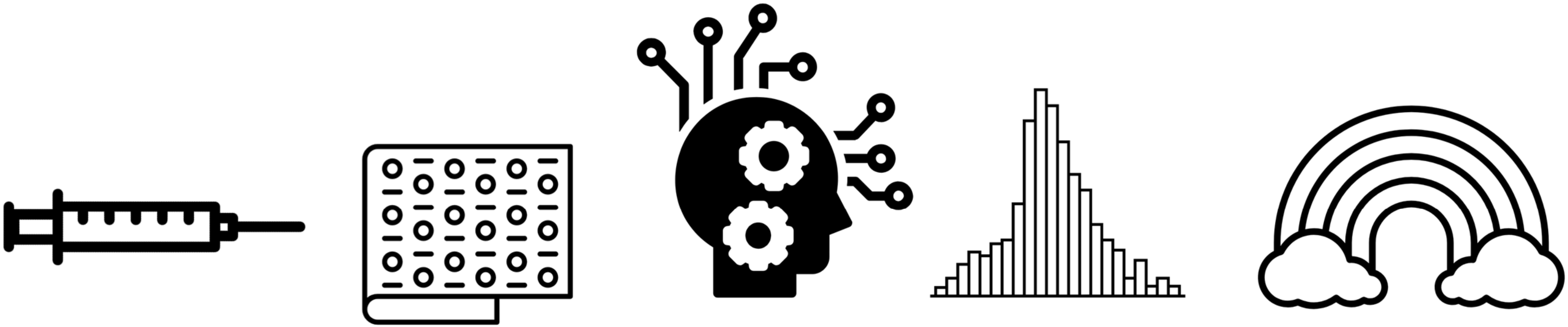


reddit2rehab

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

Shawn Syms

- **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?

It's challenging to run addiction-recovery initiatives, whether inpatient or outpatient rehab or even supervised injection sites with on-site access to treatment and counselling. Reviewing the thoughts and feelings of our patients and clients, often through their creative output, is one key part of the recovery process.

We are GetBetter, a [hypothetical] non-profit data-science consultancy with experience in the recovery sector, and we can help. We know that in counselling, clients are not always truthful about their problem substance use because of guilt, shame or stigma. So we have developed a machine learning model that can help determine if a client is still actively using substances or actively working on their recovery.

The model has a success rate of over 90 percent, which is higher than many subjective treatment modalities and less invasive and confrontational than other compliance tools such as urinalysis. And all the hard work happens behind the scenes in the model itself. Today we will walk you through the development process and show you how easy it is for you to implement and use the model, empowering you to leave this room with a powerful new tool in your arsenal.

At GetBetter, we believe that everyone, no matter what their circumstances, deserves their best chance for health, wellness, productivity and happiness—and we know that you do, too. Let's work together to help make that a reality.

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?

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?

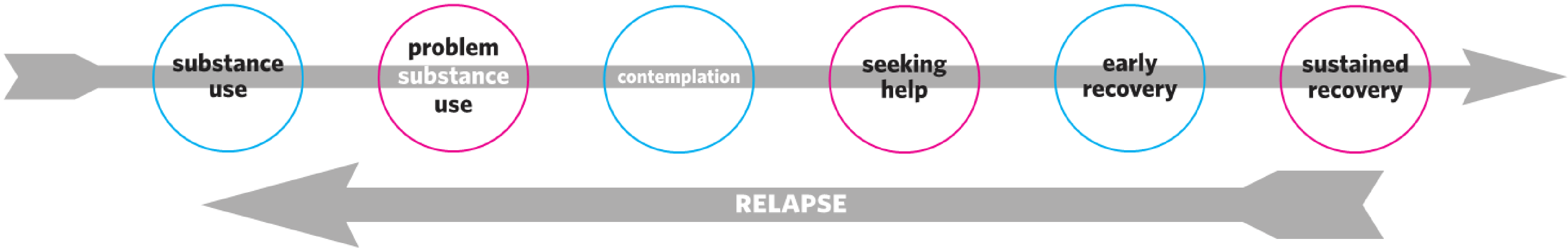
To find out, I constructed an NLP model that took as training input the content from over 3,000 preprocessed reddit posts, half originating in four subreddits dedicated to discussion of active substance use and half from four subreddits focused on recovery from drug & alcohol addiction.

I ran the model using a train-test split, then used another 600 reddit posts as unknown data for the model to classify as active drug use or active recovery. The project would be considered successful if the model can properly identify 90 percent or more of the posts.

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

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

data

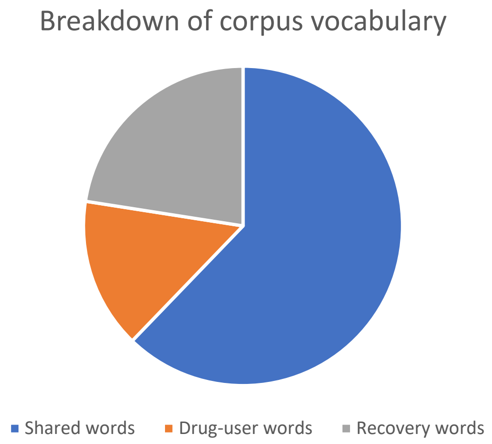


Challenge

- 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

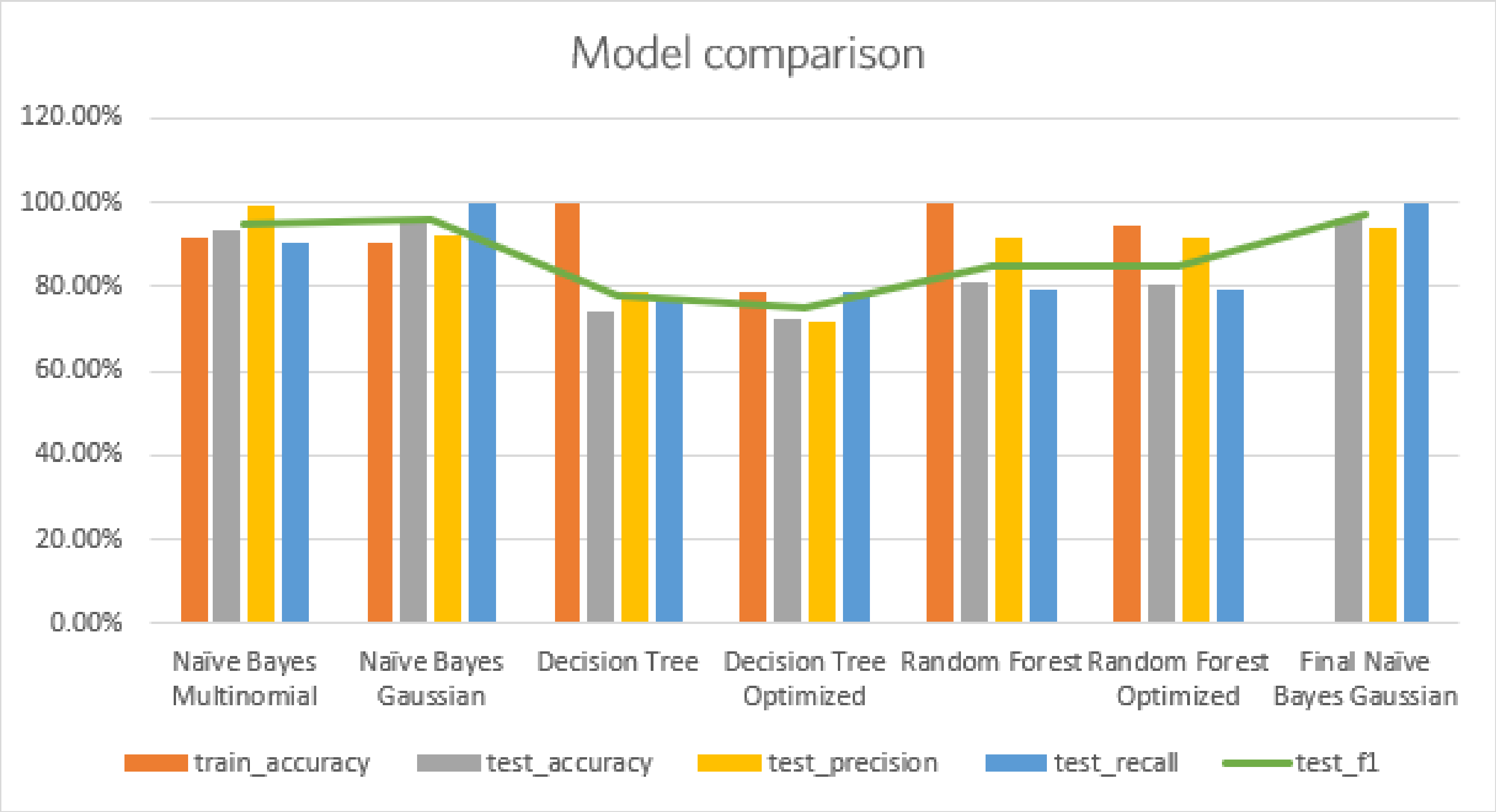


I compared four different models to see which would perform best:

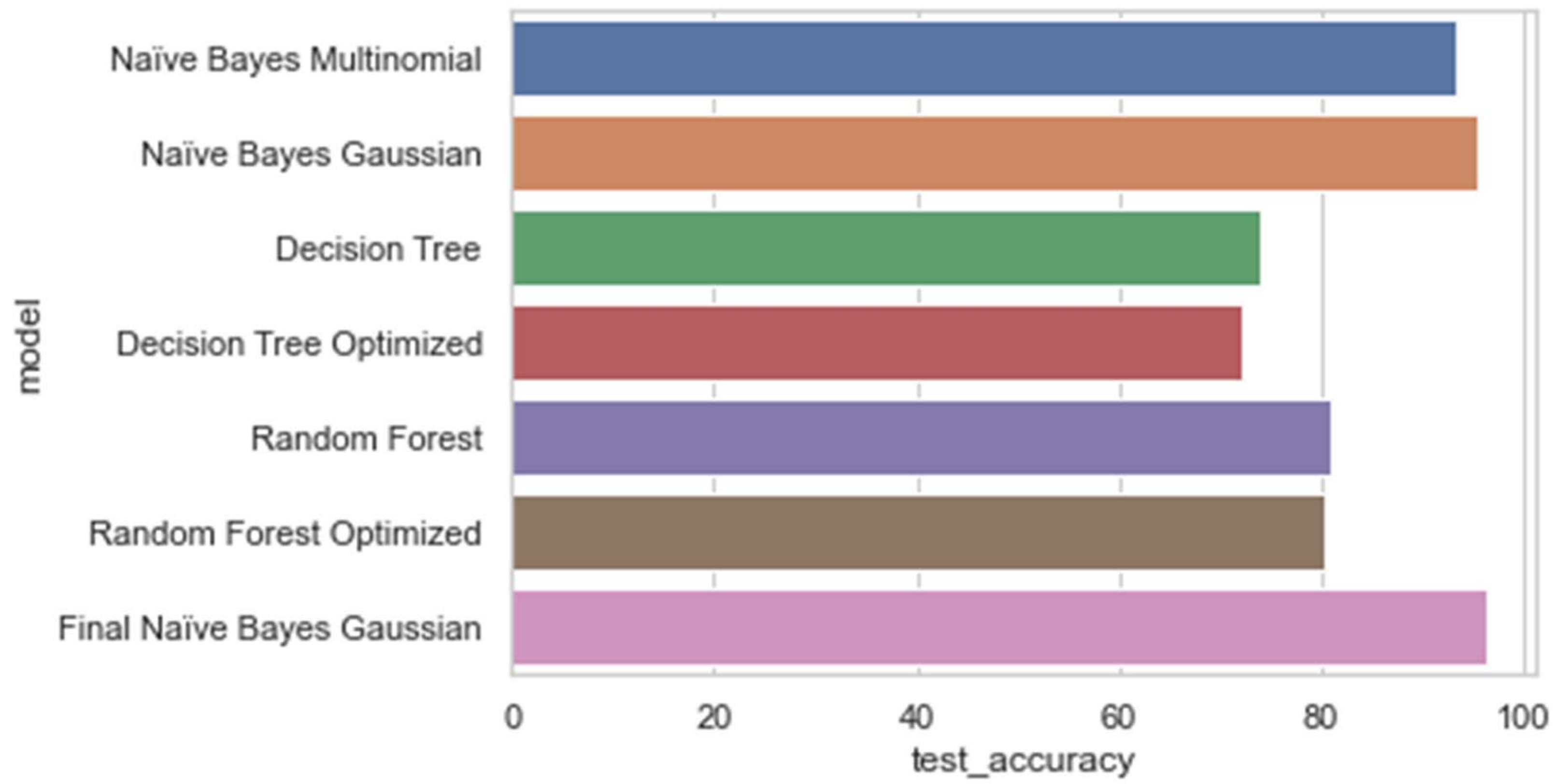
- Naïve Bayes Multinomial
- Naïve Bayes Gaussian
- Decision Tree
- Random Forest (both with default settings and with hyperparameters performance-tuned)

	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

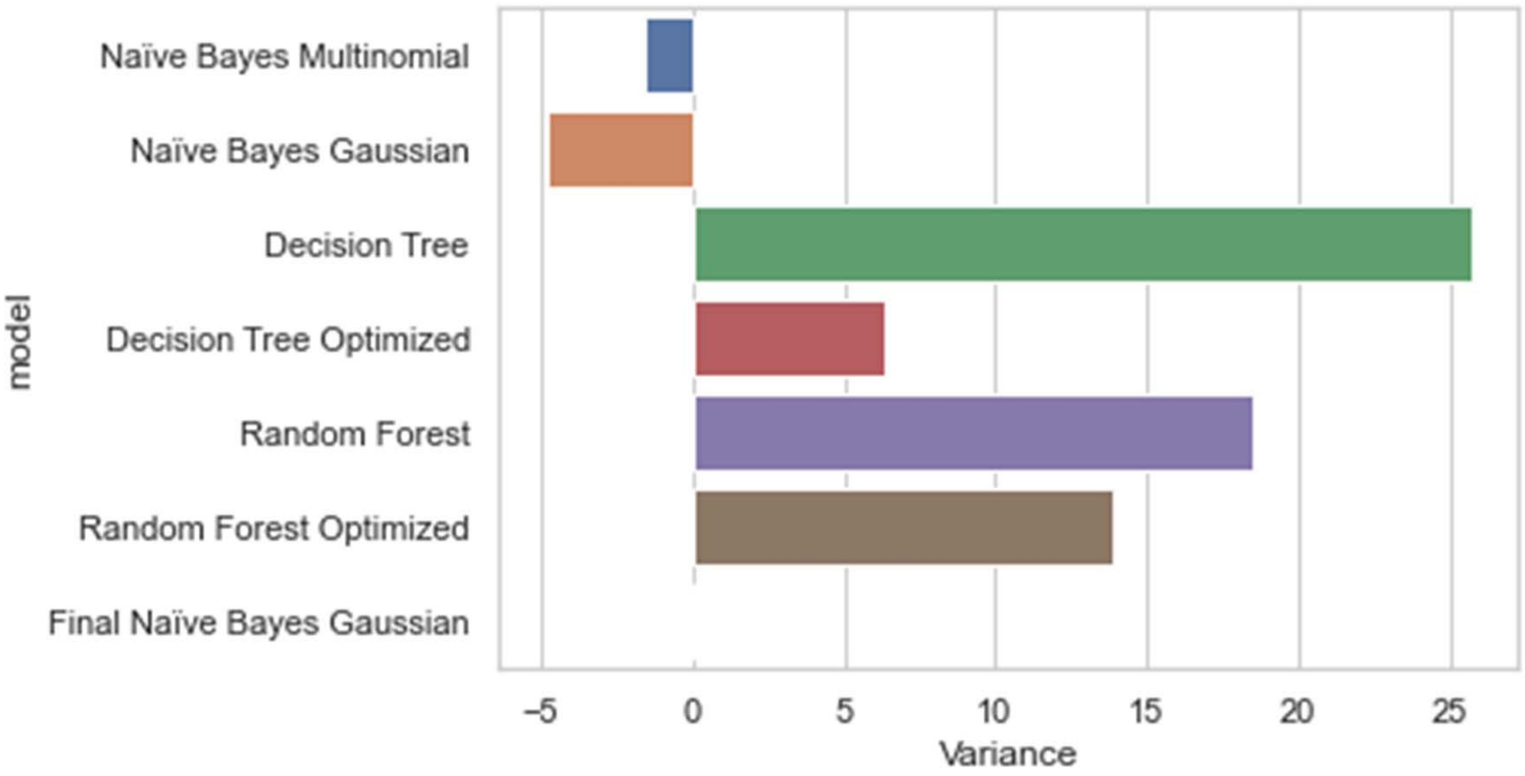
models: performance



models: metrics



models: accuracy



models: variance between training & unknown data

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: naïve bayes models

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: decision tree (& 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

models: random forest (& optimized)

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%

successful model

The model's performance statistics indicate this project is viable. Next steps could include expanding the number of inputs exponentially to train the model further, designing further experiments involving other social media, and working with recovery advocates and stakeholders to determine the most appropriate intervention that the model could potentially support.