Exploring Online Drug Reviews for Depression Treatment using Text Mining Models

Project Introduction by Grace Luan

Introduction: Mental Health Awareness and Depression Treatment

- **Gender Representation**:

In some studies, the proportion of female reviewers for depression treatment drugs can be as high as 70% to 80%, depending on the platform and sample size.

- **Age Distribution**:

Reviewers span a wide age range, from teenagers to older adults. However, there may be peaks in review activity among individuals in their 30s to 50s, reflecting the typical age of onset for depression and seeking treatment.

- **Satisfaction Ratings**:

Satisfaction ratings can vary greatly, but on average, around 60% to 70% of reviewers may express satisfaction with their depression treatment, while the remainder may report dissatisfaction or neutral feelings.

- **Side Effect Reporting**:

A significant portion of reviews (often around 30% to 40%) may mention specific side effects experienced while taking depression treatment drugs, with common side effects including nausea, dizziness, weight changes, and sexual dysfunction.

WebMD® health services

Better Information. Better Health.

User Reviews for Aricept Oral

Comments & ratings on the side effects, benefits, and effectiveness of Aricept Oral.

View Free Coupon >

Full Drug Information | **Reviews (236)**

Show ratings & reviews for

All Conditions (236 reviews)

2.9 Overall Rating

Share Your Experience

Effectiveness

5 stars	44
4 stars	31
3 stars	42
2 stars	26
1 star	93

Ease of Use

5 stars	85
4 stars	46
3 stars	49
2 stars	14
1 star	42

Satisfaction

5 stars	45
4 stars	28
3 stars	36
2 stars	24
1 star	103

Most voted positive review

89 People found this comment helpful

My husband was prescribed Aricept back in November and it's improved his memory/confusion... I noticed in about 2 weeks that he wasn't as confused and had better recall. I know it's not a cure but it helps and it's very obvious right now.

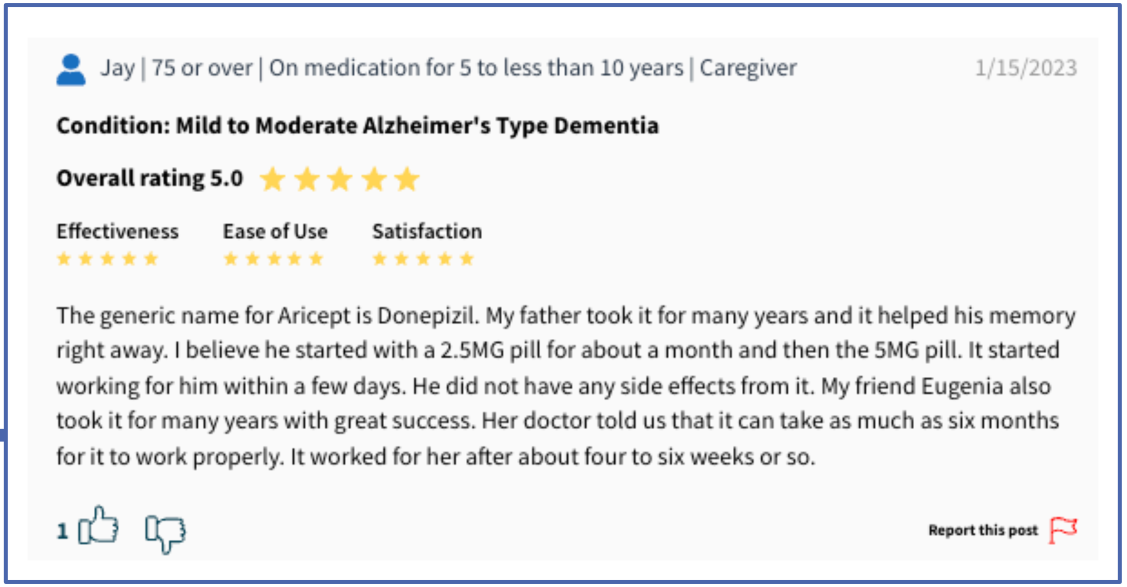
Most voted negative review

16 People found this comment helpful

I have had sever tremors ever sinse in legs and hips. I'm now in a wheel chair.Not sure if this is the cause. I woke up one morning could not walk.????

WebMD Dataset

12 columns, 362,807 rows, csv file



Age	Condition	Date	Drug	DrugId	EaseofUse	Effectiven	Reviews	Satisfactio	Sex	Sides	UsefulCou
75 or over	Stuffy Nose	9/21/14	25dph-7.5pe	146724	5	5	I'm a retired physician and	5	Male	Drowsiness, dizziness , dry mouth /nose/throat, headache ,	0
25-34	Cold Symptoms	1/13/11	25dph-7.5pe	146724	5	5	cleared me right up even w	5	Female	Drowsiness, dizziness , dry mouth /nose/throat, headache ,	1
65-74	Other	7/16/12	warfarin (bul	144731	2	3	why did my PTINR go from	3	Female		0
75 or over	Other	9/23/10	warfarin (bul	144731	2	2	FALLING AND DON'T REALIS	1	Female		0
35-44	Other	1/6/09	warfarin (bul	144731	1	1	My grandfather was prescr	1	Male		1
55-64	Other	7/19/08	warfarin (bul	144731	4	4	help heart condition opera	4	Male		0
25-34	Birth Control	6/15/17	wymzya fe	163180	5	5	Haven't gotten pregnant so	2	Female	Nausea , vomiting , headache , bloating , breast tenderness,	0
45-54	Disease of Ovaries with Cysts	1/30/17	wymzya fe	163180	5	5	I have take this for 5 years	5	Female	Nausea , vomiting , headache , bloating , breast tenderness,	0
25-34	Acne	4/27/16	wymzya fe	163180	4	2		2	Female	Nausea , vomiting , headache , bloating , breast tenderness,	1
55-64	Stuffy Nose	10/29/12	12 hour nasa	9800	4	2	The 12 hour spray only wo	2	Male	Temporary burning, stinging, dryness in the nose, runny nose	0
65-74	Other	3/15/16	pyrogallol cry	12112	5	5	Excellent in reducing inlam	5	Male		0
19-24	Birth Control	11/17/18	lyza	164750	5	5	Taking Lyza made me break	2		Nausea , vomiting , headache , bloating , breast tenderness,	0
	Birth Control	7/3/18	lyza	164750	2	1	This stuff needs to be	1	Female	Nausea , vomiting , headache , bloating , breast tenderness,	0
19-24	Birth Control	3/6/18	lyza	164750	2	3	I usually have zero to little	1	Female	Nausea , vomiting , headache , bloating , breast tenderness,	0
35-44	Birth Control	2/13/18	lyza	164750	5	5	I was concerned about star	5		Nausea , vomiting , headache , bloating , breast tenderness,	0
25-34	Birth Control	12/9/17	lyza	164750	2	2	The birth control was very	1		Nausea , vomiting , headache , bloating , breast tenderness,	1
25-34	Birth Control	10/7/17	lyza	164750	1	1	LYZA BIRTH CONTROL	1	Female	Nausea , vomiting , headache , bloating , breast tenderness,	1

	Age	Condition	Drug	Reviews	Satisfaction	Sex
3611	55-64	Anxiousness associated with Depression	loxapine		2	Female
3612	25-34	Anxiousness associated with Depression	loxapine	i like the fact i canwake up and feel like i c...	1	Female
3613	45-54	Anxiousness associated with Depression	loxapine	a very dangerous drug especially that should b...	1	Male
3618	55-64	Anxiousness associated with Depression	loxapine	MY HUSBAND HAS TAKEN THIS MEDICATION FOR BIPOL...	5	Female
7596	19-24	Manic-Depression	lithobid	Lithium gets a bad rap but for me it has worke...	3	
...	...	...	...	...	...	...
361498	45-54	Depression	celexa		5	Female
361499	25-34	Bipolar Depression	celexa		4	Female
361501	13-18	Depression	celexa	Didn't help what-so-ever for my depression and...	1	Female
361504	35-44	Depression	celexa	I suffer from depression and noticed that when...	5	Female
361505	19-24	Depression	celexa	I have depression, generalized anxiety, and bu...	4	Female

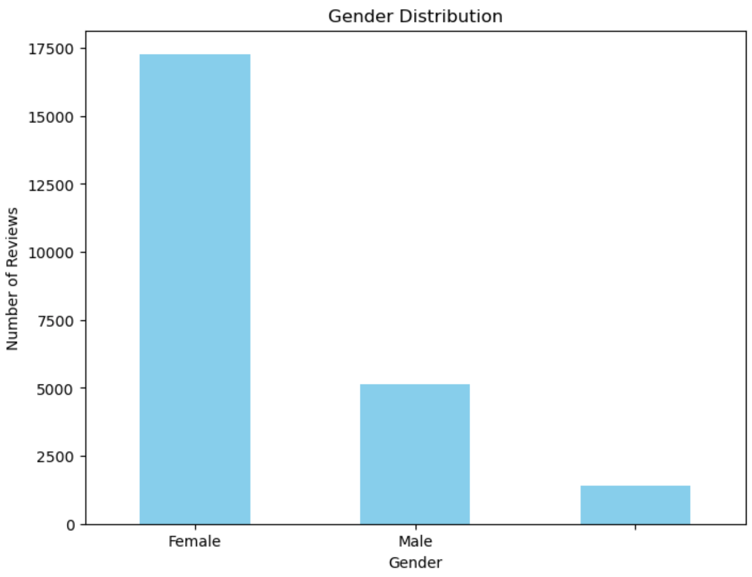
23798 rows × 6 columns

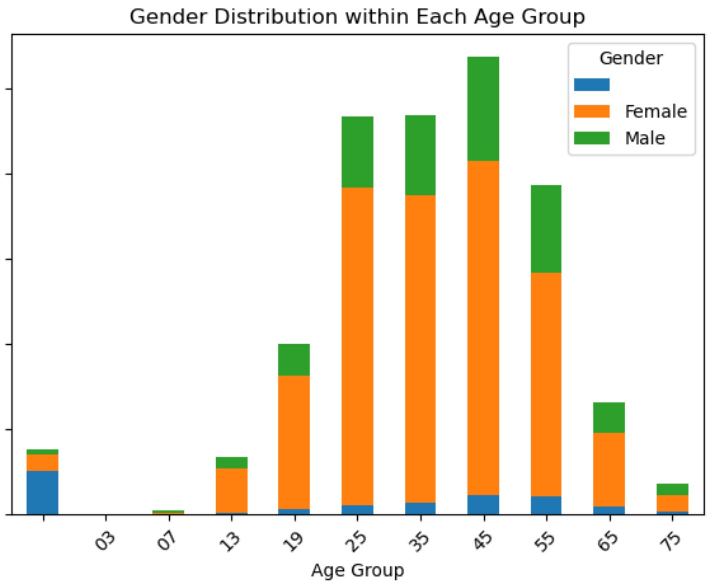
WebMD Dataset

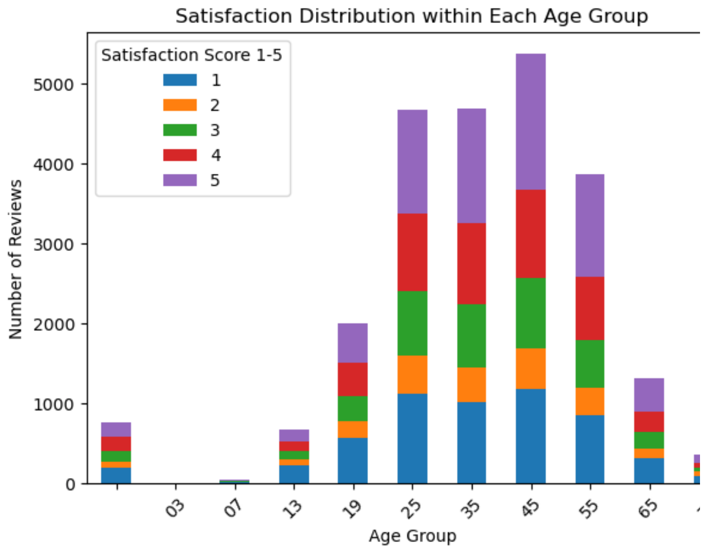
Reviews on Depression Treatment

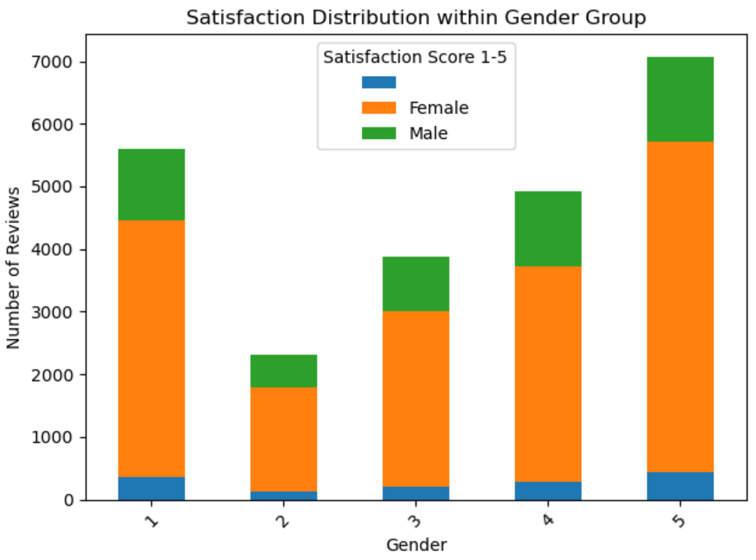
EDA

- Age Distribution
- Gender Disparity
- Satisfaction Score









	aa	aadhd	aand	aback	abated	abates	abdomen	abdominal	abilify	abilities	...	zoned	zonk	zonked	zoo	zprexa	zydis	zydus
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4995	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4996	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4997	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4998	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4999	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

5000 rows × 9084 columns

Topic #0	Topic #1	Topic #2
drug	lithium	taking
pain	life	depression
blood	son	feel
rash	wondering	like
medication	kill	medication
body	thyroid	drug
know	problems	years
heart	worst	effects
depakote	legs	day
therapy	caused	time
times	leg	started
lamictal	im	just
pressure	house	better
dr	extreme	weight
severe	brain	sleep
caused	liver	anxiety
hate	levels	life
year	causes	months

Topic #0	Topic #1	Topic #2
generic	pain	taking
lithium	rash	depression
drug	medication	feel
life	dry	effects
years	mouth	medication
insurance	caused	drug
blood	legs	like
taking	heart	years
brand	body	day
wellbutrin	reaction	just
lexapro	muscle	time
lamictal	chest	started
saved	syndrome	sleep
went	lamictal	weight
depressed	minutes	better
switched	extreme	anxiety
high	severe	life
hair	100	months

CountVectorization & LDA

Synthetic Minority Over-sampling Technique (SMOTE)

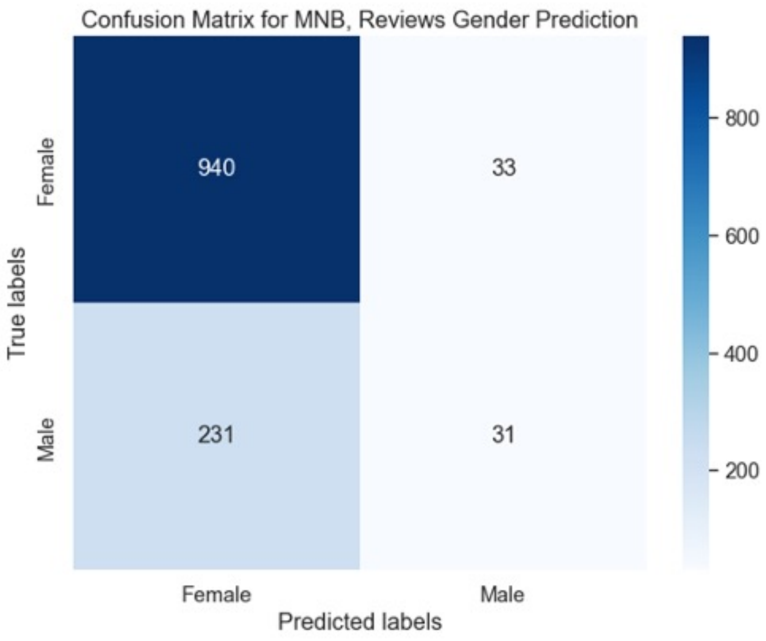
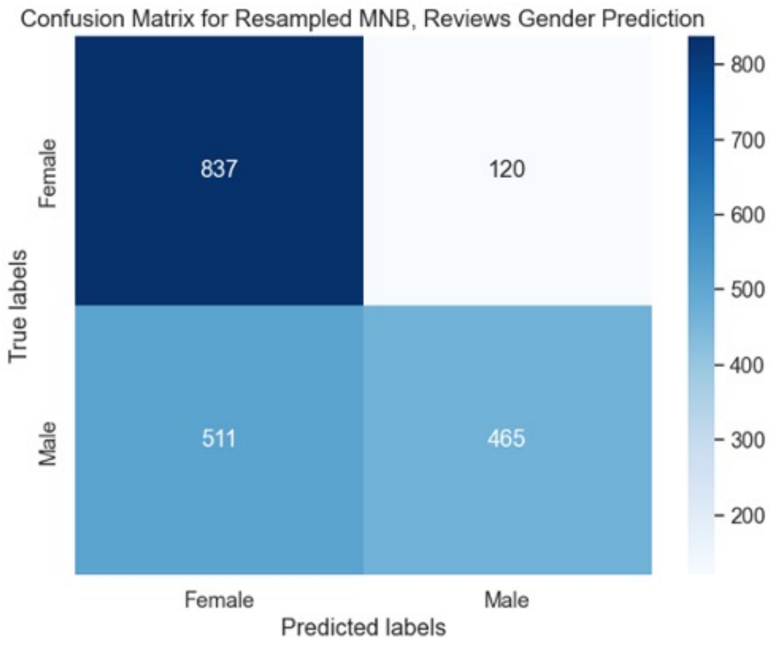
	aa	aand	aback	abated	abates	abdomen	abdominal	abilify	abilities	ability	...	zonk	zonked	zoo	zprexa	zydis	zydus	zyloft	zyprexa	zyprexia	Gender
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Male
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4109	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
4110	0	0	0	0	0	0	0	0	0	1	...	0	0	0	0	0	0	0	0	0	Male
4111	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
4112	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Female
4113	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	Male

4114 rows × 8784 columns

```python
from imblearn.over_sampling import SMOTE
smote = SMOTE()
x = DFreviewgender
y = list(Label1)
data_resampled, label_resampled = smote.fit_resample(x,y)
data_resampled #increased datasize due to oversampling, more synthetic samples for the minority class
```

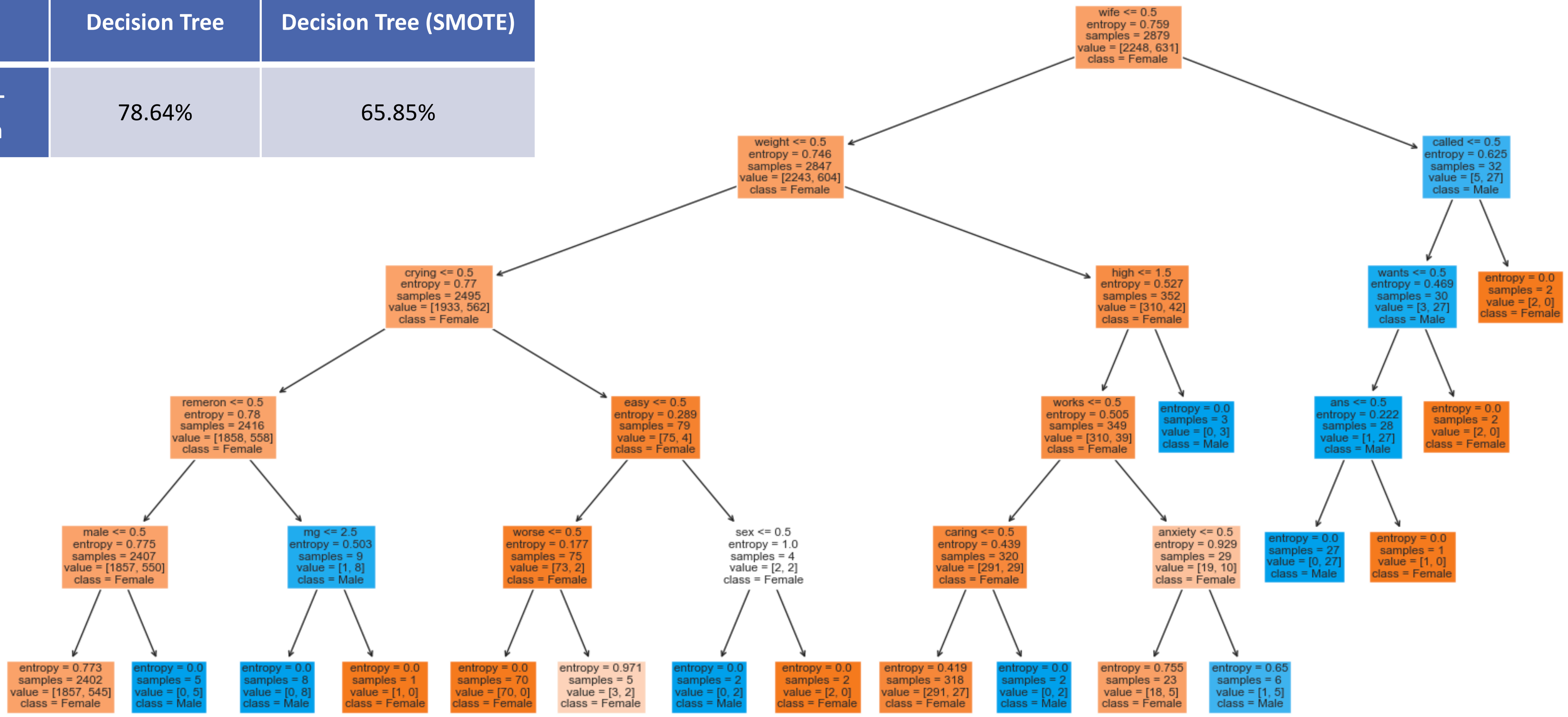
	aa	aand	aback	abated	abates	abdomen	abdominal	abilify	abilities	ability	...	zoned	zonk	zonked	zoo	zprexa
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6437	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
6438	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
6439	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
6440	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
6441	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

6442 rows × 8783 columns

Model	Naïve Bayes	Naïve Bayes (SMOTE)
Confusion Matrix		
10K Cross-validation	77.94%	66.29%

Naïve Bayes Analysis & Result

Model	Decision Tree	Decision Tree (SMOTE)
10K Cross-validation	78.64%	65.85%



Decision Tree Analysis & Result

SVM

Model	SVM C=50, Linear	SVM C=20, Linear	SVM C=70, Linear	SVM C=10, RBF	SVM C=50, RBF	SVM C=80, RBF	SVM (SMOTE) C=80, RBF	SVM C=10, POLY	SVM C=40, POLY	SVM C=80, POLY
10K Cross-validation	69.33 %	69.95 %	69.22 %	78.08 %	78.12 %	78.33 %	74.56%	78.08 %	78.08 %	78.08 %

```
#linear #linear, C=20, p=10
SVM_Model2=SVC(C=20, kernel='linear', degree=10, gamma="auto")
SVM_Model2.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG2 = cross_val_score(SVM_Model2, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG2.mean()))
#linear, C=70, p=10
SVM_Model3=SVC(C=70, kernel='linear', degree=10, gamma="auto")
SVM_Model3.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG3 = cross_val_score(SVM_Model3, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG3.mean()))
```

```
# RBF
SVM_Model4=SVC(C=10, kernel='rbf', verbose=True, gamma="auto")
SVM_Model5=SVC(C=50, kernel='rbf', verbose=True, gamma="auto")
SVM_Model6=SVC(C=80, kernel='rbf', verbose=True, gamma="auto") #BEST

SVM_Model4.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG4 = cross_val_score(SVM_Model4, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG4.mean()))

SVM_Model5.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG5 = cross_val_score(SVM_Model5, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG5.mean()))

SVM_Model6.fit(RGTrain_data, RGTrain_label) #BEST
scores_SVM_RG6 = cross_val_score(SVM_Model6, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG6.mean()))
```

```
# POLY

SVM_Model7=SVC(C=10, kernel='poly',degree=3,gamma="auto", verbose=True)
SVM_Model8=SVC(C=10, kernel='poly',degree=3,gamma="auto", verbose=True)
SVM_Model9=SVC(C=10, kernel='poly',degree=3,gamma="auto", verbose=True)

SVM_Model7.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG7 = cross_val_score(SVM_Model7, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG7.mean()))

SVM_Model8.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG8 = cross_val_score(SVM_Model8, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG8.mean()))

SVM_Model9.fit(RGTrain_data, RGTrain_label)
scores_SVM_RG9 = cross_val_score(SVM_Model9, RGTrain_data, RGTrain_label, cv = 10, scoring='accuracy')
print('Average cross-validation score: {:.4f}'.format(scores_SVM_RG9.mean()))
```

Conclusion

- SVM RBF model, particularly with SMOTE, emerges as the most suitable.
- Understanding gender's role in depression treatment reviews is crucial.
- Gender-specific patterns offer insights for tailored interventions.
- Implications for healthcare quality improvement discussed.