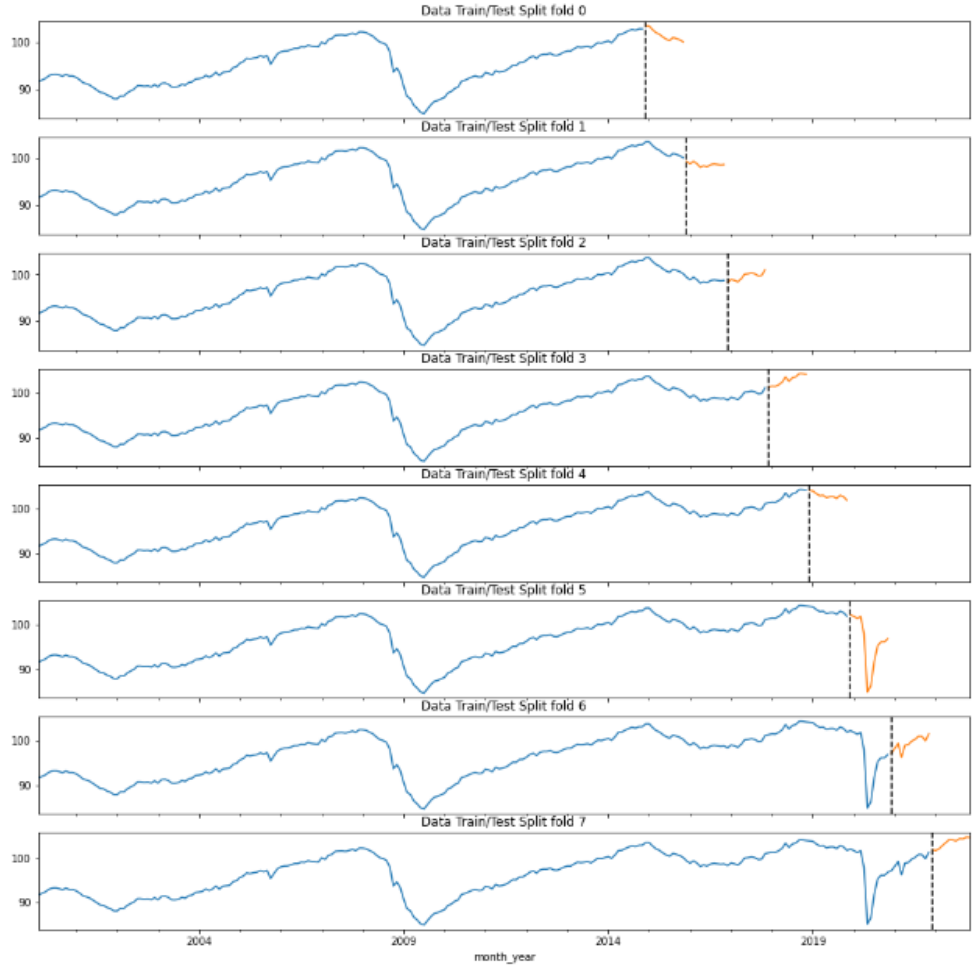
Nowcasting Project- INDPRO

Presentation by **Thien Ly**
January 2022

Training Process

List of models:

- Linear Regression
- Elastic Net
- Lasso
- Support Vector Machine Regression
- Random Forest
- XGBoost
- LightGBM
- Bayesian ARD regression



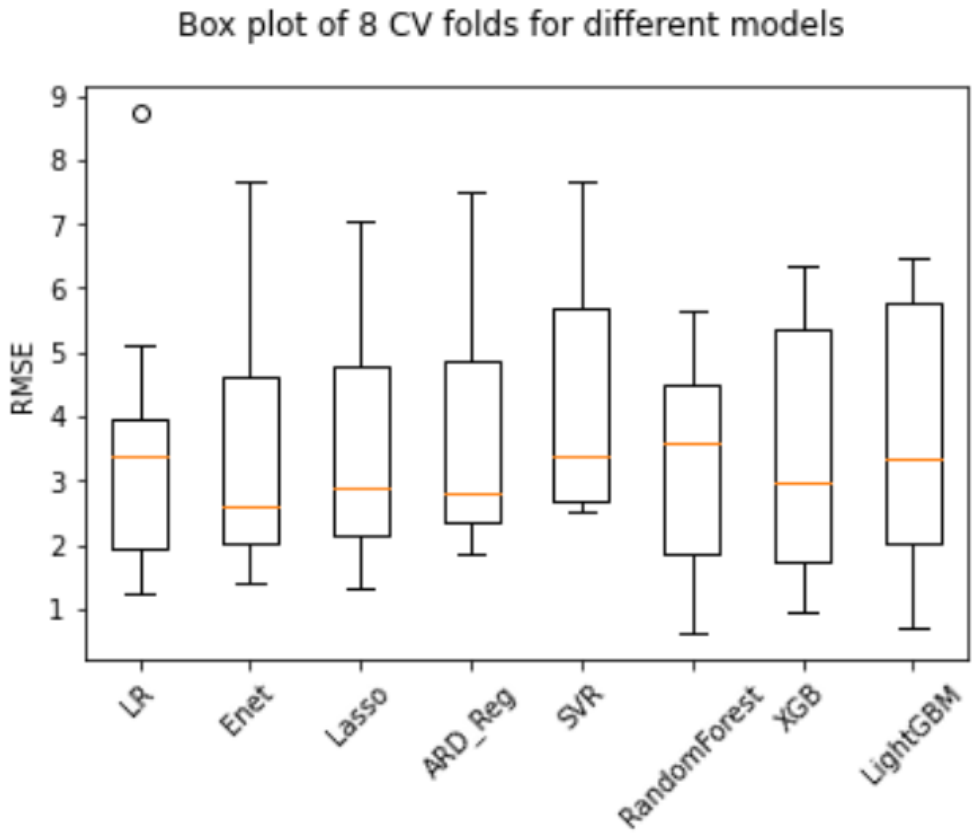
Validation strategy:

8 folds time-series split, each validation set is 12 consecutive months.

Hyperparameters:

Default from each model, second step to make final decision, three best models' hyperparameters are tuned by GridSearchCV.

Validation results

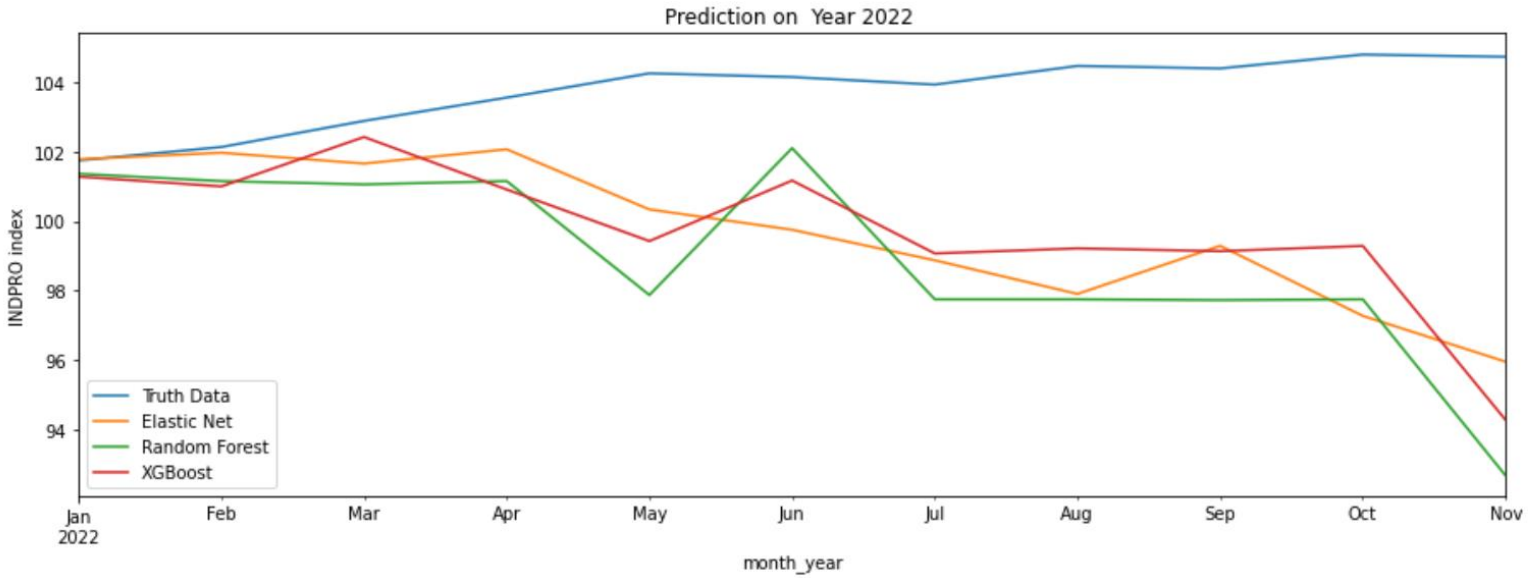


Models	Mean RMSE	Mean MAE
Linear Regression	3.61	3.13
Elastic Net	**3.48**	**2.99**
Lasso	3.55	3.01
Bayesian ARD	3.85	3.31
Support Vector Reg	4.33	3.93
Random Forest	**3.28**	**2.71**
XGBoost	**3.38**	**2.79**
LightGBM	3.65	3.07

Average RMSE and average MAE of 8 folds cross-validation results.

> We choose 3 models: Elastic Net, Random Forest, XGBoost to the next round of hyperparameter tuning and prediction on actual values.

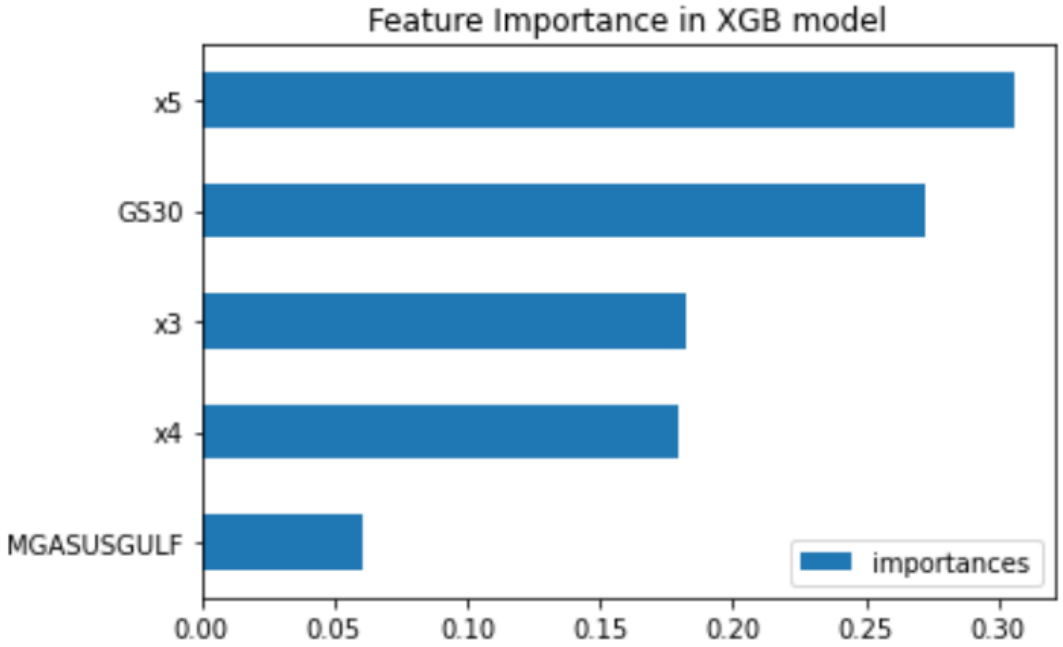
Chosen models results

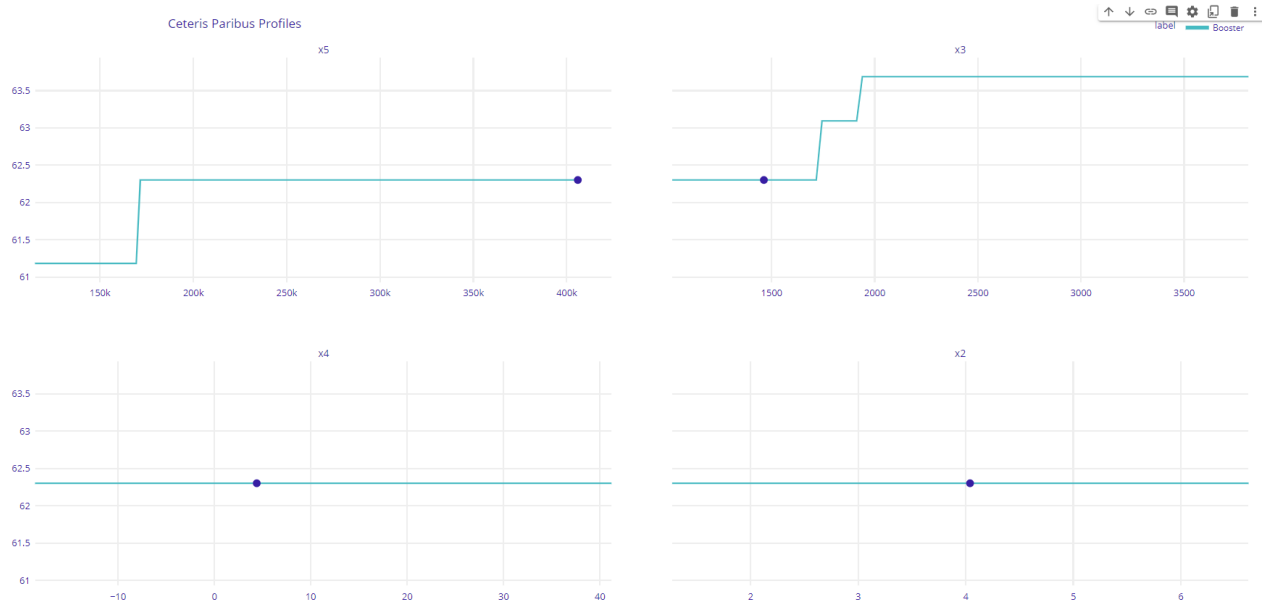


Models	RMSE
Elastic Net	4.934
Random Forest	5.873
XGBoost	4.865

> We choose our model is: XGBoost.

Interpretability





> XGBoost suggests the score for each variable, the higher the score, more useful that feature is in the construction of model.

>Ceteris Paribus method explained the prediction for the final month in case we keep other variables and change value of one left variable.
> Other explainable AI method can be thought of: Shapley Value, Feature Importance Permutation.