**ANONYMIZED COVID PATIENTS' OUTCOME ANALYSIS & PREDICTION – FINAL REPORT**

**1.0 Introduction**

COVID-19 pandemic disrupted lives unevenly, with certain sociodemographic groups being overexposed to the pandemic. Despite prior studies' focus on forecasting the epidemiological outcomes using empirical and data-driven models, a general statistical that incorporates societal variables is missing. We propose to jointly develop a modeling and visualization framework that jointly makes predictions of COVID-19 health outcomes with respect to epidemiological and sociodemographic data. The impact of this work is modeling and visualizing the impact of COVID-19 on disproportionately exposed groups in the US.

**2.0 Problem Definition [Objectives & Motivations]**

COVID-19 is a severe acute respiratory syndrome (SARS) caused by the SARS-CoV-2 virus [15]. Globally, the COVID-19 pandemic infected 435M people (78.8M in the US) and claimed 5.95M deaths (946K in the US) as of 2/28/2022. Despite the indiscriminate spread of the virus, socio-demographics cause disproportionate exposure of certain communities [9,10]. The COVID-19 pandemic is unique in terms of the heavy role of technology in aiding public health decisions, where personal devices, social media, and interactive maps played a large part in social distancing and vaccine policy [11]. These analytic techniques have also successfully helped determine the impact of factors like symptoms and demographics on the spread of the disease [16]. In this project, we are motivated to explore the impact of powerful descriptive & predictive analytics on the well-being and safety of the population during the pandemic. We will create an easy-to-use web app for visualization & prediction of COVID-19 cases outcomes concerning epidemiological, geo- & demographic drivers for the US.

**3.0 Survey**

We studied papers that used predictive models to provide diagnostic testing tools based on demographic and lab testing data [3] or to identify individuals' risk of severe complications based on health history survey data [2]. Tiwari, A. et al. (2021) also built a predictive model that highlights the existing disparities in COVID-19 vulnerability among different subpopulations based on race and financial status. These studies work with publicly available datasets from CDC, DHS, Johns Hopkins University, and other institutions. On dissemination, we studied the building blocks of web apps, user interface overview, and characteristics of the REST APIs [5] and the types, attributes, and benefits of cloud computing [4].

**4.0 Proposed Method**

**4.1 Intuition**

Our chosen dataset consists of demographic features and health history, exposure, and severity of status features of anonymized patients. These features provide a broader context and insights into the occurrence and development of cases in different populations. By visualizing cases and their severity geographically and by relations between patient demographics and medical history data to COVID-19 outcomes, our approach provides better insights into the vulnerability at the patient level or geographical area. Combined with a severity risk assessment predictive model, those insights can be an additional tool for research institutes and decision-makers.

**4.2 Details of the Approaches**

We used the data of 63M patients from CDC [12] with geographical, demographic, and health features. We enriched this data with other geographical datasets providing information on vaccination and the current pandemic status. We built two main types of deliverables: First, a descriptive dashboard exploring patients' outcomes by geography and demographic features in the form of a Tableau dashboard and a predictive model on patient outcomes. We packaged the dashboards & models in the web app published on Heroku.

Our implemented list of innovations is:

1. **Data Handling:** Integrating data from varying spatial scales and granularity.
2. **Modeling:** Two-tiered model for predicting the probability of COVID-19 health outcome and associated deterministic outcome. Model evaluation metrics ensure rare event accuracy.
3. **Visualization:** Informatics on patient outcome and the underlying model's performance metrics.
4. **Dissemination:** Easy to use app with < 3-clicks to an insight.

*1. Data handling approach*

We found that it contained many missing values in several features by exploring the cases dataset. Different features also had a variety of data types, including numeric, categorical, and timestamp. Also, there were three target outcomes that the model tried to predict. Those include whether or not the patient has been hospitalized, moved to the ICU, or died. The dataset posed a challenge as, in some cases, there would be missing values for some of the three targets while only 2, 1, or none have information (yes or no). We did our data pre-processing in 2 rounds.

The first round includes dropping rows with all targets missing as those do not inform the model of any possible outcome, which the model needs to predict. Next, we identified the features that had more than 40% missing and removed those features from further consideration. Those will not be informative in most of the dataset, and imputing them with any method will significantly bias the data. Next, since the resulting dataset was still significantly large, we could afford to drop the rows with missing values (in the feature columns) without having too little data to work with.

The second round of processing included separating the date timestamp column into two numeric columns for month and year (since ML models expect only numeric type). Since we needed to include all columns numerically, the one-hot encoding county feature would have generated thousands of new columns. Instead, we joined our cases data with vaccination data on the county FIPS code and substituted the county column with the county's population, metro status (whether in the metro area or not), and Social Vulnerability Index (SVI). The other categorical features (like state, sex, age group, etc.) were 1-hot encoded. The resulting dataset contained around 20M rows and 74 features (77 including targets).

Those are detailed in the table below:

|  |  |
| --- | --- |
| Round 1 Preprocessing | Round 2 Preprocessing |
| * Dropped rows with all targets being missing | * Replaced date with month and year |
| * Removed features with >40% missing | * Joined with vax data to replace the county |
| * Dropped rows with missing values in features | * 1-hot encoding categorical features |

*2. Modeling approach*

We decided to develop two models that achieve different tasks: A single outcome model that predicts the most probable combination of outcomes and a multi-outcome model that predicts the probability of each outcome occurring to the patient. The reason behind developing two models is two-fold: Firstly, if both models perform well concerning our evaluation metrics, they can be used in a nested fashion where the multi-outcome model predicts probabilities of hospitalization, ICU need, and death risk, and the single outcome model can predict specific outcome categories for the most probable outcome. Secondly, we can easily switch to the other model if we run into computational bottlenecks with one of the models, such as run-time, model generalization, or accuracy. Lastly, having deterministic and probabilistic results makes the results approachable to a wide audience, especially since non-expert users may perceive probabilities differently. We narrowed our search to Logistic Regression, Decision Trees, Boosted Trees, and SVM classifiers.

*3. Visualizations approach*

The team decided to use a combination of web application (using Node.Js framework) and Tableau for visualization & dissemination of information. Our approach included the following:

* Uses Tableau for descriptive visualizations to provide insights into the data & model performance. Tableau is easy & time efficient to develop compared to tools like D3 or Python.
* Used the web application to allow users to interact with the model and embedded the Tableau dashboards within the web app using the embedded code provided by Tableau Public Server for one packaged final product.
* Used wireframe & prototype tools like Figma to visualize & design the application pre-dev.

*4. Final dissemination approach (the Web App)*

We have built a web application on the Heroku cloud application platform ([link](https://cse-6242-team12.herokuapp.com/)) using Node.js as the backend technology. The application has a home page and 3 pages for Tableau dashboards and outcome prediction. The model takes user inputs and predicts the outcome in the Get Prediction webpage. The ML models were converted into Joblib objects and deployed within the web application to enable real-time predictions. The Heroku server has 1 GB of RAM. Due to server limitations, ML models can take up to 15 seconds to predict.

**5.0 Experiments & Evaluation**

*Model Evaluation Metrics:*

We tested the efficiency and accuracy of different classifiers for predicting COVID-19 health outcomes using the following metrics: F1 for training and testing, one-versus-one (OVO) accuracy for categorical outcomes, and one-versus-all (OVA) accuracy for categorical outcomes. F1 informs the overall accuracy of the model. In contrast, OVO and OVA inform the predictive capability of different categories in the presence of unbalancing in the training categories, some categories being observed more frequently.

Our team had the following questions to solve for the successful completion of this project:

1. How to explore the data? What tools to use to handle the size? How to perform data wrangling?
2. How to aggregate the millions of rows for visualizations? What tools should we use?
3. How do we visualize the data & model's performance? How can we make it interactive for users?
4. How do we define the ML problem for outcome prediction? Should we build one model or more?

**5.2 Experiments Description**

*Data Exploration*

The main dataset, CDC's individual case data with geography information, is 9 Gigabytes in size. It has 63 million rows and 19 columns. We loaded the data to a commercial MySQL cloud instance hosted on digitalocean.com for exploratory analysis. We decided to discontinue the instance due to budget constraints. We then experimented with Google collab notebook Pro+ service and found it can host the data and be more cost-effective and collaborative.

*Data Aggregation*

In addition to the main dataset, we wanted to use the vaccination data for enrichment which has over 1M rows and 66 columns. Any visualization tool would struggle to work with the raw data at individual case grain. We decided to perform some tests on aggregations to see if we could shrink the dataset without losing the information. Tableau was being considered for visualization, and Tableau Public Server has a 10M row limit per connection. Moreover, Tableau's performance struggles with more data. Due to loss issues with aggregations [23], we designed our experiment as follows:

* Filtered data to one month & performed clean upon this data (e.g., NaN, 'Unknowns' handling).
* Performed Group-By & Case Count aggregation across all columns to see size shrinkage achieved.
* Checked smaller dataset for loss of information & features.

This test was performed on Jan 2022 data which had 14M rows. Using the steps above, we could get this dataset down to 486K rows without losing the total case count (96.5% size reduction). This enabled us to perform these actions on the full dataset to bring 63M rows down to 6.8M, which were successfully loaded to Tableau Desktop.

Another aggregation was needed on the ML models data, time-series of actual and predicted classes for three stochastic models, and one big dataset of actuals & predictions for the deterministic model. Across train & test datasets, the three stochastic models had anywhere from 34K to 2.1M rows (combined 3.2M rows). To calculate classification model perf metrics, the dataset was converted to True Positive, True Negative, False Positive & False Negative counts and aggregated down to a mere 24-row data frame. Similar aggregation was performed on the 24.5M rows of the deterministic model and added to the same dataset.

*Tableau Dashboards Development*

The concern was the dashboard's performance for the descriptive visualizations dashboard using the aggregated 6.8M rows of data. Tableau is typically pretty good with query time, even for several million rows. We slowly built the dashboard and monitored performance as we went along. We kept the number of clicks needed to get an insight to < 3 to keep the query time manageable.

For the 2nd dashboard visualizing the ML models' performance, we used the aggregated data as mentioned above, and with the number of rows being low, the dashboard performed well.

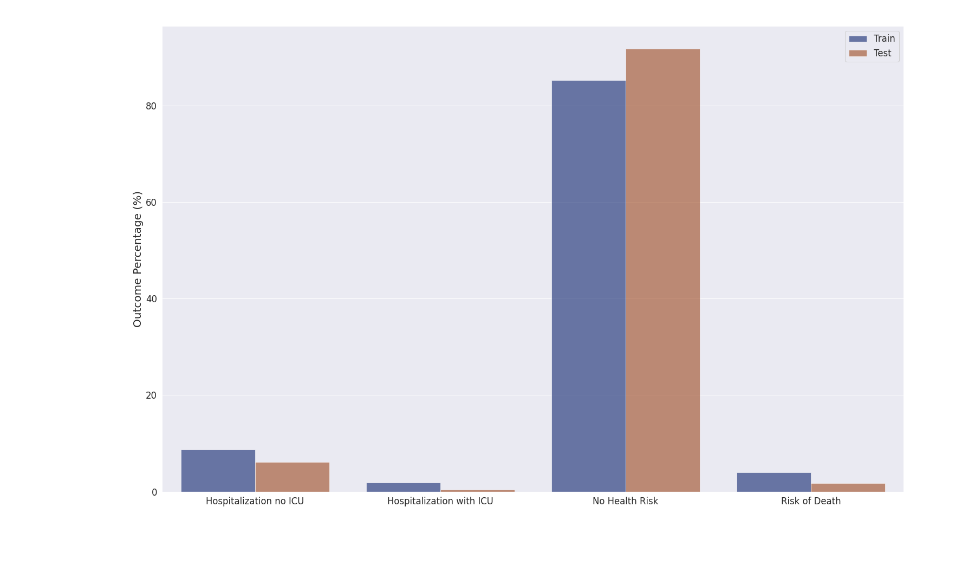
*Visualizations & Model Interaction*

To paint a picture of the descriptive visualizations (for insights, trends & model performance) and model interaction, we decided to build a full working prototype of the web application in Figma. This prototype allowed us to experiment with the app layout without compromising the project timeline. The final working prototype of the app is available in Figma ([link](https://www.figma.com/proto/U2cieCsx1zauvFY2QQUfLe/Team12_Wireframe?node-id=36%3A2&scaling=min-zoom&page-id=0%3A1&starting-point-node-id=1%3A3) here).

*Machine Learning Modeling*

The analytical engine of our project answers the following question: "Given a patient's characteristics such as age, location, and gender, what are the expected health outcomes given prior data?". This problem has two components. The first is "given the time of the year, what are the probabilities of three main health outcomes: hospitalization, death, no infection"? The second is, "irrespective of the time of the year, what is the most likely health outcome for a patient?"

The deterministic, most-probable health outcome model uses all of the data except for time-related predictors to predict one of four categories: No Health Risk, Hospitalization, Hospitalization with ICU need, and Risk of Death. A 75-25 training-test split is implemented for model selection and parameter optimization. The categories in the underlying training are below:

Our team has tried support vector machine (SVM), logistic regression, random forest, and boosted decision trees. We ran into memory issues with the scikit-learn implementation of the SVM classifier. Even after variable scaling and category encoding, the model ran out of memory. The logistic regression model results in severe multicollinearity due to a high number of one-hot-encoded categorical variables. Thus, random forest and boosted decision trees are evaluated vigorously using grid search. We focused on the number of trees and maximum depth for both models with the addition of the gamma parameter for the boosted decision tree. Our simulations showed that boosted decision trees with an alpha of 1 and max depth of 10 with 500 trees obtain the highest overall, one-versus-rest, and one-versus-one accuracy for all health outcomes that are being predicted. The optimal random forest model was close to the boosted tree model with an average accuracy of 0.91 (compared to boosted tree's accuracy of 0.918). We provide pictorial summaries of our results in the Appendix.

For the probability of outcomes model, the team has initially considered a single model with multiple outputs. That model posed a strict requirement of training labels being fully defined for a given data point (all outcomes being non-missing for a patient). This requirement has radically reduced the dataset to train the model as in the original set; most cases had outcomes that are partially defined. Therefore, the team considered an alternative approach where three different models were trained separately for each outcome. Each individual set was stratified, and a random forest classifier was trained due to the advantages of tree-based models discussed above. Each random forest had a max depth of 18 and 100 estimators. These values maximized model performance on all outcomes while still fitting in the space allocated to models in the deployed application. The table below details the test scores achieved by the models:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outcome | Train Acc | Train F1 | Test Acc | Test F1 |
| Hospitalization | 77.88% | 77.63% | 77.50% | 77.25% |
| ICU | 84.93% | 85.85% | 82.78% | 83.91% |
| Death | 93.81% | 94.13% | 93.70% | 94.03% |

Here, we notice that the model that predicts death is the most accurate while the one that predicts hospitalization is the least of the 3. One reason could be that hospitalization is a more general and less extreme outcome, and a wide variety of groups could be hospitalized, even for a very short time. ICU, and more extreme death, are more extreme events. Only a few distinctive groups of people (e.g., elderly, poor, minority patients) could go to the ICU or die. Hence, these outcomes were more predictable by our random forest models.

Some cases could have contradicting predictions between the probability models and the distinct outcomes model. For example, a patient could be predicted to have high chances of being hospitalized and low chances of survival using the 3-probability model. In contrast, the distinct outcomes model predicts no health risk. This might happen since the 3-probability model was trained with down-sampled data of equal outcome counts (positive and negative) and may exaggerate certain group exposure randomly (fatal cases of 30-year-old people may have been sampled strongly). The input to the distinct outcomes model, on the other hand, did not apply to resample, and it may represent groups more accurately. Also, when the 3-probability model exaggerates the severity of the outcomes, this indicates an underlying selection bias in the dataset where only extreme cases were reported in various groups of people. For example, most 30 years old, people who contracted the virus might not have reported their cases, and the data would not represent such demographic sufficiently. In this case, the 3-probability model may provide a slightly skewed prediction.

**6.0 Plan of Activities & Success Metrics Update**

Graphical user interface, application, table

Description automatically generatedThe intended timeline for this project was to finish all work by the end of April 22. The Gantt chart with an update on how the tasks progressed is shown below:

All team members contributed a similar amount of effort with the following success metrics. An update on the metrics is given below:

1. Completing key milestones per the schedule above *(****Completed.*** *We had a slight delay in one task but overall, the project followed the plan)*.

2. Accuracy of the predictive model will be performed using recall rates for varying test/train ratios *(****Completed****. All Modeling metrics are available in the app and summarized in section 5)*.

3. The web design should preserve the Three-click rule [19] *(****Completed.*** *The wireframe navigates the user to key pages in <3 clicks. Tableau Dashboards also mostly get to a meaningful insight in < 3 clicks.)*.

4. The number of statistics displayed per information pop-up should follow 7±2 [20]: *(****Completed****)*.

**7.0 Conclusions & Discussion**

Among all methods, we experimented with decision-tree-based methods that have the best balance between explainability and accuracy. Our prediction metrics allowed us to fine-tune models to capture rare health outcomes. With both 3-probability and distinct outcomes models, severe patient outcomes were highly predictable. The model performance dashboard helped the team understand model limitations, analyze metrics, and suggest new improvements such as more representative sampling of data (groupwise and timewise). We found that random forest was the best choice for the 3-probability model, whereas boosted decision trees provided the highest accuracy for the distinct outcome model. The final web application ([link](https://cse-6242-team12.herokuapp.com/)) nicely packaged all components, dashboards, and picked models to enable the user to truly explore the COVID data and get individual COVID case predictions.

**References**

1. [Tiwari, Anuj, Arya V. Dadhania, Vijay Avin Ragunathrao, and Edson R.A. Oliveira. "Using Machine Learning to Develop a Novel COVID-19 Vulnerability Index (C19vi)."](https://www.nature.com/articles/s41598-021-93126-7) *[Science of The Total Environment](https://www.nature.com/articles/s41598-021-93126-7)* [773 (2021): 145650.](https://www.nature.com/articles/s41598-021-93126-7) [[https://doi.org/10.1016/j.scitotenv.2021.145650](https://www.nature.com/articles/s41598-021-93126-7)](https://doi.org/10.1016/j.scitotenv.2021.145650)[.](https://www.nature.com/articles/s41598-021-93126-7)
2. [DeCaprio, Dave, Joseph Gartner, Carol J. McCall, Thadeus Burgess, Kristian Garcia, Sarthak Kothari, and Shaayaan Sayed. "Building A Covid-19 Vulnerability Index."](https://www.nature.com/articles/s41598-021-93126-7) *[Journal of Medical Artificial Intelligence](https://www.nature.com/articles/s41598-021-93126-7)* [3 (2020): 15–15.](https://www.nature.com/articles/s41598-021-93126-7) [[https://doi.org/10.21037/jmai-20-47](https://www.nature.com/articles/s41598-021-93126-7)](https://doi.org/10.21037/jmai-20-47)[.](https://www.nature.com/articles/s41598-021-93126-7)
3. [Zhang, Jimmy, Tomi Jun, Jordi Frank, Sharon Nirenberg, Patricia Kovatch, and Kuan-lin Huang. "Prediction of Individual COVID-19 Diagnosis Using Baseline Demographics and Lab Data."](https://www.nature.com/articles/s41598-021-93126-7) *[Scientific Reports](https://www.nature.com/articles/s41598-021-93126-7)* [11, no. 1 (2021).](https://www.nature.com/articles/s41598-021-93126-7) [[https://doi.org/10.1038/s41598-021-93126-7](https://www.nature.com/articles/s41598-021-93126-7)](https://doi.org/10.1038/s41598-021-93126-7)[.](https://www.nature.com/articles/s41598-021-93126-7)
4. [Charan, N. Ram Ganga, S. Tirupati Rao, and P. V. S. Srinivas. "Deploying an Application on the Cloud."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[International Journal of Advanced Computer Science and Applications](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [2, no. 5 (2011).](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
5. [Kuuskeri, Janne. "Experiences on a design approach for interactive web applications." In](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[2nd USENIX Conference on Web Application Development (WebApps 11)](https://www.sciencedirect.com/science/article/pii/S004896972100718X)*[. 2011.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
6. [Quiroz-Juárez, Mario A., Armando Torres-Gómez, Irma Hoyo-Ulloa, Roberto de J. León-Montiel, and Alfred B. U’Ren. "Identification of high-risk COVID-19 patients using machine learning."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Plos one](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [16, no. 9 (2021): e0257234.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
7. [Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al . 2020. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine (2020).](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
8. [World Health Organization et al. 2022. Coronavirus disease (COVID-19): situation report, 182. (2022).](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
9. [Golestaneh, Ladan, Joel Neugarten, Molly Fisher, Henny H. Billett, Morayma Reyes Gil, Tanya Johns, Milagros Yunes et al. "The association of race and COVID-19 mortality."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[EClinicalMedicine](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [25 (2020): 100455.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
10. [Jin, Jin, Neha Agarwala, Prosenjit Kundu, Benjamin Harvey, Yuqi Zhang, Eliza Wallace, and Nilanjan Chatterjee. "Individual and community-level risk for COVID-19 mortality in the United States."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Nature medicine](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [27, no. 2 (2021): 264-269.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
11. [Reeves, J. Jeffery, Hannah M. Hollandsworth, Francesca J. Torriani, Randy Taplitz, Shira Abeles, Ming Tai-Seale, Marlene Millen, Brian J. Clay, and Christopher A. Longhurst. "Rapid response to COVID-19: health informatics support for outbreak management in an academic health system."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Journal of the American Medical Informatics Association](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [27, no. 6 (2020): 853-859.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
12. ["Covid-19 Case Surveillance Public Use Data with Geography." Centers for Disease Control and Prevention. Accessed February 28, 2022. https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data-with-Ge/n8mc-b4w4.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
13. [Ding, Yongmei, and Liyuan Gao. "An evaluation of COVID-19 in Italy: A data-driven modeling analysis." Infectious Disease Modelling 5 (2020): 495-501.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
14. [Zoabi, Yazeed, Shira Deri-Rozov, and Noam Shomron. "Machine learning-based prediction of COVID-19 diagnosis based on symptoms." npj digital medicine 4, no. 1 (2021): 1-5.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
15. [Li, Lixiang, Zihang Yang, Zhongkai Dang, Cui Meng, Jingze Huang, Haotian Meng, Deyu Wang et al. "Propagation analysis and prediction of the COVID-19."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Infectious Disease Modelling](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [5 (2020): 282-292.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
16. [Khanam, Fahima, Itisha Nowrin, and M. Rubaiyat Hossain Mondal. "Data visualization and analyzation of COVID-19." Journal of Scientific Research and Reports 26, no. 3 (2020): 42-52.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
17. [Healey CG, Simmons SJ, Manivannan C, Ro Y (2021) Visual analytics for the coronavirus COVID-19 pandemic. Big Data 3:X, 1–20, DOI: 10.1089/big.2021.0023.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
18. [Healey, Christopher G., Susan J. Simmons, Chandra Manivannan, and Yoonchul Ro. 2022. "Visual Analytics for the Coronavirus COVID-19 Pandemic."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Big Data](https://www.sciencedirect.com/science/article/pii/S004896972100718X)*[.](https://www.sciencedirect.com/science/article/pii/S004896972100718X) [[https://doi.org/10.1089/big.2021.002](https://www.sciencedirect.com/science/article/pii/S004896972100718X)](https://doi.org/10.1089/big.2021.002)
19. [Zeldman, Jeffrey.](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Taking your talent to the web: A guide for the transitioning designer](https://www.sciencedirect.com/science/article/pii/S004896972100718X)*[. New Riders Publishing, 2001.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
20. [Migliore, Michele, Gaspare Novara, and Domenico Tegolo. "Single neuron binding properties and the magical number 7."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Hippocampus](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [18, no. 11 (2008): 1122-1130.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
21. [Allouche, Omri, Asaf Tsoar, and Ronen Kadmon. "Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS)."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Journal of applied ecology](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [43, no. 6 (2006): 1223-1232.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
22. [Gorodkin, Jan. "Comparing two K-category assignments by a K-category correlation coefficient."](https://www.sciencedirect.com/science/article/pii/S004896972100718X) *[Computational biology and chemistry](https://www.sciencedirect.com/science/article/pii/S004896972100718X)* [28, no. 5-6 (2004): 367-374.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)
23. [Pollet, Thomas V., Gert Stulp, S. Peter Henzi, and Louise Barrett. "Taking the aggravation out of data aggregation: A conceptual guide to dealing with statistical issues related to the pooling of individual‐level observational data." American Journal of Primatology 77, no. 7 (2015): 727-740.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)

**[Appendix I : Mathematical Details of Evaluation Metrics](https://www.sciencedirect.com/science/article/pii/S004896972100718X)**

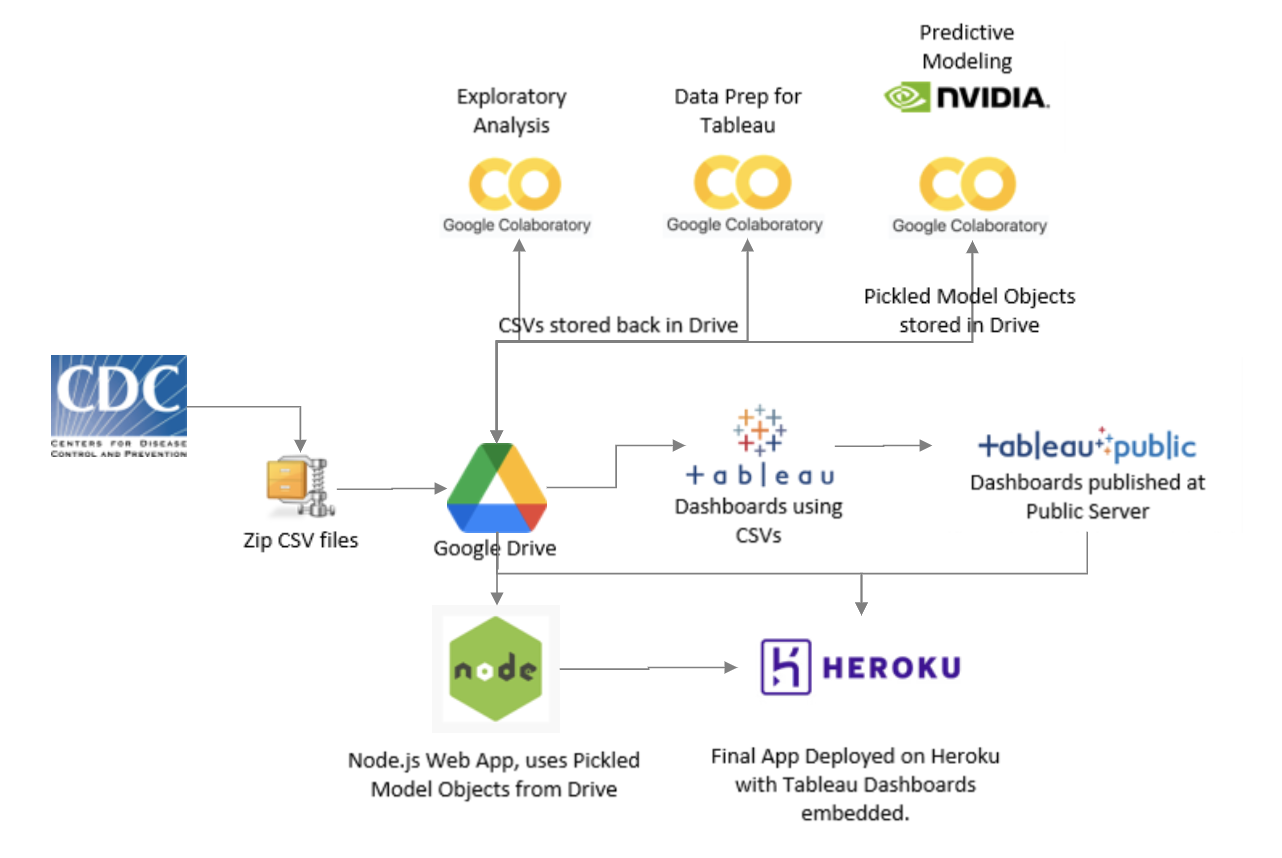
[Comparison metrics are defined in terms of the confusion matrix elements, True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), as below:](https://www.sciencedirect.com/science/article/pii/S004896972100718X)

|  |  |
| --- | --- |
|  | [(1)](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |
|  | **[(2)](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |
|  | **[(3)](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |
|  | **[(4)](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |

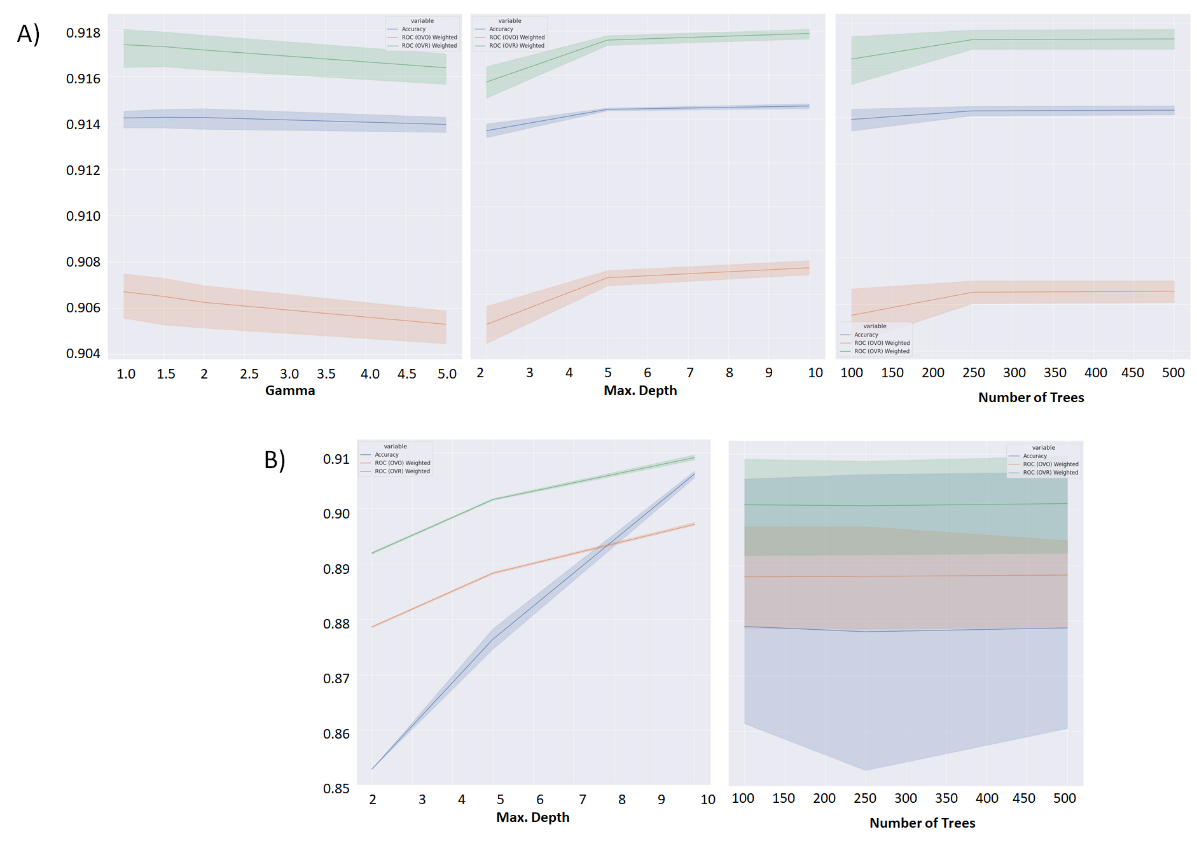
**[Appendix II : Derived Health Outcome Categories for the Single Response Model](https://www.sciencedirect.com/science/article/pii/S004896972100718X)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **[Category](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[Hosp.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[ICU](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[Death](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[Category](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[Hosp.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[ICU](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | **[Death](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |
| **[No health Risk](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |  |  | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) | **[Death Risk](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |  |  | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |
| **[Hospitalization Risk](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |  |  | **[Death Risk: Hosp.](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |  | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |
| **[ICU Need](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |  | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |  | **[Death: ICU](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |  | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |
| **[ICU + Bed Need](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) | [X](https://www.sciencedirect.com/science/article/pii/S004896972100718X) |  | **[Death Risk: Hosp + ICU](https://www.sciencedirect.com/science/article/pii/S004896972100718X)** |  |  |  |

**[Appendix II: Web App Architecture:](https://www.sciencedirect.com/science/article/pii/S004896972100718X)**

[](https://www.sciencedirect.com/science/article/pii/S004896972100718X)

**[Appendix III: Single Category Model Comparison:](https://www.sciencedirect.com/science/article/pii/S004896972100718X)**

[](https://www.sciencedirect.com/science/article/pii/S004896972100718X)

1. [Diagnostics for Boosted Regression Tree Model Search B) Diagnostics for Random Forest Model Search](https://www.sciencedirect.com/science/article/pii/S004896972100718X)