STA130 Final Presentation

By: Selina, Polyna, Angel, Flora, Matthew.

Model Performance Metrics: 3 Different Models.

-XGboost

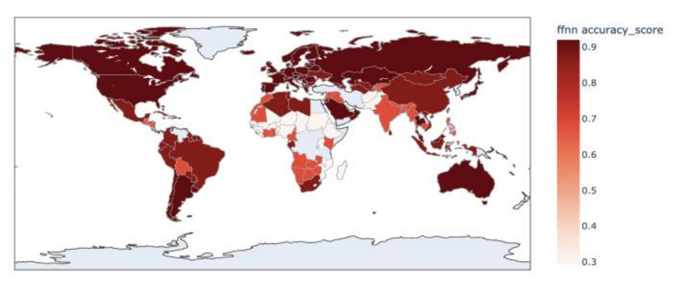
-FFNN

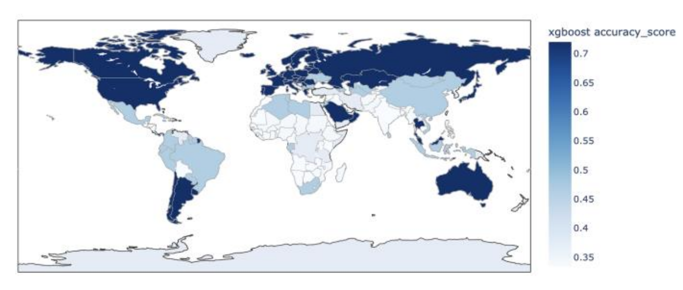
-Transformer

Differences Between the 3 Models

There are 3 models that we are currently using to interpret our data, the FFNN model, XGboost, and the Transformer model. These 3 models each have strengths and weakness when it comes to comparing different indicators using 3 alternative metrics; Accuracy, Sensitivity, and lastly, specificity. The goal of this presentation is to determine which of the 3 models FFNN, XGboost, and transformer, is best for measuring the 3 metrics, Accuracy, measuring how close the results are from their true value. sensitivity, measuring true positives and false negatives. And specificity, measuring true negatives. To do this, we modelled 2 different indicator categories, the human development index and Income then analyzed the results.

Models for HDI Accuracy Scores



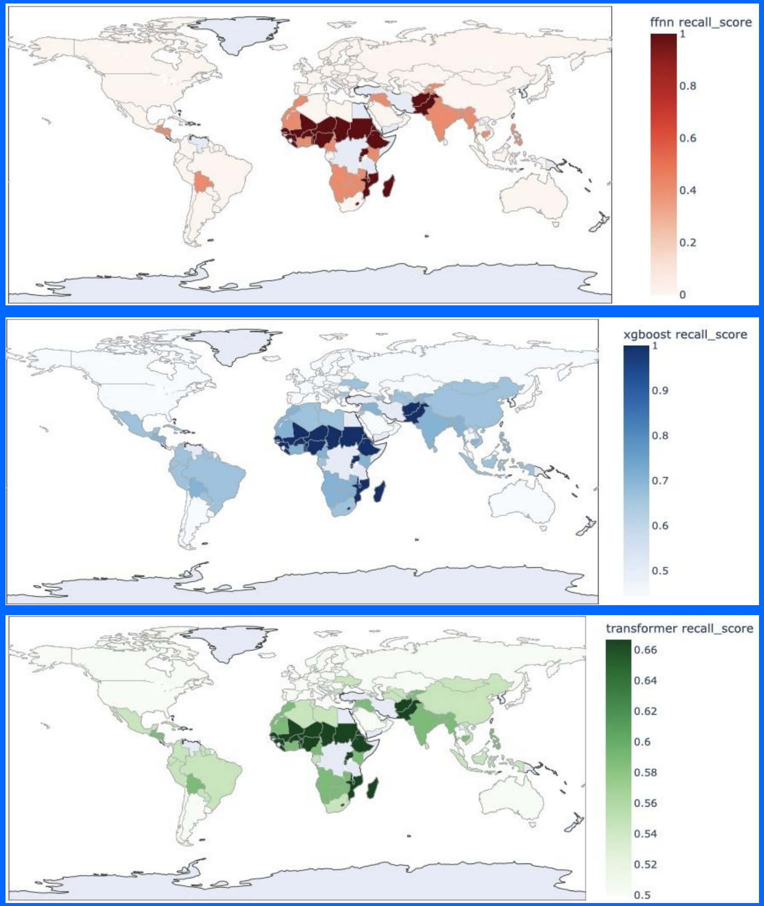


model	hdr_hdicode	accuracy_score
ffnn	High	0.869
xgboost	High	0.464
transformer	High	0.548
ffnn	Low	0.294
xgboost	Low	0.338
transformer	Low	0.529
ffnn	Medium	0.679
xgboost	Medium	0.333
transformer	Medium	0.603
ffnn	Very High	0.920
xgboost	Very High	0.720
transformer	Very High	0.720

Analyzing our Models: Accuracy

- From looking at our maps and tables from the previous slide, this is what we have noticed:
- Maximum accuracy score from FFNN is the highest, with a score of 0.92. while the maximum score for the XGboost and Transformer are tied with a score of 0.72.
- Judging from the increments of the scales used in each model, we find that:
  - FFNN has the greatest range of 0.6
  - XGboost is a close second with 0.35
  - Transformer has the smallest range of 0.15
- Transformer with a medium hdr_hdicode is greater than its high hdr_hdicode, while XGboost with a low, medium hdr_hdicode is roughly the same
- The accuracy is quite spread out throughout the countries, although It seems that countries higher up in the maps have a high accuracy.
- From this analysis, we can assume that FFNN is the best model for determining the accuracy score for the hdi, since it has the highest accuracy score.

Models for HDI Sensitivity Score



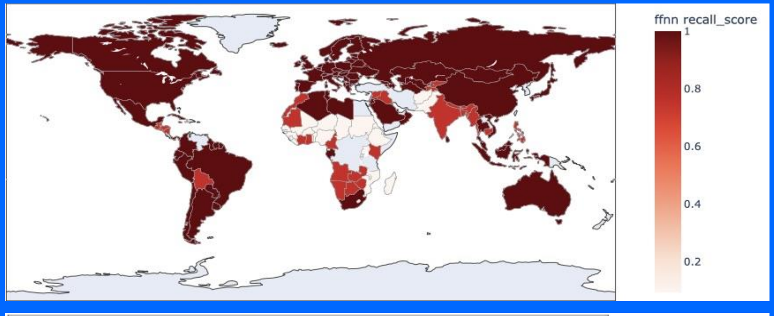
model	hdr_hdicode	recall_score
ffnn	High	0.000
xgboost	High	0.667
transformer	High	0.545
ffnn	Low	1.000
xgboost	Low	1.000
transformer	Low	0.667
ffnn	Medium	0.412
xgboost	Medium	0.706
transformer	Medium	0.588
ffnn	Very High	0.000
xgboost	Very High	0.444

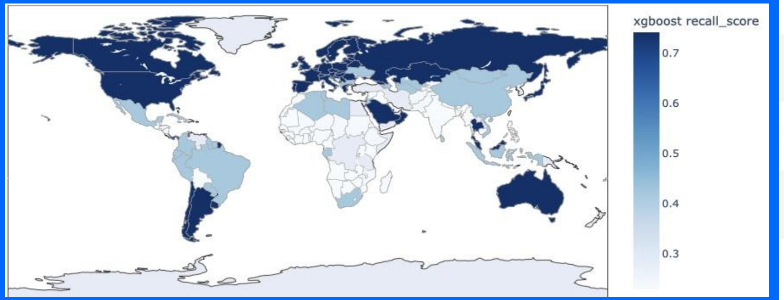
Analyzing our Models: Sensitivity

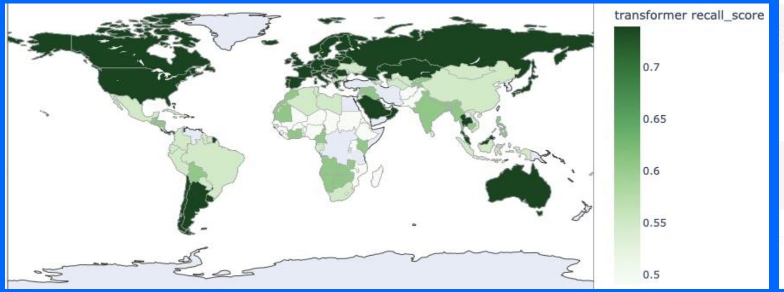
From our analysis, we have noticed:

- The highest sensitivity score for the FFNN and XGboost model is 1, and the Transformer score is 0.667
- In terms of sensitivity in our maps, we find that countries focused on Northern Africa and the Middle East have the highest sensitivity scores.
- From the data table, we have identified that overall, the XGboost mode has the highest score in terms of sensitivity (judging based on the average).
- Finally, we can assume that the XGboost model is the best for measuring the sensitivity for HDI. The XGboost model has the highest sensitivity score, meaning that it is the model that provides the least 'false negative' results. The XGboost model has the least falsely indicated tests for when something is not present, when it is present.

Models for HDI Specificity Score







model	hdr_hdicode	recall_score
ffnn	High	1.000
xgboost	High	0.420
transformer	High	0.548
ffnn	Low	0.094
xgboost	Low	0.237
transformer	Low	0.491
ffnn	Medium	0.754
xgboost	Medium	0.230
transformer	Medium	0.607
ffnn	Very High	1.000
xgboost	Very High	0.741
transformer	Very High	0.739

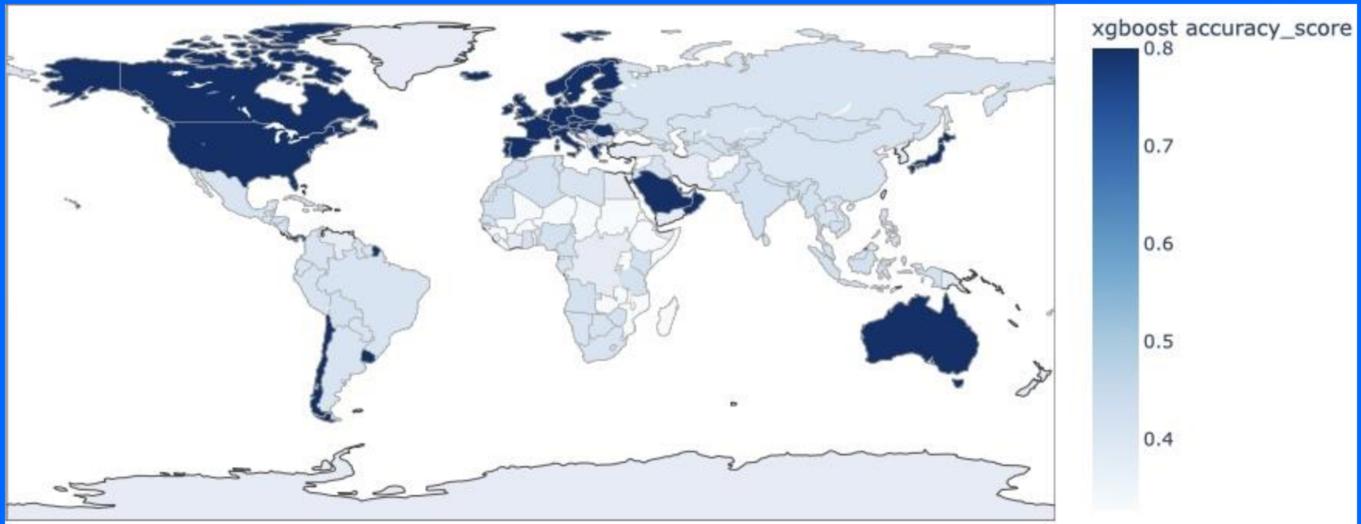
Analyzing our Models: Specificity

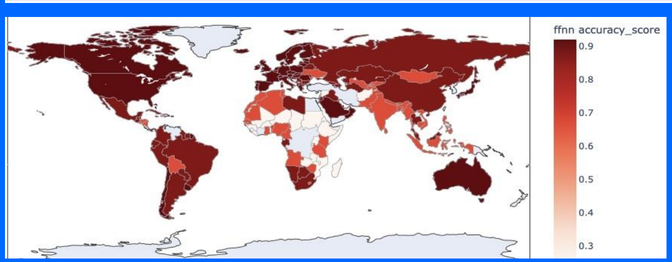
From our analysis, we have noticed:

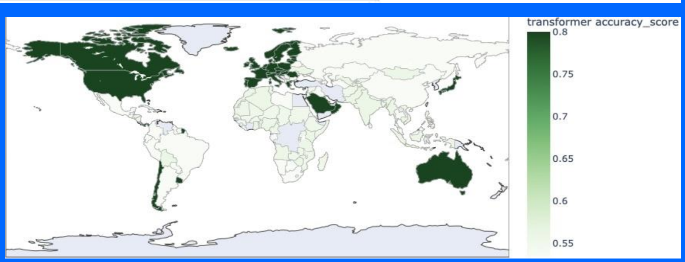
- FFNN model has the highest score for specificity for HDI
- The highest score for XGboost is 0.741, and the maximum score for transformer is 0.739. the 2 scores are very similar with only a 0.002 difference.
- The FFNN model is also the one with the highest range from the scores, with a range of around 0.9, followed by XGboost with a range of 0.5, and Transformer with a range of around 0.2.
- The maps of specificity for the three models are similar to the maps of accuracy, with the countries higher up and in the corners of the map having the highest specificity scores.

From the maps and table, we have noticed that the sensitivity and specificity could be inversely related. If one of the models have a high sensitivity score, it seems that the specificity score is lower. Which makes sense, since we know that specificity tests for any 'true negatives' while sensitivity tests for 'false negatives' in a set of data. We have assumed that the FFNN is the best for measuring the specificity, as it has the highest score.

Models for Income Group: Accuracy Scores







model	wbi_income_group	accuracy_score
ffnn	High income	0.920
xgboost	High income	0.800
transformer	High income	0.800
ffnn	Low income	0.259
xgboost	Low income	0.328
transformer	Low income	0.552
ffnn	Lower middle income	0.667
xgboost	Lower middle income	0.422
transformer	Lower middle income	0.559
ffnn	Upper middle income	0.874
xgboost	Upper middle income	0.408
transformer	Upper middle income	0.534

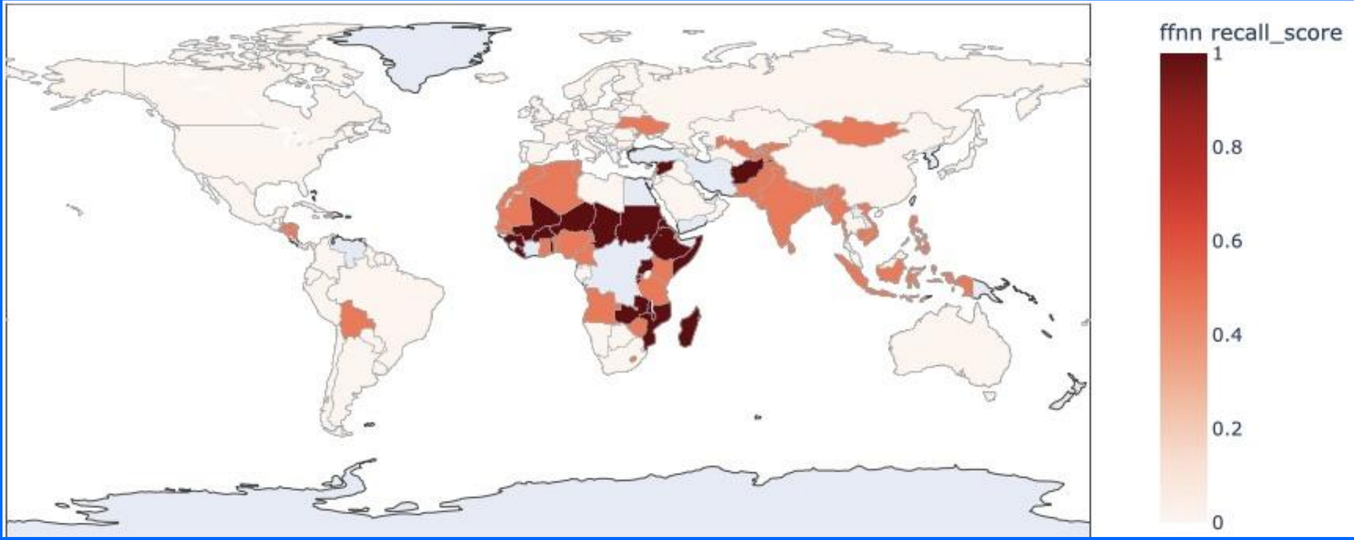
Analyzing our Models: Accuracy

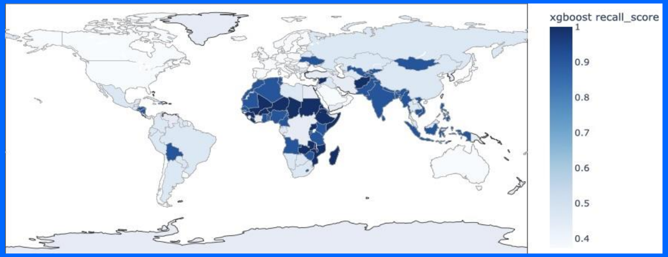
From this data we have determined that depending on the income group, the accuracy varies.

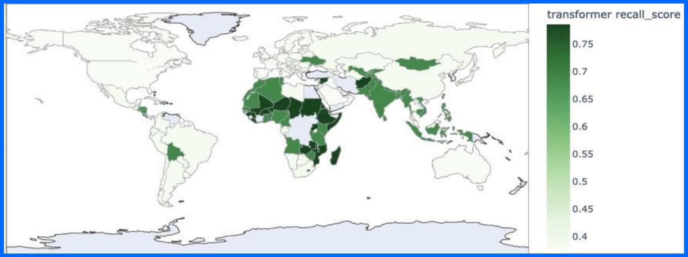
- In the low-income group, Transformer has the highest accuracy of 0.552.
- Lower middle-income group, FFNN has the highest score of 0.667.
- Upper middle-income, FFNN has the highest score of 0.874
- High-income, FFNN has the highest score of 0.920.

From the maps we find that the XGboost and Transformer models look very similar, with the same countries having the highest accuracy scores. With the FFNN model the accuracy scores are very spread out, and generally more leveled. There are a lot more countries with higher accuracy scores compared to XGboost and Transformer. We assumed that FFNN is the best model for accuracy in terms of income groups, since FFNN has the highest accuracy score out of 3 of the 4 different income groups.

Models for Income Group: Sensitivity Scores







model	wbi_income_group	recall_score
ffnn	High income	0.000
xgboost	High income	0.375
transformer	High income	0.375
ffnn	Low income	1.000
xgboost	Low income	1.000
transformer	Low income	0.786
ffnn	Lower middle income	0.474
xgboost	Lower middle income	0.909
transformer	Lower middle income	0.684
ffnn	Upper middle income	0.000
xgboost	Upper middle income	0.467
transformer	Upper middle income	0.385

Analyzing our Models: Sensitivity

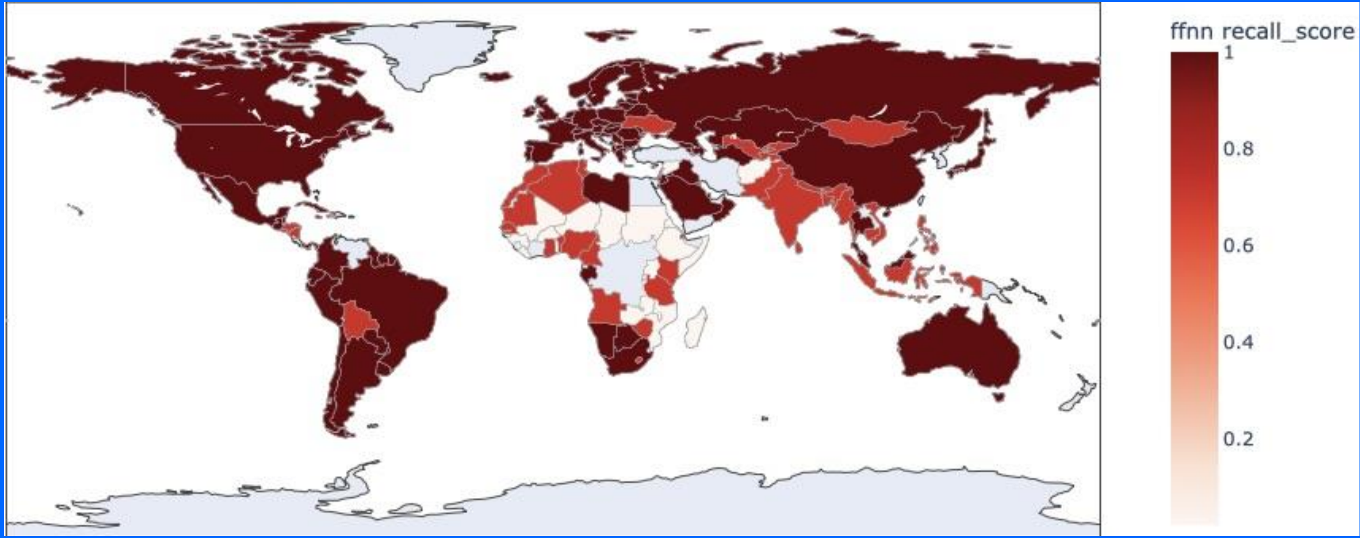
For determining sensitivity, the scores between each model varies depending on the income group.

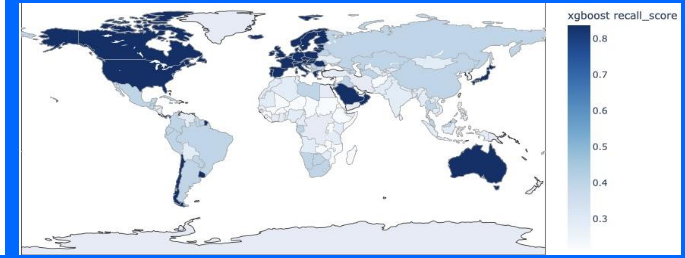
- ❖ For the low-income group in our data, XGboost and FFNN ties between the highest score for sensitivity with both their scores being 1.000. while Transformer has a score of 0.786.
- ❖ The lower middle-income group, XGboost has the highest sensitivity score of 0.909.
- ❖ Upper middle income, the highest is XGboost with a score of 0.467.
- ❖ High income group, XGboost and Transformer ties at being the highest. Both with a score of 0.375, while FFNN is 0.000.

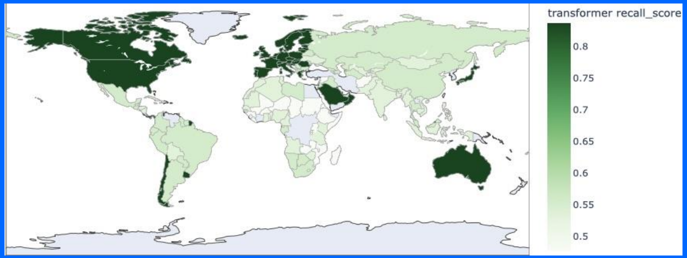
After analyzing the maps, we have found that all 3 maps look very similar, with the highest sensitivity scores focused on Africa. Lastly, we have determined that XGboost is the model for determining sensitivity. As it has the highest sensitivity score for all the income groups, even after tying with some of the other models.

Models for Income Group: Specificity Scores

model	wbi_income_group	recall_score
ffnn	High income	1.000
xgboost	High income	0.837
transformer	High income	0.837
ffnn	Low income	0.023
xgboost	Low income	0.220
transformer	Low income	0.477
ffnn	Lower middle income	0.711
xgboost	Lower middle income	0.287
transformer	Lower middle income	0.530
ffnn	Upper middle income	1.000
xgboost	Upper middle income	0.398
transformer	Upper middle income	0.556







Analyzing our Models: Specificity

Using the table to determine our specificity, we can see that our score between each model varies depending on the income group.

- In the lower-income group, the model with the highest specificity score is the Transformer model, with a score of 0.477.
- Lower middle-income group, FFNN has the highest specificity score of 0.711.
- Upper middle-income group, FFNN has the highest specificity score of 1.000.
- In the high-income group, FFNN once again has the highest specificity. With a score of 1.000.

From our analysis of the maps, we find that the XGboost and Transformer models have very similar maps. The highest specificity scores are usually focused in more "Eurocentric" places like Europe, North America and Australia. However, Chilly and a small region of Africa also have high specificity score. The FFNN map is very different from the 2 other models. With no specific focus on certain countries. The FFNN map is lot more "evened" out through the map.

We have previously assumed that sensitivity and specificity have an inverse relationship. And we can see that this also the case for the income group models. With models scoring high in sensitivity having low scores in specificity. And from this Data we have determined that FFNN is the best model for measuring Specificity, as it scored the highest in terms of specificity for 3 out of the 4 different income groups.

Making Assumptions

Following the analysis from the maps and models, we will now assume and hypothesize that determining a model that is best is quite difficult, since each model are best at measuring different metrics; sensitivity, accuracy, and specificity, depending on the category we want to measure.

To further prove our hypothesis of "the indicators an observation belongs to does matter", we did a linear regression model and hypothesis testing.

Proving our Assumption from Data

OLS Regression Results

Dep. Variable:	error	**R-squared:**	0.294
Model:	OLS	**Adj. R-squared:**	0.285
Method:	Least Squares	**F-statistic:**	34.51
Date:	Thu, 07 Dec 2023	**Prob (F-statistic):**	1.60e-72
Time:	00:20:17	**Log-Likelihood:**	523.74
No. Observations:	1092	**AIC:**	-1019.
Df Residuals:	1078	**BIC:**	-949.5
Df Model:	13		
Covariance Type:	nonrobust		

Omnibus:	3.204	**Durbin-Watson:**	1.421
Prob(Omnibus):	0.201	**Jarque-Bera (JB):**	3.520
Skew:	0.027	**Prob(JB):**	0.172
Kurtosis:	3.273	**Cond. No.**	12.6

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.3831	0.008	48.403	0.000	0.368	0.399
predicts1	0.2348	0.028	8.501	0.000	0.181	0.289
wbi_income_group_Low income	0.1049	0.019	5.453	0.000	0.067	0.143
wbi_income_group_Low income X predicts1	-0.2070	0.036	-5.780	0.000	-0.277	-0.137
wbi_income_group_Lower middle income X predicts1	-0.1848	0.031	-5.986	0.000	-0.245	-0.124
xgboost	-0.1183	0.016	-7.284	0.000	-0.150	-0.086
fsi_category_Alert X xgboost	0.2304	0.026	8.702	0.000	0.178	0.282
fsi_category_Sustainable X transformer	-0.1140	0.027	-4.276	0.000	-0.166	-0.062
fsi_category_Warning X xgboost	0.1913	0.020	9.544	0.000	0.152	0.231
hdr_region_LAC X xgboost	0.1279	0.021	5.953	0.000	0.086	0.170
wbi_income_group_Lower middle income X transformer	0.0962	0.017	5.701	0.000	0.063	0.129
wbi_income_group_Lower middle income X xgboost	0.0832	0.021	4.027	0.000	0.043	0.124
wbi_income_group_Upper middle income X transformer	0.0543	0.017	3.204	0.001	0.021	0.088
xgboost X predicts1	-0.2336	0.028	-8.225	0.000	-0.289	-0.178

Proving our Assumption from Data

After testing different combinations, we concluded on the previous model with indicators as such. Since our goal was to have a model with little correlation so that we have less multicollinearity, our condition number of 12.6 is great.

Now focusing on the null hypothesis, which states that the independent variable has no association with the dependent variable. Our null hypothesis of "the indicators an observation belongs to does not matter" is tested with the p-values from the model. For all our indicators, the maximum p-value we got was 0.001, which means overall, we have strong evidence against the null hypothesis.

The conclusion would be that the null hypothesis is incorrect and the categories an observation belongs to does matter when determining which metric works best. This also ties into the proof of our hypothesis.

Ethical Consideration in “Conflict Escalation Predictions”

	Predict conflict	Predict no conflict
Conflict occur	(TP, true positive) Prepared for conflict	(FN, false negative) Unprepared for conflict
Conflict didn’t occur	(FP, false positive) Waste of prepared resources	(TN, True negative) No waste of resources

TP: The country would be prepared to distribute resources and deal with the conflict
TN: The country wouldn’t waste resources to prepare for a potential conflict
FP: The country would waste resources to prepare for a non-existent conflict
FN: The country would be unprepared to deal with the conflict and aid its citizens during this time

Conclusion

In conclusion, The null hypothesis is proven to be false, since it is evident that all 3 models have their own strengths and weaknesses depending on the category they are measuring. All three models serve as tools for forecasting the escalation of conflicts. The first graph utilizes a Feed Forward Neural Network. The second graph employs the XGBoost algorithm (Extreme Gradient Boosting). Lastly, the third graph uses a transformer model. It is remarkable that all three methods mentioned above are neural networks, which use text data alone to identify potential conflict zones. Despite using different data sampling methods and modeling approaches, these neural network-based graphs show remarkably similar results.

Upon thorough analysis of the provided information, it becomes evident that these neural network-based graphs appeared to be great at analyzing geographical escalations. XGBoost, FFNN and transformer can potentially be adapted for forecasting conflict escalation as a binary classification task, their direct comparison in terms of sensitivity and specificity might be challenging. However, despite being similar, none of these graphs can be used as the best model for conflict escalation prediction. These results can be used by researchers on this topic in the future.