Instagram Reach Analysis and Prediction

1. Introduction

This project focuses on analyzing Instagram post reach and developing a machine learning model to predict impressions based on user engagement metrics. The analysis covers:

- Sources of impressions (home, hashtags, explore, etc.)
- Content analysis (captions and hashtags)
- Relationships between key metrics (likes, comments, shares, etc.)
- Development of a regression model for reach prediction.

---

2. Data Cleaning and Preprocessing

- **Missing Values**: Null values were present in all columns and were removed to ensure data quality.
- **Data Types**: Insights into column data types were extracted to confirm compatibility with numerical analyses.

---

3. Insights from Analysis

3.1 Distribution of Impressions

- **Home**: Most impressions came from followers but showed a limited daily reach.
- **Hashtags**: Posts reached many new users, but not all hashtags were effective.
- **Explore Section**: The reach from Instagram's recommendation system was minimal, suggesting limited visibility through this channel.

3.2 Source Distribution

A donut chart revealed the percentage of impressions from different sources:

- 50% from followers (Home)
- 38.1% from hashtags
- 9.14% from explore
- 3.01% from other sources

---

4. Content Analysis

Word clouds of post captions and hashtags highlighted frequently used terms, revealing trends in content creation.

---

5. Relationship Analysis

- **Likes vs. Impressions**: A strong positive correlation indicated that more likes lead to higher reach.
- **Comments vs. Impressions**: Comments had little to no impact on reach.
- **Shares vs. Impressions**: Shares contributed to reach but were not as impactful as likes.
- **Saves vs. Impressions**: A strong linear relationship suggested that saves significantly boost reach.

---

6. Conversion Rate Analysis

- **Formula**: (Follows/Profile Visits)×100(\text{Follows} / \text{Profile Visits}) \times 100(Follows/Profile Visits)×100
- **Result**: A high conversion rate of 41% indicated effective engagement.

---

7. Machine Learning Model

7.1 Model Selection

- **Algorithm**: Passive Aggressive Regressor
- **Features**: Likes, Saves, Comments, Shares, Profile Visits, Follows
- **Target**: Impressions

7.2 Training and Evaluation

- **Performance Metrics**:
    - Mean Cross-Validation Score: [insert value from code output]
    - Mean Squared Error: [insert value from code output]
    - R-squared Score: [insert value from code output]

7.3 Hyperparameter Tuning

GridSearchCV was used to optimize hyperparameters, improving model performance.

---

8. Prediction

The trained model was used to predict impressions for new posts, demonstrating its utility in forecasting reach based on engagement metrics.

---

9. Conclusion

This project successfully analyzed Instagram post reach, providing insights into effective strategies for maximizing visibility. The machine learning model offers a practical tool for predicting impressions, aiding data-driven decision-making for social media content strategies.