## Random Forest :

**Feature Importance:**

Based on the above feature importance scores , we can see which feature contributes more in predicting the likelihood of continuing to use LinkedIn.

**Q5\_2 [All things considered,is it likely that I will continue using LinkedIn]:**

The feature Q5\_2 has the highest importance score of 0.197 (19.7%), which means it is the most important factor in predicting if someone will continue using LinkedIn.This question asks how likely the respondent is to keep using LinkedIn in the future, based on their own personal feelings and opinions.

**Q5\_5 [Linkedin properly satisfies user needs]:**

The question Q5\_5 asks how well LinkedIn meets the user's needs and it has an importance score of 0.158, making it the second most important factor in predicting continued usage after Q5\_2. This feature suggests that how often someone uses LinkedIn and their familiarity with the platform are important factors in determining whether they will continue using it.

**Q5\_8 [I really on linkedin when I have specific information need]:**

This feature has an importance score of 0.130, which means it is the third most important factor in predicting continued usage. This feature appears to be related to how easy it is to find specific information on LinkedIn, indicating that the convenience and efficiency of LinkedIn's functions are important factors in whether someone will continue using it.

On the other hand,Q5\_9[linkedin provides convenient functions],Q5\_3 [I value information what linkedIn offers,with the help of which I get what I needed in any situation],Q5\_4 [using linkedIn helps me to feel accepted by others.] have a moderate contribution of predicting outcome of person's staying on linkedIn whereas other features such as Q5\_6 [linkedin is part of my everyday activity],Q5\_7 [I would be upset if linked in shut down] are least important features to consider for prediction.

**Part 1: Random Forests – Binary**

1. **Random Forest:**

According to the above results we can evaluate the model's performance.

**Accuracy:**

The model's overall accuracy on the test set is 0.79, which means it correctly classified 79% of the samples.

**Error rate:**

The error rate is 0.21, which means the model incorrectly classified 21% of the samples.

**Precision:**

The precision score is 0.75, which means when the model predicted that a user will continue using LinkedIn, it was correct 75% of the time out of all the positive samples classified by model.

**Sensitivity:**

The sensitivity score is 0.61, which means the model correctly identified 61% of the users who will continue using LinkedIn.

**Specificity:**

The specificity score is also 0.77, which means the model correctly identified 77% of the users who will not continue using LinkedIn.

**AUC:**

The AUC score is 0.78, which is an overall measure of the model's ability to distinguish between the users who will continue using LinkedIn and those who will not.

**Cross-entropy:**

The cross-entropy score is 7.29, which measures the difference between the predicted probabilities and the actual labels. A lower value indicates better performance.

1. **Extra Tree Classifier:**

**Accuracy:** 79% of the predictions made by the model are correct.

**Error rate:** 21% of the predictions made by the model are incorrect.

**Precision:** When the model predicts a positive outcome, it is correct 75% of the time.

**Sensitivity:** The model correctly identifies 61.29% of the positive cases in the dataset.

**Specificity:** The model correctly identifies 77% of the negative cases in the dataset.

**AUC:**

The model has an area under the curve (AUC) of 78%, which indicates its overall ability to distinguish between positive and negative cases.

**Cross-entropy:**

The average cross-entropy loss for the predictions made by the model is 7.29.

**Extra Tree Vs Random Forest**

Based on the results, both the random forest and extra tree models had similar accuracy, precision, sensitivity, specificity, AUC, and cross-entropy. However, the random forest model performed slightly better than the extra tree model in terms of accuracy, AUC, and cross-entropy. The random forest model had an accuracy of 0.7887, while the extra tree model had an accuracy of 0.7887 as well. The precision of both models were the same at 0.75. The sensitivity and specificity of both models were also the same at 0.6129 and 0.7742, respectively. However, the AUC of the random forest model was slightly higher than that of the extra tree model, with values of 0.7871 and 0.7871, respectively. Finally, the cross-entropy of the random forest model was lower than that of the extra tree model, with values of 7.2970 and 7.2970, respectively.

Therefore, based on these results, it can be concluded that the random forest model performed slightly better than the extra tree model, although the difference is not significant.

## **Random Forest Vs Logit Model**

1) The random forest model had an accuracy of 0.8, while the logistic regression model had an accuracy of 0.8429. Therefore, logistic regression model performed better than the random forest model in terms of accuracy.

2) The error rate of the logistic regression model was lower (0.1571 vs 0.2) than that of the random forest model.

3) The precision of the logistic regression model was higher (0.7879 vs 0.7667) than that of the random forest model.

4) The sensitivity of both models were the same (0.6129), while the specificity of the logistic regression model was higher (0.8667 vs 0.7667) than that of the random forest model.

5) The AUC of the logistic regression model was higher (0.8458 vs 0.7958) than that of the random forest model.

6) The cross-entropy of the logistic regression model was lower (5.4276 vs 6.9078) than that of the random forest model.

##### **Overall, the logistic regression model appears to be more accurate, precise, and better at identifying true negatives than the random forest model. However, the random forest model may be better at identifying true positives.**