

prediction result based on survey

Y\_PREDICT Y\_xgb\_PREDICT Y\_rf\_PREDICT Actual GDP (normalized)

2016 1.065349 0.301881 0.010663 0.881876

2017 1.301549 0.671907 0.292508 1.060189

2018 1.169575 -0.407999 -0.308963 1.283080

2019 1.313972 -0.471359 -0.475997 1.483681

2020 1.837124 -0.309756 0.066742 1.416814

Random forest fit MSE: 0.017696450768226374

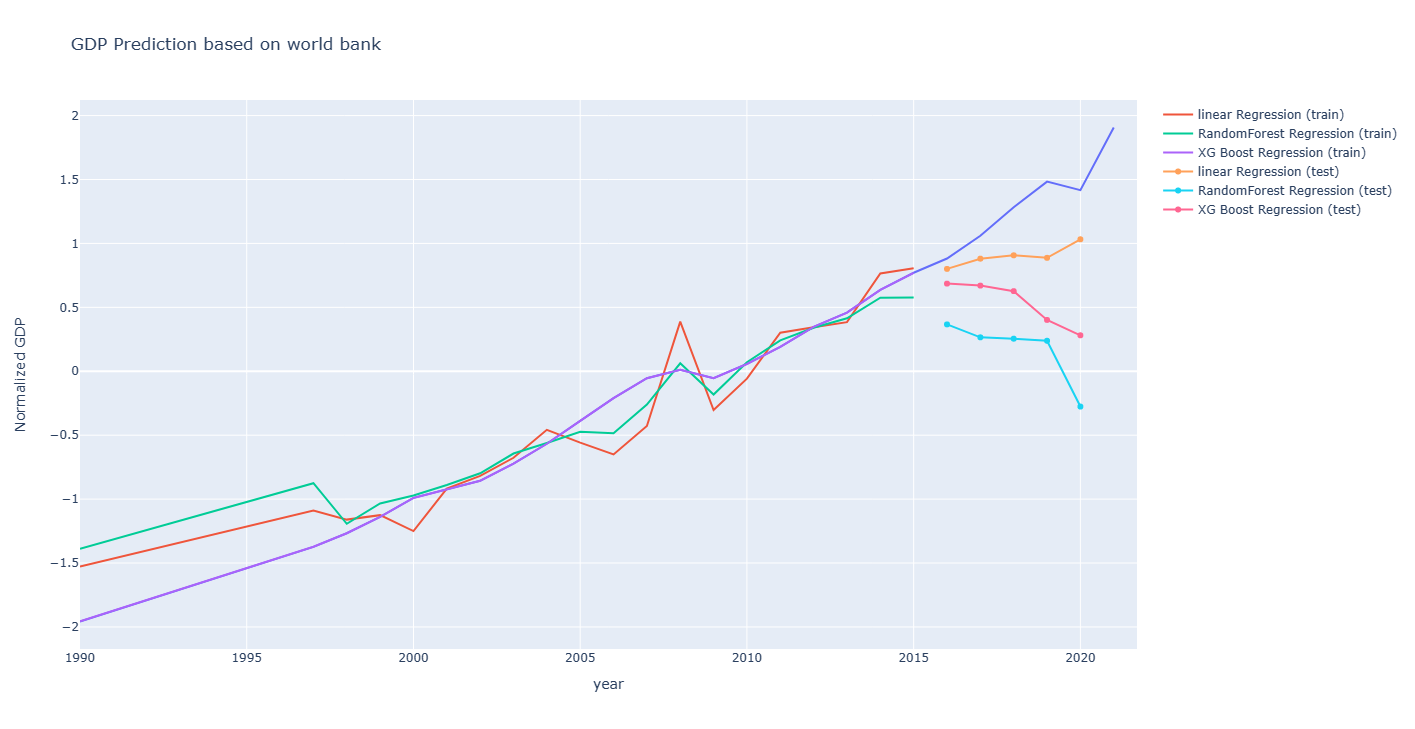
XGBoost fit MSE: 3.7131708546349683e-07

Linear regression n\_MSE: 0.06205238714832588

Random forest MSE: 1.9091957059298448

XGBoost MSE: 2.0300266869267167

Linear regression prediction: 2.0300266869267167



prediction result:

Y\_PREDICT Y\_xgb\_PREDICT Y\_rf\_PREDICT Actual GDP (normalized)

2016 0.800698 0.685643 0.366586 0.881876

2017 0.880749 0.670015 0.265550 1.060189

2018 0.907112 0.626763 0.254851 1.283080

2019 0.887332 0.401534 0.238580 1.483681

2020 1.032350 0.281414 -0.276164 1.416814

Linear regression fit n\_MSE: 0.04872908864363885

Random forest fit MSE: 0.03947574564539995

XGBoost fit MSE: 2.646786878949603e-07

Linear regression n\_MSE: 0.13671717619756904

Random forest MSE: 1.2741367094303135

XGBoost MSE: 0.616334083743745

Linear regression prediction: 0.616334083743745

(Fit: means the training part)

Analysis

Three different regression model:

As the result shows, the prediction performs bad especially for Random forest and XGboost. Insufficient data might be the main reason and cause the simplest method, linear regression model, to surpass others.

Two data set:

Sorry, I don’t find possible reason to explain what makes those two dataset perform differently.