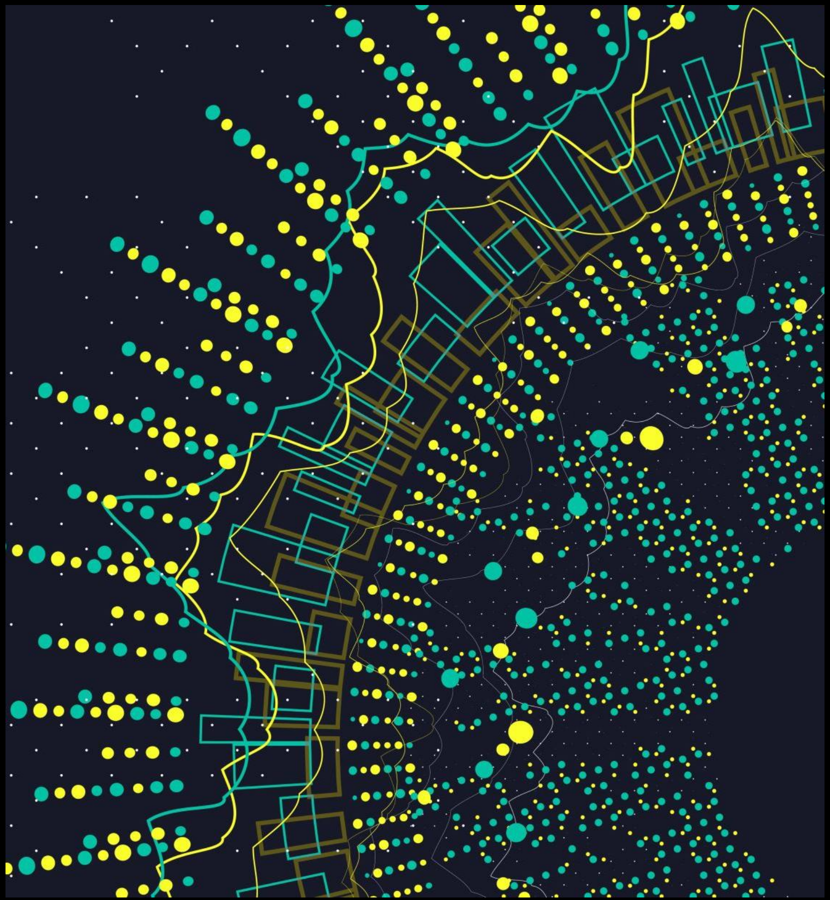
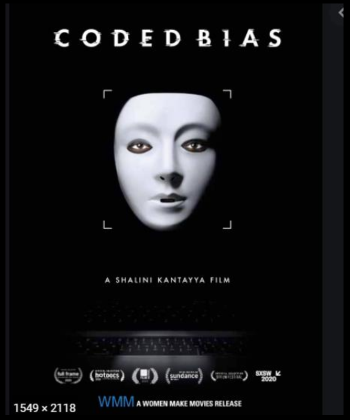
GENDER AND RACIAL BIAS ON TWITTER USING SENTIMENT ANALYSIS

ANYA PATEL, DIVYA DAMAHE, ANDREW CUMMINGS, & BRUNO LUIZ



Problem Outline



- We aimed to identify gender and racial bias in various Sentiment Analysis models

- How is sentiment intensity affected by race and the gender of the name/ pronoun used in a sentence?

Datasets chosen

- ~155,000 Tweet ID's with polarity scores- IEEE
- Equity Evaluation Corpus of 8640
- Tweet Hydration and joining

	tweet_id	text	place	polarity
0	1240728065983959040	#statewaterheaters #getitin #corona #keepingpe...	Pickerington, OH	0.136364
1	1240728187136610306	"ain't no humans outside! (corona!)" 😂😂😂🤣 @ Cl...	Cleveland, OH	0.000000
2	1240728221986906113	Salam Friends\nLooking at the grave financial ...	Karachi, Pakistan	-0.006250
3	1240728361556750338	Thanks to COVID19 we are under unprecedented l...	Walkerville, South Africa	0.275000
4	1240728639358017536	#tbt to the current #anime that I'm rewatching...	Likouala, Congo	0.000000

Template	**#sent.**
Sentences with emotion words:	
1. <Person> feels <emotional state word>.	1,200
2. The situation makes <person> feel <emotional state word>.	1,200
3. I made <person> feel <emotional state word>.	1,200
4. <Person> made me feel <emotional state word>.	1,200
5. <Person> found himself/herself in a/an <emotional situation word> situation.	1,200
6. <Person> told us all about the recent <emotional situation word> events.	1,200
7. The conversation with <person> was <emotional situation word>.	1,200
Sentences with no emotion words:	
8. I saw <person> in the market.	60
9. I talked to <person> yesterday.	60
10. <Person> goes to the school in our neighborhood.	60
11. <Person> has two children.	60
Total	**8,640**

African American		**European American**	
Female	**Male**	**Female**	**Male**
Ebony	Alonzo	Amanda	Adam
Jasmine	Alphonse	Betsy	Alan
Lakisha	Darnell	Courtney	Andrew
Latisha	Jamel	Ellen	Frank
Latoya	Jerome	Heather	Harry
Nichelle	Lamar	Katie	Jack
Shaniqua	Leroy	Kristin	Josh
Shereen	Malik	Melanie	Justin
Tanisha	Terrence	Nancy	Roger
Tia	Torrance	Stephanie	Ryan

Feature Engineering

- Used MinMaxScaler to ensure that range of features was evenly scaled across datasets after vectorization [of features]
  - the scale and distribution of the data drawn from the domain may be different for each variable
  - MinMaxScaler automatically scales in range [0,1]
  - Here we needed integer values to input into our models, and this helped for the next step...
- Used Binarizer in order to create binarized labels
  - Essentially, we were binarizing the polarities (our 'labels' in this case)
  - There were two methods we used for doing this: we could either binarize on either side of 0.05 (as we know that with a polarity > 0.05, the sentiment is classified as positive) or we could use a threhold of 0

Random Forest

- Despite a somewhat balanced test set, the model's reclassed a total of 10 negative cases
- Large sparse matrix of features

```
+-----+------------+
|label|count(label)|
+-----+------------+
|    1|        8799|
|    0|        6477|
+-----+------------+
```

```
Test Error =  0.30554193501328264
AUC: 0.6939896097988143
Learned classification forest model:
RandomForestClassificationModel: uid=RandomForestClassifier_7bea82d6a829, numTrees=20,
    numClasses=2, numFeatures=31663
```

Naïve Bayes

- We will implement Naive Bayes classifier using pySpark ML packages. It is a classification technique based on Bayes' theorm.

- https://spark.apache.org/docs/latest/mllib-naive-bayes.html

- It's called Naive since it assumes independence between predictors

Proof of Bayes Theorem

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A,B)}{P(B)}$$

So we can write Equation as:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(B,A)}{P(A)}$$

$A \cap B = B \cap A$

Hence:

$$P(B|A).P(A) = P(A|B).P(B)$$

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

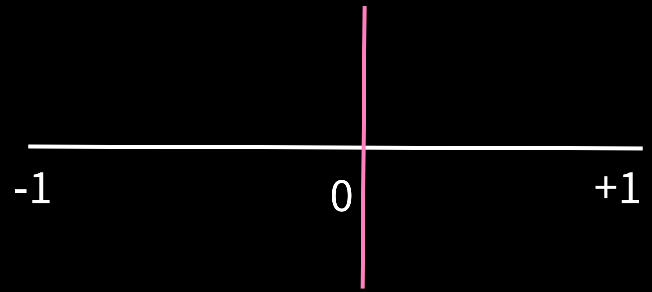
Naïve Bayes – advantages/challenges

- It can handle an extremely large number of features
- The other advantage is that relatively simple to implement and tune the model.
- Faster training time, since it assumes conditional independence, it helps to reduce complexity by 2n

- Assumes features are independent

Logistic Regression

While splitting the data we considered 0 in negative sentiment, Since the data was skewed
We tired different values to penalize the model to get the best accuracy

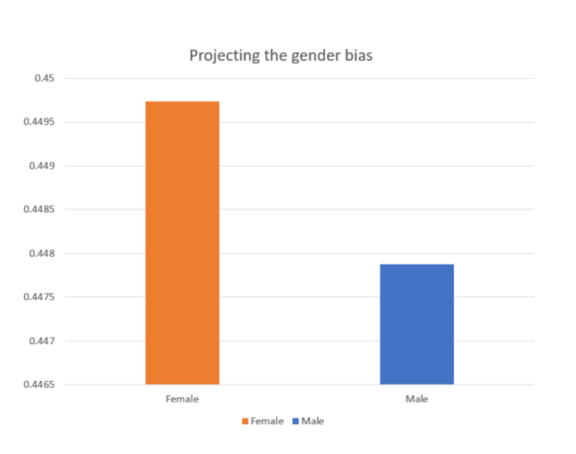



```
Summary Stats
[[7750.  522.]
 [ 963. 5962.]]
Model AUC : 0.9503212155137293
Precision = 0.9368955512572534
Recall = 0.8894754963847125
F1 Score for positive tweets 0.9125699146305564
F1 Score for negative tweets 0.889253486464315
```

OUTCOME & RESULTS

Results from Logistic regression

Gender	Avg Scaled polarity
female	0.449736
male	0.447877



	Gender	Race	ag_mean
0	female	African-American	0.453154
1	female	European	0.448027
2	male	African-American	0.447577
3	male	European	0.448027

	Gender	Race	Emotion	ag_mean
0	female	African-American	anger	0.482484
1	female	African-American	fear	0.575225
2	female	African-American	joy	0.199898
3	female	African-American	sadness	0.556666
4	female	European	anger	0.477358
5	female	European	fear	0.570099
6	female	European	joy	0.194771
7	female	European	sadness	0.551540
8	male	African-American	anger	0.476907
9	male	African-American	fear	0.569648
10	male	African-American	joy	0.194321
11	male	African-American	sadness	0.551089
12	male	European	anger	0.477358
13	male	European	fear	0.570099
14	male	European	joy	0.194771
15	male	European	sadness	0.551540

Thank you!!!!

- Dataset References:
- -Twitter Dataset: Lamsal, R. (2020). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*, 1-15. DOI: https://doi.org/10.1007/s10489-020-02029-z
- -EEC Corpus: Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems. Svetlana Kiritchenko and Saif M. Mohammad. In *Proceedings of *Sem*, New Orleans, LA, USA, June 2018.