

Major Project June Batch 3

Google Classroom code: ML063B18

Answers

1. What are the most common words used by Males and Females?

Ans: The most common word used by male is “the” and most common word used by Female is “and”.

2. How significant are the colors used by the users in classifying gender?

Ans: The colors used for Link_color and Sidebar_color in their profile are very much significant in identifying whether the person is male or female. The Link color “0084B4” and the Sidebar_colour “CODEED” are most commonly used colors by both the gender.

Summary of the Major Project:

The given problem (dataset) was based on Twitter Tweets count and based on various features we can classify whether the person is male or female. Three classification algorithms were used namely Logistic Regression, k-Nearest Neighbor(kNN) and Random Forest to find out which model suits the best.

Out of the 3 classification algorithms **Random Forest model with an accuracy of approximately 59%** is found to be the best suited model for the above mentioned problem. The other 2 models (**Logistic Regression, kNN**) accuracies were found to be approximately **52% and 55%** respectively. In all the three models, Gender was taken to be the dependent variable and unit_id , tweet_count, Linkcolor, Text1, Sidebar1 were taken to be the Independent variables.