Predicting COVID-19 Cases in California

Annie Adams, Anum Damani, Andrew Bissell, Eric Cha

Abstract & Motivation

<u>Problem:</u>
Given the past data, can we build predictive models that forecast the future of the pandemic so that we can see one step ahead and prepare accordingly? Additionally, which data is highly relevant to the prediction and how should that affect our policies?

<u>Motivation:</u>
To better understand the scale of the infectious disease and better guide our decision making.

<u>Uses:</u>
Can better inform health legislators of future spikes, informs them to take preemptive action.

<u>Challenges:</u>
i.e. COVID constantly mutating, people holding opinions on lockdown/mask procedures, negligent legislation, travel, etc.

Technical Details

* **covidcast** data module, 5 signals pulled from “Change Healthcare” & “Doctor Visits”

* We first appropriately split our data using **TimeSeriesSplit()** to maintain the chronological order of our data. Then our training consisted of feeding data into a **Decision Tree Regressor** model and a **Support Vector Machine** for Regression (or **SVR**) model. We used cross validation across 5 folds and then produced a final model test score.

* With the Decision Tree model we experimented with several **max_depth** values to see if we could improve our test results. Similarly, with the SVR model we experimented with the different **kernels** types to see which yielded the best results.

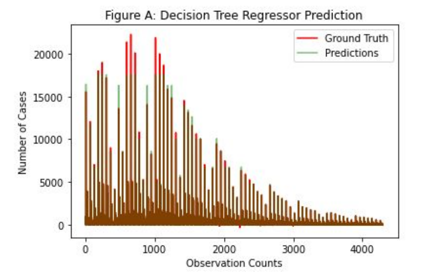

Figure A: Decision Tree Regressor Prediction

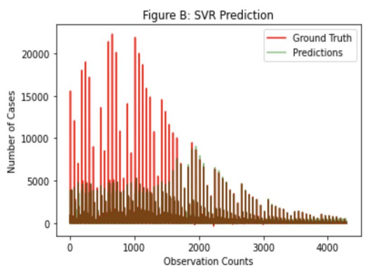

Figure B: SVR Prediction

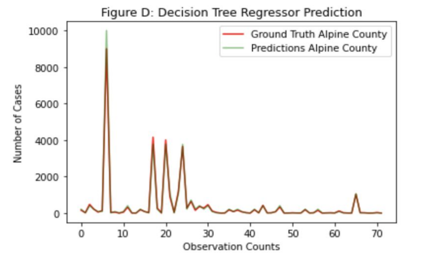

Figure D: Decision Tree Regressor Prediction

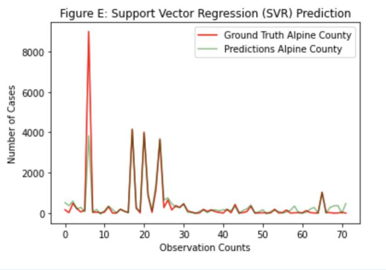

Figure E: Support Vector Regression (SVR) Prediction

Conclusion

* To us, getting all of the signals was the easiest part of the project.

* We achieved better than we expected; going into this project we did not anticipate to develop our results the way we did - our original performances/accuracies did not make us optimistic of our progress.

* We ran into MANY difficulties along the way: having to practically redo each step from scratch, troubleshooting nearly inexplicable errors, interpreting our results to figure out if we were even going into the right direction.

Future Work

* We can choose more diverse signals.
* We can work with different metrics besides R^2 to evaluate our model performances.
* We make distinctions between time periods with/ without vaccines to make better predictions
* We can repeat this process for another state in the US and compare with California.

Coordination

* The completion of this project depended on a collaborative effort between all group members.

References

<u>Delphi Epidata API</u>
* Change Healthcare
* Doctor Visits

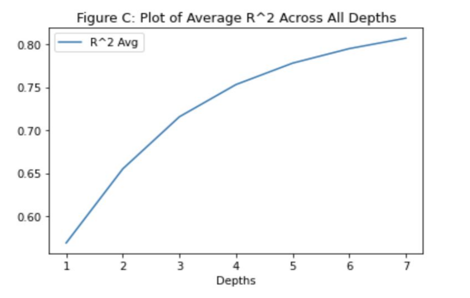

Figure C: Plot of Average R^2 Across All Depths