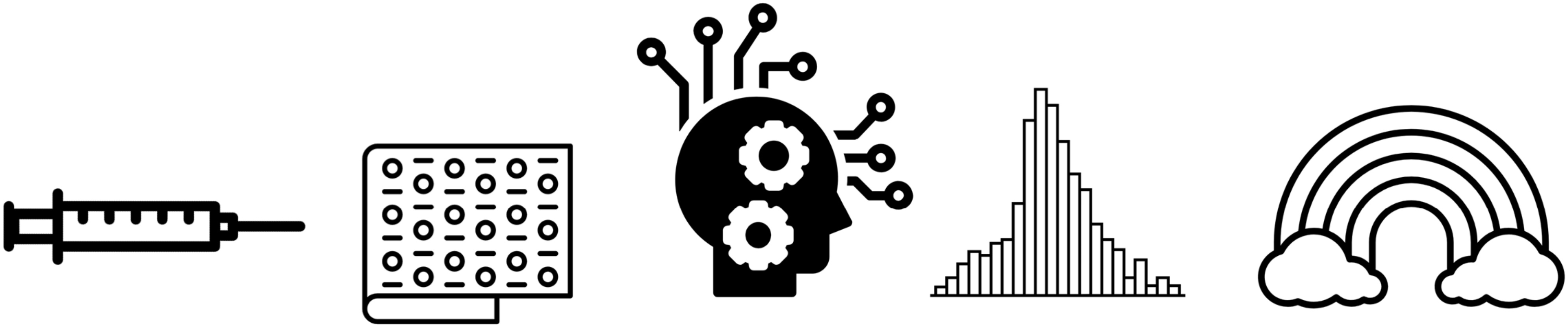


from here to recovery

Using a Machine Learning Model to predict the transition
from active substance use to recovery from addiction

Shawn Syms, 14 April 2020

- **executive summary**
- **problem statement**
- **data**
- **context**
- **models**
- **recommendations**

agenda

What if you could tell where clients are in the recovery lifecycle based on their writing?

Lw*v#fkdoohqjlqj#wr#uxq#dgglfwlrq0uhfryhu|#lqlwldwlyhv/#zkhwkhu#lqsdwlhqw#ru#rxwsdwlhqw#uhkde#ru#hyhq#vxshuylvhg#lqmhfwlrq#vlwhv#zlwk#rq0vlwh#dffhvv#wr#wuhdwphqw#dqg#frxqvholqj1#Uhylhzlqj#wkh#wkrxjkw#dqg#ihholqjv#ri#rxu#sdwlhqw#dqg#folhqwv/#riwhq#wkurxjk#wkhlu#fuhdwlyh#rxwsxw/#lv#rqh#nh|#sduw#ri#wkh#uhfryhu|#surfhvv1

Zh#duh#JhwEhwwhu/d#qrq0surilw#gdwd0vflhqfh#frqvxowdqf|#zlwk#h{shulhqfh#lq#wkh#uhfryhu|#vhfwru/#dqg#zh#fdq#khos1#Zh#nqrz#wkdw#lq#frxqvholqj/#folhqwv#duh#qrw#dozd|v#wuxwkixo#derxw#wkhlu#sureohp#vxevwdqfh#xvh#ehfdxvh#ri#jxlow/#vkdph#ru#vwljpd1#Vr#zh#kdyh#ghyhorshg#d#pdfklqh#ohduqlqj#prgho#wkdw#fdq#khos#ghwhuplqh#li#d#folhqw#lv#vwloo#dfwlyho|#xvlqj#vxevwdqfhv#ru#dfwlyho|#zrunlqj#rq#wkhlu#uhfryhu|1

Wkh#prgho#kdv#d#vxffhvv#udwh#ri#ryhu#<3#shufhqw/#zklfk#lv#kljkhu#wkdq#pdq|#vxemhfwlyh#wuhdwphqw#prgdolwlhv#dqg#ohvv#lqydvlyh#dqg#frqiurqwdwlrqdo#wkdq#rwkhu#frpsoldqfh#wrrov#vxfk#dv#xulqdo|vlv1#Dqg#doo#wkh#kdug#zrun#kdsshqv#ehklqg#wkh#vfhqhv#lq#wkh#prgho#lwvhoi1#Wrgd|#zh#zloo#zdon#|rx#wkurxjk#wkh#ghyhorsphqw#surfhvv#dqg#vkrz#|rx#krz#hdv|#lw#lv#iru#|rx#wr#lpsohphqw#dqg#xvh#wkh#prgho/#hpsrzhulqj#|rx#wr#ohdyh#wklv#urrp#zlwk#d#srzhuixo#qhz#wrro#lq#|rxu#duvhqdo1

DwJhwEhwwhu/zh#eholhyh#wkdw#hyhu|rqh/#qr#pdwwhu#zkdw#wkhlu#flufxpvwdqfhv/#ghvhuyhv#wkhlu#ehvw#fkdqfh#iru#khdowk/#zhoojhvv/#surgxfwlylw|#dqg#kdsslqhvvÈ dqg#zh#nqrz#wkdw#|rx#gr/#wrr1#Ohw*v#zrun#wrjhwkhu#wr#khos#pdnh#wkdw#d#uhdolw|1

executive summary

Can we successfully build a natural language processing (NLP) binary-classification model that will distinguish between writing, in the form of reddit posts, by active substance users vs people in recovery from addiction? How accurate, and how generalizable, could it be?

Wr#ilqg#rxw/#zh#frqvwuxfwhg#dq#QOS#prgho#wkdw#wrrn#dv#wudlqlqj#lqsxw#wkh#frqwhqw#iurp#ryhu#6/333#suhsurfhvvhg#uhgglw#srvwv/#kdoi#ruljlqdwlqj#lq#rxu#vxeuhgglw#ghglfdwhg#wr#glvfxvvlrq#ri#dfwlyh#vxevwdqfh#xvh#dqg#kdoi#iurp#rxu#vxeuhgglw#irfxvhg#rq#uhfryhu|#iurp#guxj#)#dofrkro#dgglfwlrq1#Zh#udq#wkh#prgho#xvlqj#d#wudlq0whvw#vsolw/#wkhq#xvhg#dqrwkhu#933#uhgglw#srvwv#dv#xqnqrzq#gdwd#iru#wkh#prgho#wr#fodvvli|#dv#dfwlyh#guxj#xvh#ru#dfwlyh#uhfryhu|1#Wkh#surmhfw#zloo#eh#frqvlghuhg#vxffhvvixo#li#wkh#prgho#fdq#surshuo|#lghqwli|#<3#shufhqw#ru#pruh#ri#wkh#srvwv1

problem statement

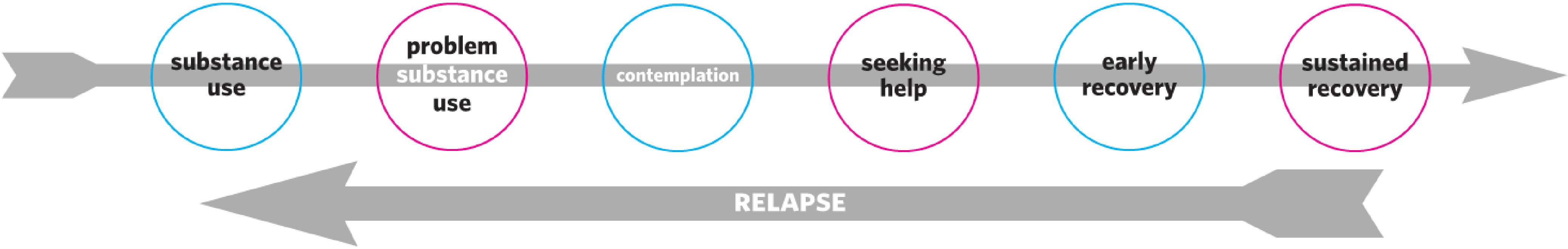
Can we successfully build a natural language processing (NLP) binary-classification model that will distinguish between writing, in the form of reddit posts, by active substance users vs people in recovery from addiction? How accurate, and how generalizable, could it be?

Wr#lqg#rxw#z h#frqvwuxfwhg#dq#QOS#p rgho#wkdw#wrrn#dv#wudlqlqj#lqsxw#wkh#frqwhqw#iurp #ryhu#6/333#suhsurfhvvhg# uhgglw#srvwv/#kdoi#ruljlqdwlqj#lq#irxu#vxeuhgglwv#ghglfdwhg#wr#glvfxvvlrq#ri#dfwlyh#vxevwdqfh#xvh#dqg#kdoi#iurp #irxu# vxeuhgglwv#irfxvhg#rq#uhfryhu|#iurp #guxj#) #dofrkro#dgglfwlrq1#

Z h#udq#wkh#p rgho#xvlqj#d#wudlq0whvw#vsolw# wkhq#xvhg#dqrwkhu#933#uhgglw#srvwv#dv# xqnqrz q#gdwd#iru#wkh#p rgho#wr#fodvvli|# dv#dfwlyh#guxj#xvh#ru#dfwlyh#uhfryhu|1# Wkh#surmhfw#z loo#eh#frqvlghuhg# vxffhvvixo#li#wkh#p rgho#fdq#surshuo|# lghqwli|#<3#shufhqw#ru#p ruh#ri#wkh#srvwv1

- r/recovery
- r/stopdrinking
- r/opiatesrecovery
- r/redditorsinrecovery

- r/drugs
- r/stims
- r/opiates
- r/drinking

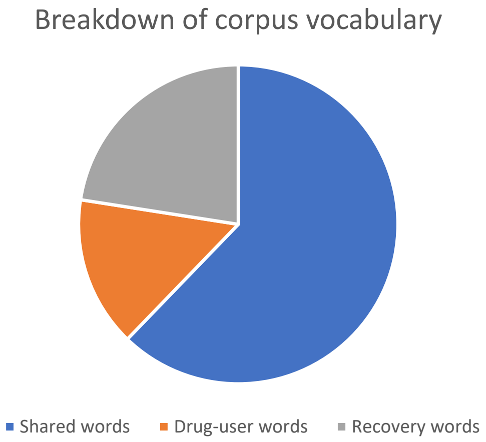


Challenge

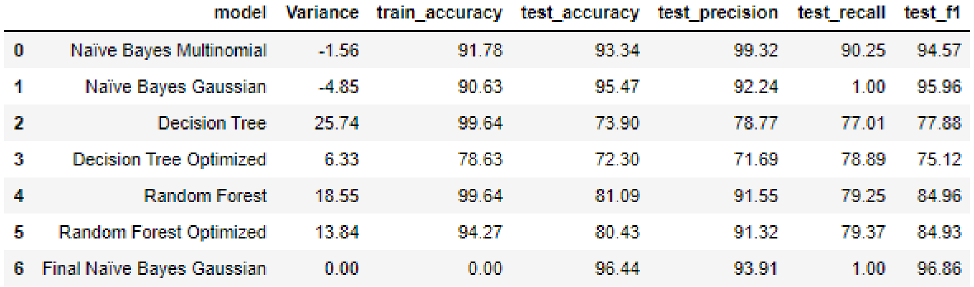
- Since relapse is a part of recovery, we are talking about one community vs two separate groups
- There is a lot of shared vocabulary between both groups (62% of all post vocabulary is shared between both sets of subreddits)

Opportunity

- 15% of vocabulary existed only in drug groups
- 23% of vocabulary existed only in recovery groups

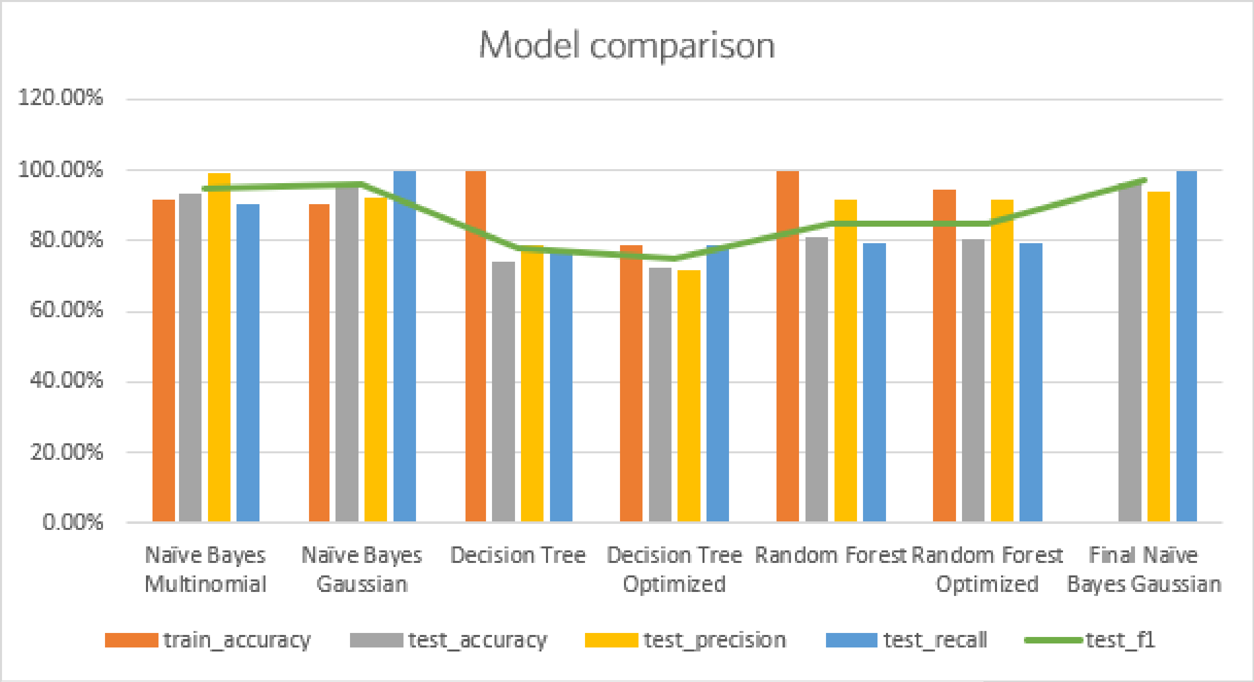


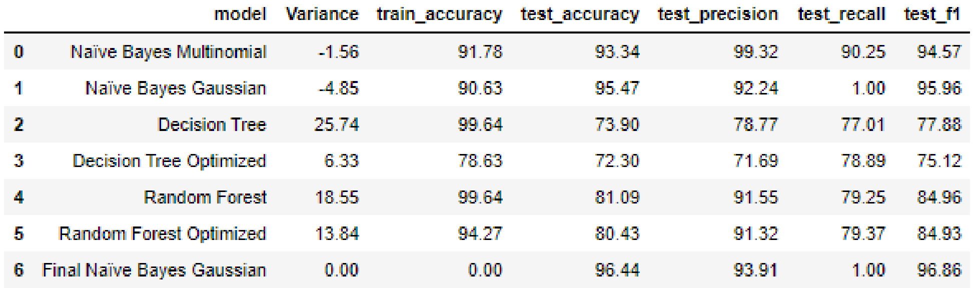
models: overview



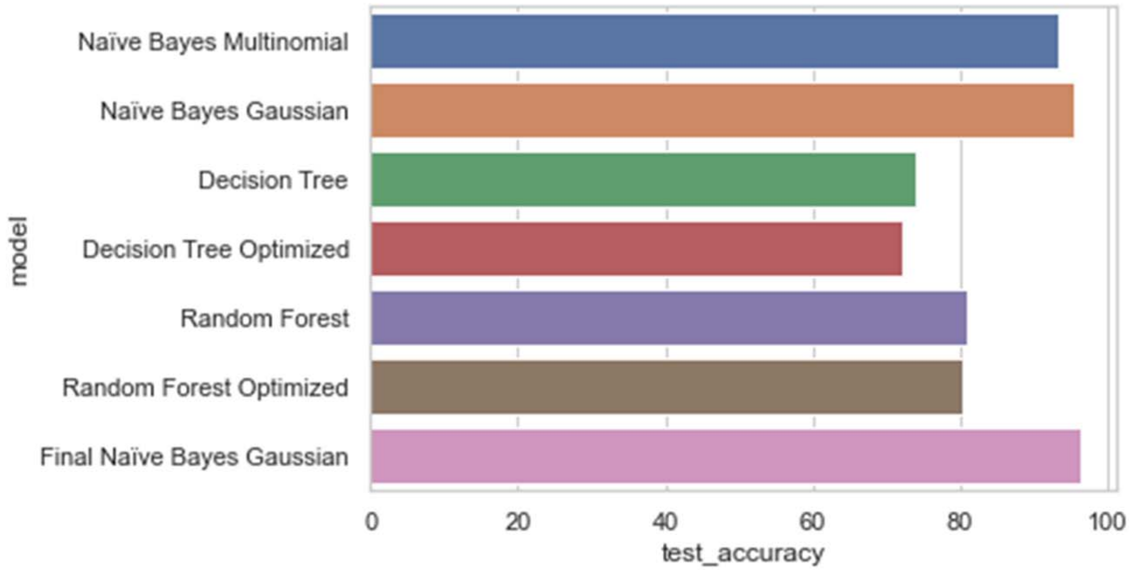
	model	Variance	train_accuracy	test_accuracy	test_precision	test_recall	test_f1
0	Naïve Bayes Multinomial	-1.56	91.78	93.34	99.32	90.25	94.57
1	Naïve Bayes Gaussian	-4.85	90.63	95.47	92.24	1.00	95.96
2	Decision Tree	25.74	99.64	73.90	78.77	77.01	77.88
3	Decision Tree Optimized	6.33	78.63	72.30	71.69	78.89	75.12
4	Random Forest	18.55	99.64	81.09	91.55	79.25	84.96
5	Random Forest Optimized	13.84	94.27	80.43	91.32	79.37	84.93
6	Final Naïve Bayes Gaussian	0.00	0.00	96.44	93.91	1.00	96.86

models: overview

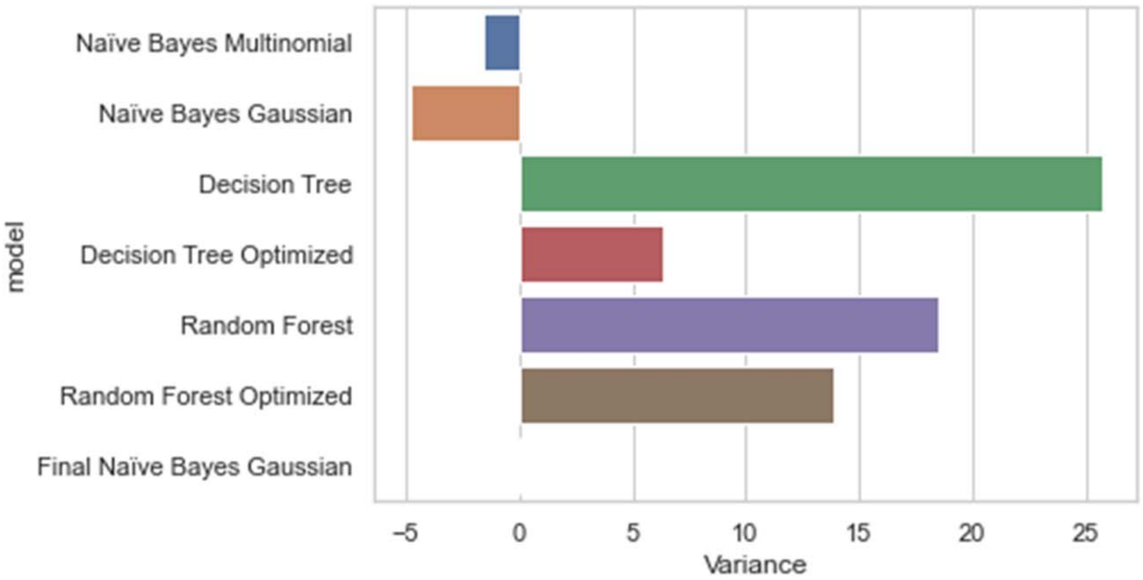




	model	Variance	train_accuracy	test_accuracy	test_precision	test_recall	test_f1
0	Naïve Bayes Multinomial	-1.56	91.78	93.34	99.32	90.25	94.57
1	Naïve Bayes Gaussian	-4.85	90.63	95.47	92.24	1.00	95.96
2	Decision Tree	25.74	99.64	73.90	78.77	77.01	77.88
3	Decision Tree Optimized	6.33	78.63	72.30	71.69	78.89	75.12
4	Random Forest	18.55	99.64	81.09	91.55	79.25	84.96
5	Random Forest Optimized	13.84	94.27	80.43	91.32	79.37	84.93
6	Final Naïve Bayes Gaussian	0.00	0.00	96.44	93.91	1.00	96.86



models: overview



models: overview

models: naïve bayes models

Naïve Bayes (Multinomial)

tn	fn
266	47
3	435
fp	tp

Naïve Bayes (Gaussian)

tn	fn
313	0
34	404
fp	tp

Naïve Bayes (Multinomial)
(train-test split)

accuracy (train)	91.78%
accuracy (test)	93.34%
variance	-1.56%
precision	99.32%
recall	90.25%
f1	94.57%

Naïve Bayes (Gaussian)
(train-test split)

accuracy (train)	90.63%
accuracy (test)	95.47%
variance	-4.84%
precision	92.24%
recall	100.00%
f1	95.96%

- both models performed well and were slightly underfit
- Multinomial had a lower variance, but it was edged out by Gaussian in all other metrics

models: decision tree (& optimized)

Decision Tree

tn	fn
210	103
93	345
fp	tp

Optimized Decision Tree

tn	fn
229	84
124	314
fp	tp

Decision Tree
(train-test split)

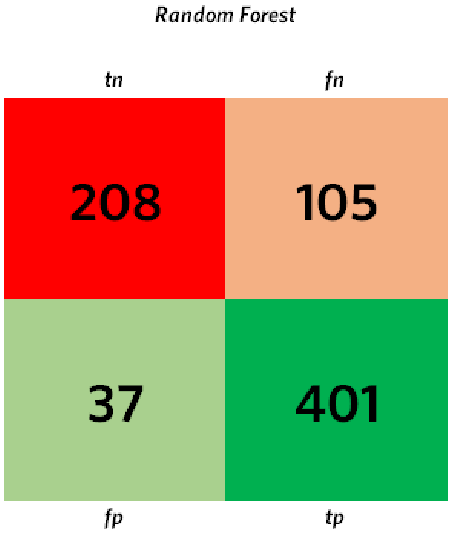
accuracy (train)	99.64%
accuracy (test)	73.90%
variance	25.74%
precision	78.77%
recall	77.01%
f1	77.88%

Optimized Decision Tree
(train-test split)

accuracy (train)	78.63%
accuracy (test)	72.30%
variance	6.33%
precision	71.69%
recall	78.89%
f1	75.12%

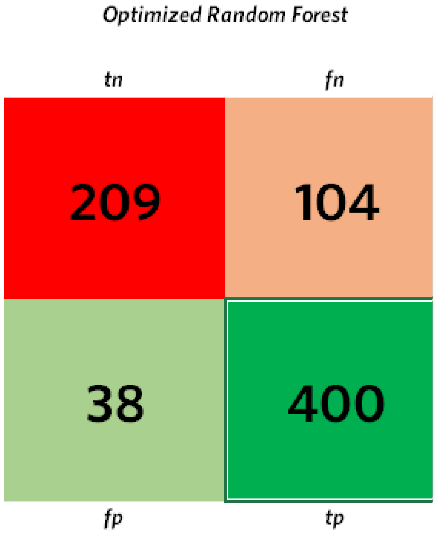
- the default decision-tree model was dramatically overfit at over 25%
- optimization using GridSearchCV helped reduce overfitting but at the significant expense of accuracy

models: random forest (& optimized)



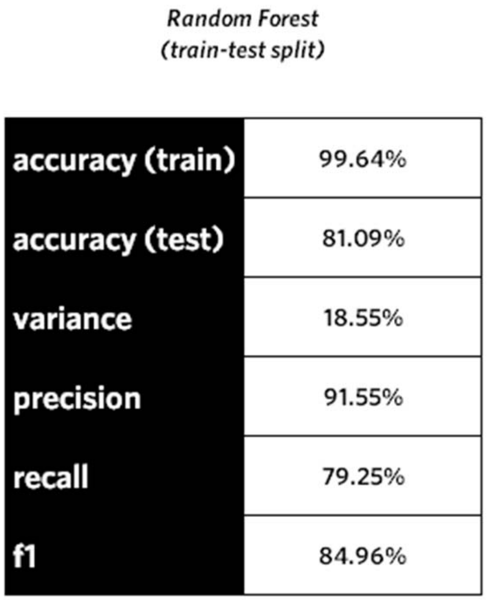
Random Forest

tn	*fn*
208	105
37	401
fp	*tp*



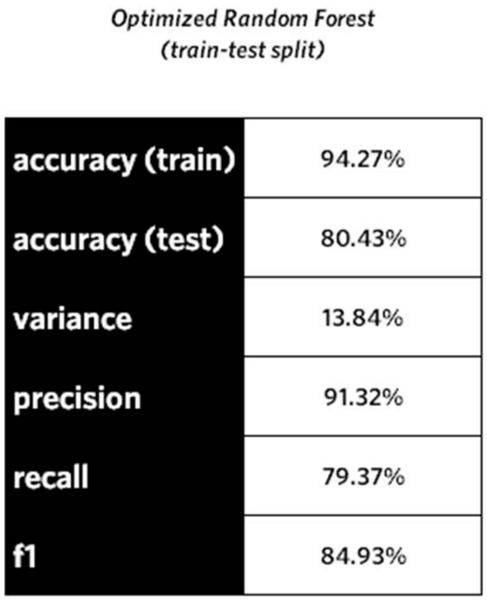
Optimized Random Forest

tn	*fn*
209	104
38	400
fp	*tp*



Random Forest
(train-test split)

accuracy (train)	99.64%
accuracy (test)	81.09%
variance	18.55%
precision	91.55%
recall	79.25%
f1	84.96%

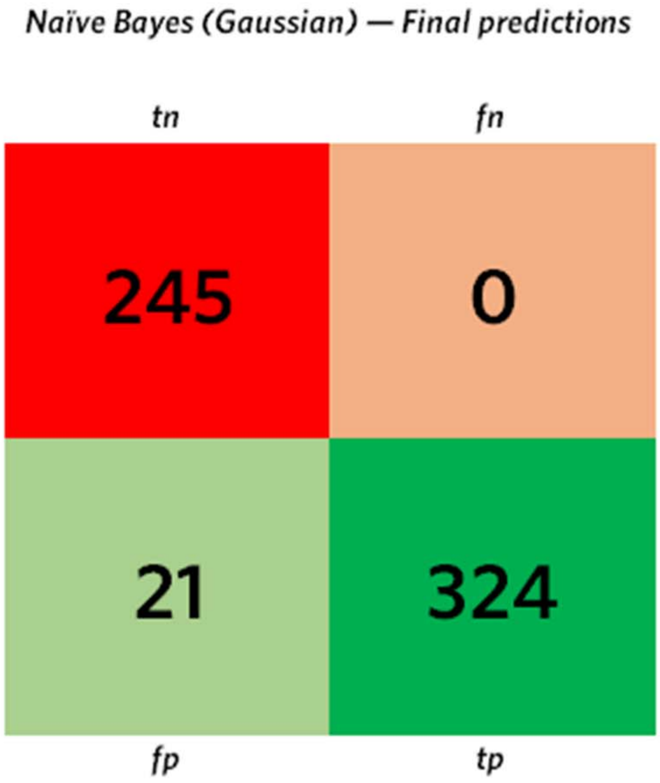


Optimized Random Forest
(train-test split)

accuracy (train)	94.27%
accuracy (test)	80.43%
variance	13.84%
precision	91.32%
recall	79.37%
f1	84.93%

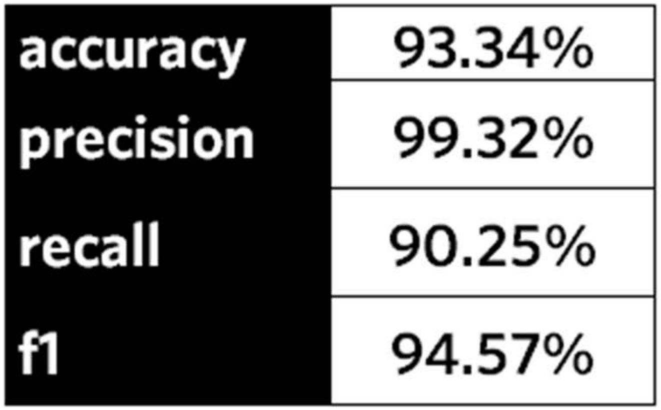
- the default model was significantly overfit
- optimization helped, but overall performance compared unfavourably with the NB models

successful model



Naïve Bayes (Gaussian) — Final predictions

tn	fn
245	0
21	324
fp	tp

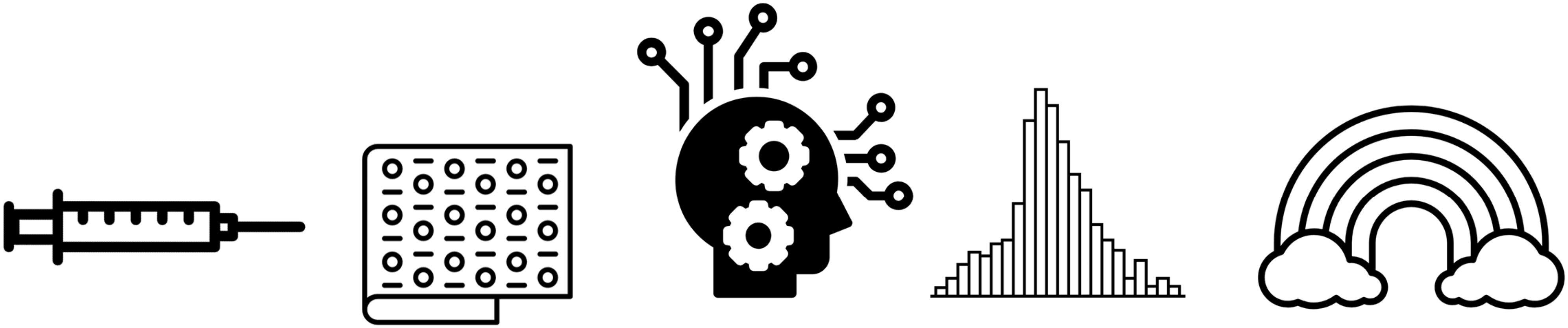


accuracy	93.34%
precision	99.32%
recall	90.25%
f1	94.57%

- Highest accuracy of all models
- Smallest variance of all models
- Slightly underfit (4.86%

We conclude from our model statistics that this project is indeed viable and meets our success criteria. It is our recommendation that the model immediately be adopted in all treatment settings that involve a creative-writing component, from group work to art-therapeutic workshops to individualized counselling in both in-patient and out-patient settings. We believe that our model can be extended to analyze all social media postings by individual clients, either in an outpatient setting or with proper consent upon entry into in-patient rehabilitation stay, and this implementation will be the next focus of our research.

recommendations



questions