**Generalized Linear Models (GLM)**

GLM was used since the response value (retweets) was not continuous and was truncated at 0. Next steps include checking predictors/x variables for collinearity, applying variable selection or shrink methods, and implementing cross-validations.

Call:

glm(formula = retweets ~ isretweeted + length + favorites + ishealth +

ispandemic + isvirus + isemergency + isdeaths + iswho + iscdc +

isnih + isdisease + isquarantine + isrecover + isban + iscoronavirus +

iscovid19 + iswash + isracist + isasian + ischinese + isinfectious +

ch.confirmed + it.confirmed + ot.confirmed + us.confirmed +

ch.deaths + it.deaths + ot.deaths + us.deaths, data = covid19current)

Deviance Residuals:

Min 1Q Median 3Q Max

-139722 -7264 -2056 2839 187104

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.299e+07 5.800e+05 22.387 < 2e-16 \*\*\*

isretweetedTRUE 6.546e+03 3.160e+02 20.713 < 2e-16 \*\*\*

length -5.138e+01 6.945e+00 -7.398 1.40e-13 \*\*\*

favorites 8.145e-01 4.204e+00 0.194 0.846379

ishealth -4.176e+02 5.016e+02 -0.833 0.405118

ispandemic -1.525e+03 9.141e+02 -1.669 0.095180 .

isvirus 1.358e+03 3.725e+02 3.646 0.000267 \*\*\*

isemergency -4.448e+03 1.301e+03 -3.419 0.000630 \*\*\*

isdeaths -3.387e+03 8.674e+02 -3.905 9.46e-05 \*\*\*

iswho 2.319e+04 5.130e+02 45.218 < 2e-16 \*\*\*

iscdc -3.676e+03 1.008e+03 -3.645 0.000268 \*\*\*

isnih -6.797e+03 2.430e+03 -2.797 0.005167 \*\*

isdisease -4.519e+03 1.164e+03 -3.883 0.000103 \*\*\*

isquarantine -3.845e+03 8.403e+02 -4.575 4.77e-06 \*\*\*

isrecover -3.071e+03 1.830e+03 -1.678 0.093328 .

isban 1.502e+03 7.469e+02 2.011 0.044285 \*

iscoronavirus 5.704e+03 4.150e+02 13.744 < 2e-16 \*\*\*

iscovid19 -4.556e+03 2.076e+02 -21.950 < 2e-16 \*\*\*

iswash -2.697e+03 8.856e+02 -3.046 0.002322 \*\*

isracist -8.455e+03 3.713e+03 -2.277 0.022795 \*

isasian -3.421e+03 2.978e+03 -1.149 0.250661

ischinese 4.201e+03 6.736e+02 6.237 4.51e-10 \*\*\*

isinfectious -2.736e+03 2.172e+03 -1.259 0.207878

ch.confirmed -1.806e+02 8.150e+00 -22.157 < 2e-16 \*\*\*

it.confirmed -1.475e+01 6.582e-01 -22.414 < 2e-16 \*\*\*

ot.confirmed 1.124e+01 3.724e-01 30.190 < 2e-16 \*\*\*

us.confirmed -5.654e+01 4.096e+00 -13.802 < 2e-16 \*\*\*

ch.deaths 4.929e+02 2.713e+01 18.167 < 2e-16 \*\*\*

it.deaths 3.917e+01 5.369e+00 7.296 3.02e-13 \*\*\*

ot.deaths -6.725e+01 1.311e+01 -5.131 2.89e-07 \*\*\*

us.deaths -2.440e+02 1.523e+02 -1.603 0.109003

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 458727180)

Null deviance: 2.3711e+13 on 41999 degrees of freedom

Residual deviance: 1.9252e+13 on 41969 degrees of freedom

AIC: 956870

Number of Fisher Scoring iterations: 2

**Regression Tree**

Regression tree:

tree(formula = retweets ~ isretweeted + length + favorites +

ishealth + ispandemic + isvirus + isemergency + isdeaths +

iswho + iscdc + isnih + isdisease + isquarantine + isrecover +

isban + iscoronavirus + iscovid19 + iswash + isracist + isasian +

ischinese + isinfectious + ch.confirmed + it.confirmed +

ot.confirmed + us.confirmed + ch.deaths + it.deaths + ot.deaths +

us.deaths, data = covid19current)

