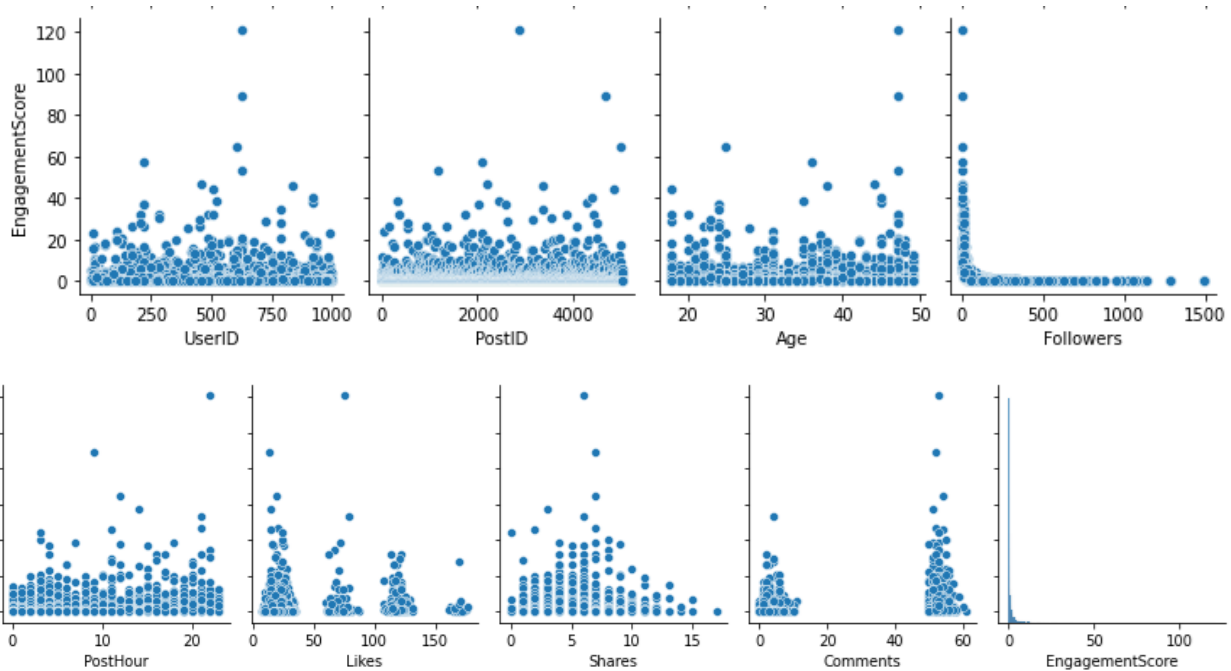


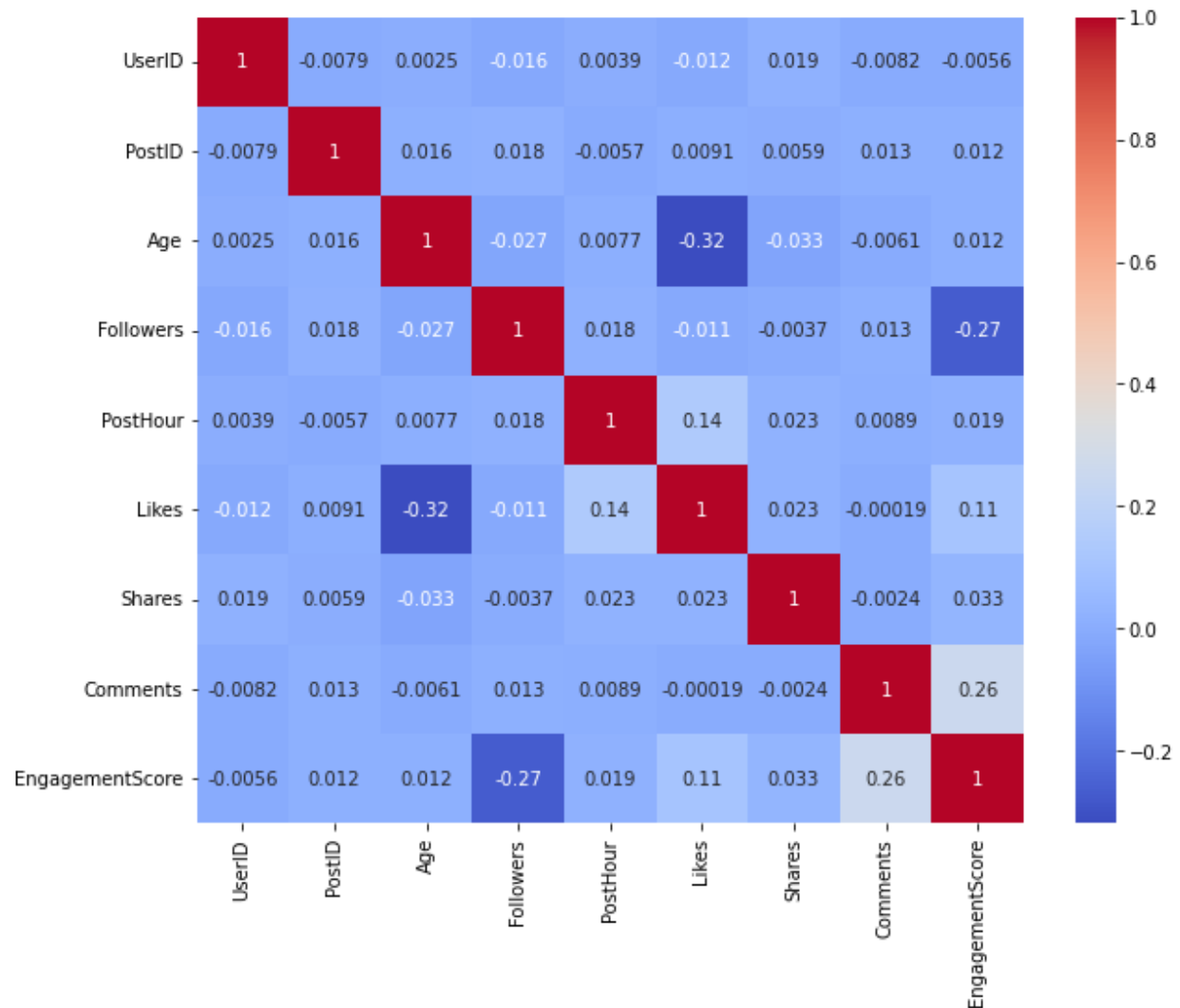
Test 1 - Are you a good detective?

Data : y_posts.csv

- 12 columns (UserID , PostID , Age , Country, Followers, PostType, PostHour ,PostWeekday ,Likes , Shares, Comments , EngagementScore)
- No null values

Data Correlations:





The Engagement Score appears to decrease as the number of Followers increases, indicating a strong negative correlation. Conversely, there is a strong positive correlation with Comments and a moderate positive correlation with Likes, meaning that the Engagement Score tends to rise with an increase in these interactions. A possible formula for calculating the Engagement Score might be:

$$\text{EngagementScore} = a * \text{Comments} + b * \text{Likes} - c * \text{Followers} + d$$

Can we accurately predict the Engagement Score for newly created posts? Let's find out.

I applied predictive models like Random Forest and Gradient Boosting to our training data (4000 points) and evaluated them on test data (1000 points).

Results:

Random Forest Performance:

MAE: 0.1046

MSE: 0.1050

R2: 0.9896

Gradient Boosting Performance:

MAE: 0.1420

MSE: 0.0799

R2: 0.9921

Both models predict the Engagement Score accurately, with Gradient Boosting performing slightly better. These results look promising, but let's further verify by using cross-validation on different training sets.

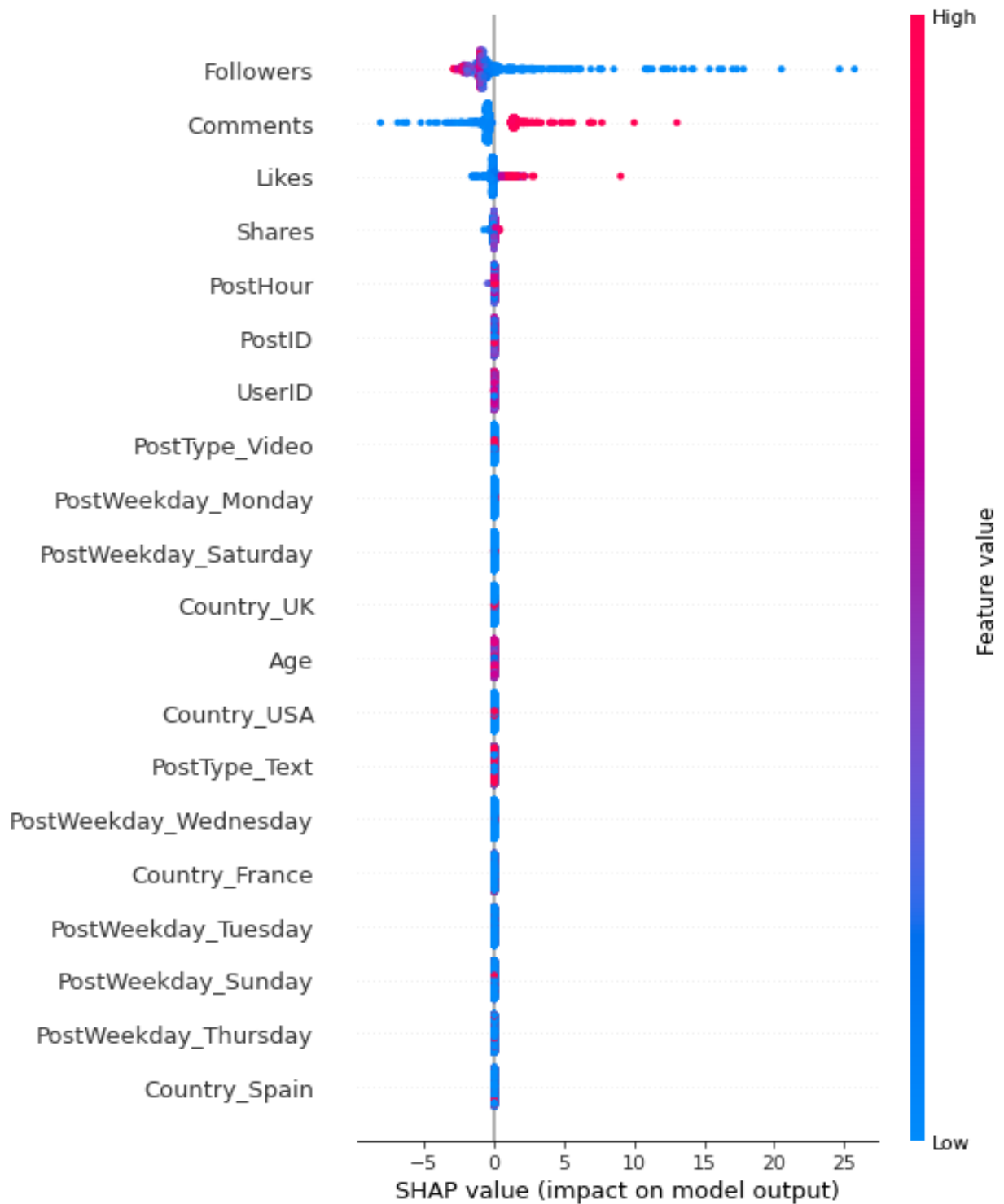
Cross-Validation Results:

R2 Scores: [0.9805, 0.9363, 0.9579, 0.9914, 0.9940]

Mean R2 Score: 0.9720

These results confirm that the models, especially Gradient Boosting, provide consistent and accurate predictions for the Engagement Score.

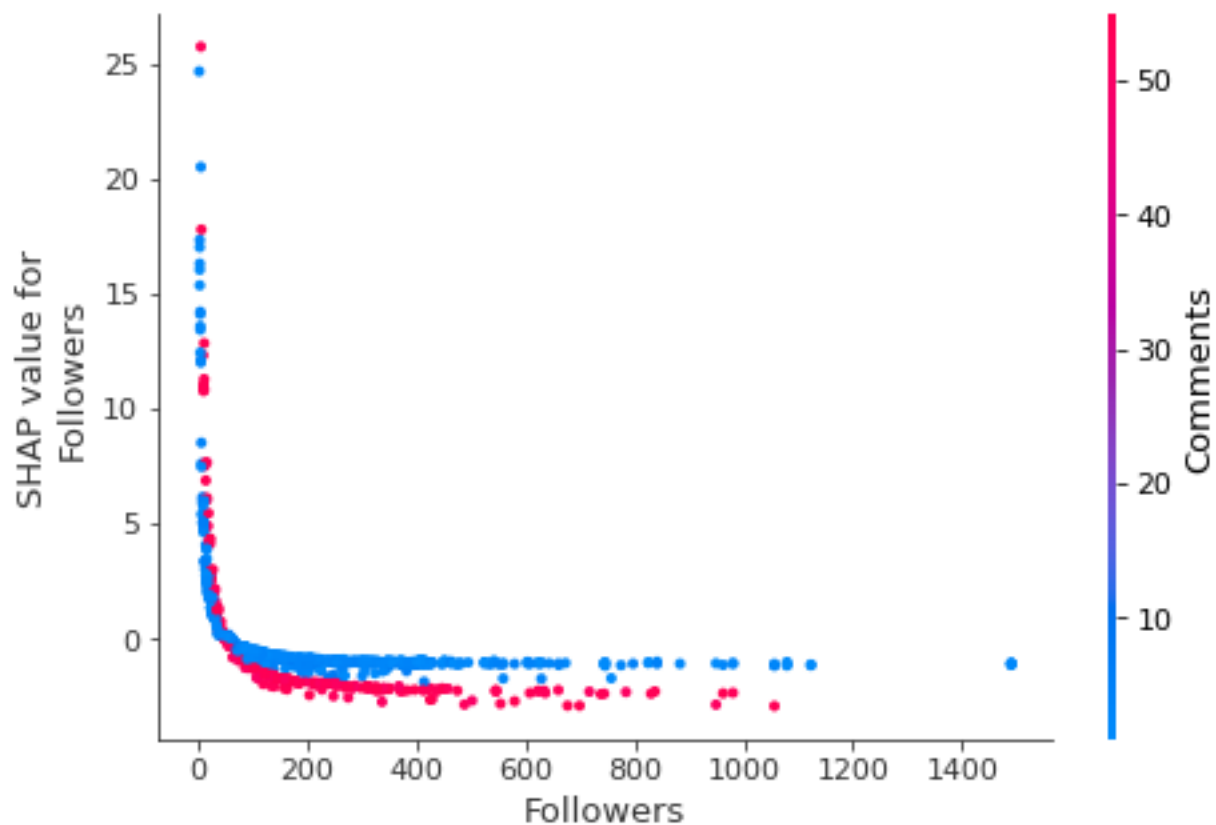
From the **Shap analysis**, it is possible to confirm once again that features that influence the most the model prediction are: followers, comments and Likes.



High Followers counts are linked with lower engagement scores, as indicated by negative SHAP values.

More comments lead to higher engagement scores and positive SHAP values. Likes positively influence the engagement score but not as strongly as comments. Suggesting that while likes are important, comments might be a better measure of engagement.

The SHAP dependence plot for "Followers":



The color gradient shows that even when the number of comments is high (red dots), the SHAP values for "Followers" remain negative when followers are high. This means that while comments are a strong positive indicator of engagement, the negative impact of having a high number of followers still dominates.

In other words, although high comments contribute positively to the engagement score, the overall influence of having a large follower base still reduces the engagement score due to lower relative engagement.