CSE400 Assignment 3 report

Huahao Shang
hushang@syr.edu
SUID: hushang

1.
(a) DEM party votes fraction for Autauga County is 0.27

(b) 3018 rows in total

(c)

]:

	fips	county_x	state_x	state_code_x	male	female	median_age	population	female_percentage	lat_x	...	StaffPC	FoodInsc	state	county	candidat
0	1001	Autauga County	Alabama	AL	10.20	10.25	37.80	10.92	51.32	32.53	...	6.94	13.40	Alabama	Autauga County	Joe Bide
1	1003	Baldwin County	Alabama	AL	11.52	11.58	42.80	12.25	51.38	30.73	...	6.59	12.30	Alabama	Baldwin County	Joe Bide
2	1005	Barbour County	Alabama	AL	9.52	9.40	39.90	10.16	46.87	31.87	...	6.79	23.20	Alabama	Barbour County	Joe Bide
3	1007	Bibb County	Alabama	AL	9.41	9.25	39.90	10.02	46.06	33.00	...	6.83	15.80	Alabama	Bibb County	Joe Bide
4	1009	Blount County	Alabama	AL	10.26	10.28	40.80	10.96	50.67	33.98	...	7.75	11.00	Alabama	Blount County	Joe Bide

5 rows x 38 columns

dInsc	state	county	candidate	party	total_votes	won	votes	DEM_fraction
13.40	Alabama	Autauga County	Joe Biden	DEM	7503	False	27770	0.27
12.30	Alabama	Baldwin County	Joe Biden	DEM	24578	False	109679	0.22
23.20	Alabama	Barbour County	Joe Biden	DEM	4816	False	10518	0.46
15.80	Alabama	Bibb County	Joe Biden	DEM	1986	False	9595	0.21
11.00	Alabama	Blount County	Joe Biden	DEM	2640	False	27588	0.10

(d)

```
dem_draction mean:  0.3226172432069653
dem_draction std:  0.15146763111955566
```

(e)
Training length: 2359
Testing length: 590

2
(a)

```
Mean absolute error is
0.23230412410529497
Fraction MAE is
0.029922886992137243
```

(b)

```
Mean absolute error is
0.24090281080820003
Fraction MAE is
0.031030476155621944
```

(c)

```
Mean absolute error is
0.5243525234124582
Fraction MAE is
0.0675413807763534
```

(d)

```
Mean absolute error is
0.21113042927317346
Fraction MAE is
0.027195522249441 61
```

(e)

```
Mean absolute error is
0.19102519690402223
Fraction MAE is
0.02460578520344717
```

3.
(a)

:

	cases	predict LR log_case	predict Lasso log_case	predict Bagging log_case	predict RF log_case	predict Gard Boos log_case
0	8.84	8.62	8.58	8.27	8.61	8.70
1	7.07	6.30	6.41	6.88	6.26	6.27
2	7.97	7.81	7.65	7.65	7.73	7.81
3	6.71	6.70	6.83	7.00	6.81	6.82
4	7.98	7.46	7.48	7.43	7.61	7.68

(b)

```
LR Mean absolute error is
3302.862809748904
LR ratio MAE is
456.3681064653794
Lasso Mean absolute error is
3200.329211059047
Lasso ratio MAE is
442.2006805144576
Bagging Mean absolute error is
2224.4940224014936
Bagging ratio MAE is
307.36611943143447
RF Mean absolute error is
3312.335424841095
RF ratio MAE is
457.6769708239409
GB Mean absolute error is
3362.73916571482
GB ratio MAE is
464.6414319918155
```

(c)

```
for number of cases
LR sAMPLE 0.9956366946744619
Lasso sAMPLE 0.9954975305357755
Bagging sAMPLE 0.9935351679632073
RF sAMPLE 0.9956491185616423
GB sAMPLE 0.9957140539547416

for number of log cases
LR sAMPLE 0.014578333457077552
Lasso sAMPLE 0.016216942731567312
Bagging sAMPLE 0.034120573607639974
RF sAMPLE 0.014630249389655183
GB sAMPLE 0.013037112223976032
```

(d)
Bagging has the best predictions by looking at the ratio of prediction. But Linear regression, Lasso and Bagging score about 94, Random forest scores about 98, gradient boosting scores about 97. Bagging selects all features to make the prediction, this may cause the prediction error ratio be the least among those five algorithms.


```
for number of cases
LR sAMPLE 0.9956366946744619
Lasso sAMPLE 0.9954975305357755
Bagging sAMPLE 0.9935351679632073
RF sAMPLE 0.9956491185616423
GB sAMPLE 0.9957140539547416

for number of log cases
LR sAMPLE 0.014578333457077552
Lasso sAMPLE 0.016216942731567312
Bagging sAMPLE 0.034120573607639974
RF sAMPLE 0.014630249389655183
GB sAMPLE 0.013037112223976032
```

```
LR Mean absolute error is
3302.862809748904
LR ratio MAE is
456.3681064653794
Lasso Mean absolute error is
3200.329211059047
Lasso ratio MAE is
442.2006805144576
Bagging Mean absolute error is
2224.4940224014936
Bagging ratio MAE is
307.36611943143447
RF Mean absolute error is
3312.335424841095
RF ratio MAE is
457.6769708239409
GB Mean absolute error is
3362.73916571482
GB ratio MAE is
464.6414319918155
```