2020 Covid-19 Pandemic

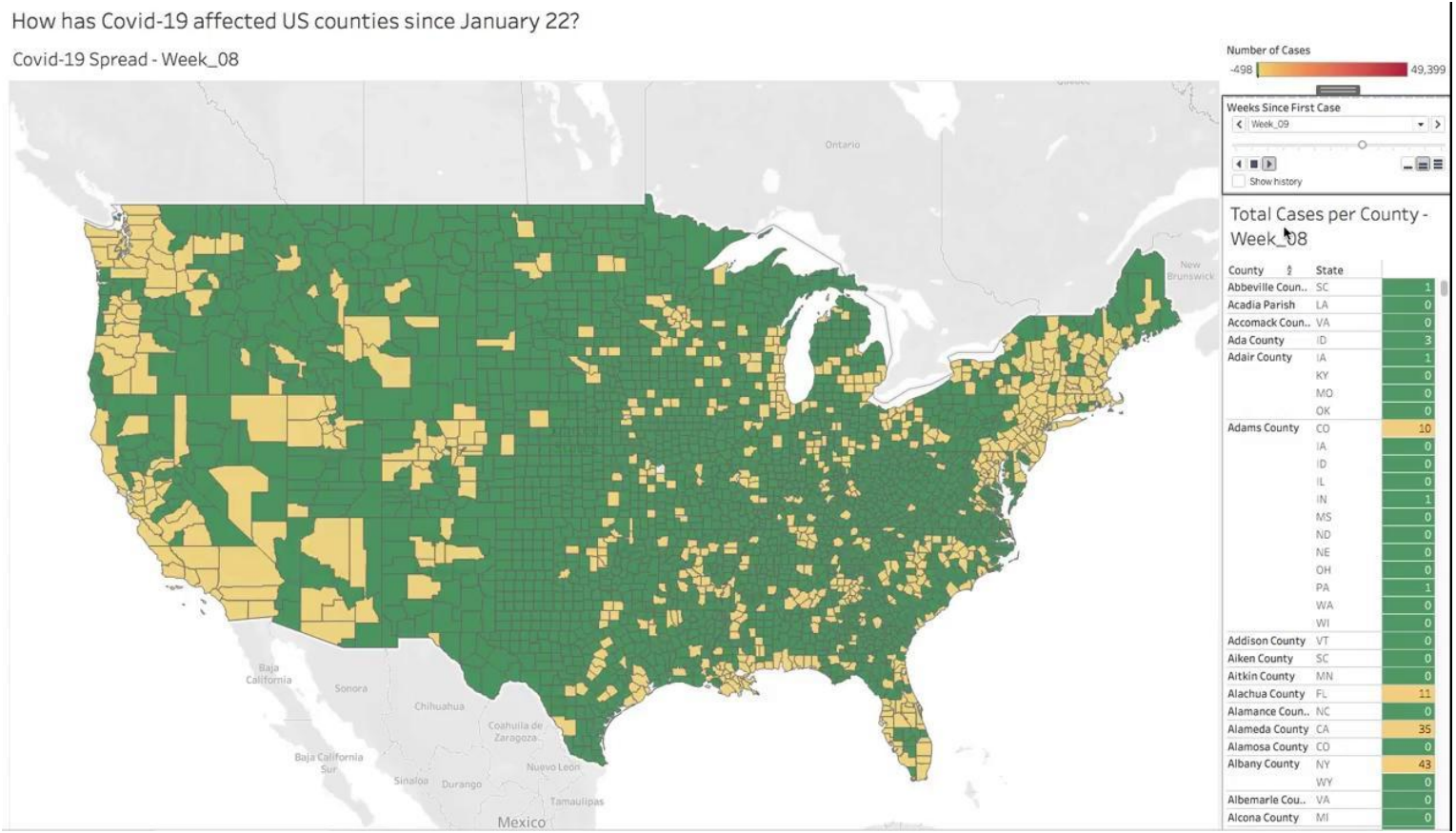
Which U.S. counties are at risk of being the next hotspots?

Darien Mitchell-Tontar
May 6, 2020

Which U.S. counties are at risk the next two weeks?

Specifically, which counties will have over *100 cases* **and** see a *weekly increase* in cases of over 50%.

Where has the virus spread?



County	State	
Abbeville Coun..	SC	1
Acadia Parish	LA	0
Accomack Coun..	VA	0
Ada County	ID	3
Adair County	IA	1
	KY	0
	MO	0
	OK	0
Adams County	CO	10
	IA	0
	ID	0
	IL	0
	IN	1
	MS	0
	ND	0
	NE	0
	OH	0
	PA	1
	WA	0
	WI	0
Addison County	VT	0
Aiken County	SC	0
Aitkin County	MN	0
Alachua County	FL	11
Alamance Coun..	NC	0
Alameda County	CA	35
Alamosa County	CO	0
Albany County	NY	43
	WY	0
Albemarle Cou..	VA	0
Alcona County	MI	0

Data Collection

Tracked cases from Jan 22 to April 27

- Up to date covid-19 data from CDC.
- County demographic data from US Census.
- Hospital Data from Johns Hopkins University.

- Covid data from April 13 - 27 was used to determine the whether a county was labeled at risk.

Population Density and Recent Cases are strong predictors for counties at risk

Random Forest was the best model

- Over sampling for class imbalance
- Parameter tuning
- Engineered additional data

Other strong predictors include:

- Total cases and deaths per capita.
- Days since first case.
- Percentage of population who are black or hispanic (nearly twice as high in at risk counties).

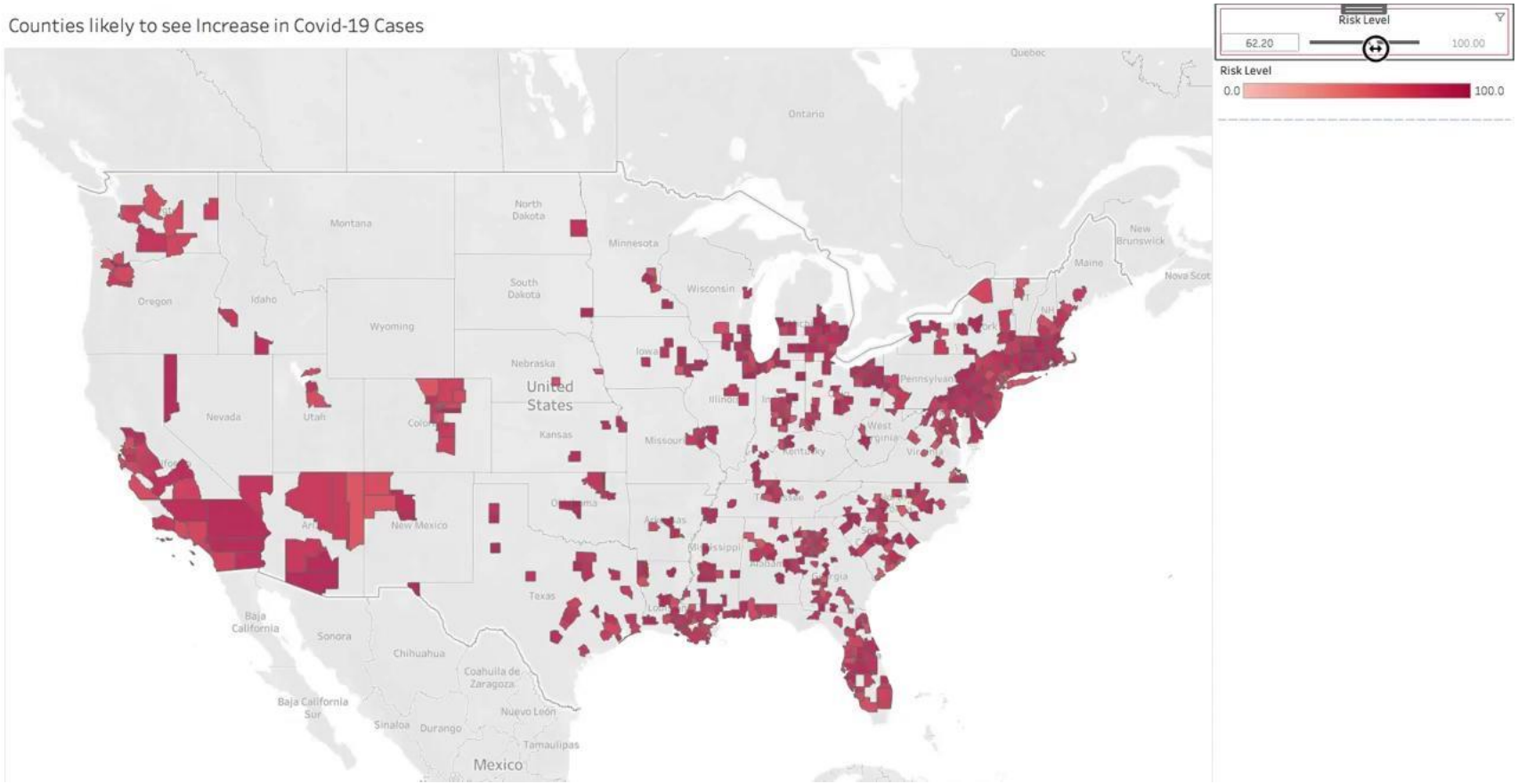
Interesting observation: Testing rate was 10 times higher on average in counties labeled not to be at risk.

Model predicts 64% of at risk counties correctly

Primarily used F1 (0.354) to determine the best model

	Actually at Risk	Not at Risk
Predicted at Risk	14	43
Predicted Not at Risk	8	564

Virus spreading in counties near previous hotspots



Thank you

Appendix A: Tables

at_risk	ppl_popdensity	total_cases	total_deaths	cases_wk_before	days_since_case1	inc_med_hh	ppl_bl_pct	ppl_hisp_pct
0	233.768353	184.682765	8.359907	59.020840	18.910685	50768.109825	8.484433	8.187244
1	910.672373	591.872881	22.288136	250.745763	30.796610	59350.483051	15.496780	10.866695

	total_cases	total_deaths
count	3141.000000	3141.000000
mean	199.979943	8.883158
std	1458.708553	86.804029
min	0.000000	0.000000
25%	2.000000	0.000000
50%	9.000000	0.000000
75%	41.000000	1.000000
max	36765.000000	2672.000000

Appendix B: The target

	County	State	Week_1	Week_2	Week_3	Week_4	Week_5	Week_6	Week_7	Week_8	Week_9	Week_10	Week_11	Week_12	Week_13	Week_14
1	Autauga County	AL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.0	10.0	17.0	25.0	33.0	39.0
2	Baldwin County	AL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	5.0	25.0	59.0	102.0	143.0	168.0
3	Barbour County	AL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.0	14.0	30.0	35.0
4	Bibb County	AL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.0	11.0	23.0	33.0	42.0
5	Blount County	AL	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	6.0	11.0	18.0	31.0	34.0

County	State	Week_1	Week_2	Week_3	Week_4	Week_5	Week_6	Week_7	Week_8	Week_9	Week_10	Week_11	Week_12	Week_13	Week_14
Autauga County	Alabama	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	inf	6.00	0.714286	0.916667	0.304348	0.300000
Baldwin County	Alabama	0.0	NaN	NaN	NaN	NaN	NaN	NaN	inf	3.0	3.75	1.210526	1.071429	0.413793	0.365854
Barbour County	Alabama	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	0.00	1.500000	2.666667	1.545455	0.250000
Bibb County	Alabama	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.60	1.666667	1.125000	0.882353	0.312500
Blount County	Alabama	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.60	1.000000	0.600000	0.625000	0.307692

Appendix C: All features

```
Index(['countyfips', 'county_name', 'state', 'region', 'Northeast', 'Other',
       'South', 'West', 'coastal', 'population', 'ppl_popdensity',
       'county_beds', 'total_cases', 'total_deaths', 'days_since_case1',
       'days_since_death1', 'county_hospitalizations', 'county_testing_rate',
       'cases_wk_before', 'deaths_wk_before', 'pcnt_inc_wk10', 'pcnt_inc_wk11',
       'pcnt_inc_wk12', 'inc_pov_rate', 'inc_med_hh', 'inc_per_cap',
       'jobs_serv', 'jobs_manu', 'jobs_agro', 'jobs_unem', 'jobs_const',
       'jobs_trans', 'ppl_ov65_pct', 'ppl_un18_pct', 'ppl_nohs_pct',
       'ppl_college_pct', 'ppl_wht_pct', 'ppl_hisp_pct', 'ppl_bl_pct',
       'at_risk'],
      dtype='object')
```

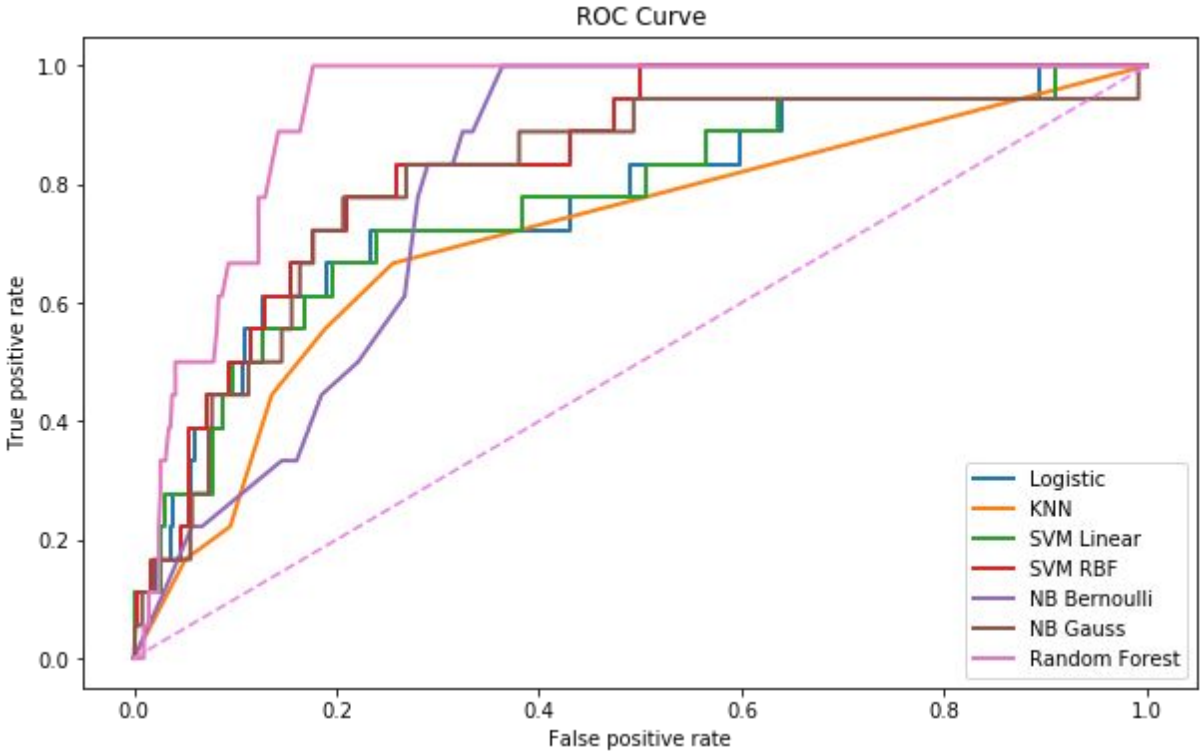
Appendix D: Best Features

```
selected_feat= X_train_new.columns[(rfm_fe.get_support())]
selected_feat
```

```
Index(['ppl_popdensity', 'county_beds', 'total_cases', 'total_deaths',
       'tot_cases_percap', 'tot_deaths_percap', 'days_since_death1',
       'cases_wk_before', 'cases_deaths'],
      dtype='object')
```

at_risk	ppl_popdensity	total_cases	total_deaths	cases_wk_before	days_since_case1	inc_med_hh	ppl_bl_pct	ppl_hisp_pct
0	233.768353	184.682765	8.359907	59.020840	18.910685	50768.109825	8.484433	8.187244
1	910.672373	591.872881	22.288136	250.745763	30.796610	59350.483051	15.496780	10.866695

Appendix E: Model Selection - ROC Curve



Appendix F: Confusion Matrix

0.35443037974683544