Asylum Applications in the EU

Tuesday Dec 5th, 2023

Alexander Portland

Brown University

https://github.com/AlexanderPortland/migr

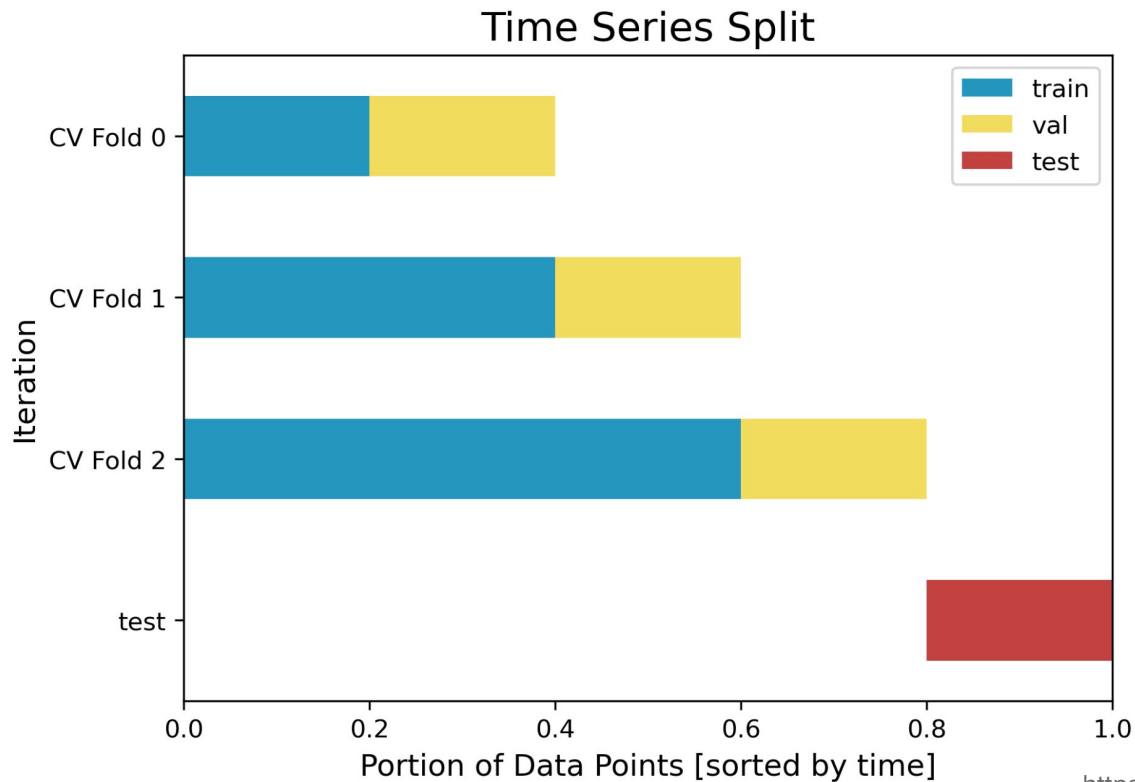# How many people will be granted asylum next year? Why?

(a Regression Question)

Eurostat Asylum Decision Dataset
'Migr_asydcfstq'

- Citizenship

- Sex

- Age Range

- Country of Asylum

- Timing of Application (by quarter)

https://github.com/AlexanderPortland/migr

Cross Validation

CV Strategy

Time Series Split

The Models

Model	Hyperparameters
Ridge	alpha
Ridge (w/ poly features)	alpha, poly_degree
[!] Random Forest	max_depth, max_features, *non deterministic*
LinearSVR	C *(gamma not needed for linear)*
[!] XGBoost	reg_alpha, reg_lambda, *non deterministic*

The Models

Model	Hyperparameters
Ridge	alpha
Ridge (w/ poly features)	alpha, poly_degree
[!] Random Forest	max_depth, max_features, *non deterministic*
LinearSVR	C *(gamma not needed for linear)*
[!] XGBoost	reg_alpha, reg_lambda, *non deterministic*

Lasso

KNN

SVR (rbf)
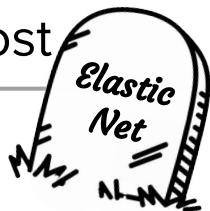
Elastic Net <- Other attempted models

BASELINES

Best RMSE Test Score

Last Year's Value	10.344
Always Zero	25.796
Always Mean of y_Train	25.790

MODELS

XGBoost	9.815	(1.190 std)
LinearSVR	9.003	
RandomForestRegressor	7.479	(0.391 std)
Ridge	8.303	
Ridge (w/ polynomial features)	6.833	

Results

Test Score by Model

BASELINES
- Last Year's Value
- Always Zero
- Always Mean of y_Train

MODELS
- XGBoost
- LinearSVR
- RandomForestRegressor
- Ridge
- Ridge (w/ polynomial features)

BEST BASELINE

RMSE [people]

Results

Interpretation

Linear Model Interpretation

	importance	abs_importance
std__TOTAL_POS - lag 1 quarter	39.104998	39.104998
std__TOTAL_APPS	30.798702	30.798702
std__TOTAL_APPS - lag 1 quarter	-27.984294	27.984294
std__TOTAL_POS - lag 3 quarters	12.156198	12.156198
std__TOTAL_POS - lag 2 quarters	-12.040153	12.040153
std__TOTAL_APPS - lag 2 quarters	7.662631	7.662631
std__TOTAL_POS - lag 4 quarters	-5.107131	5.107131
std__TOTAL_APPS - lag 3 quarters	-5.030410	5.030410
onehot__citizen_SY	1.869084	1.869084
std__TOTAL_APPS - lag 4 quarters	-0.783363	0.783363

Ridge Linear Regression

	importance	abs_importance
std__TOTAL_POS - lag 1 quarter	31.977728	31.977728
std__TOTAL_APPS	30.598162	30.598162
std__TOTAL_APPS - lag 1 quarter	-21.003031	21.003031
std__TOTAL_APPS - lag 2 quarters	-9.452198	9.452198
std__TOTAL_POS - lag 2 quarters	8.204418	8.204418
std__TOTAL_APPS - lag 3 quarters	5.849496	5.849496
std__TOTAL_POS - lag 3 quarters	-5.138919	5.138919
std__TOTAL_APPS - lag 4 quarters	-3.013767	3.013767
std__TOTAL_POS - lag 4 quarters	1.627679	1.627679
onehot__citizen_SY	1.204743	1.204743

Ridge (w/ Polynomial Features)

Linear Model Interpretation

	importance	abs_importance
std__TOTAL_POS - lag 1 quarter	31.977728	31.977728
std__TOTAL_APPS	30.598162	30.598162
std__TOTAL_APPS - lag 1 quarter	-21.003031	21.003031
std__TOTAL_APPS - lag 2 quarters	-9.452198	9.452198
std__TOTAL_POS - lag 2 quarters	8.204418	8.204418
std__TOTAL_APPS - lag 3 quarters	5.849496	5.849496
std__TOTAL_POS - lag 3 quarters	-5.138919	5.138919
std__TOTAL_APPS - lag 4 quarters	-3.013767	3.013767
std__TOTAL_POS - lag 4 quarters	1.627679	1.627679
onehot__citizen_SY	1.204743	1.204743
onehot__citizen_AL	-0.553468	0.553468
onehot__citizen_XK	-0.402028	0.402028
onehot__citizen_PK	-0.325702	0.325702
onehot__citizen_RS	-0.321574	0.321574
onehot__citizen_NG	-0.297984	0.297984
onehot__citizen_AF	-0.269620	0.269620
onehot__citizen_RU	-0.268432	0.268432
onehot__geo_DE	-0.261527	0.261527
onehot__citizen_ER	0.231218	0.231218
onehot__citizen_MK	-0.208984	0.208984

HISTORICAL TRENDS
All values from last year used
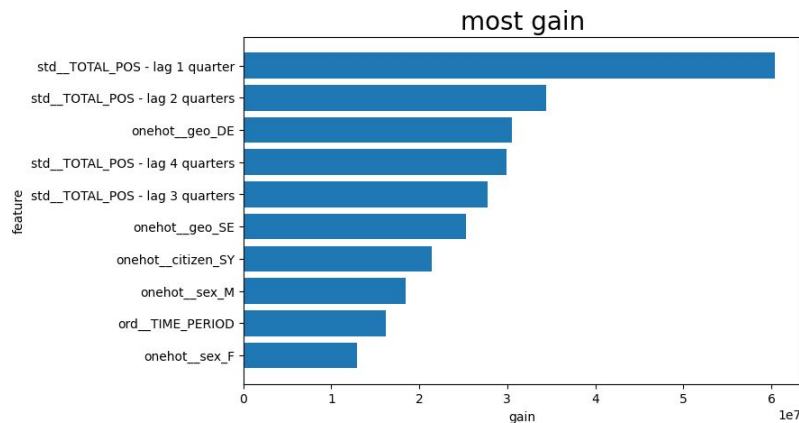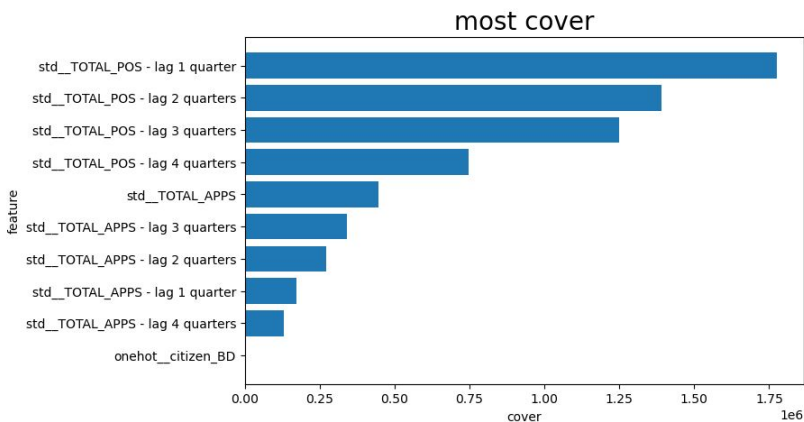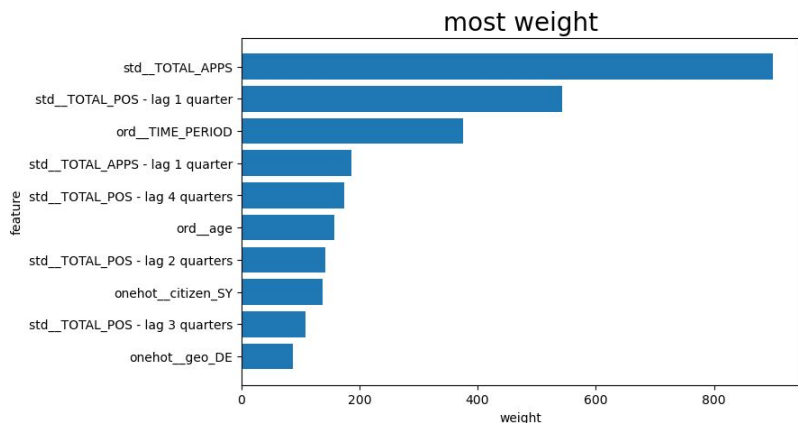Model general global flow of migrants

SPECIFIC INFORMATION
How demographics shape acceptance rates
e.g. Syrian citizens are predicted to have **1.2 more acceptances** each quarter than avg,
While Germany seems to grant **0.26 fewer acceptances** each quarter than average

Ridge (w/ Polynomial Features)

https://github.com/AlexanderPortland/migr

XGBoost Interpretation
With the 5 importance metrics

most weight

most gain

^^shows how much the feature improves the model

most cover

TAKEAWAYS
Age very prevalent (weight)

Used as secondary consideration at nodes?

Sex has bigger role (gain)

https://github.com/AlexanderPortland/migr

XGBoost Interpretation
With mean shap value*

Force Plot (i=31)

Force Plot (i=55)

Force Plot (i=64282)

mean(|SHAP value|) (average impact on model output magr

https://github.com/AlexanderPortland/migr
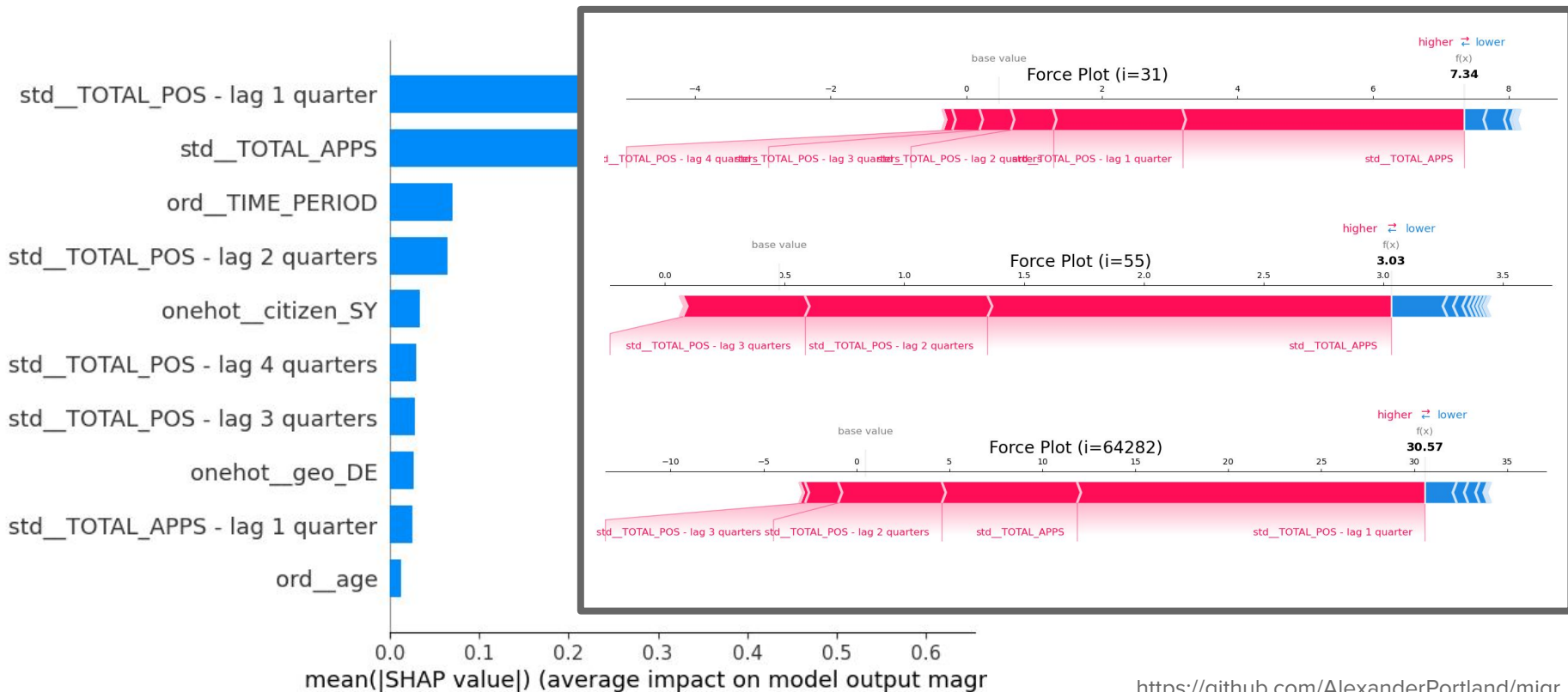
Outlook

TO IMPROVE INTERPRETABILITY…

Build XGBoost full force plots (current shap values only calculated on subset of test for performance)

WITH MORE PREDICTIVE POWER / TIME

Retrain XGBoost with more hyperparameters

Train more Random Forests with more max_layers

More specific location features (categorical -> numeric)

Try non-linear SVR kernels?

Thank You!