Supplementary Material

Tuning of the Xg Boost Model

The following three parameters were used in Xg Boost.

Nrounds
This parameter controls the maximum number of iterations. For classification, it is similar to the number of trees to grow.

Eta
Range is (0,1)
This parameter controls the rate at which our model learns patterns in data. After every round, the parameter shrinks the feature weights to reach the optimum. Lower eta leads to slower computation. It must be supported by an increase in nrounds. Typically, eta lies between 0.01 and 0.3.

max_depth
Range is (0, Inf)
It controls the depth of the tree. The larger the depth, the more complex the model becomes with higher chances of overfitting. There is no standard value for max_depth. Larger datasets require deep trees to learn the rules from data.

To tune these parameters, we tried different combinations using an automated program in R and chose the one which produced the minimum MAPE on test data. The program iteratively stepped through different values of the tuning parameters within a specified range while calculating the MAPE. The ranges and step increment values of the different parameters were:

Nrounds: 10 to 25 [ step increment = 1]
Eta: 0 to 1 [step increment = 0.01]
Max_depth: 3 to 10 [step increment = 1]

The final set of tuning parameters considered were: nrounds = 24, eta = 0.2, and max_depth = 5.

Citations

R packages xgboost, randomForest, rpart, and neuralnet were used for developing extreme gradient boosting, random forest, CART, and neural network models, respectively.

Citation: XGBoost:
https://CRAN.R-project.org/package=xgboost

Citation: randomForest:
https://CRAN.R-project.org/doc/Rnews/

Citation: rpart:
https://CRAN.R-project.org/package=rpart

Citation: neuralnet
https://CRAN.R-project.org/package=neuralnet