

Project 3:

Web APIs & Classification (Dating Apps)



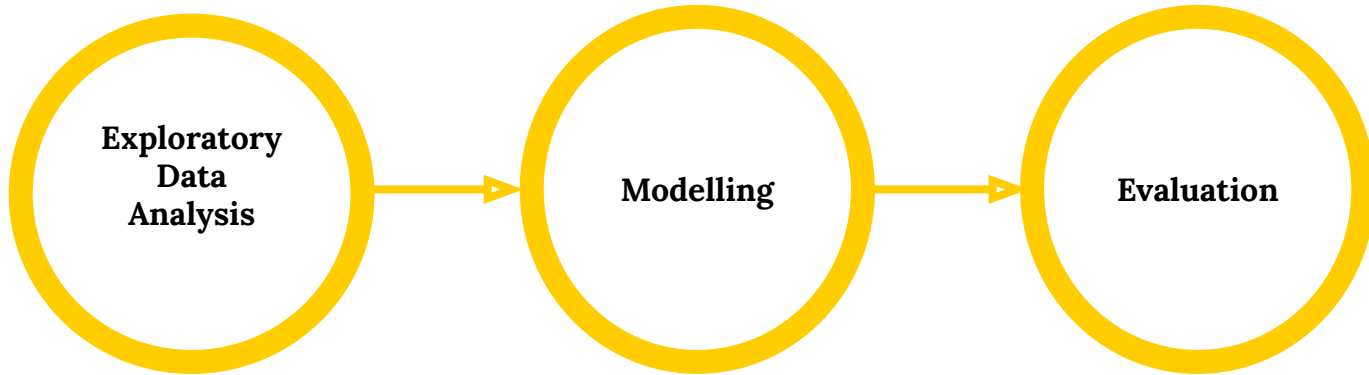


Problem Statement

- Using predictive modelling, how can we best classify posts from 2 subreddits (/r/bumble & /r/coffeemeetsbagel)?
- How can this data benefit dating apps?



Presentation Flow



1

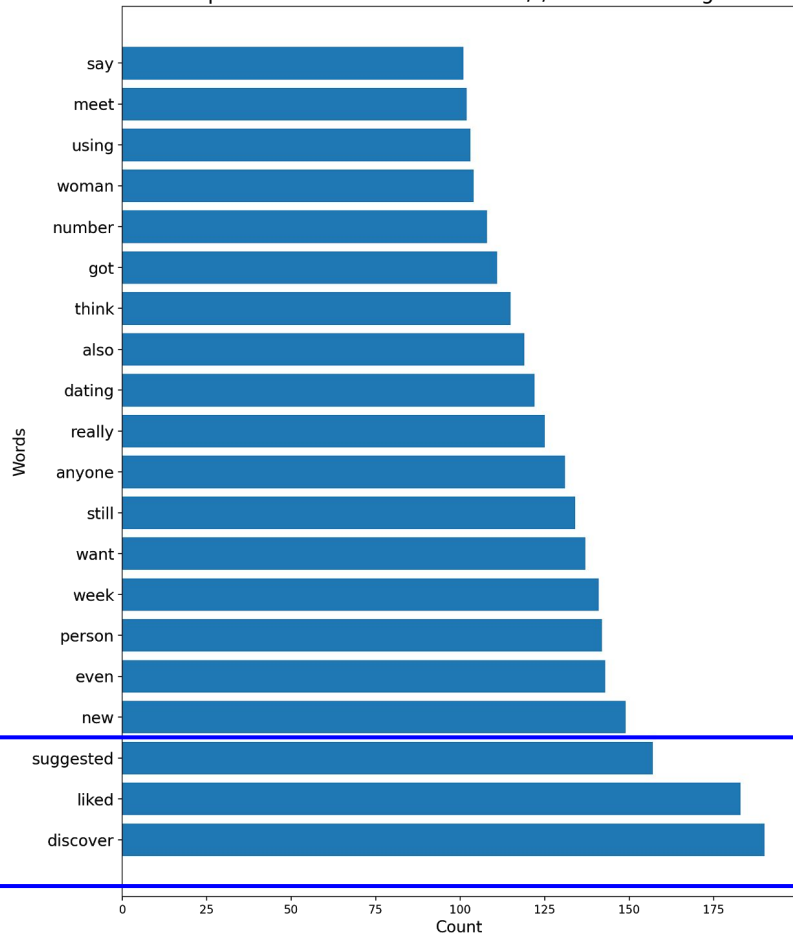
Exploratory Data Analysis

Putting the date...into dat(e)-a

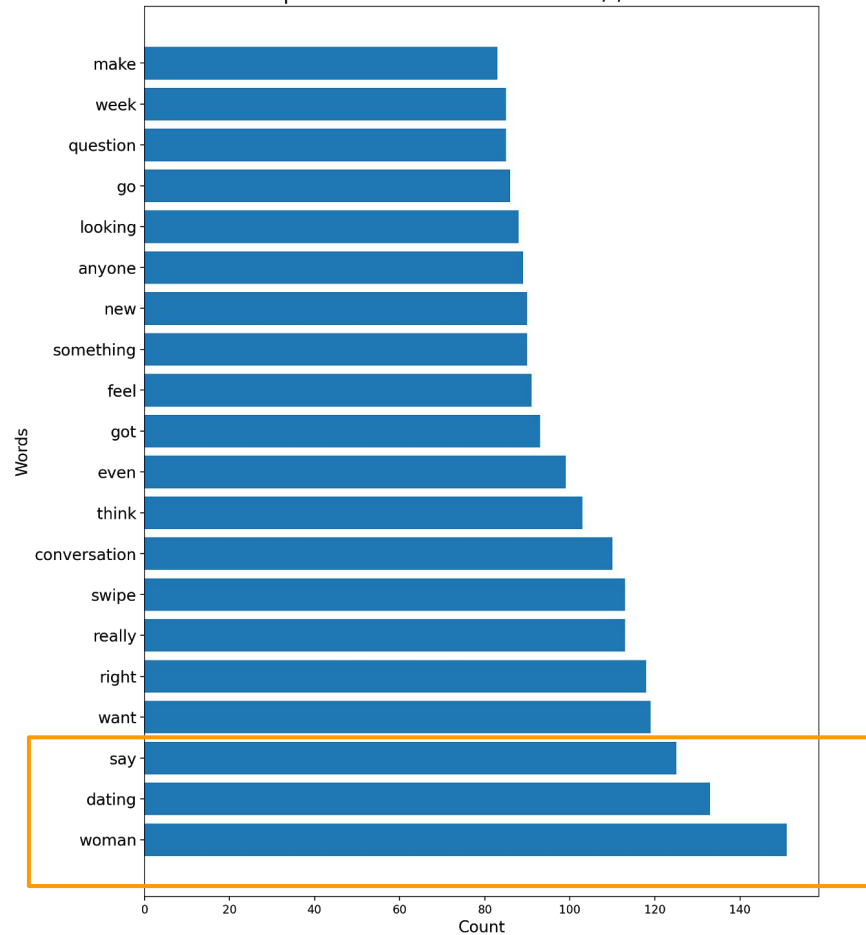
/r/coffeemeetsbagel



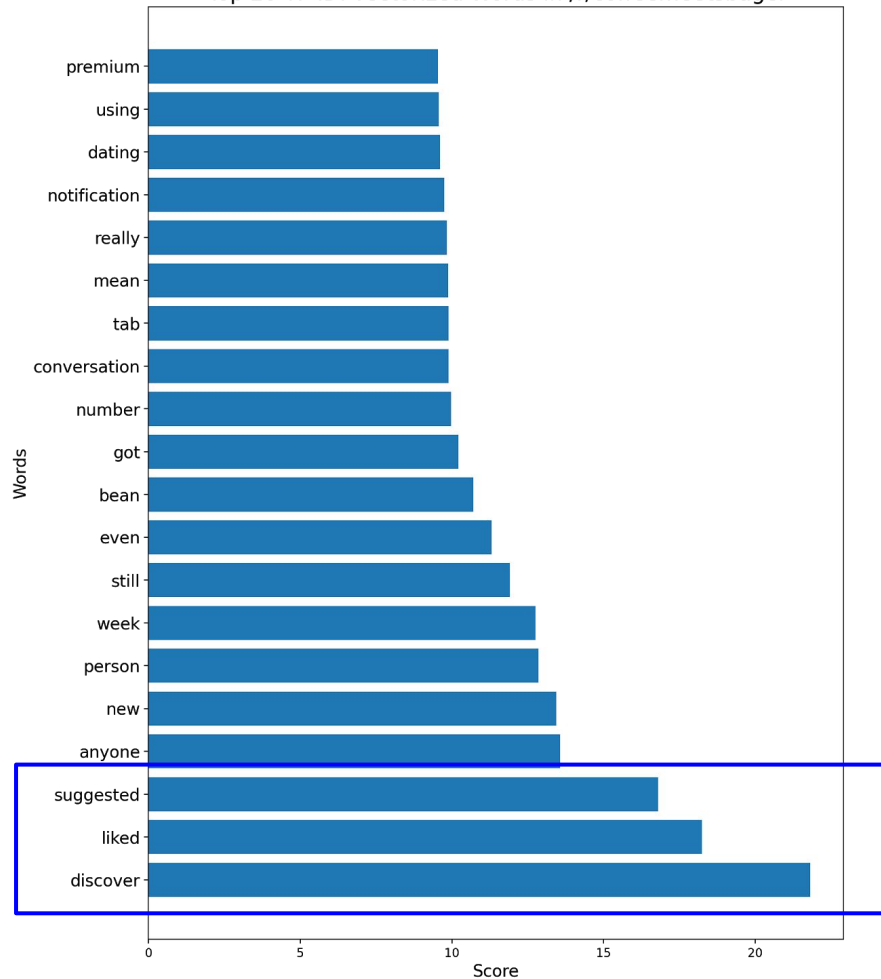
Top 20 CountVectorized Words in /r/coffeemeetsbagel



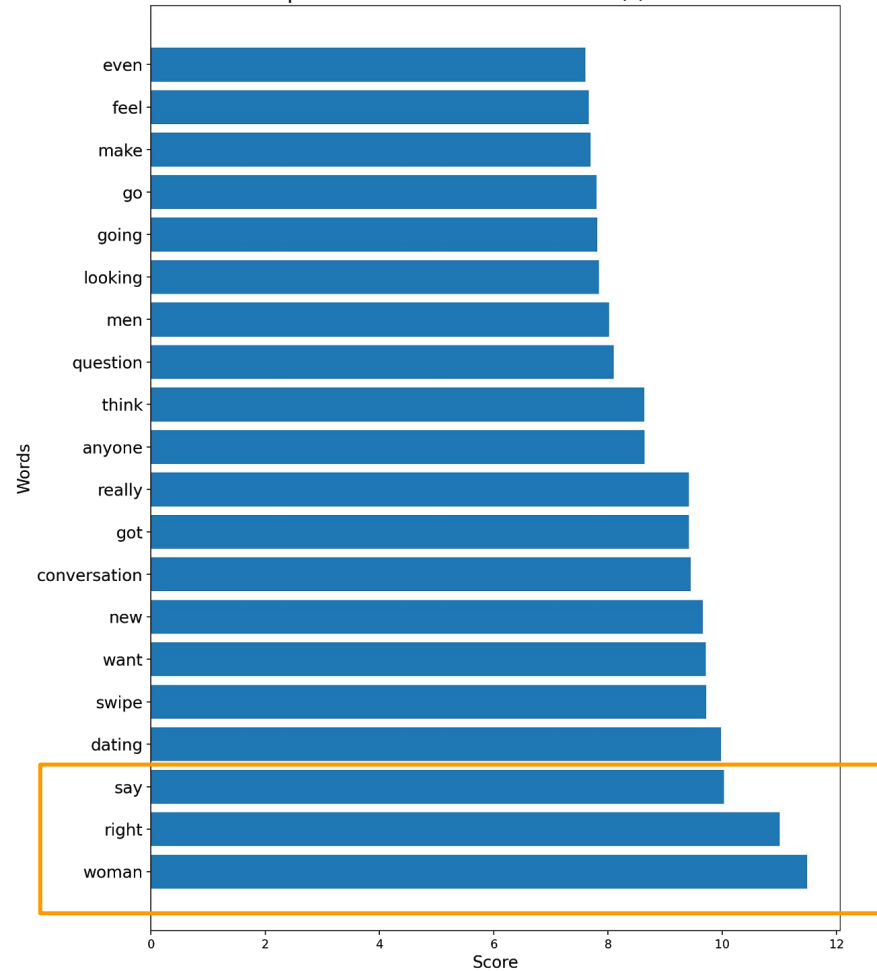
Top 20 CountVectorized Words in /r/bumble



Top 20 TF-IDFVectorized Words in /r/coffeemeetsbagel



Top 20 TF-IDFVectorized Words in /r/bumble



2

Modelling



Modelling Results

Model	Train Score (CountVectorized)	Test Score (CountVectorized)	Train Score (Tf-IdfVectorized)	Test Score (Tf-IdfVectorized)
Log Reg	0.740	0.703	0.733	0.718
Naive Bayes	0.770	0.736	0.757	0.712
Random Forest	0.735	0.700	0.749	0.683
KNN	0.651	0.635	0.752	0.739
SVM	0.745	0.701	0.747	0.712



Naive Bayes Model

	Predicted /r/bumble	Predicted /r/coffeemeetsbagel
Actual /r/bumble	115	51
Actual /r/coffeemeetsbagel	53	175

Correctly classified

- 76.8% of actual CMB posts
- 77.4% of predicted CMB posts
- 69.3% of actual Bumble posts



Metrics	Percentage, %
Accuracy	73.6
Misclassification Rate	26.4
Recall	76.8
Precision	77.4
Specificity	0.693



Misclassification Analysis

● Predicted Bumble, Actual CMB

advice hey could please check rate bio create feedback much appreciated http imgur com eerjg bio looking inspire pick guitar hold hand st
rolling make believe cooking skill acceptable really caring person four houseplant none dead yet moved part town cool place definitely ta
lk thanks everyone wish best

need advice hey facetime met online soon talking week already worried making bad impression facetime never really made past par
t tip make sure still want go next week spring break

help service helping f month start paid premium service however cancelled early january saw still getting charged emailed help
service got told remove payment detail continued charged emailed reply second charged service want anyone help

work please help example jane log tuesday john popup john login monday tuesday log wednesday wednesday john liked jane amp x b
happens case jane liked tues liked wed show swipe deck disappear since timing jane receive notification wednesday jane loses du
e timing make sense jane login every week rack swiping since woman drive confused



Misclassification Analysis

● Predicted CMB, Actual Bumble

help new online dating help dating far got bit nerdy experience new thing movie video game tabletop rpgs anime satisfying curiosity talk academic big issue appreciate broad sense humor drive self improvement frankly long honest thanks advance

advice recently divorced hello used met ex husband really nervous dating online divorced love insight met late early divorced liked successful age range successful gotten divorced insight helpful thank

never got single hit several month got nothing appropriate post help

need suggestion texting long paragraph texting really nice sweet often sends really mean really long paragraph appreciate honestly little overwhelming also considering responded asap set tone let without rude discouraging also havent met hate give bad impression



beware military scam caught scammer told army deployed west africa texted immediately feeling scammer since experienced many to ld positive experience military men respond asked told lie verify mentioned could nigerian scammer soon sent immediately unmatched immediately reported fake careful old especially person tell military deployed foreign country wish better job verifying experience many scammer site



Strongest Predictors

	words	coefs
1332	liked	-5.125009
748	discover	-5.176523
2335	suggested	-5.212401
1667	person	-5.240181
1541	new	-5.298168
2770	want	-5.318269
842	even	-5.359726
2822	week	-5.370364
2261	still	-5.414087
150	anyone	-5.425323

/r/coffeemeetsbagel

2211	split	-10.601473
348	bio problem	-10.601473
921	fake bot	-10.601473
2214	spotlight	-10.601473
341	big question	-10.601473
943	feel uncomfortable	-10.601473
337	bi	-10.601473
336	bff	-10.601473
314	bee	-10.601473
315	beeline	-10.601473

/r/bumble



Evaluation

- High number of similarities in misclassified posts
- Users seeking general feedback on dating, relationships, profiles, etc
- Complaints & scammers



Recommendations

- CMB
 - Has more special features (i.e discover, suggested, like)
 - Improve functionalities of existing features
- Bumble
 - Lacks distinguishing features
 - Implement more unique functions
- Both
 - Customer service and app security



Thanks!

Any questions ?