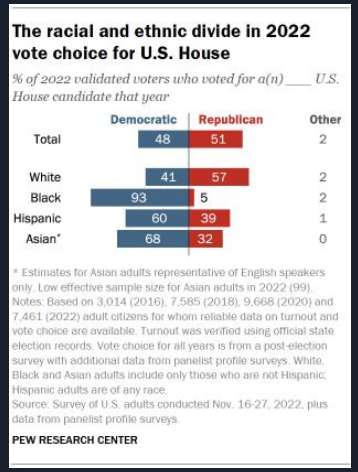
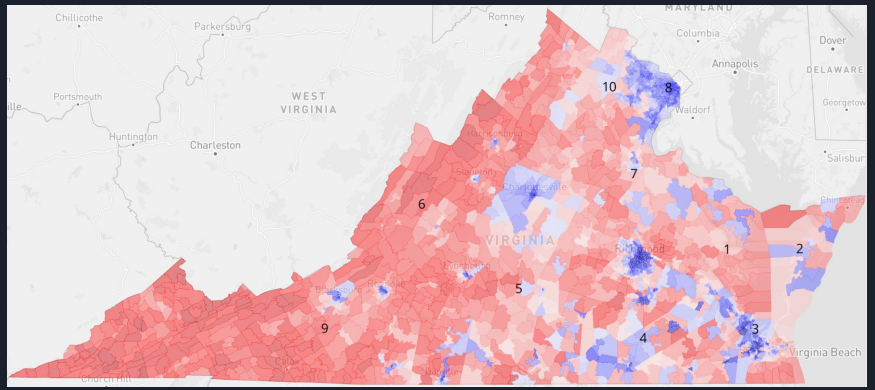
Predicting Electoral Results Using Census and RDH Data

Deven Hagen and Vibby Janardhan

Project Goal

- Electoral precincts - individual polling stations
- Predicting electoral results is a huge market
  - Polymarket, PredictIt
- Using fundamental data rather than polling





Dataset

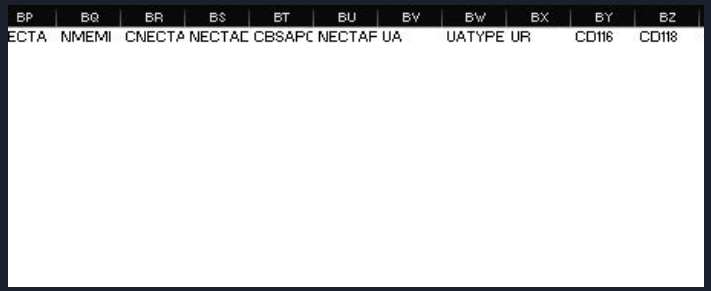
- Source (https://redistrictingdatahub.org/dataset/virginia-block-county-and-vtd-pl-94-171-2020-official)
- Very Large
- Class Transformation
- 463 Attributes and 2465 Instances

STATEFP20	COUNTYFP20	VTDST20	GEOID20	VTDI20	NAME20	NAMELSAD20	LSAD20	MTFCC20	FUNCSTAT20	ALAND20	AWATER20	INTPTLAT20	INTPTLON20	FILEID	STUSAB	SUMLEV	GEOVAR
51	149	105	51149000105	A	Harrison	Harrison Voting Distri	V2	G5240	N	116568311	529708	37.1616583	-77.1821536	PLST	VA	700	0
51	149	202	51149000202	A	Blackwater	Blackwater Voting Di	V2	G5240	N	78672067	7529647	37.2398647	-77.1958255	PLST	VA	700	0
51	149	205	51149000205	A	Jefferson Park	Jefferson Park Voting	V2	G5240	N	30903103	1085514	37.263106	-77.3372307	PLST	VA	700	0
51	149	201	51149000201	A	Bland	Bland Voting District	V2	G5240	N	37721412	4337040	37.2545759	-77.2430783	PLST	VA	700	0
51	149	104	51149000104	A	Rives	Rives Voting District	V2	G5240	N	35489092	50496	37.2080484	-77.3084528	PLST	VA	700	0
51	149	204	51149000204	A	Courts Bldg.	Courts Bldg. Voting D	V2	G5240	N	22526572	7005262	37.2728157	-77.2692636	PLST	VA	700	0
51	149	203	51149000203	A	Brandon	Brandon Voting Distr	V2	G5240	N	135987153	20361457	37.2540131	-77.0730038	PLST	VA	700	0
51	149	102	51149000102	A	Templeton	Templeton Voting Di	V2	G5240	N	163930478	941327	37.077782	-77.3048884	PLST	VA	700	0
51	149	101	51149000101	A	Richard Bland	Richard Bland Votin	V2	G5240	N	26547216	0	37.1421074	-77.3749943	PLST	VA	700	0
51	149	103	51149000103	A	Union Branch	Union Branch Voting	V2	G5240	N	38886837	904747	37.1582155	-77.3103155	PLST	VA	700	0
51	595	401	51595000401	A	Precinct 4-1	Precinct 4-1	0	G5240	N	4160228	78930	36.708765	-77.5500243	PLST	VA	700	0
51	595	201	51595000201	A	Precinct 2-1	Precinct 2-1	0	G5240	N	997929	0	36.6992806	-77.5387812	PLST	VA	700	0
51	595	501	51595000501	A	Precinct 5-1	Precinct 5-1	0	G5240	N	3765223	9104	36.7036103	-77.5213484	PLST	VA	700	0
51	595	301	51595000301	A	Precinct 3-1	Precinct 3-1	0	G5240	N	1651252	0	36.689912	-77.5469783	PLST	VA	700	0
51	595	701	51595000701	A	District 7	District 7	0	G5240	N	864079	0	36.688873	-77.5342971	PLST	VA	700	0

Data Preprocessing

- Attribute Deletion
  - Derived Attributes
- Instance Deletion
- Normalization
  - Election results
  - Racial demographics
  - Population
  - Housing
  - Area
- 32 Attributes and 2411 Instances

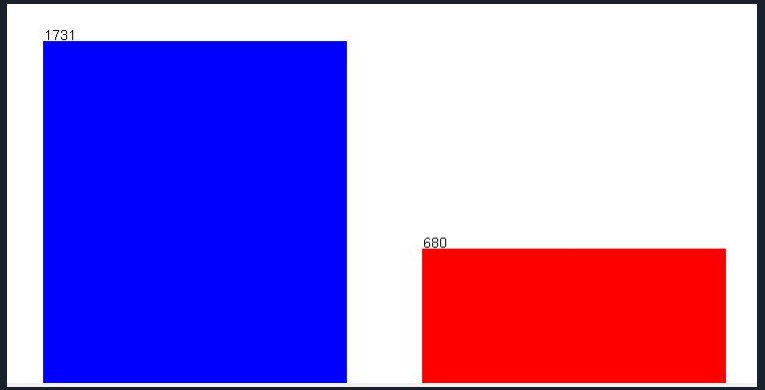
I	J
MTFCC20	FUNCSTAT
G5240	N
G5240	N
G5240	N
G5240	N



BP	BQ	BR	BS	BT	BU	BV	BW	BX	BY	BZ
ECTA	NMEMI	CNECTA	NECTAC	CBSAPC	NECTAF	UA	UATYPE	UR	CD116	CD118

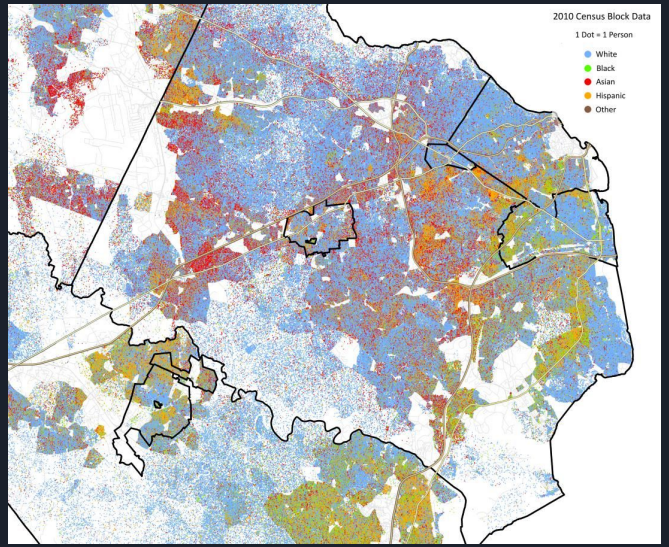
Class Distribution

- Created class labels based on number of voters for each party in the 2020 election
- Removed tied precincts
- Distribution: 71.8% Republican, 28.2% Democratic



Attribute Selection

- 4 Attribute Selection Techniques
  - OneRAttributeEval
  - CorrelationAttributeEval
  - InfoGainAttributeEval
  - ReliefFAttributeEval
- Our chosen attributes
- Similarities
  - Electoral Data
- Differences
  - Geographic and Race Data



OneR

```
Ranked attributes:
94.8569   27 PRES16DPERC
94.7325   23 USSEN18DPERC
93.3637   26 LTGOV17DPERC
93.1564   24 AG18DPERC
92.7001   25 GOV17DPERC
91.7876   31 PRES12DPERC
91.6217   29 GOV13DPERC
91.4144   28 AG13DPERC
89.5064   30 LTGOV13DPERC
85.9394    5 VANWHTALN
84.6537   14 TNWHALN
79.4276   15 TNBLKALN
77.6856    6 VANBLKALN
77.022     1 AREALAND
76.1095    4 VAHISPANIC
75.4873   13 TAHISPANIC
75.3214    2 AREAWATR
73.8698   17 TNASIANALN
72.5425    8 VANASANALN
71.2982   12 TAPERSONS
71.2153   19 TNOTHRALN
70.7175   22 TAHOCCUPID
69.8465    3 VAPERSONS
68.8096    9 VANNHPOALN
68.6437    7 VANAIANALN
68.3119   18 TNNHPOALN
67.8142   16 TNAIANALN
67.7727   10 VANORALN
67.4824   11 VANM2RACES
67.3165   21 TAHOUSING
66.1551   20 TN2MRACES
```

Correlation

```
Ranked attributes:
 0.7888   27 PRES16DPERC
 0.7865   26 LTGOV17DPERC
 0.7824   24 AG18DPERC
 0.7807   31 PRES12DPERC
 0.7802   28 AG13DPERC
 0.7797   29 GOV13DPERC
 0.7765   25 GOV17DPERC
 0.7758   23 USSEN18DPERC
 0.7473   30 LTGOV13DPERC
 0.6876    5 VANWHTALN
 0.6648   14 TNWHALN
 0.5316   15 TNBLKALN
 0.4888    6 VANBLKALN
 0.4346    4 VAHISPANIC
 0.4191   13 TAHISPANIC
 0.3609    1 AREALAND
 0.3413   17 TNASIANALN
 0.3145   19 TNOTHRALN
 0.3082    8 VANASANALN
 0.3075    3 VAPERSONS
 0.3009   12 TAPERSONS
 0.2394   20 TN2MRACES
 0.2107   10 VANORALN
 0.1813   11 VANM2RACES
 0.1606   18 TNNHPOALN
 0.1324    9 VANNHPOALN
 0.0742   21 TAHOUSING
 0.0579    2 AREAWATR
 0.0387    7 VANAIANALN
 0.0287   22 TAHOCCUPID
 0.0174   16 TNAIANALN
```

InfoGain

0.6916	23	USSEN18DPERC
0.6796	27	PRES16DPERC
0.6722	25	GOV17DPERC
0.6669	24	AG18DPERC
0.6639	26	LTGOV17DPERC
0.6253	31	PRES12DPERC
0.6228	28	AG13DPERC
0.6164	29	GOV13DPERC
0.5746	30	LTGOV13DPERC
0.4144	5	VANWHTALN
0.3728	14	TNWHALN
0.2569	1	AREALAND
0.2382	15	TNBLKALN
0.2034	6	VANBLKALN
0.1572	4	VAHISPANIC
0.1487	2	AREAWATR
0.1407	13	TAHISPANIC
0.1061	17	TNASIANALN
0.101	8	VANASANALN
0.0831	19	TNOTHRALN
0.0673	3	VAPERSONS
0.0639	12	TAPERSONS
0.0427	10	VANORALN
0.0384	20	TN2MRACES
0.038	22	TAHOCCUPID
0.0259	18	TNNHPOALN
0.0237	9	VANNHPOALN
0.0226	11	VANM2RACES
0	21	TAHOUSING
0	7	VANAIANALN
0	16	TNAIANALN

ReliefF

0.206555	23	USSEN18DPERC
0.183269	27	PRES16DPERC
0.180663	26	LTGOV17DPERC
0.178971	25	GOV17DPERC
0.175165	24	AG18DPERC
0.123077	28	AG13DPERC
0.113753	29	GOV13DPERC
0.110577	30	LTGOV13DPERC
0.10999	31	PRES12DPERC
0.082153	5	VANWHTALN
0.075585	14	TNWHALN
0.061844	15	TNBLKALN
0.056405	6	VANBLKALN
0.04872	1	AREALAND
0.041586	20	TN2MRACES
0.038933	17	TNASIANALN
0.037806	22	TAHOCCUPID
0.036286	8	VANASANALN
0.027004	11	VANM2RACES
0.02451	13	TAHISPANIC
0.024192	4	VAHISPANIC
0.020693	12	TAPERSONS
0.01809	3	VAPERSONS
0.014764	10	VANORALN
0.014508	9	VANNHPOALN
0.012837	21	TAHOUSING
0.009908	18	TNNHPOALN
0.009466	19	TNOTHRALN
0.000832	16	TNAIANALN
0.00083	7	VANAIANALN
-0.000177	2	AREAWATR

Our Chosen Attributes

- Combination of each selection method
  - USSEN18DPERC, AG18DPERC, GOV17DPERC, LTGOV17DPERC, PRES16DPERC, AG13DPERC, GOV13DPERC, LTGOV13DPERC, PRES12DPERC, VANBLKALN, AREALAND

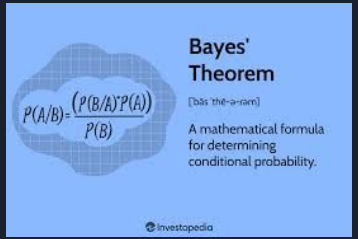
AREALAND	VANWHTALN	VANBLKALN	USSEN18DPERC	AG18DPERC	GOV17DPERC	LTGOV17DPERC	PRES16DPERC	AG13DPERC	GOV13DPERC	LTGOV13DPERC	PRES12DPERC	WIN_PARTY
0.116568	0.760894	0.139665	0.239468	0.219072	0.21883	0.205656	0.237258	0.316568	0.248521	0.393491	0.299484	R
0.078672	0.707191	0.202622	0.342271	0.333975	0.329828	0.317919	0.293536	0.330208	0.300306	0.420063	0.339479	R
0.037721	0.668154	0.210554	0.390798	0.351495	0.359974	0.350365	0.341407	0.350951	0.309257	0.423186	0.352407	R
0.035489	0.419699	0.298236	0.578216	0.573723	0.578871	0.574344	0.538462	0.573333	0.539474	0.623529	0.551509	R
0.022527	0.48436	0.381043	0.549313	0.514329	0.522846	0.515813	0.501065	0.510061	0.484321	0.570423	0.528642	R
0.135987	0.768549	0.169435	0.315488	0.314031	0.323661	0.329621	0.299509	0.335784	0.305825	0.424691	0.326367	R
0.16393	0.751823	0.164731	0.262866	0.245961	0.255344	0.244165	0.238056	0.31178	0.287724	0.37669	0.301431	R
0.026547	0.705838	0.203593	0.398542	0.34104	0.333333	0.325145	0.317497	0.367107	0.340708	0.474852	0.354031	R
0.038887	0.626954	0.282636	0.44164	0.430037	0.42844	0.428304	0.405617	0.474006	0.457372	0.541117	0.479739	R
0.00416	0.41226	0.496394	0.535032	0.43662	0.437931	0.43554	0.531317	0.506452	0.52322	0.602649	0.530612	R
0.001651	0.402821	0.492163	0.45679	0.440678	0.413502	0.405983	0.445714	0.395161	0.416667	0.585062	0.470297	R
0.001239	0.530928	0.360825	0.450355	0.381107	0.362745	0.388889	0.409091	0.397924	0.405405	0.566434	0.42446	R
0.002935	0.581509	0.043777	0.656716	0.596072	0.598131	0.605646	0.563649	0.555246	0.529086	0.600561	0.532468	R
0.00228	0.563664	0.032687	0.700523	0.668194	0.671208	0.66819	0.615273	0.601385	0.558824	0.626131	0.57537	R
0.002018	0.638917	0.028426	0.682975	0.651079	0.646515	0.650179	0.610406	0.602904	0.569257	0.63503	0.571257	R
0.004435	0.559259	0.044907	0.680664	0.633929	0.641905	0.636999	0.598681	0.587189	0.568614	0.632184	0.565117	R
0.082556	0.908634	0.015088	0.297041	0.298438	0.288596	0.277821	0.229406	0.303872	0.297678	0.352743	0.325872	R

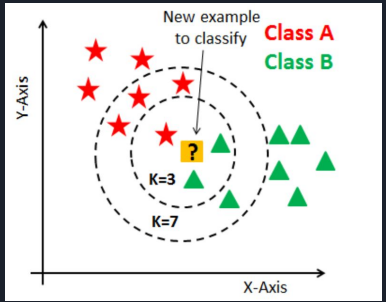
Classifiers

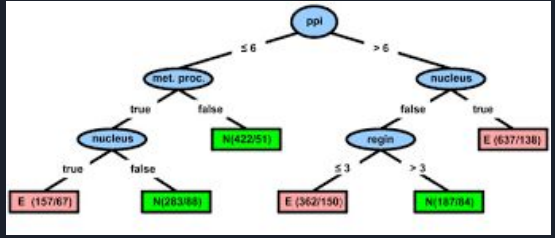
- J48 decision tree
- Naïve Bayes
- OneR Classifier
- K-nearest neighbor

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

IF Outlook = Sunny THEN PlayGolf = Yes
IF Outlook = Overcast THEN PlayGolf = Yes
IF Outlook = Rainy THEN PlayGolf = No







Accuracy Table

	OneR Attribute Eval	Correlation	InfoGain	ReliefF	Our Chosen Attributes
J48	0.954791	0.944836	0.951058	0.942762	0.944836
OneR Classifier	0.948569	0.948569	0.948569	0.948569	0.948569
K-nearest neighbor	0.941518	0.941103	0.933637	0.941103	0.9382
Naïve Bayes	0.92866	0.929905	0.92949	0.934467	0.92949

Results

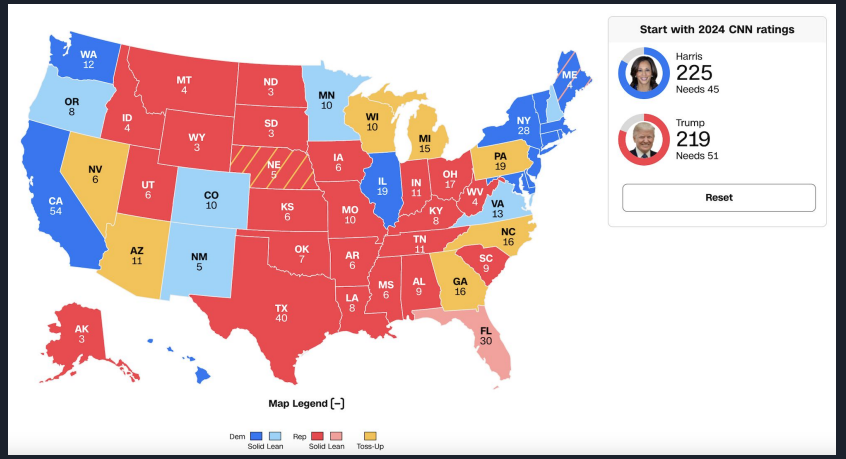
- K-fold value = 10
- OneR Classifier had identical results
- Accuracy better metric than False Positives and Negatives
- Consistent 92-95% values

Best Model: OneRAttribute + J48 Decision Tree Classifier

```
Correctly Classified Instances        2302               95.4791 %
Incorrectly Classified Instances       109                4.5209 %
Kappa statistic                          0.8875
Mean absolute error                      0.0589
Root mean squared error                  0.1951
Relative absolute error                 14.5386 %
Root relative squared error             43.3486 %
Total Number of Instances             2411

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.973    0.093    0.964      0.973    0.969      0.888    0.961     0.966     R
                 0.907    0.027    0.931      0.907    0.919      0.888    0.961     0.905     D
Weighted Avg.    0.955    0.074    0.955      0.955    0.955      0.888    0.961     0.949

=== Confusion Matrix ===

    a    b   <-- classified as
 1685   46 |    a = R
   63  617 |    b = D
```

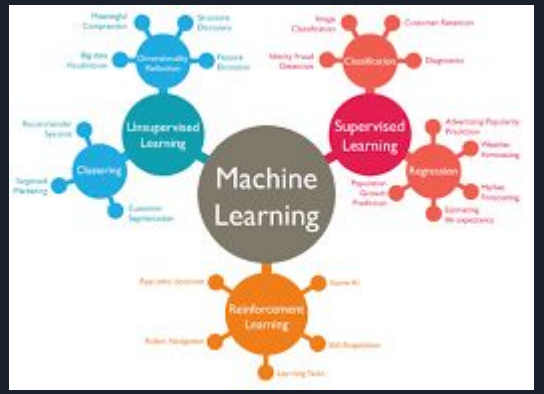
Conclusion

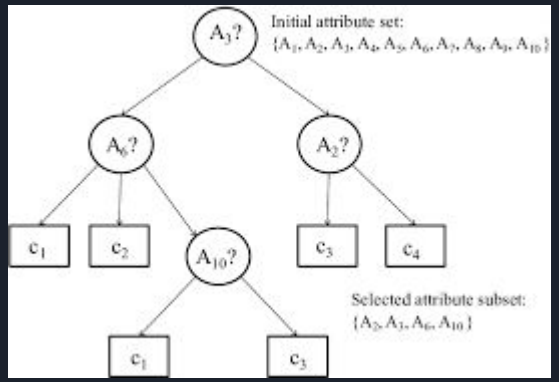
- Previous elections, racial data are good predictors
- No difference between Voting Age vs Total Population
- Increasing geographic polarization



Future Research

- More Attribute Selection Algorithms/Classifiers
- Different K-Fold values
- Expand to other states
  - Use counties instead of precincts?





Thank you!

Any Questions?