NFL / 49ers

REDDITS BEST SEASON

BUSINESS CASE:

Build a classification model to predict the 49ers subreddit over the NFL subreddit for social media marketing purposes.

FITTING MODELS:

	LOGR: TFIDF VECTORIZER	LOGR: COUNT-VECTORIZER	BERNOULLI : NAIVE BAYES	KNN MODEL	RANDOM FORREST: COUNT-VECTORIZER	SVC: TFIDF VECTORIZER
SCORE						
BALANCE ACCURACY						
SENSITIVITY						
SPECIFICITY						
PRECISION						

FITTING MODELS:

	LOGISTIC: TFIDF VECTORIZER	LOGISTIC: COUNT-VECTORIZER	BERNOULLI : NAIVE BAYES	KNN MODEL	RANDOM FORREST: COUNT-VECTORIZER	SVC: TFIDF VECTORIZER
SCORE	.910 / .778	.950 / .788	.703 / 0.676	.719 / .576	.976 / .739	.930 / .775
BALANCE ACCURACY	0.787	0.772	0.667	0.635	0.722	0.775
SENSITIVITY	0.825	0.84	0.984	0.97	0.892	0.785
SPECIFICITY	0.749	0.704	0.35	0.30	0.552	0.766
PRECISION	0.777	0.75	0.616	0.581	0.669	0.733

SENTIMENT CHECKS:

r/49ERS

Negative : 4.8%

Positive : 14.37%

Neutral : 80.83%

Compound : .207

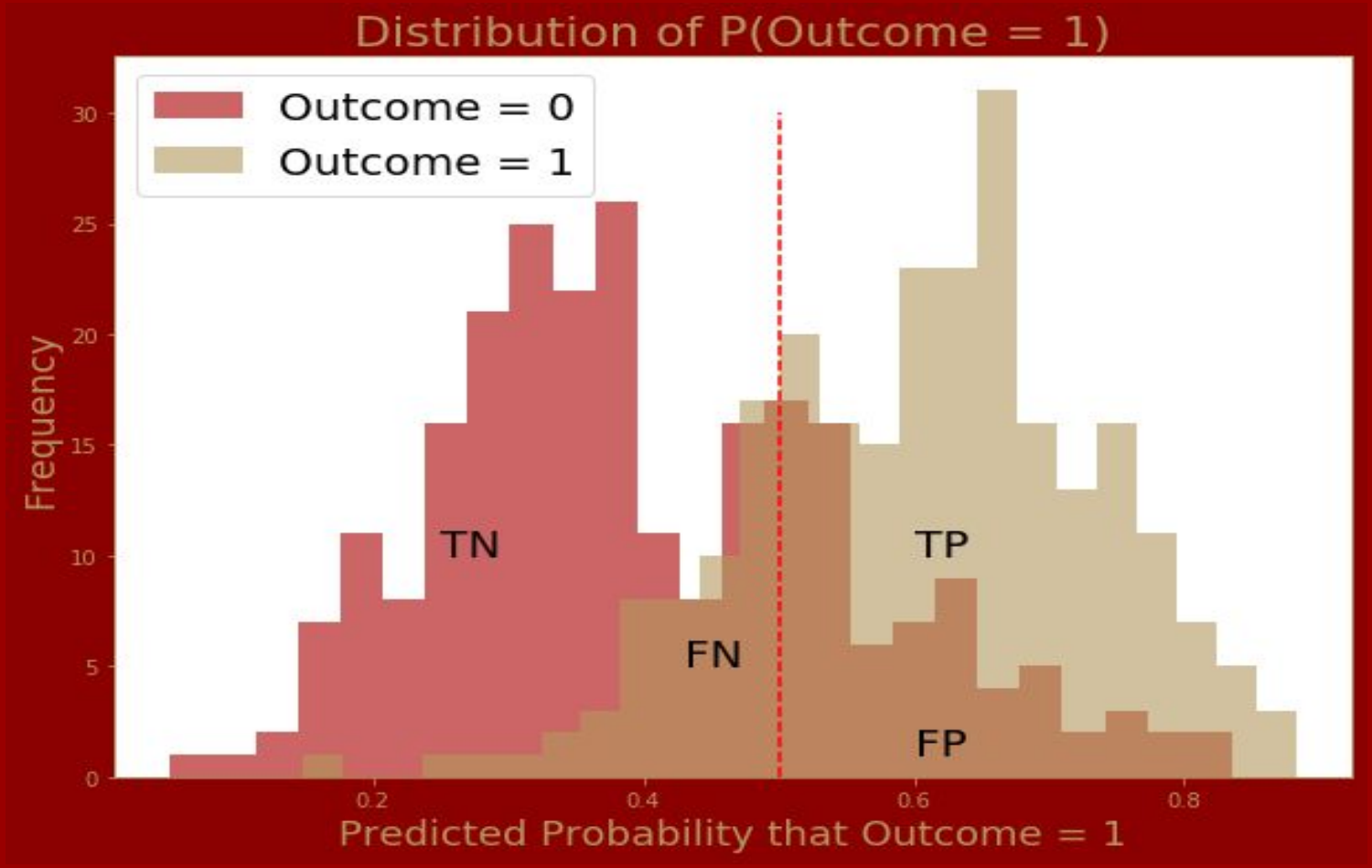
r/NFL

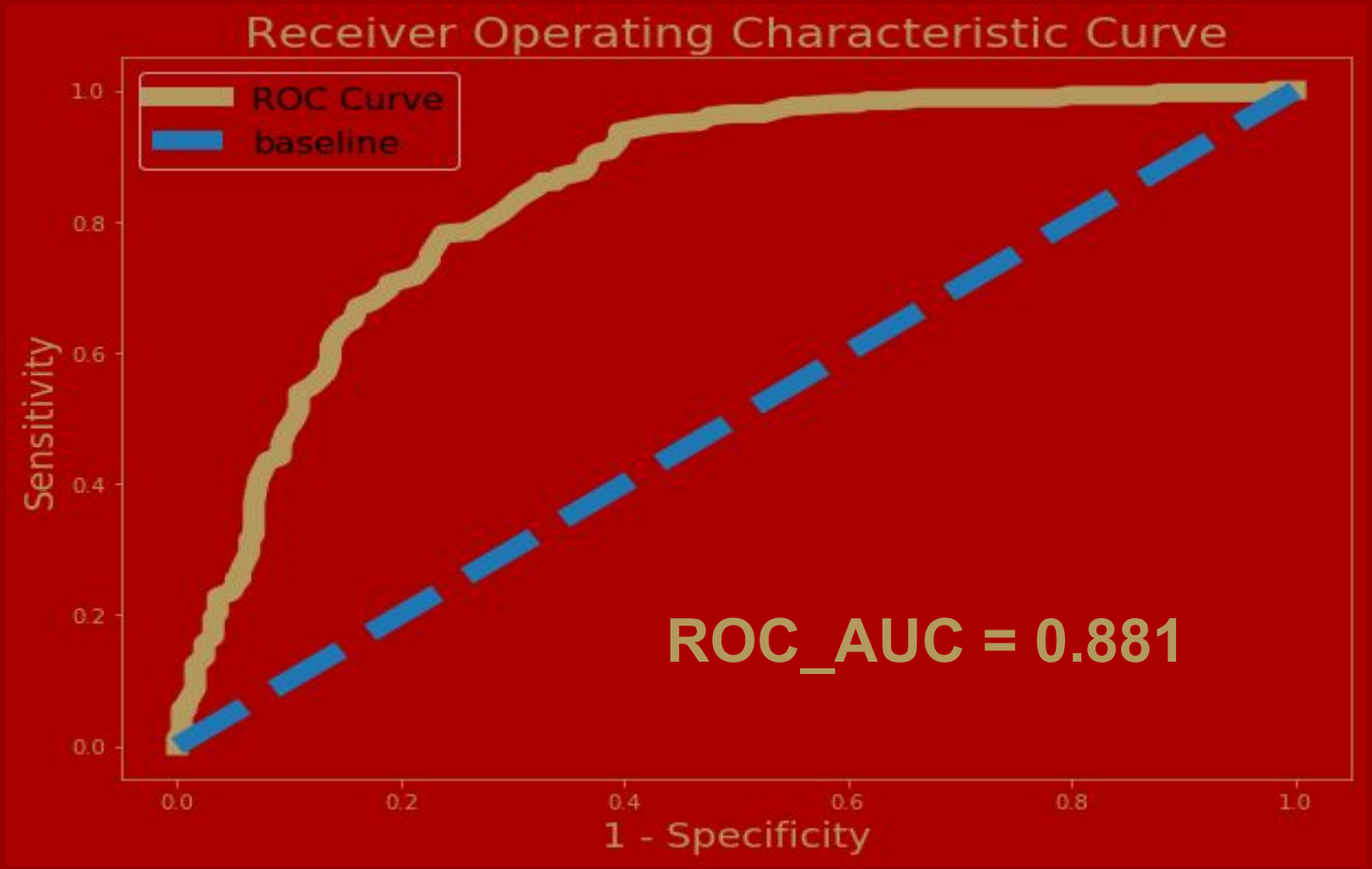
Negative : 4.61%

Positive : 13.61%

Neutral : 81.68%

Compound : .282


Distribution of P(Outcome = 1)
Outcome = 0
Outcome = 1
Frequency
TN
FN
TP
FP
0
5
10
15
20
25
30
0.2
0.4
0.6
0.8
Predicted Probability that Outcome = 1


Receiver Operating Characteristic Curve
ROC Curve
baseline
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
ROC_AUC = 0.881


Top 20 Features from Each Class
Weighted Coefficient
3
2
1
0
−1
−2
NFL
49ers
nfl
pro
20
players
brady
patriots
favorite
line
steelers
richard
quarterback
cowboys
worst
vikings
hacked
kansas
jones
lost
hall
halftime
willis
fuck
hat
faithful
meme
patch
excellence
kyle
smorgasbord
49er
network
looks
shanahan
days
bosa
nip
niners
kittle
jimmy
49ers
Feature

Final Thoughts

Continue improving/exploring models:
...feature engineering
...lemmatization
Run model on other football subreddits

Remember: It's not if they'll buy it's where they are at...if we find the loyal fans...they'll do the rest

REFERENCES:

https://towardsdatascience.com/natural-language-processing-and-sports-subreddits-d470e8bfc2e1

https://medium.com/@aneesha/visualising-top-features-in-linear-svm-with-scikit-learn-and-matplotlib-3454ab18a14d