Crossfit
OR
Powerlifting

Round 1
NEXT ROUND
WHY? WHAT?
CATEGORY: PLACE
INCORRECT
SPIN
Win
Bankrupt
Cheer
Boo
Silence

Data

Scraped: 6,000 posts

Clean: 2097

Models I Employed

Logistic Regression

KNN

Decision Tree

Random Forest

Naive Bayes

Support Vector Machine

Gradient Boosting

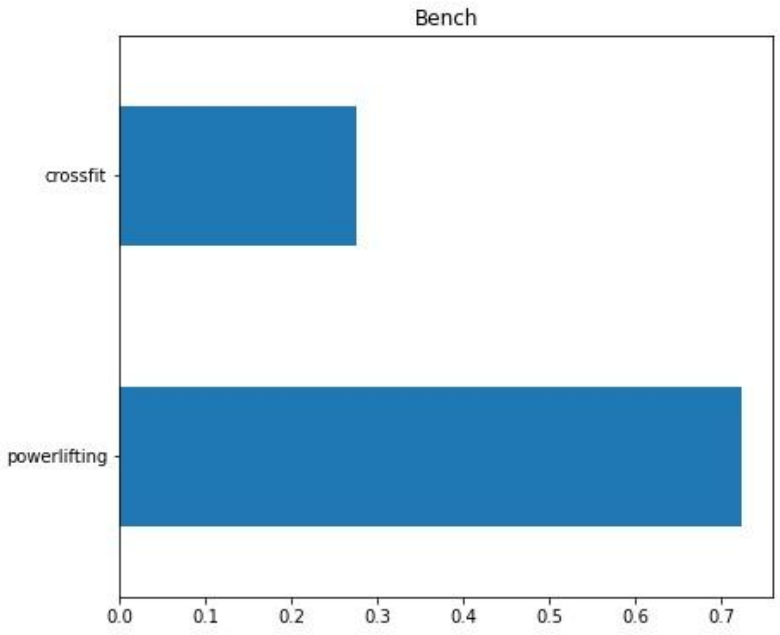
Model Comparison

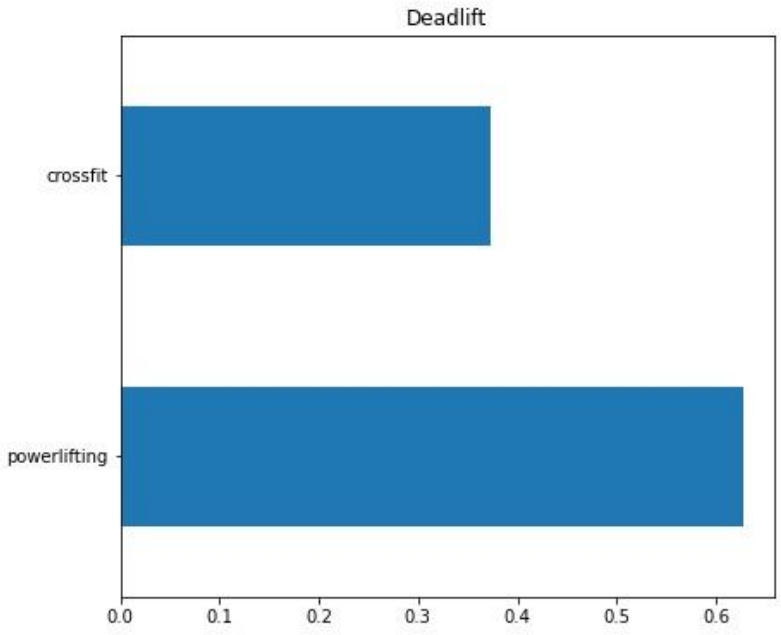
	train f1	test f1	overfit	score spread
Gradient	0.964784	0.954048	Y	0.010735
Decision Trees	0.954895	0.945856	Y	0.009039
Random Forest	0.953413	0.942060	Y	0.011353
SVM	0.951965	0.940152	Y	0.011813
Naive Bays	0.944635	0.936508	Y	0.008127
Logistic Regression	0.965771	0.935800	Y	0.029971
KNN	0.924382	0.914826	Y	0.009555

Findings: Top 5 Important Keywords

	importance
powerlifting	0.277527
bench	0.144171
crossfit	0.128027
deadlift	0.117946
??????????????????	

Mention of ‘bench’ and ‘deadlift’





Findings: Top 5 Important Keywords

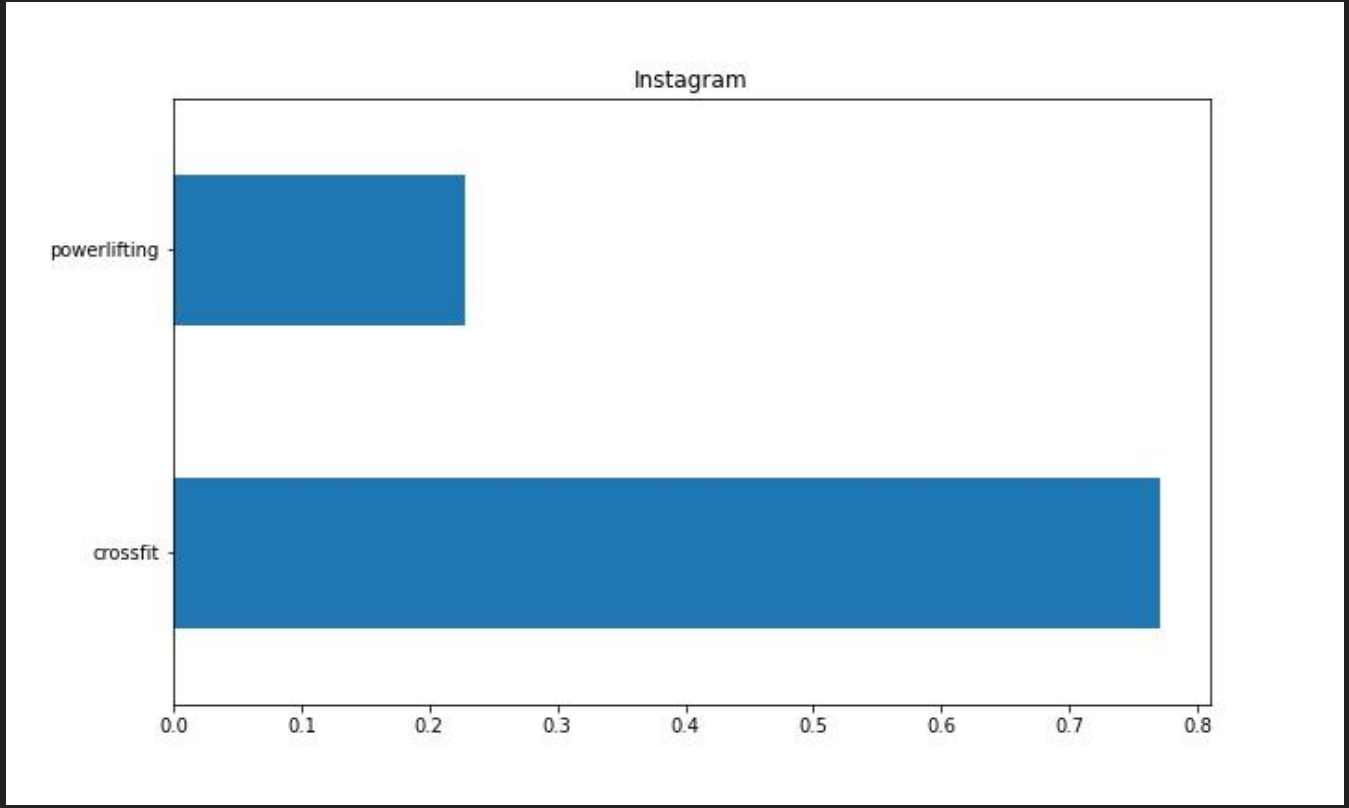
	importance
powerlifting	0.277527
bench	0.144171
crossfit	0.128027
deadlift	0.117946

??????????????????

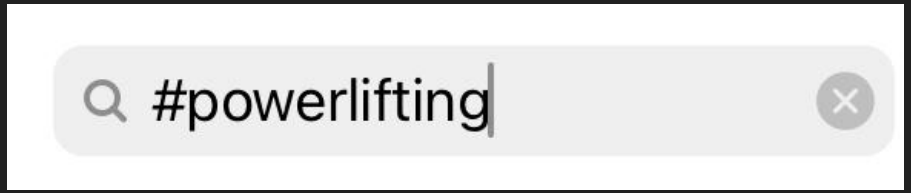
Findings: Top 5 Important Keywords

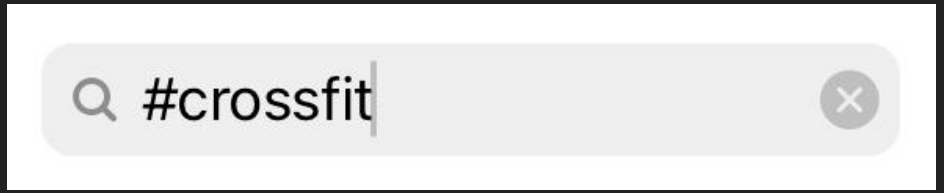
	importance
powerlifting	0.277527
bench	0.144171
crossfit	0.128027
deadlift	0.117946
instagram	0.033289

Mention of ‘Instagram’

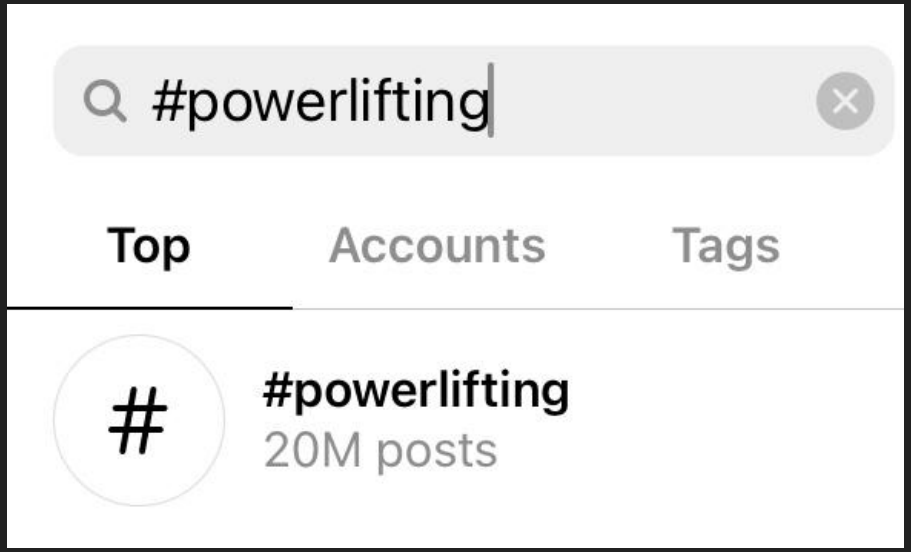


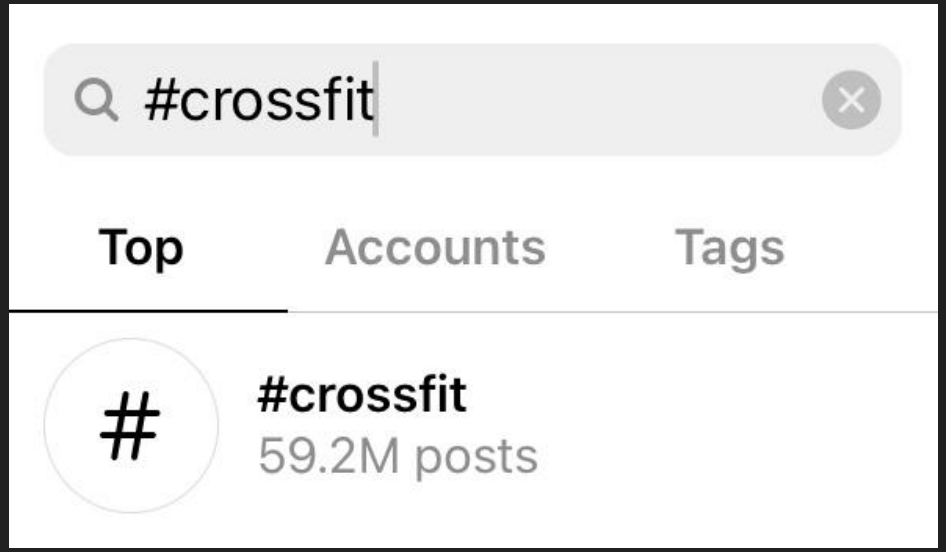
Cross Check





Cross Check


#powerlifting
Top
Accounts
Tags
#
#powerlifting
20M posts


#crossfit
Top
Accounts
Tags
#
#crossfit
59.2M posts

BULLDOGS
1 2 3 4 5 6 7 8
Now What?

Crossfit
OR
Powerlifting

Thank you all