Task 1 - Understand the Algorithm

Determine how the EngagementScore is calculated based on various user and post features. Explore what factors influence this score.

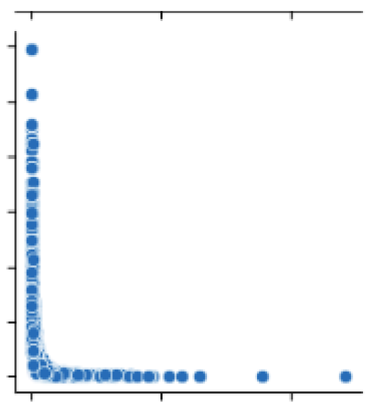

Figure 1.1 - Followers to engagement score

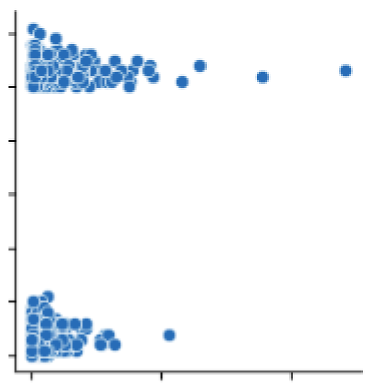

Figure 1.2 - Comments to Engagement Score

Immidietally, as we can see, the fewer followers you have - the higher the score you get, and the more comments your post gets - the bigger the score will be.

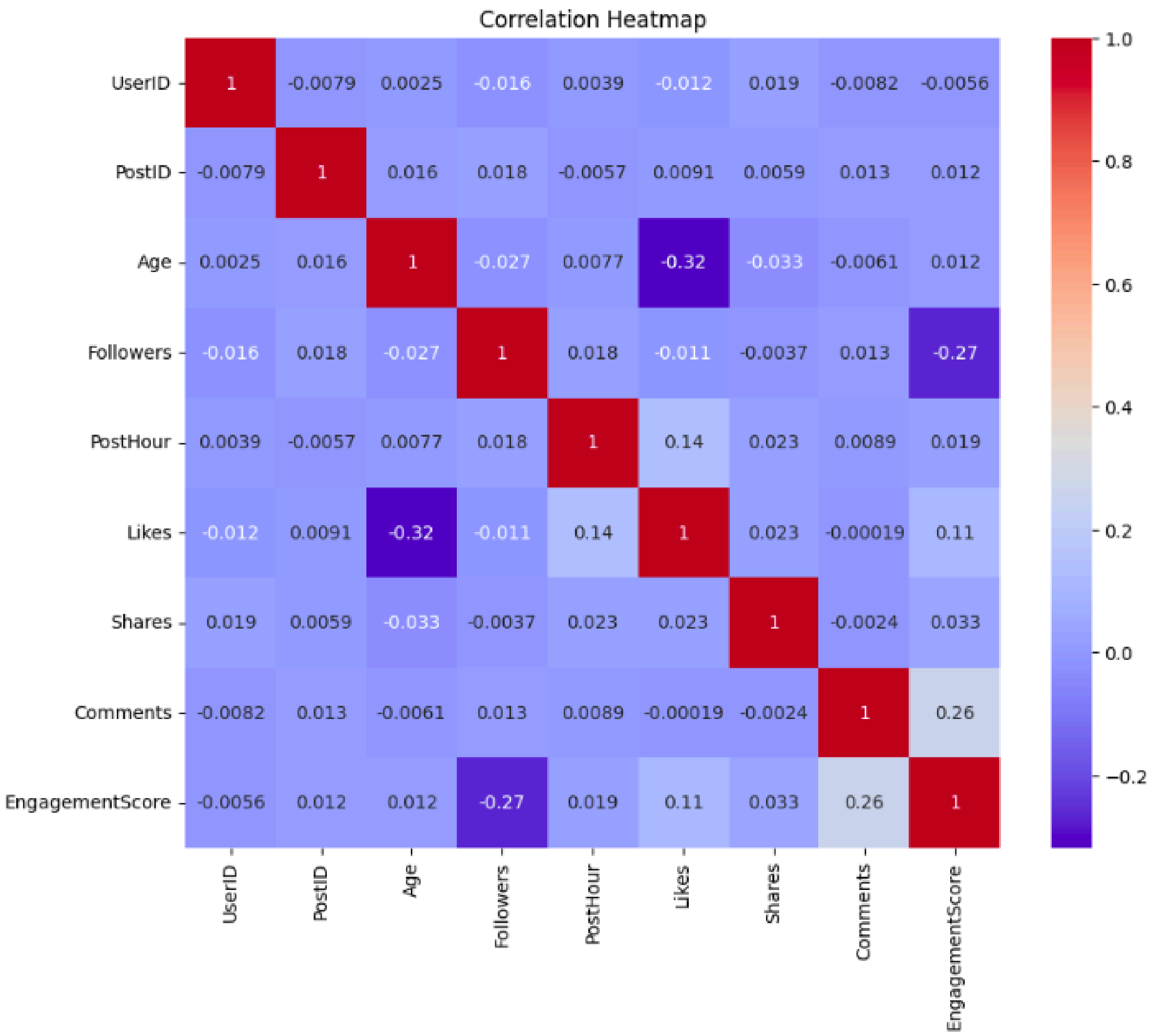

Figure 1.3 - Correlation Heatmap

As we can see in Figure 1.3, the number of comments, followers, and likes have the biggest correlation.

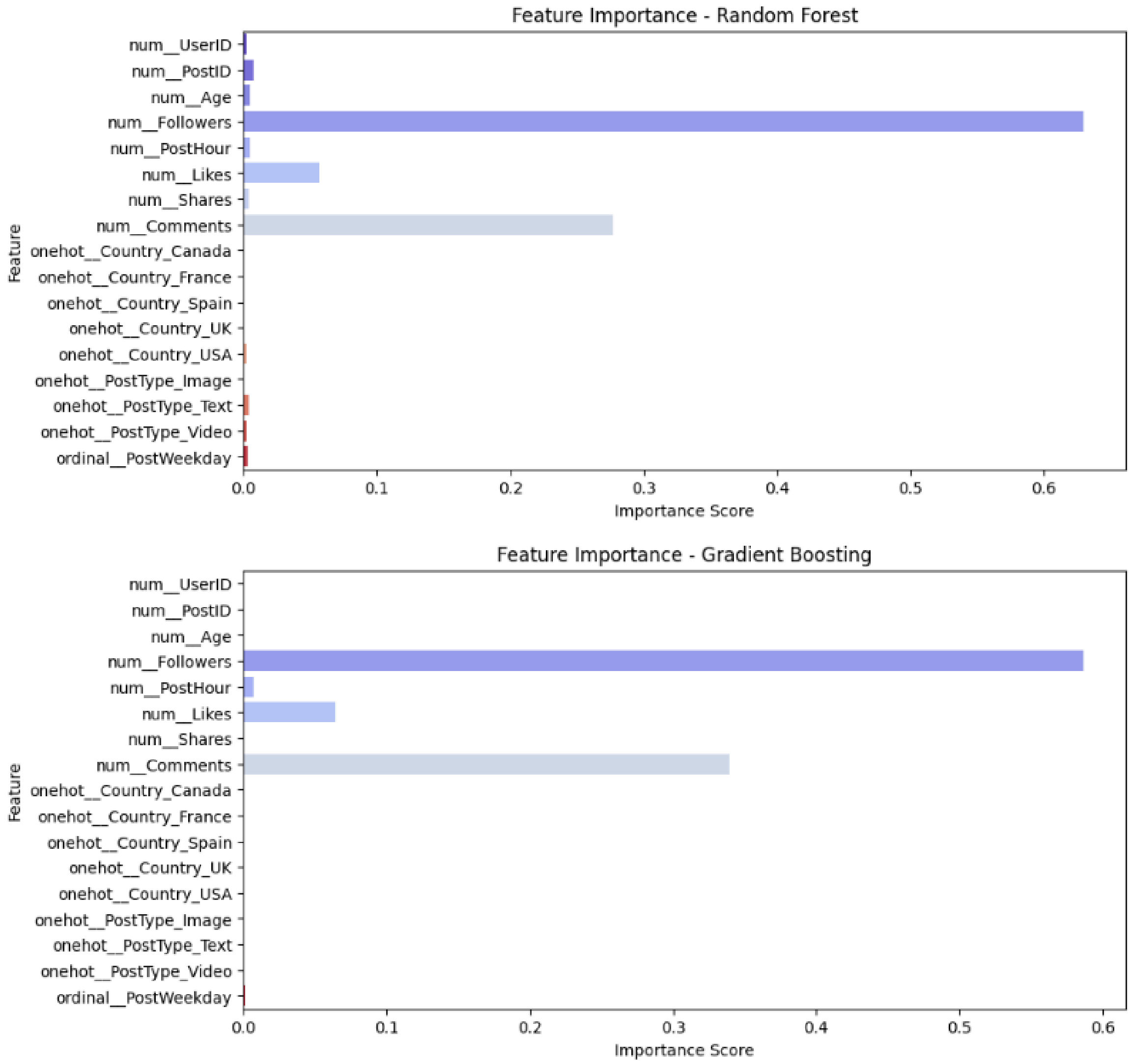

Figure 1.4 - Feature Importances

As we can see, the number of followers, likes and comments are the most important features in predicting the Engagement Score.

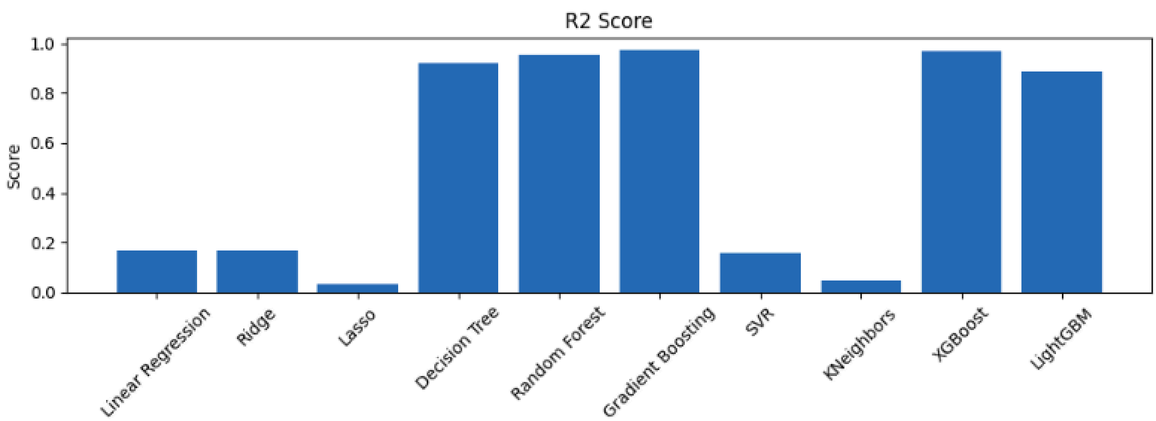

Figure 1.5 - R2 score for each model

As we can see, the formula of the Engagement Score is not linear, since LinearRegression, Ridge, and Lasso have low r2 scores compared to other models.

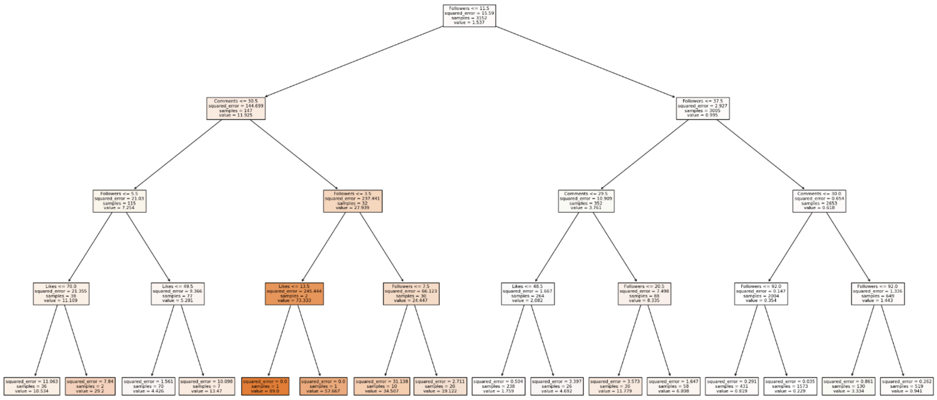

Figure 1.6 - Random tree structure

Random tree structure indicates that the fewer followers you have, the more comments you receive - the bigger your score.

$$\text{Engagement Score} = 16.5156 \times \frac{Likes^{0.3275} \times Comments^{0.4684}}{(Followers+1)^{1.1289}}$$

Figure 1.7 - Engagement score formula

Formula created using **scipy.optimize.curve_fit.** It shows such results:

Mean absolute Error (MAE)	**0.2957553568760472**
R-squared	**0.9600977375803951**

Task 2 Predict the Unpredictable

Assess if it's possible to accurately predict the EngagementScore for newly created posts.

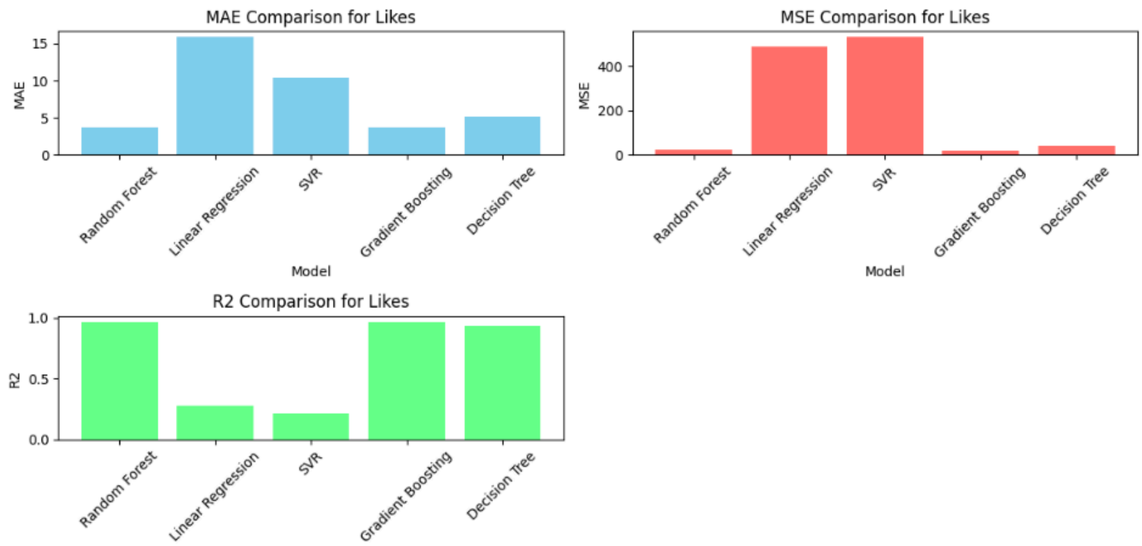

Figure 2.1 - Metrics for Likes Prediction

As we can see, Likes are well-predicted in this case.

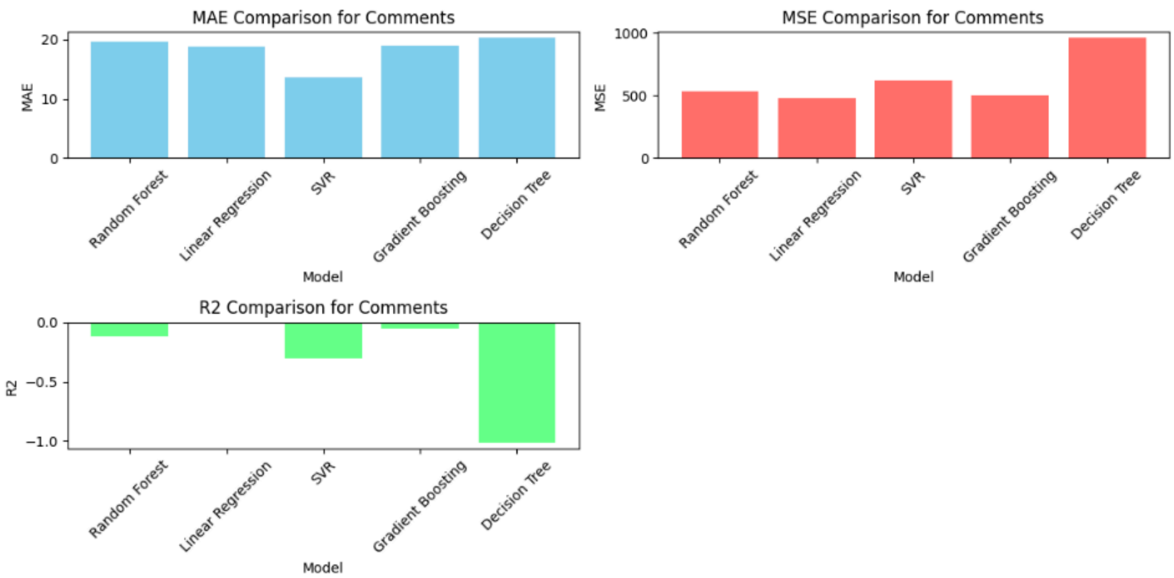

Figure 2.2 - Metrics for Comments Prediction

As we can see, comments are not predicted.

So, we can predict likes, but can not predict comments, and should have as least followers as possible to increase the Engagement Score.

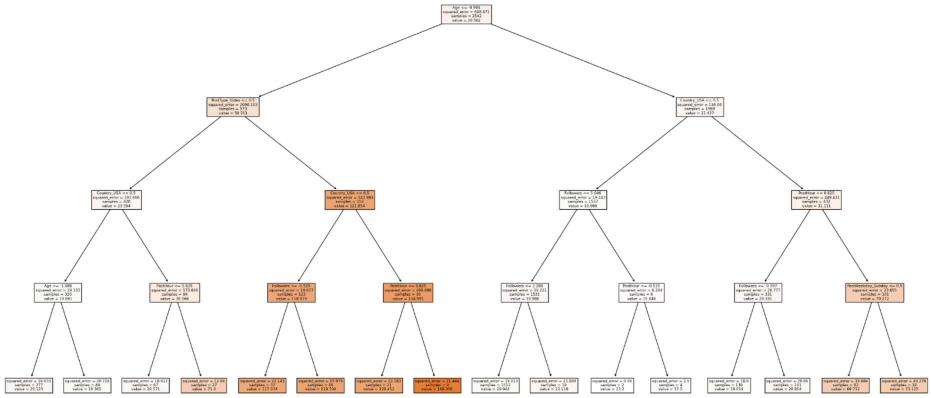

Figure 2.3 - Random tree structure Likes prediction

As we can see from this tree, we have the most likes when posting videos in the US, when the post hour is > 0.825, posted preferably on Sunday.

Summary

Task 1: Understand the Algorithm

Findings:

1. EngagementScore is influenced by factors such as followers, comments, and likes.
2. Fewer followers correlate with a higher score, and more comments lead to a larger score.
3. Features such as the number of comments, followers, and likes show a significant correlation.
4. The number of followers, likes, and comments are the most important features in predicting EngagementScore.
5. The Engagement Score formula is not linear, as indicated by low R2 scores for linear models.
6. Random tree structure shows that fewer followers and more comments result in a higher score.

7. The Engagement Score formula generated through curve fitting shows high accuracy.
8.

Task 2: Predict the Unpredictable

Findings:

1. Likes can be predicted accurately based on given metrics.
2. Comments, however, cannot be accurately predicted.
3. Posts with fewer followers tend to have higher EngagementScore.
4. Random tree analysis for Likes prediction highlights factors like posting videos in the US, posting on Sundays, and specific post hours for maximizing likes.
5. In conclusion, while likes can be predicted reliably, comments are more challenging to predict, and having fewer followers can boost EngagementScore.