MMR Vaccination Rates

Final Project Report

DATS 6103-10, Group Four

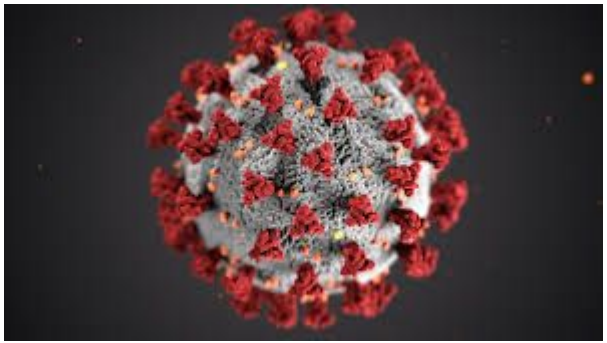
Benjamin Lee

Kristin Levine

Russell Moncrief

History of MMR Vaccine

- MMR Vaccine since 1971 -- prevents measles, mumps, and rubella
- WHO recommends a 95% vaccination rate to prevent disease
- Scientist in France -- using MMR vaccine as backbone of their coronavirus vaccine
  -



Dataset from Wall Street Journal Article

What's the Measles Vaccination Rate at Your Child's School?

Compare the immunization rate at schools across the U.S., according to data collected by The Wall Street Journal

By Dylan Moriarty, Taylor Umlauf, Brianna Abbott

Published Oct.3, 2019 at 12:19 p.m. ET

Summary of features in dataset

Attribute	Description	Optional?
state	School's state	
county	School's county	Y
district	School's district	Y
name	School's name	
type	Whether a school is public, private, charter	Y
enroll	Enrollment*	Y
mmr	School's Measles, Mumps and Rubella (MMR) vaccination rate	Y
overall	School's overall vaccination rate	Y
xmed	Percentage of students exempted from vaccination for medical reasons	Y
xper	Percentage of students exempted from vaccination for personal reasons	Y
xrel	Percentage of students exempted from vaccination for religious reasons	Y

Data Available from Following States:

2017 - 2018 School Year	2018 - 2019 School Year	
Colorado	Arizona	North Carolina
Connecticut	Arkansas	Ohio
Minnesota	California	Oklahoma
Montana	Florida	Oregon
New Jersey	Idaho	Rhode Island
New York	Illinois	Tennessee
North Dakota	Iowa	Texas
Pennsylvania	Maine	Vermont
South Dakota	Massachusetts	Virginia
Utah	Michigan	Wisconsin
Washington	Missouri	

Descriptions of Algorithms Used

- State as Target
  - KNN
- Vaccination Rate as Target
  - Decision Tree
  - Random Forest
  - KNN
- Regression

Data Cleaning for State as Target

- Imputing values using the state-specific mean for numerical features

	state	enroll	mmr	overall	xrel	xmed	xper
0	Arizona	72.66	92.70	NaN	NaN	2.00	7.53
1	Arkansas	509.20	80.52	NaN	NaN	NaN	NaN
2	California	70.51	95.59	94.19	NaN	3.11	NaN
3	Colorado	53.20	94.42	87.97	3.22	2.30	8.02
4	Connecticut	NaN	96.77	NaN	3.20	1.96	NaN
5	Florida	89.51	NaN	92.53	3.69	2.04	NaN
6	Idaho	NaN	NaN	82.44	11.80	5.05	NaN
7	Illinois	350.80	97.07	NaN	1.96	0.59	NaN
8	Iowa	276.39	NaN	95.81	NaN	NaN	NaN
9	Maine	72.61	92.55	NaN	3.60	3.55	9.37
10	Massachusetts	NaN	96.92	NaN	NaN	3.00	NaN
11	Michigan	61.57	NaN	92.12	NaN	6.54	NaN
12	Minnesota	60.19	90.27	NaN	NaN	1.95	6.29
13	Missouri	NaN	91.14	NaN	NaN	NaN	NaN
14	Montana	231.81	93.87	NaN	NaN	1.04	NaN
15	New Jersey	52.43	NaN	95.31	6.00	3.27	NaN
16	New York	NaN	95.94	NaN	7.51	0.97	NaN
17	North Carolina	64.03	NaN	96.79	4.86	1.95	NaN
18	North Dakota	38.00	92.84	NaN	3.39	6.84	6.48
19	Ohio	66.33	90.37	87.85	NaN	4.45	NaN
20	Oklahoma	NaN	NaN	90.89	NaN	4.13	NaN
21	Oregon	56.24	93.91	91.28	NaN	0.47	3.92
22	Pennsylvania	72.09	96.83	NaN	NaN	NaN	NaN
23	Rhode Island	50.80	NaN	94.87	NaN	4.14	NaN
24	South Dakota	32.77	95.03	NaN	NaN	NaN	NaN
25	Tennessee	75.58	NaN	94.95	3.09	1.45	NaN
26	Texas	NaN	93.11	NaN	NaN	NaN	NaN
27	Utah	488.26	95.20	NaN	0.31	0.36	4.62
28	Vermont	157.60	94.58	91.58	NaN	NaN	NaN
29	Virginia	72.22	NaN	93.80	3.13	1.93	NaN
30	Washington	NaN	88.64	83.70	2.67	3.08	7.68
31	Wisconsin	NaN	NaN	85.81	5.08	NaN	7.75

- Imputing again using the mean of all values for numerical features

```
state          0
name           0
type       27174
city       17339
county      5158
enroll         0
mmr            0
overall        0
xrel           0
xmed           0
xper           0
dtype: int64
```

Data Cleaning for MMR Rate as Target

- Got rid of schools with no enrollment data
- Combined MMR and Overall vaccination rates, used MMR if given and Overall if no MMR was available
- Set target rate >= 95; also did >= 90
- Mean vaccination rate for each state, county, and city
- Totaled different types of exemption rates
- Filled NaN values with zero
- Selected features to use: state_mean, city_mean, county_mean, type_of_school, enroll, xtotal
- Target: either at_least_95 or at_least_90

Data Cleaning for Regression

- Impute missing values with means of column
- Drop non-numeric variables
- Subset by state for regression

```
Missing Values:
index          0
state          0
year        4681
name           0
type       27174
city       17339
county      5158
district   39009
enroll     12844
mmr            0
overall        0
xrel       34270
xmed       33439
xper       40000
dtype: int64
```

```
Missing Values:
index          0
state          0
year        4681
name           0
type       27174
city       17339
county      5158
district   39009
enroll         0
mmr            0
overall        0
xrel           0
xmed           0
xper           0
dtype: int64
```

KNN -- State as Target

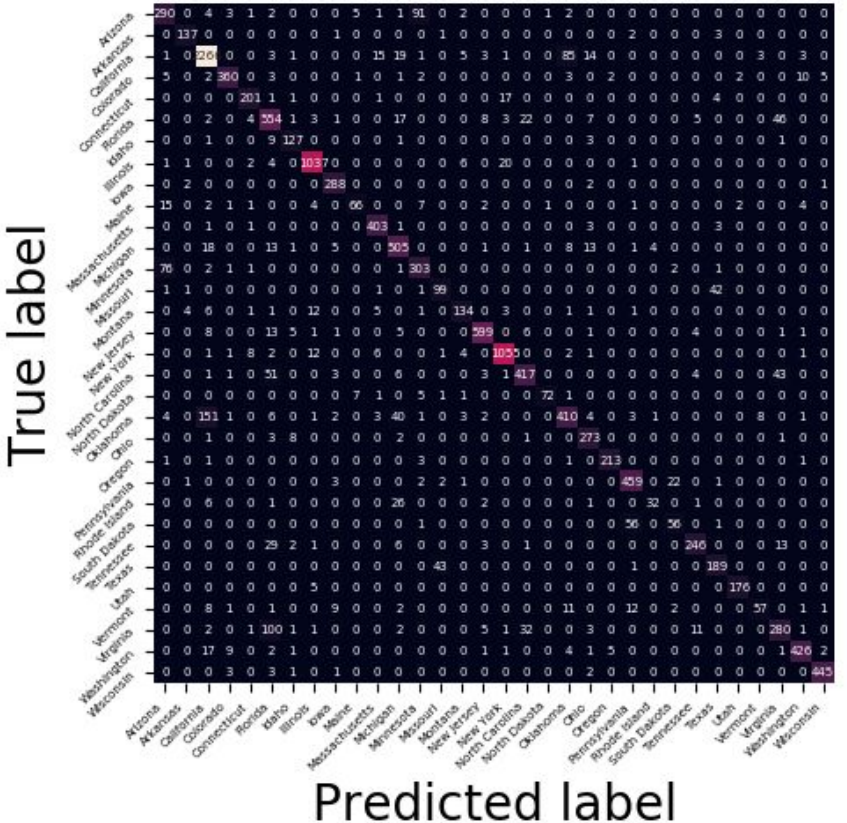
Accuracy: 87.44%

Cross-Validation score and GridSearch revealed best choice for K

```
cv_scores = cross_val_score(model, X, Y, cv=5)
param_grid = {'n_neighbors': np.arange(1, 25)}
knn_gscv = GridSearchCV(model2, param_grid, cv=5)
knn_gscv.fit(X, Y)
knn_gscv.best_params_
```

Classification Report:

	precision	recall	f1-score	support
0	0.74	0.72	0.73	403
1	0.94	0.95	0.94	144
2	0.91	0.94	0.92	2420
3	0.94	0.91	0.93	396
4	0.91	0.89	0.90	225
5	0.69	0.82	0.75	673
6	0.85	0.89	0.87	142
7	0.96	0.97	0.97	1072
8	0.92	0.98	0.95	293
9	0.84	0.62	0.71	106
10	0.92	0.98	0.95	412
11	0.80	0.89	0.84	570
12	0.72	0.78	0.75	387
13	0.67	0.68	0.68	145
14	0.86	0.79	0.82	170
15	0.95	0.93	0.94	645
16	0.96	0.96	0.96	1094
17	0.87	0.79	0.83	530
18	0.97	0.82	0.89	88
19	0.78	0.64	0.70	641
20	0.83	0.94	0.88	289
21	0.97	0.97	0.97	220
22	0.85	0.93	0.89	491
23	0.86	0.46	0.60	69
24	0.68	0.49	0.57	114
25	0.91	0.82	0.86	301
26	0.77	0.81	0.79	233
27	0.98	0.97	0.98	181
28	0.84	0.54	0.66	105
29	0.73	0.64	0.68	440
30	0.95	0.91	0.93	470
31	0.98	0.98	0.98	455



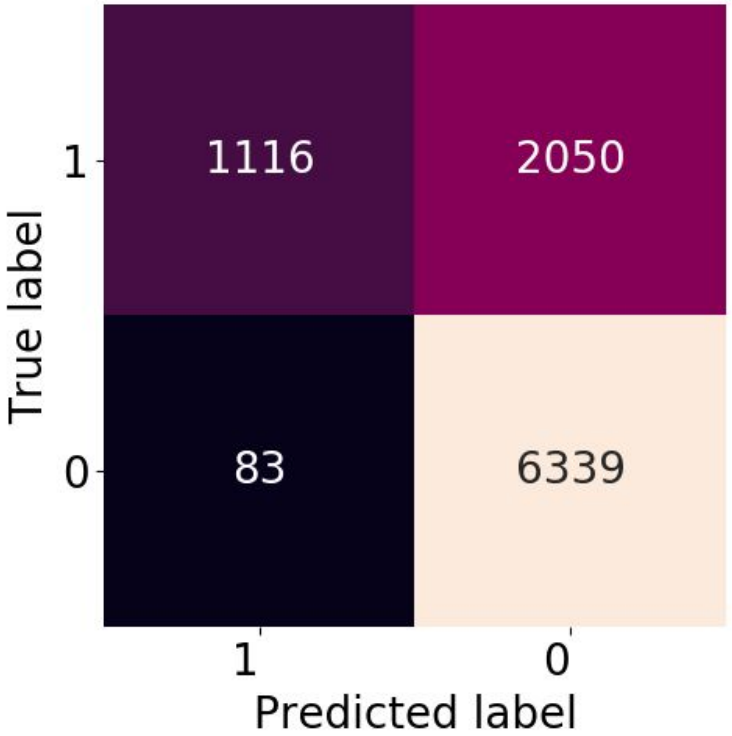
Decision Tree -- 95% MMR vaccination target

Results Using Entropy:

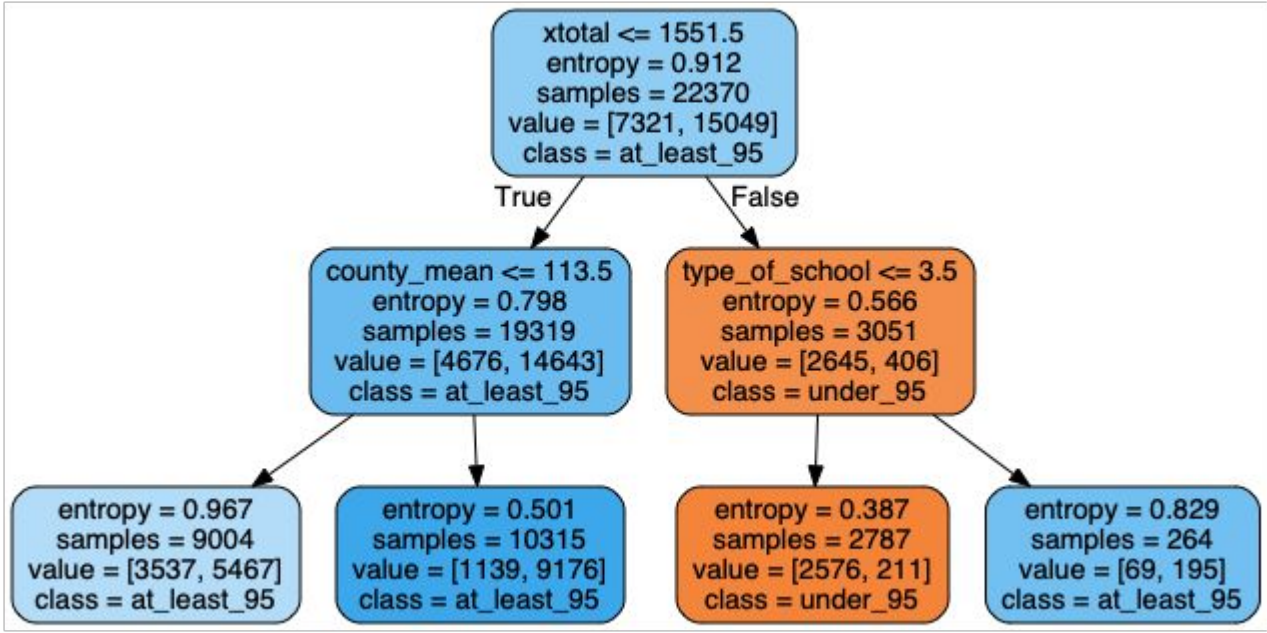
Classification Report:

	precision	recall	f1-score	support
0	0.93	0.35	0.51	3166
1	0.76	0.99	0.86	6422
accuracy			0.78	9588
macro avg		0.84	0.67	0.68	9588
weighted avg		0.81	0.78	0.74	9588

Accuracy : 77.75344180225282



Decision Tree -- 95% MMR vaccination target



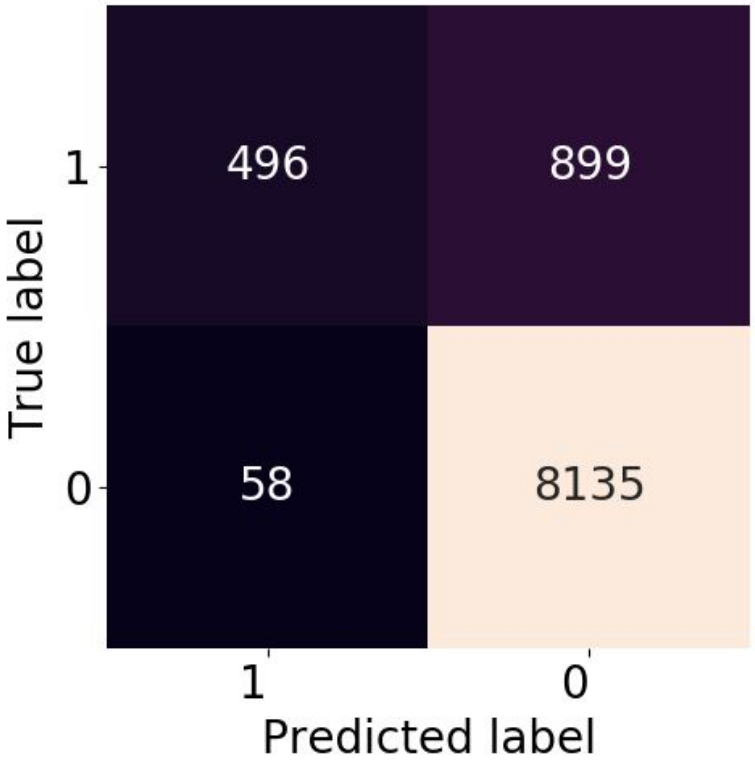
Decision Tree -- 90% MMR vaccination target

Results Using Entropy:

Classification Report:

	precision	recall	f1-score	support
0	0.90	0.36	0.51	1395
1	0.90	0.99	0.94	8193
accuracy			0.90	9588
macro avg		0.90	0.67	0.73	9588
weighted avg		0.90	0.90	0.88	9588

Accuracy : 90.01877346683355



Random Forest -- 95% MMR vaccination target

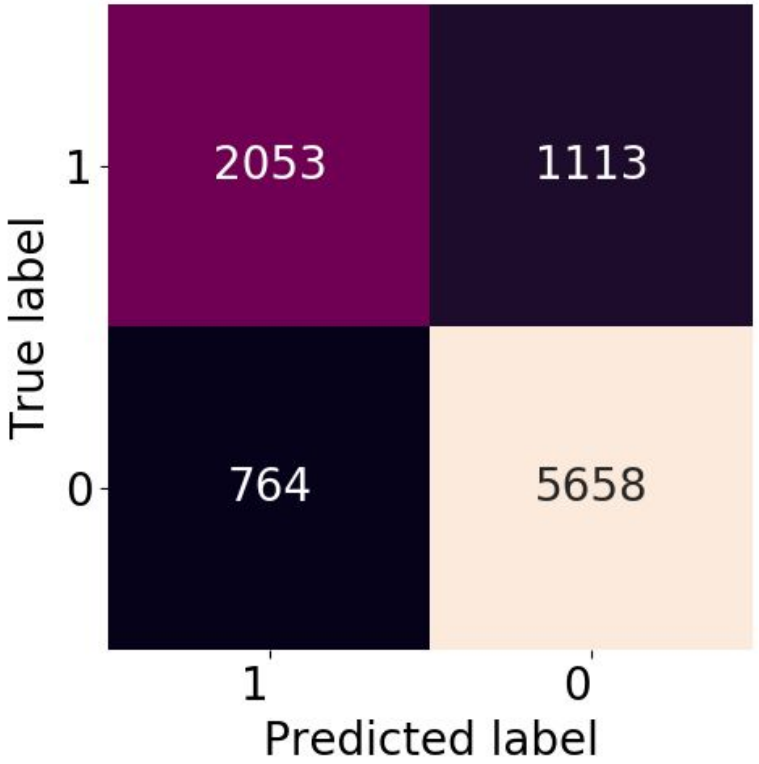
Results Using All Features:

Classification Report:

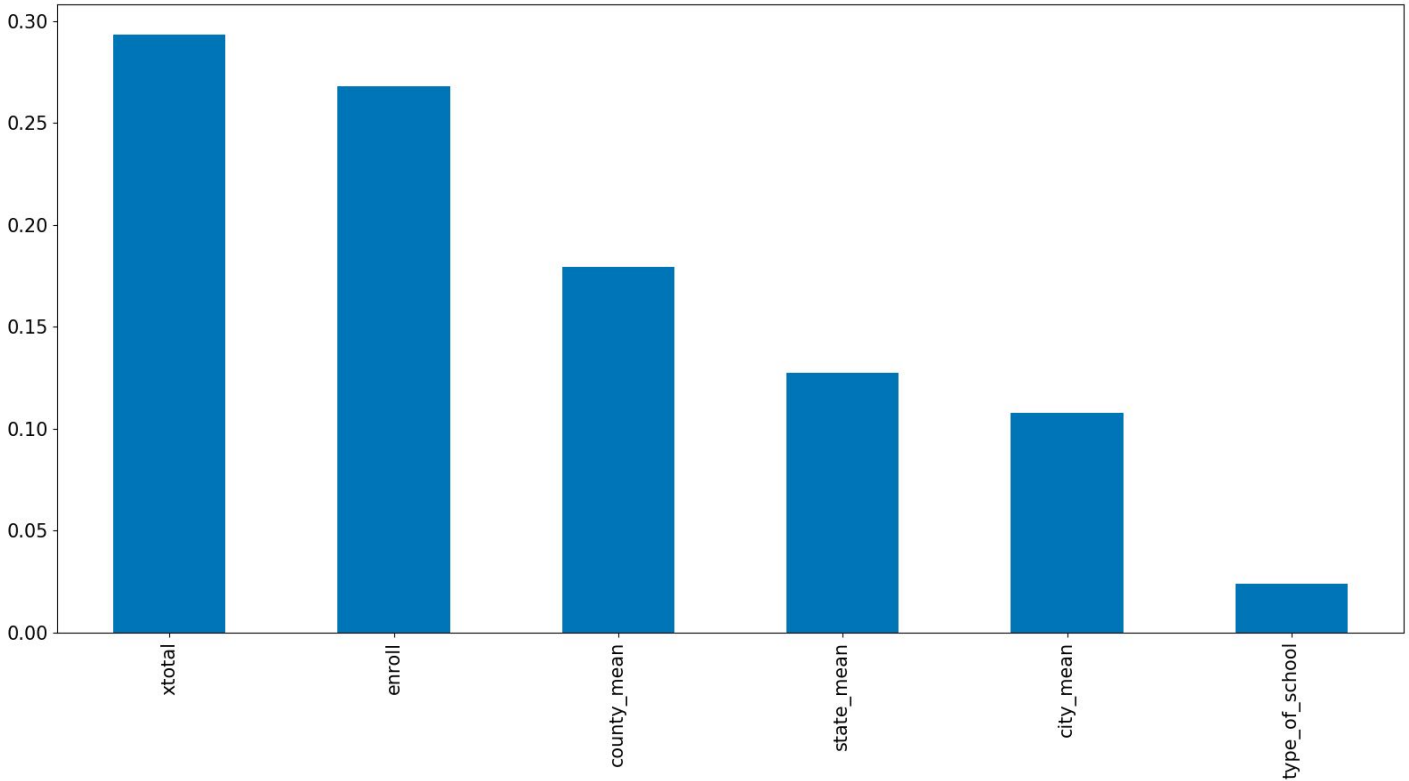
	precision	recall	f1-score	support
0	0.73	0.65	0.69	3166
1	0.84	0.88	0.86	6422
accuracy			0.81	9588
macro avg	0.79	0.77	0.78	9588
weighted avg	0.80	0.81	0.80	9588

Accuracy : 80.70504797663747

ROC_AUC : 86.5299946114637



Random Forest

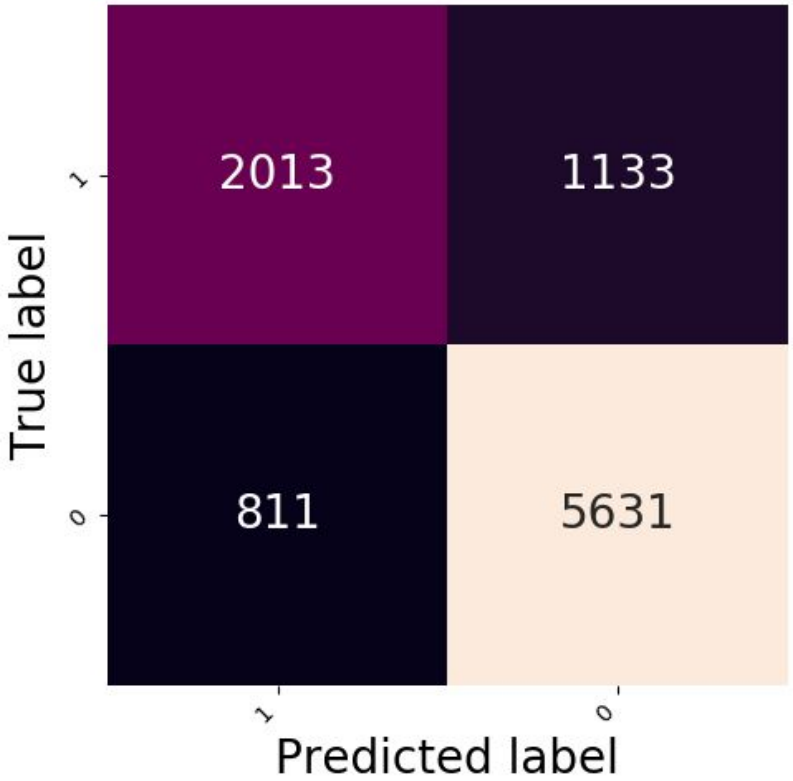


KNN -- 95% MMR vaccination target

Classification Report:

	precision	recall	f1-score	support
0	0.71	0.64	0.67	3146
1	0.83	0.87	0.85	6442
accuracy			0.80	9588
macro avg	0.77	0.76	0.76	9588
weighted avg	0.79	0.80	0.79	9588

Accuracy : 79.72465581977471



Comparing Accuracy of All Three Models

	95% Threshold	**90% Threshold**
Decision Tree	77.8	90.0
Random Forest	80.7	89.4
KNN	79.7	89.4

Linear Regression Results

Model

$$MMR = Enroll + Overall + Xrel + Xmed + Xper$$

Vermont

- 96.6 % model score

	Coefficient
enroll	-7.432885e-04
overall	9.415009e-01
xrel	-9.819041e-33
xmed	0.000000e+00
xper	2.182009e-32

Mean Absolute Error: 2.2785468001960116
Mean Squared Error: 9.546673263958125
Root Mean Squared Error: 3.089769127937899

	Actual	Predicted
41518	100.000000	101.140927
41564	100.000000	102.379223
41782	92.372881	91.932427
41557	100.000000	102.382940
41807	90.361446	92.121864
...	...	...
41618	98.333333	99.120474
41670	97.435897	99.804568
41632	98.148148	98.865424
41731	95.683453	98.225253
41575	99.288256	100.508470

[70 rows x 2 columns]
The score of the model: 0.966

Oregon

- 93% model score

	Coefficient
enroll	1.068379e-04
overall	9.834023e-01
xrel	5.551115e-17
xmed	-1.466286e-01
xper	1.406847e-01

Mean Absolute Error: 1.6435932013181025
Mean Squared Error: 10.018183429125447
Root Mean Squared Error: 3.1651514069828393

	Actual	Predicted
36347	97.116844	96.141501
36338	97.202797	97.310538
36165	99.337748	94.200416
36182	98.920863	99.016867
36541	95.238095	96.539395
...	...	...
36718	92.950392	93.740052
36840	80.530973	78.571371
36417	96.509240	94.228215
36655	93.989071	95.057575
36390	96.782842	98.233714

[147 rows x 2 columns]
The score of the model: 0.93

Montana

- 2.7% model score

	Coefficient
enroll	-0.002925
overall	0.000000
xrel	0.000000
xmed	-10.282346
xper	0.000000

Mean Absolute Error: 32.40150632638283
Mean Squared Error: 1477.8715426690649
Root Mean Squared Error: 38.44309486330497

	Actual	Predicted
25140	0.00	68.356510
25085	0.00	68.356510
25067	60.14	67.872564
24926	95.00	68.650577
24665	100.00	68.638877
...	...	...
24841	97.33	68.050980
25147	0.00	68.356510
25060	68.97	29.611823
25144	0.00	68.356510
25128	0.00	68.356510

[114 rows x 2 columns]
The score of the model: 0.027

Summary and Conclusion

- Number of exemptions most important feature
- State laws (whether or not they allow religious and personal exemptions)
- Look at schools with LOWER vaccination rates
- Track down data from other states
- Linear model suitable for prediction given complete data