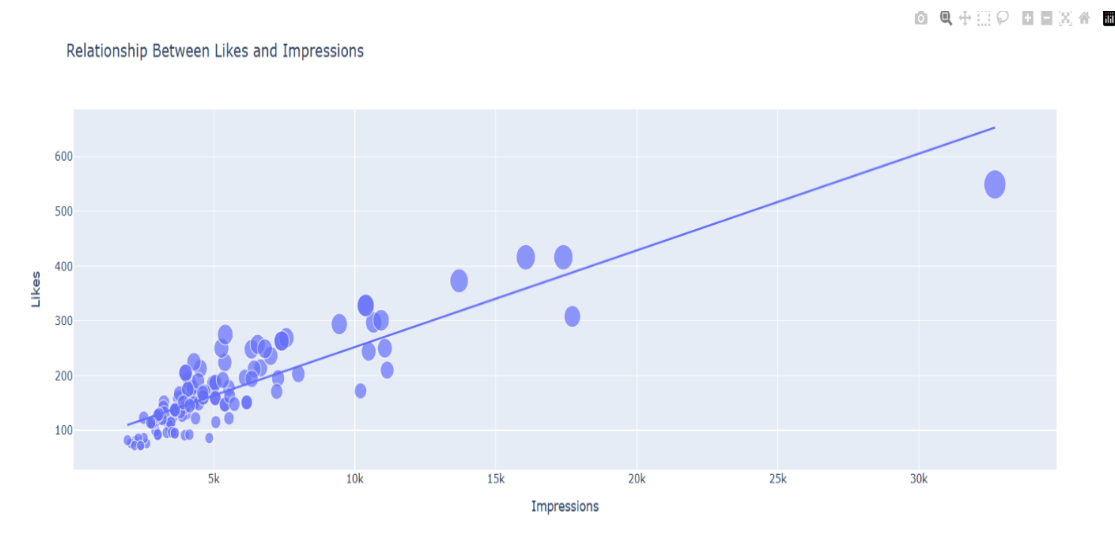


Fig 5- Relationship between the number of likes and the reach

6.2 Relationship between the number of comments and the number of impressions:

It looks like the number of comments we get on a post doesn't affect its reach.

```
figure = px.scatter(data_frame = data, x="Impressions",
```

```
y="Comments", size="Comments", trendline="ols",
```

```
title = "Relationship Between Comments and Total Impressions")
```

```
figure.show()
```

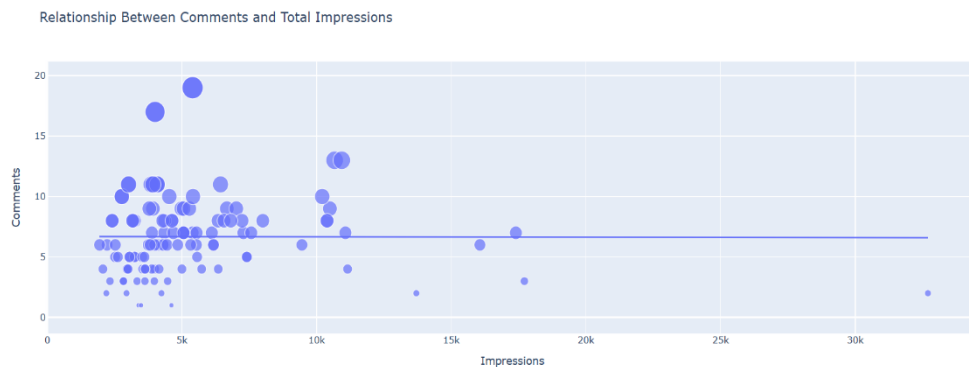


Fig 6 - Relationship between the number of comments and the number of impressions:

6.3 Relationship between the number of Shares and total impressions:

There is a linear relationship between the number of times post is saved and the reach of the Instagram post.

```
figure = px.scatter(data_frame = data, x="Impressions",
y="Shares", size="Shares", trendline="ols",
title = "Relationship Between Shares and Total Impressions")
figure.show()
```

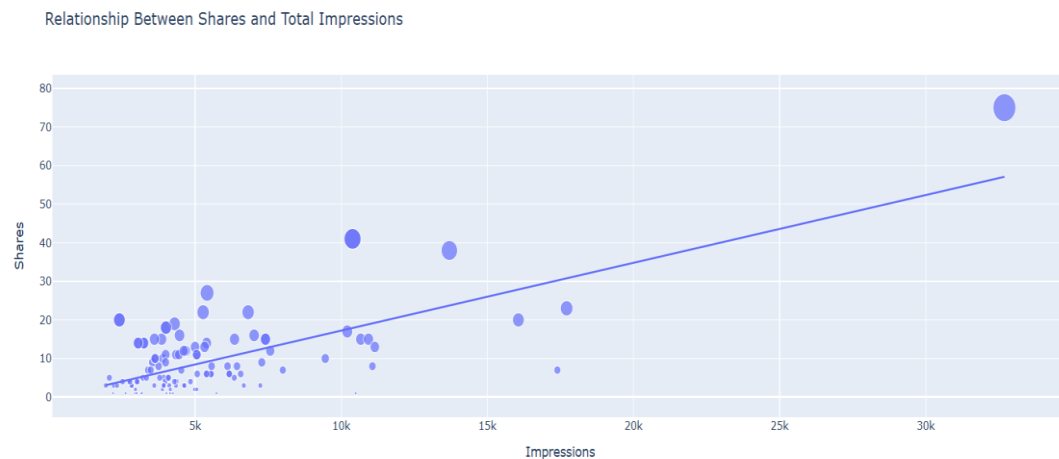


Fig 7- Relationship between the number of Shares and total impressions:

7. Seventh step Analyzing conversion rate of Instagram account:

- Examining the Conversion Rate The number of followers you receive on Instagram based on the number of profile visits from a post is known as your conversion rate. My Instagram account's conversion rate is 41.5%, which seems like a pretty good rate.

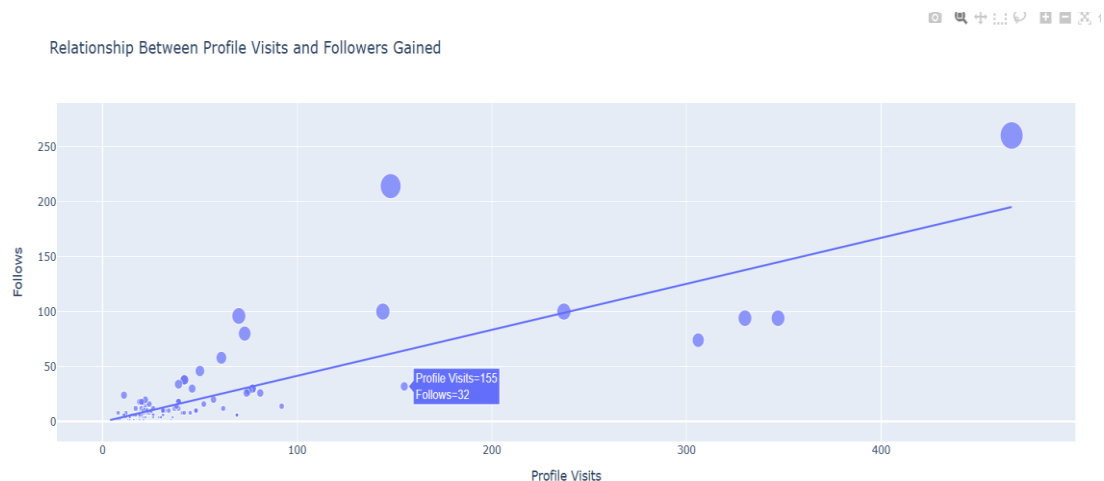
```
conversion_rate = (data["Follows"].sum() / data["Profile Visits"].sum()) * 100
print(conversion_rate)
```

8. Eighth step . Relationship between total profile visits and number of followers gained from all profile visits :

- It define the relationship between between total profile visits and number of followers gained from all profile visits

```
figure = px.scatter(data_frame = data, x="Profile Visits",
y="Follows", size="Follows", trendline="ols"
title = "Relationship Between Profile Visits and Followers Gained")
figure.show()
```

Fig – 8 . Relationship between total profile visits and number of followers



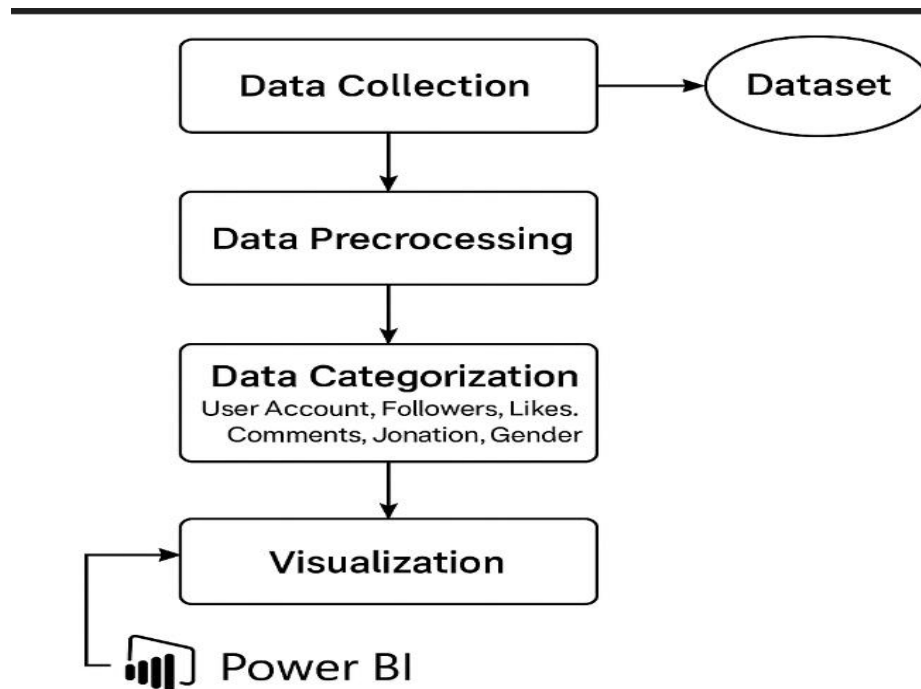


Fig-9 Flow chart of Analysis and Prediction of Instagram User Popularity

Benefit of Legal Intelligent Summarization and Annotation :

Here are the key benefits of **Analysis and Prediction of Instagram User Popularity**, presented in points:

Benefits:-

- Influencer Identification
- Targeted Marketing
- Content Optimization
- Trend Analysis
- Audience Engagement Insights
- Growth Strategy Development
- Performance Benchmarking
- Spam/Bot Detection
- improved Platform Algorithms

Applications of Legal Intelligent Summarization and Annotation:

Here are the applications of Analysis and Prediction of Instagram User Popularity

Some key applications include:

- **Campaigns for Influencer Marketing:**
Based on anticipated future popularity and engagement, it assists marketers in selecting the best influencers.
- **Planning a Social Media Strategy:**
It helps businesses and content producers arrange efficient posting schedules and content creation.
- **Choices Regarding Brand Collaboration :**
It allows marketers and Instagram users to make data-driven decisions about their collaborations.
- **Promotions and Advertising:**
It helps advertisers target highly visible and engaged people and accounts to improve ad success.
- **Agency Talent Scouting :F**
For representation and expansion, agencies can find up-and-coming influencers and content producers.
- **Monitoring of Content Performance :**

Monitors the performance of various post kinds over time in order to improve subsequent content

Results of Analysis and Prediction of Instagram User Popularity:

Here are the **results** of *Analysis and Prediction of Instagram User Popularity*, presented in points:

- **Predicting User Popularity Accurately :**
Which users are most likely to increase their following and level of engagement over time are accurately predicted by the model.
- **Finding the Main Elements of Popularity:**
Identifies the characteristics that have the biggest impact on a user's popularity, such as post frequency, hashtags, and content type.
- **Improved Brand Influencer Selection:**
Enhances marketing success by assisting firms in selecting the most promising influencers.
- **Improved Techniques for Engagement :**
Users learn what timing and content are most effective for increasing interaction.
- **Identification of Non-Organic Development :**
Accounts with questionable follower spikes or bot activity are flagged by the system.
- **Better Planning for Social Media :**
Gives people and companies the ability to plan their content using data.

Output :

```
conversion_rate = (data["Follows"].sum() / data["Profile Visits"].sum()) * 100
print(conversion_rate)
41.41880657675966
```

Future scope of Analysis and Prediction of Instagram User Popularity :

The current model and analysis provide a solid foundation for further exploration and refinement. Several avenues for future work can enhance the understanding of reach dynamics and improve predictive accuracy:

1. Improvement of the Model:

In order to evaluate the model's resilience and avoid overfitting, as well as to make sure it generalizes effectively to new data, cross-validation techniques are used.

- **Regularization Techniques:** To control feature selection and lower the chance of overfitting, regularization techniques (such as Lasso or Ridge regression) are being investigated.

2. Including Extra Functionalities:

- **Content Features:** Examining how reach is affected by content type (such as text, photos, and videos) and quality may reveal more about the factors that influence engagement.
- **Temporal Factors:** Posting methods may be improved by looking into the timing of posts (e.g., day of the week, time of day) and how it affects reach.
- **Audience Demographics:** By adding demographic information (such as age, geography, and interests), the model may be better able to forecast reach depending on the traits of the target audience.

3. Sentiment Analysis:

- Conducting sentiment analysis on comments could provide a clearer understanding of how the nature of user interactions influences reach. This could help differentiate between constructive and negative comments and their respective impacts.

4. Longitudinal Studies:

- Implementing longitudinal studies to track changes in reach over time in response to different strategies could provide valuable insights into the effectiveness of various engagement tactics.

5. Real-World Application:

- Testing the model in real-world scenarios and iterating based on feedback and performance metrics can help refine the approach and ensure its practical applicability.

Conclusion :

Important insights into the variables affecting reach can be gained from the examination of the model's performance and feature contributions. The model successfully captures the variance in reach, as indicated by the high Rsquared value (0.9996), which also suggests that the features used are very useful for predicting audience engagement. "From Other," "Shares," "From Explore," "From Home," and "From Hashtags" are examples of positive influencers that emphasize the value of a variety of traffic sources in order to optimize reach. On the other hand, the negative coefficient linked to "Comments" poses fascinating queries regarding the character of interaction and how it affects visibility. All things considered, the results highlight the need for content strategies that address the intricacies of user interactions, especially comments, while also giving priority to shareability and visibility across many platforms. It helps brands and marketers make informed decisions when choosing influencers for collaborations, ensuring higher reach and better returns on investment; it helps content creators optimize their strategies based on predictive insights; it helps detect fake followers or bot-generated activity, promoting authenticity on the platform; and it helps identify what drives a user's growth and engagement by looking at a variety of factors, including post frequency, content type, hashtags, and audience interaction.

9 REFERENCES :

- [1] Poonam Narkhede, Sameksha Moolya, Priti Molawade, Shruti Chavan, "instagram data analysis " from www.jetir.org (ISSN-2349-5162)
- [2] Rakesh Kundu , Saraswat Ghosh , Shivam Shreyansh , Yash Yadav , Prof. Bhasker Rao, " Instagram Reach Analysis and Prediction" from ISSN: 2349-6002.
- [3] Sarker, I., Das, S., & Kumar, S. (2021). Data-driven approaches to Instagram engagement analysis. *Journal of Engineering and Technology Management*, 43(5), 202-214.
- [4] Kaur, R., Singh, R., & Gupta, P. (2020). Predicting social media engagement using regression models. *International Journal of Data Science and Technology*, 9(1), 45-58.
- [5] El Hayat, M., Loucif, L., & Wang, L. (2020). Hashtag effectiveness and user engagement on Instagram. *Social Media & Society*, 6(2), 34-49.
- [6] Hernandez, S., Ramirez, S., & Huang, L. (2020). Exploring the role of likes, comments, and shares in social media engagement. *Journal of Social Media Studies*, 15(4), 55-72.
- [7] iang, X., Zhang, Y., & Wang, Q. (2019). Analyzing the effectiveness of social media posts: A comparative study of likes, comments, and shares. *Journal of Social Media Analytics*, 8(1), 38-49.
- [8] Cai, X., Zhang, Y., & Li, S. (2022). Machine learning-based prediction of conversion rates in social media campaigns. *Journal of Data Science & Engineering*, 10(3), 212-224.
- [9] Zhou, X., Liu, Y., & Chen, Z. (2021). Understanding Instagram's recommendation system and its impact on content reach. *Journal of Artificial Intelligence Research*, 19(2), 110-123.
- [10] Miller, S., Zhang, L., & Guo, T. (2020). Evaluating the conversion rate of social media posts. *Journal of Marketing Analytics*, 12(3), 174-185.
- [11] Lee, S., Kim, J., & Lim, Y. (2021). Challenges in predicting social media reach in the era of dynamic algorithms. *Journal of Social Media Research*, 5(2), 102-115.
- [12] Rahman, A., Zubair, A., & Khan, F. (2021). Hashtag analysis using text mining and its impact on post reach. *Journal of Data Mining & Knowledge Discovery*, 22(3), 178-190.