



Home



Moments



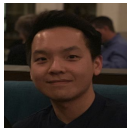
Notifications



Messages



Tweet



DanielLin

@sf20_ds18

Following

Predicting User Gender from Profile

05/06/2020

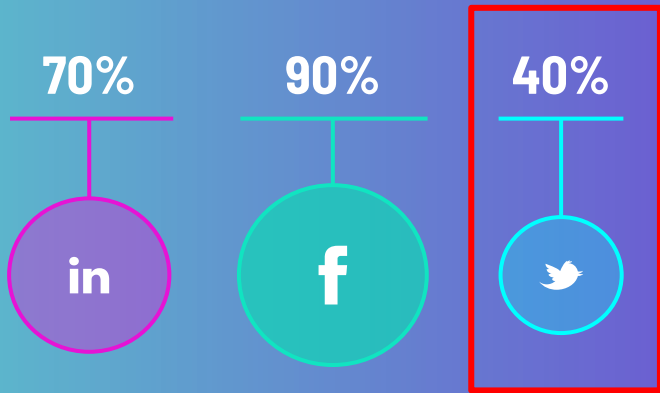


Why do we care?

- Targeted ads for user based on gender
 - Makeup for female
 - Video games for male
- If users link their social media profile:
 - Determine gender distribution of users for product
 - Change product to fit user needs



TOP SOCIAL NETWORKING SITES



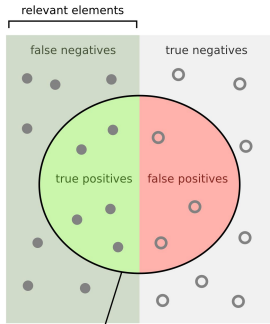
What's in a User Profile?



The image shows a screenshot of LeBron James' Twitter profile. At the top is a large header image featuring a collage of photos, including LeBron in his Lakers jersey, him with his family, and various community events. Below this is a circular profile picture of LeBron James. To the right of the profile picture is a blue 'Follow' button. Under the profile picture, his name 'LeBron James' is displayed with a verified account checkmark, followed by his handle '@KingJames'. Below the handle, his bio reads: 'EST. AKRON - ST.V/M Class of '03 LeBronJamesFamilyFoundation.org #IPROMISE'. Further down, it says 'Amongst La Familia!' with a link to 'LeBronJames.com' and 'Joined March 2009'. At the bottom of the profile section, it shows '188 Following' and '45.9M Followers'. Below this are four tabs: 'Tweets', 'Tweets & replies', 'Media', and 'Likes'.

- Data from Kaggle
 - User Name
 - Text
 - Description
 - Tweet Count
 - Retweets
 - Favorite Number
 - Sidebar Color
 - Gender

Metrics



How many selected items are relevant?

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

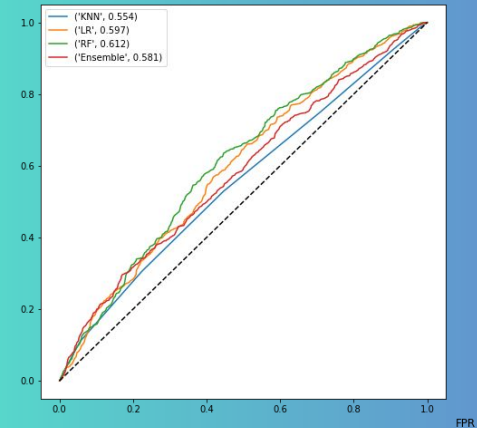
How many relevant items are selected?

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- Precision:
 - False Positive: predicted as female but actually is male
 - Males most likely won't buy makeup
- Recall:
 - False Negative: predicted as male but actually is female
 - Missed chance to promote product
- ROC/AUC Score:
 - Can adjust threshold

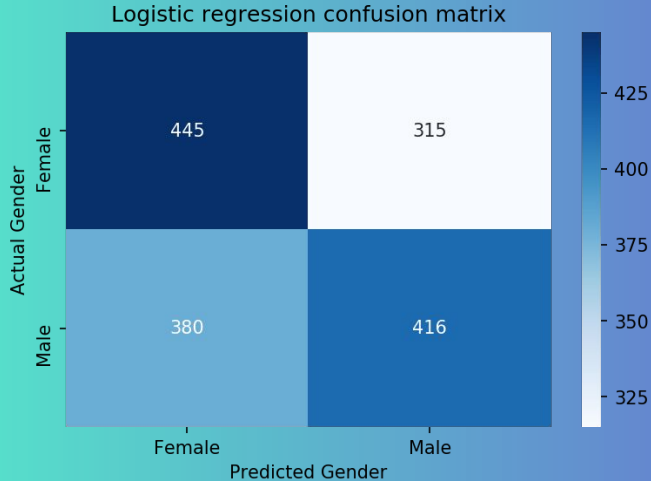
Modeling

ROC Curves for Different Classification Models



Model	Test ROC/AUC score
Random Forest	0.607
Logistic Regression	0.59
Ensemble	0.58
KNN	0.55

Results



- ROC/AUC score: 0.59
- High false negative

Future Work

- Gather more data
 - Number of followers, people user follow
- Use other platforms to confirm gender
 - Instagram, FaceBook



Thank You!

- Questions?