International Debt Statistics and Projection

Introduction

Our aim is to predicts how much debt in terms of percentage of GDP a country can pay back on its external, international loans from the World Bank or IMF. We have incorporated the country's historical debt statistics and various socio-economic factors to assess its ability to repay external loans. The end goal of our model is to identify the root causes of economic instability affecting a country, which, in turn, affects its ability to repay a loan.

The insights from this model will enable organizations like the World Bank to work with countries to mitigate loan defaulting by sending monetary aid, choosing which projects to fund, etc. These insights could be used by government organizations to make better decisions regarding external borrowing and allow them to monitor their ongoing debt. Our model can also be used by NGOs to help them target their interventions by addressing the root cause of the problem.

Data

Country	Year	Debt Percent of GDP	ANER; Female	Above Proficiency;PISA 2000 for grade 15Y using MPL Level 2 for math	Above Proficiency;TIMSS 2003 for grade 4 using MPL Low (400 points) for math	Above Proficiency;TIMSS 2007 for grade 4 using MPL Low (400 points) for math	Above Proficiency;TIMSS 2019 for grade 4 using MPL Low (400 points) for math	Adjusted net attendance rate, one year before the official primary entry age, adjusted gender parity index (GPIA)	...	Hospital beds (per 1,000 people)	Num neon de
Belgium	2015	88.82606755	..	..	..	..	..	..	..	5.83	
Belgium	2016	89.15005173	3.776007	..	..	..	..	..	..	5.76	
Belgium	2017	86.98393439	..	..	..	..	..	..	..	5.66	
Belgium	2018	84.9649278	..	..	..	..	..	..	..	5.62	
Belgium	2019	82.8536807	..	..	..	..	..	..	..	5.58	
...	...	...	...	...	...	...	...	...	...	...	
Zimbabwe	2018	51.00144461	..	..	..	..	..	..	..	..	12
Zimbabwe	2019	82.33805679	..	..	..	..	..	0.9735	..	..	12
Zimbabwe	2020	84.44771597	..	..	..	..	..	..	..	..	12
Zimbabwe	2021	59.80656696	..	..	..	..	..	..	..	..	12
Zimbabwe	2022	92.8237321	..	..	..	..	..	..	..	..	

Hospital beds (per 1,000 people)	Number of neonatal deaths	Physicians (per 1,000 people)	Population ages 15-19, male (% of male population)	Population ages 35-39, male	Population ages 55-59, male (% of male population)	Population ages 65 and above, male (% of male population)	Rural population growth (annual %)	School enrollment, secondary, female (% gross)	Sex ratio at birth (male births per female births)
0.170626998	NA	0.035	10.252415	261456	2.666401	2.949109	NA	NA	1.049
NA	NA	NA	10.291583	266698	2.633083	2.921745	1.616510333	NA	1.049
NA	NA	NA	10.310921	272509	2.592152	2.889481	1.694596635	NA	1.049
NA	NA	NA	10.263812	278443	2.548916	2.855172	1.748878535	NA	1.049
NA	NA	NA	10.162685	283951	2.512455	2.821868	1.797359522	NA	1.049
...	...	...	...	...	...	...	...	...	...
NA	12874	0.201	11.836914	429834	1.591278	2.877222	2.061844399	NA	1.023
NA	12638	NA	11.782546	450009	1.567252	2.902687	1.987780433	NA	1.023
NA	12421	0.189	11.662186	462417	1.555282	2.902155	1.983890973	NA	1.024
NA	12211	NA	11.512926	463351	1.549076	2.856109	1.955651308	NA	1.024
NA	NA	NA	11.378125	457281	1.554298	2.783890	1.888043327	NA	NA

The source for our dataset is mainly the world bank and U.S. Department of the Treasury.

This dataset contains around 172 countries from 1960s to 2022. It has 114 columns and 11K rows in total. The columns comprises a wide array of indicators spanning multiple socio-economic, educational, health, labor, demographic, and miscellaneous domains.

Methods

XGBoost

	Importance
Number of neonatal deaths	0.042312
Adolescent fertility rate (births per 1,000 women ages 15-19)	0.038844
Age dependency ratio, old	0.034451
Gross enrolment ratio, primary, gender parity index (GPI)	0.033064
Age dependency ratio (% of working-age population)_y	0.030751
Rural population growth (annual %)	0.029595
Sex ratio at birth (male births per female births)	0.026358
Population ages 15-19, male (% of male population)	0.023815

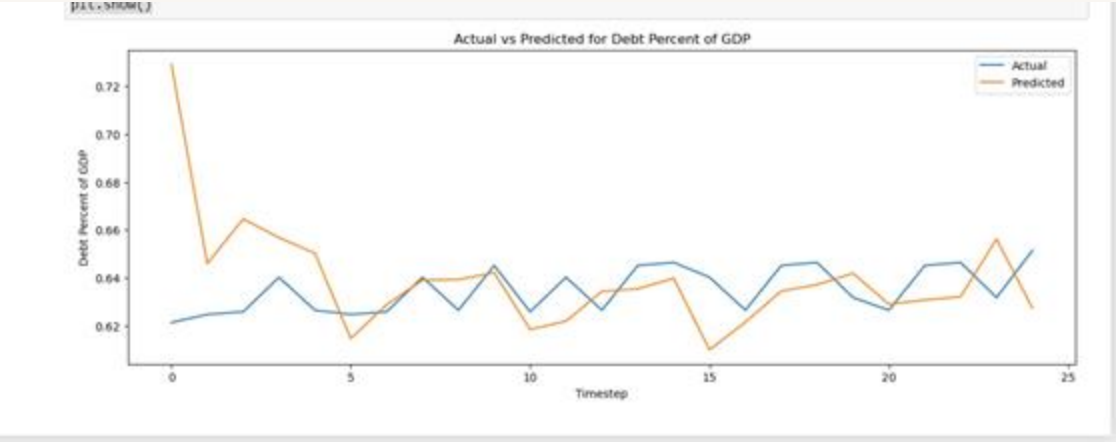
VAR

	AIC	BIC	FPE	HQIC
0	49.58	49.93	3.416e+21	49.72
1	-2.737	1.136*	0.06768	-1.225
2	-4.259	3.136	0.01958	-1.372
3	-6.514	4.401	0.004763	-2.253
4	-10.32*	4.113	0.0008466*	-4.688*

LSTM

- Masking
- Rolling Window
- Custom Error Function

Current results



0.003

Mean Absolute Error

Work remaining

Currently we have some work that we need to do to make our model better than what it is now

1. Reshape Time-Series data
2. Test a custom cube loss function
3. Try dealing with the lag by adding a lag column

Thank You!


HOW'S THE BIG DATA PROJECT COMING ALONG, HOSKINS?
© D.Fletcher for CloudTweaks.com