

Instagram Likes Prediction: Key Findings and Insights

1. Objective

The goal of this project was to predict the number of likes that an Instagram post would receive based on various factors, including engagement metrics like comments and followers, as well as the time of posting, day of the week, and the month.

2. Data and Methodology

- **Dataset:** The dataset contained 3785 Instagram posts, each with metadata such as likes, comments, follower count, and the timestamp of the post.
- **Outlier Removal:** We removed 5% of posts with unusually low engagement based on their likes-to-followers ratio.
- **Feature Engineering:** Temporal features such as the hour of the day, day of the week, and month were extracted from the timestamp.
- **Models Used:**
 - **Linear Regression:** Used as a baseline.
 - **Random Forest Regression:** Employed to improve accuracy.

3. Key Results

- **Model Performance:**
 - **Linear Regression RMSE:** 177,402 – This high error suggests the model struggled to predict accurately.
 - **Random Forest RMSE:** 94,665 – A significant improvement over the baseline.

4. Most Influential Factors (Random Forest)

The Random Forest model highlighted these top predictors for likes:

1. **Number of Comments (Importance: 0.687):**
 - The most important factor. Posts with more comments consistently received more likes, suggesting engagement drives visibility and popularity.
2. **Follower Count (Importance: 0.190):**
 - Having more followers also increased the number of likes, but the effect wasn't as strong as comments.
3. **Hour of Posting (Importance: 0.054):**
 - Timing mattered, though not as much as engagement or followers. Posting at optimal times might help slightly.
4. **Day of Week (0.028) and Month (0.042):**
 - These had minimal impact on likes but still contributed to the model's accuracy.

5. Visual Findings

The data analysis confirmed some clear trends:

- **Number of Comments vs. Likes:** A strong positive correlation. More comments usually meant more likes.
- **Follower Count vs. Likes:** A moderate positive relationship, with more followers generally leading to more likes.

6. Conclusion

- **Top Predictor:** The number of comments was by far the strongest factor in predicting likes.
- **Key Takeaway:** To maximize likes, focus on creating content that encourages comments and engagement.
- **Posting Time:** While timing had some influence, engagement through comments is the most effective strategy.