MACHINE LEARNING FOR DATA SCIENCE

PROJECT REPORT

SUBMITTED BY:

Kashaf Sajjad, Aieza Noor, Aqdas Bibi, Muhammad Shahbaz

ROLL-NO:

I24-8032, I24-8021, I24-8048, I24-8029

COURSE INSTRUCTOR:

Dr. Muhammad Ishtiaq

DATE:

June 12, 2025

ANALYZING WORLD DEVELOPMENT INDICATORS TO FIND POLICY ISSUES AND PREDICT POLICY CHANGE IMPACT

INTRODUCTION

INTRODUCTION

The World consists of developed and developing countries. As per World Data information website 152 countries (“List of 152 developing countries”) are categorized as developing countries. That represents 85.59% of the world’s population which is a significant number. This means only 14.41% of the population have a higher standard of living. A significant ratio of the world population is struggling to make ends meet. Countries are governed by their leadership using a set of policies. Policies have their impact, both, theoretically and practically. Sometimes the effect of policy changes on development indicators are immediate, sometimes they are long term. We can analyze the policies of the countries in order to understand the relationship between a policy measure and its impact on the development of that country. In this project we wish to study policies and their impacts on economies.

OBJECTIVES

Following are the objectives of our project:

- Detect changes in policy measures e.g. change in interest rate, increase or decrease in education or healthcare budget, change in policy of skill development etc.
- Analyze the change in policy measures and their impact on relevant development indicators e.g. GDP per capita, % of children enrolled in primary schools, primary literacy rate, adult literacy rate etc.
- Label the data with the type of change in policy measures and its impact.
- Train time series based model on train set and evaluate it on test set.

TARGETS

- Algorithm Selection: Determine the best deep learning algorithm for World Bank time series prediction
- Model Training: Train models using data from 1960-2010
- Model Evaluation: Test performance on unseen data from 2011-2024
- Performance Analysis: Compare predictions with actual indicators
- Visualization: Create comprehensive plots showing model performance

OVERVIEW OF DATASET

We used the World Development Indicator (“World Development Indicators | DataBank”) dataset from the World Bank Open Data repository. It provides socio-economic and development indicators across various countries, covering the period from 1960 to 2024. Each record represents a specific indicator for a country, with annual values spanning more than six decades.

Structure of Dataset

The dataset includes the following key columns:

- Country Name: Name of the country (e.g., Afghanistan).
- Country Code: A three-letter country code (e.g., AFG).
- Series Name: Description of the economic or development indicator (e.g., Access to electricity (% of population)).
- Series Code: Unique code assigned to each indicator (e.g., EG.ELC.ACCS.ZS).
- Yearly Columns (1960–2024): Annual data values corresponding to the specific indicator for each year.

Each row in the dataset corresponds to a specific country and development indicator combination.

Key Characteristics

- Time Coverage: From year 1960 to 2024.
- Geographic Scope: Global with individual country level entries.
- Indicators: Includes both economic and infrastructural development metrics like GDP per capita growth, access to electricity, agricultural land etc.
- Data Quality: Many fields for earlier years especially 1960s–1980s contain missing or unavailable values which are represented as “..” .

First 15 rows of dataset are shown:

1	Country N	Country C	Series Nai	Series Coc	1960 [YR19	1961 [YR19	1962 [YR19	1963 [YR19	1964 [YR19	1965 [YR19	1966 [YR19	1967 [YR19	1968 [YR19	1969 [YR19	1970 [YR19	1971 [YR19	1972 [YR19	1973 [YR19	1974 [YR19	1975 [YR19	1976 [YR19	1977 [YR19	1978 [YR19	1979 [YR19	19
2	Afghanist	AFG	Access to	EG.ELC.AC	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
3	Afghanist	AFG	Access to	EG.ELC.AC	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
4	Afghanist	AFG	Access to	EG.ELC.AC	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
5	Afghanist	AFG	Adjusted	NY.ADJ.NI	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
6	Afghanist	AFG	Adjusted	NY.ADJ.NI	..	..	..	..	..	..	..	..	..	..	1.67E+09	1.74E+09	1.52E+09	1.64E+09	2.04E+09	2.22E+09	2.39E+09	2.77E+09	3.1E+09	3.47E+09	
7	Afghanist	AFG	Adjusted	NY.ADJ.NI	..	..	..	..	..	..	..	..	..	..	147.7092	150.773	128.2965	135.1514	163.37	173.5771	183.1253	207.427	227.6383	253.8806	2
8	Afghanist	AFG	GDP grow	NY.GDP.M	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
9	Afghanist	AFG	GDP per c	NY.GDP.P	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
10	Afghanist	AFG	Agricultur	AG.LND.A	..	57.87836	57.95502	58.03168	58.116	58.12367	58.19266	58.22946	58.23099	58.25552	58.27086	58.31685	58.33525	58.33709	58.33878	58.33878	58.33832	58.33832	58.33832	58.33678	5
11	Afghanist	AFG	Agricultur	AG.LND.IR	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
12	Afghanist	AFG	Agricultur	TX.VAL.AG	..	..	53.75585	61.10061	54.43718	45.7609	51.71498	46.35573	38.09209	38.0508	35.824	45.65635	33.93794	26.89305	31.47573	33.67265	39.39018	36.83645	..	..	..
13	Afghanist	AFG	Agricultur	TM.VAL.A	..	..	0.966617	1.394438	1.498234	1.080156	0.672909	0.82761	1.035163	1.464899	1.460332	0.982216	1.763667	3.473725	3.11282	1.545044	1.752994	1.311936	..	..	..
14	Afghanist	AFG	Asylum-se	SM.POP.A	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
15	Afghanist	AFG	Asylum-se	SM.POP.A	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..

1	2001 [YR20	2002 [YR20	2003 [YR20	2004 [YR20	2005 [YR20	2006 [YR20	2007 [YR20	2008 [YR20	2009 [YR20	2010 [YR20	2011 [YR20	2012 [YR20	2013 [YR20	2014 [YR20	2015 [YR20	2016 [YR20	2017 [YR20	2018 [YR20	2019 [YR20	2020 [YR20	2021 [YR20	2022 [YR20	2023 [YR20	2024 [YR2024
2	9.3	14.1	19	23.8	28.7	33.5	38.4	42.4	48.3	42.7	43.2	69.1	68	89.5	71.5	97.7	97.7	93.4	97.7	97.7	97.7	85.3	85.3	..
3	..	..	2.1	7.8	15.4	19.3	25	28	36.7	30.2	29.6	60.8	60.2	86.5	64.6	97.1	97.1	91.6	97.1	97.1	97.1	81.7	81.4	..
4	74.8	76.1	77.5	78.8	74	81.6	83	89.9	85.9	82.8	86.6	95	92.2	98.7	92.5	99.5	99.5	98.8	99.5	99.5	99.5	95.9	96	..
5	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	-19.5057	..	..	..
6	..	..	..	..	..	..	..	..	1.11E+10	1.43E+10	1.66E+10	1.85E+10	1.89E+10	1.89E+10	1.86E+10	1.68E+10	1.78E+10	1.71E+10	1.77E+10	1.86E+10	1.36E+10	..	..	..
7	..	..	..	..	..	..	..	..	403.4112	505.2302	566.6681	605.752	597.3562	575.9168	548.9685	484.8906	497.406	466.5706	468.4035	475.7181	340.5232	..	..	..
8	-9.43197	28.6	8.832278	1.414118	11.22971	5.357403	13.82632	3.924984	21.39053	14.36244	0.426355	12.75229	5.600745	2.724543	1.451315	2.260314	2.647003	1.189228	3.911603	-2.3511	-20.7388	-6.24017	2.710887	..
9	-10.1195	22.02019	2.345672	-2.14821	7.383377	1.132485	11.6923	1.677279	17.0439	11.05503	-3.2133	8.279369	2.052068	-0.93998	-1.66506	-0.30012	-0.19557	-1.71374	0.856295	-5.38251	-22.5845	-7.57667	0.540656	..
10	57.94735	57.93968	58.0838	58.15127	58.1344	58.12367	58.1298	58.13287	58.13287	58.1344	58.13133	58.1298	58.12367	58.12367	58.12367	58.12367	58.12367	58.27699	58.27699	58.74155	58.74155	58.74155	..	..
11	5.662125	4.617624	7.26428	5.499895	5.83907	6.085466	5.939758	5.778563	4.842283	5.000396	5.391006	5.465	5.518333	5.742548	5.710894	6.48114	5.990504	5.122336	6.006314	6.50693	6.50693	..	..	..
12	..	..	..	..	..	..	..	9.894829	7.559673	10.76705	12.87901	0.236604	0.036992	0.047523	14.79907	16.35911	17.06363	16.6946	21.03506	..	..	..	..	..
13	..	..	..	..	..	..	..	1.476686	..	..	..	..	..	..	..	..	1.871025	1.821349	1.76946	..	..	..	..	..
14	30782	26687	26308	16983	14023	14959	16077	23177	30406	37101	39199	62304	75284	85407	258862	369072	333986	310107	251047	238791	262860	294493	296033	..
15	0	18	71	28	12	5	0	10	14	30	52	49	61	59	78	123	215	281	247	167	251	218	250	..

METHODOLOGY STEPS

Here is the methodology along with some details about the algorithms to be used.

- Split the data into train and test sets. We can include data from 1960-2010 as the train set and data from 2010+ as the test set.

- Tracking policy change, analyzing and noting them down. For this we use Change-Point Detection and Rolling Statistics methods.
- Tracking relationship between a policy change and a change in development indicator.
- Label the change in development indicator along with policy change.
- Train LSTM and TCN models on the train set.
- Evaluate the trained model on a test set and compute performance metrics.
- Visualize the training and test performance metrics to understand them better.

EXPLORATORY DATA ANALYSIS (EDA) AND PREPROCESSING

EDA STEPS

- We uploaded the world_bank.csv file to Google Colab, read it using ISO-8859-1 encoding into a DataFrame and displayed the first few rows of historical World Bank data.

Country Name	Country Code	Series Name	Series Code	1960 [YR1960]	1961 [YR1961]	1962 [YR1962]	1963 [YR1963]	1964 [YR1964]	1965 [YR1965]	...	2015 [YR2015]	2016 [YR2016]	2017 [YR2017]	2018 [YR2018]	2019 [YR2019]	2020 [YR2020]	2021 [YR2021]	2022 [YR2022]	2023 [YR2023]
Afghanistan	AFG	Access to electricity (% of population)	EG.ELC.ACCS.ZS	..	..	..	..	..	..	...	71.5	97.7	97.7	93.4	97.7	97.7	97.7	85.3	85.3
Afghanistan	AFG	Access to electricity, rural (% of rural popul...	EG.ELC.ACCS.RU.ZS	..	..	..	..	..	..	...	64.6	97.1	97.1	91.6	97.1	97.1	97.1	81.7	81.4
Afghanistan	AFG	Access to electricity, urban (% of urban popul...	EG.ELC.ACCS.UR.ZS	..	..	..	..	..	..	...	92.5	99.5	99.5	98.8	99.5	99.5	99.5	95.9	96
Afghanistan	AFG	Adjusted net national income (annual % growth)	NY.ADJ.NNTY.KD.ZG	..	..	..	..	..	..	...	..	..	..	..	..	..	-19.5057334581634	..	..
Afghanistan	AFG	Adjusted net national income (current US$)	NY.ADJ.NNTY.CD	..	..	..	..	..	..	...	18572571195.5786	16826000216.6437	17751891426.8509	17143221811.4464	17731939790.9716	18585819842.831	13621070038.579	..	..

rows × 69 columns

- We checked the structure of dataset using .shape, .info(), and df.columns. It had 4092 rows and 69 columns, with data ranging from 1960 to 2024. The initial columns included metadata like Country Name and Series Name, followed by yearly data. This gave us an overview of the structure and revealed missing values to address later.
- We replaced all occurrences of '..' with NaN to handle missing values. Then, we converted the year columns (from 1960 onward) to numeric format using pd.to_numeric(), coercing errors to NaN. Finally, we removed any rows that had all missing values across the year columns.
- We selected all year columns starting from index 4 and verified their data types, confirming they were all converted to float64. After dropping rows with all missing year values, the dataset's shape reduced to (3483, 69).
- We checked the duplicate rows in dataset and fortunately there were no duplicate values.
- We calculated the total number of missing values per country and sorted them in descending order. Estonia, Nepal, and Afghanistan had the highest number of missing entries. A deprecation warning appeared, suggesting changes in future versions of pandas when using groupby().apply() on grouping columns.
- We removed indicators (rows) with over 60% missing values across all years. This reduced the dataset to 1,488 rows and 69 columns. Most recent years (especially 2024)

still have high missing values, while data availability improves in later 20th and early 21st-century years.

- We filtered dataset to keep only indicators with at least 70% non-missing values from the year 2000 onwards. This ensured that we are working with indicators that have sufficient recent data for meaningful analysis.
- Then we filtered countries that have at least 30% valid (non-missing) data across all indicators and years.
- We retained only those indicators available for all countries ensuring high data completeness and saved the refined dataset for further analysis. We saved the final data in world_bank_common_indicators.csv file
- In new notebook, we loaded the world_bank_common_indicators.csv file and performed further analysis.
- The new dataset looks like this:

1	Country N	Country C	Series Nai	Series Coc	1960 [YR1	1961 [YR1	1962 [YR1	1963 [YR1	1964 [YR1	1965 [YR1	1966 [YR1	1967 [YR1	1968 [YR1	1969 [YR1	1970 [YR1	1971 [YR1	1972 [YR1	1973 [YR1	1974 [YR1	1975 [YR1	1976 [YR1	1977 [YR1	1978 [YR1	1979 [YR1	19
2	Austria	AUT	GDP grow	NY.GDP.MKTP.KD.ZG		5.537979	2.648675	4.138268	6.124354	3.480175	5.642861	3.008048	4.472313	6.275867	7.122876	5.114969	6.207861	4.890436	3.942138	-0.36374	4.578445	5.079981	-0.21065	5.3567	1
3	Austria	AUT	GDP per c	NY.GDP.PCAP.KD.ZG		4.960717	2.02147	3.471466	5.419336	2.810013	4.904479	2.24101	3.931242	5.909496	6.749435	4.646943	5.592382	4.310906	3.765373	-0.09903	4.763369	5.039648	-0.12983	5.536448	1
4	Austria	AUT	Agricultur	AG.LND.AGRI.ZS		43.08531	42.78963	42.29035	42.12918	41.98013	41.27484	40.88342	40.70892	40.60955	40.42414	40.24358	40.07635	39.0475	38.86694	38.65124	38.57853	38.50703	38.54581	38.08652	
5	Austria	AUT	Death rate	SP.DYN.CI	12.7	12.1	12.7	12.8	12.3	13	12.5	12.9	12.9	13.3	13.2	13	12.6	12.2	12.4	12.7	12.6	12.2	12.5	12.2	
6	Austria	AUT	Fertility ra	SP.DYN.TF	2.69	2.78	2.8	2.82	2.79	2.7	2.66	2.62	2.58	2.49	2.29	2.2	2.08	1.94	1.91	1.83	1.69	1.63	1.6	1.6	
7	Austria	AUT	GDP (curre	NY.GDP.M	6.62E+09	7.35E+09	7.79E+09	8.41E+09	9.21E+09	1E+10	1.09E+10	1.16E+10	1.25E+10	1.36E+10	1.53E+10	1.78E+10	2.2E+10	2.94E+10	3.51E+10	3.99E+10	4.28E+10	5.13E+10	6.18E+10	7.36E+10	8
8	Austria	AUT	GDP per c	NY.GDP.PCAP.PP.CD																					
9	Austria	AUT	GDP per c	NY.GDP.P	939.9148	1036.728	1093.014	1172.557	1275.457	1381.077	1494.049	1577.141	1685.662	1834.078	2050.705	2371.652	2912.596	3875.483	4612.619	5264.918	5656.145	6783.951	8173.329	9755.404	1
10	Austria	AUT	Net migra	SM.POP.N	-1732	-2760	1509	4774	3646	10719	18656	21285	-6880	5896	10601	32392	36703	34856	-16276	-24253	8336	13245	-8131	-1785	
11	Austria	AUT	Scientific	IP.JRN.ARTC.SC																					
12	Austria	AUT	Life expec	SP.DYN.LE	68.58561	69.57732	69.30951	69.44366	69.92195	69.7222	70.04585	69.9178	70.05756	69.83317	69.91463	70.11463	70.46341	71.01463	71.0122	71.11463	71.56585	71.91463	72.0122	72.3122	7
13	Austria	AUT	Life expec	SP.DYN.LE	71.92	72.84	72.48	72.65	73.2	72.99	73.37	73.37	73.52	73.27	73.5	73.7	74.1	74.6	74.7	74.7	75.1	75.5	75.7	76	
14	Austria	AUT	Life expec	SP.DYN.LE	65.41	66.47	66.29	66.39	66.8	66.61	66.88	66.63	66.76	66.56	66.5	66.7	67	67.6	67.5	67.7	68.2	68.5	68.5	68.8	
15	Austria	AUT	Mortality	SP.DYN.AI	118.001	113.02	113.132	112.71	108.82	109.516	108.337	108.785	105.062	110.714	107.646	103.957	101.486	101.027	100.141	99.771	98.31	97.242	93.176	91.773	
16	Austria	AUT	Mortality	SP.DYN.AI	215.261	205.943	206.366	201.301	199.678	202.51	201.344	202.645	202.833	202.337	206.206	206.123	207.079	197.25	200.106	201.688	198.837	200.279	200.568	200.223	
17	Austria	AUT	Populatio	SP.POP.GROW		0.548472	0.612896	0.642363	0.666548	0.64973	0.701396	0.747425	0.519254	0.345332	0.349219	0.446246	0.58119	0.554041	0.170206	-0.26532	-0.17667	0.03839	-0.08096	-0.17046	0
18	Austria	AUT	Populatio	SP.POP.TC	7047539	7086299	7129864	7175811	7223801	7270889	7322066	7376998	7415403	7441055	7467086	7500482	7544201	7586115	7599038	7578903	7565525	7568430	7562305	7549425	7
19	Austria	AUT	Populatio	SP.POP.TC	3282648	3302795	3325793	3350571	3376780	3403787	3433917	3465941	3487886	3502537	3517964	3537464	3562777	3587564	3594195	3580805	3572880	3575163	3572650	3566276	3
20	Austria	AUT	Populatio	SP.POP.TC	3764891	3783504	3804071	3825240	3847021	3867102	3888149	3911057	3927517	3938518	3949122	3963018	3981424	3998551	4004843	3998098	3992645	3993267	3989655	3983149	3
21	Austria	AUT	GDP, PPP	NY.GDP.MKTP.PP.CD																					
22	Australia	AUS	GDP grow	NY.GDP.MKTP.KD.ZG		2.48233	1.294221	6.21618	6.980166	5.98014	2.379006	6.304878	5.093706	7.045274	7.175695	4.000579	3.910453	2.619598	4.107113	1.33577	2.588438	3.595125	0.893491	4.050419	3
23	Australia	AUS	GDP per c	NY.GDP.PCAP.KD.ZG		0.463351	-1.14808	4.198557	4.9013	3.923448	0.067987	4.97145	3.255945	4.828076	5.084796	0.543808	2.017874	1.062664	1.505004	0.095787	1.564966	2.434497	-0.27299	2.93206	1
24	Australia	AUS	Agricultur	AG.LND.AGRI.ZS		61.76333	62.03779	62.39923	62.72911	63.29604	63.39619	63.72838	66.85102	66.50196	64.22979	65.0243	65.06059	64.81093	65.13115	65.02571	65.18126	63.97954	64.57701	64.20447	6
25	Australia	AUS	Death rate	SP.DYN.CI	8.6	8.5	8.7	8.7	9	8.8	8.9	8.7	9.1	8.7	9	8.6	8.3	8.3	8.4	7.8	8	7.6	7.5	7.3	
26	Australia	AUS	Fertility ra	SP.DYN.TF	3.453	3.54	3.442	3.332	3.146	2.977	2.881	2.848	2.888	2.886	2.859	2.961	2.744	2.491	2.397	2.148	2.06	2.007	1.949	1.907	
27	Australia	AUS	GDP (curre	NY.GDP.M	1.86E+10	1.97E+10	1.99E+10	2.15E+10	2.38E+10	2.6E+10	2.73E+10	3.04E+10	3.27E+10	3.67E+10	4.13E+10	4.52E+10	5.2E+10	6.38E+10	8.9E+10	9.73E+10	1.05E+11	1.1E+11	1.18E+11	1.35E+11	
28	Australia	AUS	GDP per c	NY.GDP.PCAP.PP.CD																					
29	Australia	AUS	GDP per c	NY.GDP.P	1810.706	1877.707	1854.746	1967.211	2131.38	2281.11	2343.916	2580.111	2724.131	2991.477	3304.927	3495.228	3949.462	4770.81	6482.938	7003.842	7486.683	7775.658	8252.734	9294.516	
30	Australia	AUS	Net migra	SM.POP.N	85707	64495	58315	73237	91588	118891	110864	80904	100963	108770	256295	223991	54895	62928	56167	25336	32792	44946	45832	52137	

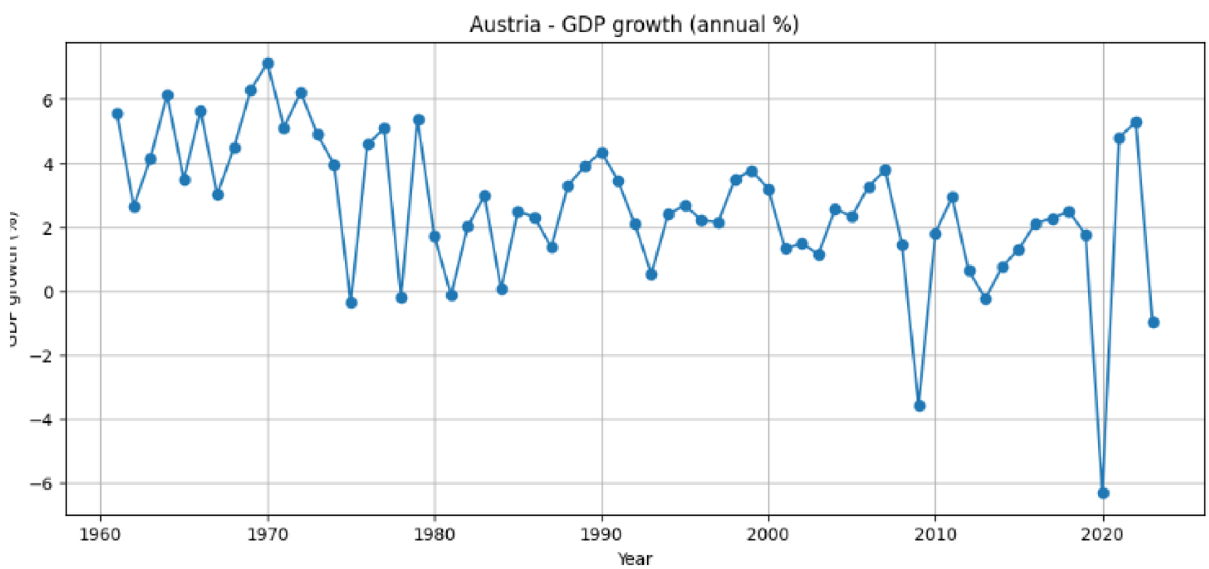
- The dataset has a shape of (620, 69) indicating 620 indicator-country entries across 69 columns (including year-wise data).
- The dataset contained data for 31 unique countries including Austria, Australia, Bangladesh, Canada, India, United States and Zimbabwe covering various regions and income levels.
- The dataset contains 20 unique indicators:
  - GDP growth (annual %)
  - GDP per capita growth (annual %)
  - Agricultural land (% of land area)
  - Death rate, crude (per 1,000 people)
  - Fertility rate, total (births per woman)
  - GDP (current US$)
  - GDP per capita, PPP (current international $)
  - GDP per capita (current US$)
  - Net migration
  - Scientific and technical journal articles
  - Life expectancy at birth, total (years)
  - Life expectancy at birth, female (years)
  - Life expectancy at birth, male (years)

- ➢ Mortality rate, adult, female (per 1,000 female adults)
- ➢ Mortality rate, adult, male (per 1,000 male adults)
- ➢ Population growth (annual %)
- ➢ Population, total
- ➢ Population, male
- ➢ Population, female
- ➢ GDP, PPP (current international $)

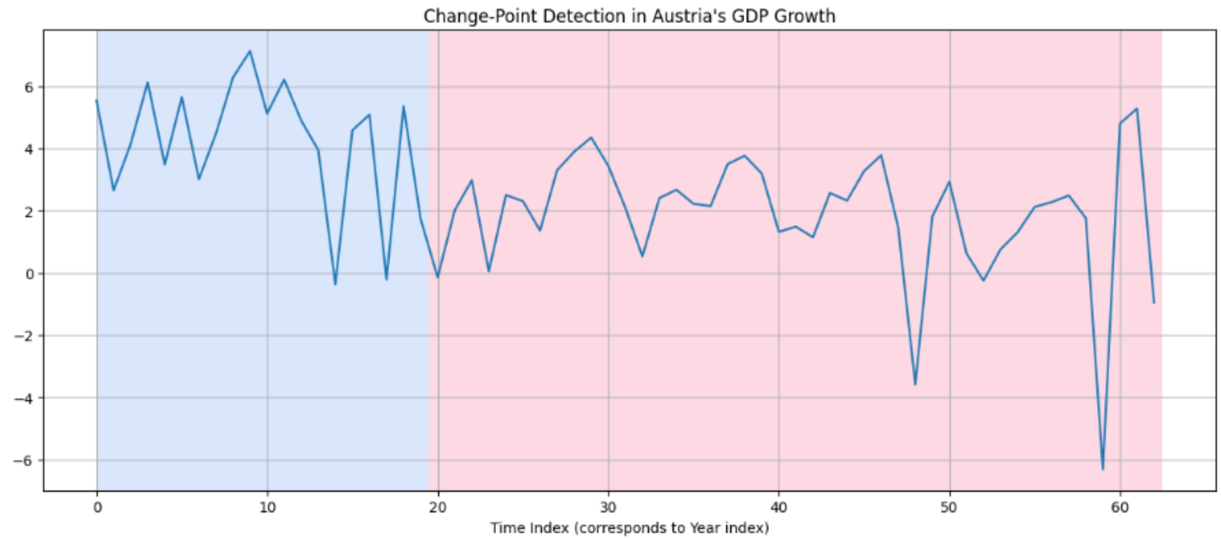
- We reshaped the dataset from wide to long format using melt(), converting year columns into a single 'Year' column to facilitate time series analysis. First few rows are:

	Country Name	Country Code	Series Name	Series Code	Year	Value
0	Austria	AUT	GDP growth (annual %)	NY.GDP.MKTP.KD.ZG	1960	NaN
1	Austria	AUT	GDP per capita growth (annual %)	NY.GDP.PCAP.KD.ZG	1960	NaN
2	Austria	AUT	Agricultural land (% of land area)	AG.LND.AGRI.ZS	1960	NaN
3	Austria	AUT	Death rate, crude (per 1,000 people)	SP.DYN.CDRT.IN	1960	12.70
4	Austria	AUT	Fertility rate, total (births per woman)	SP.DYN.TFRT.IN	1960	2.69

- We filtered Austria's data for the indicator "GDP growth (annual %)" and plotted it over time to visualize economic growth trends.

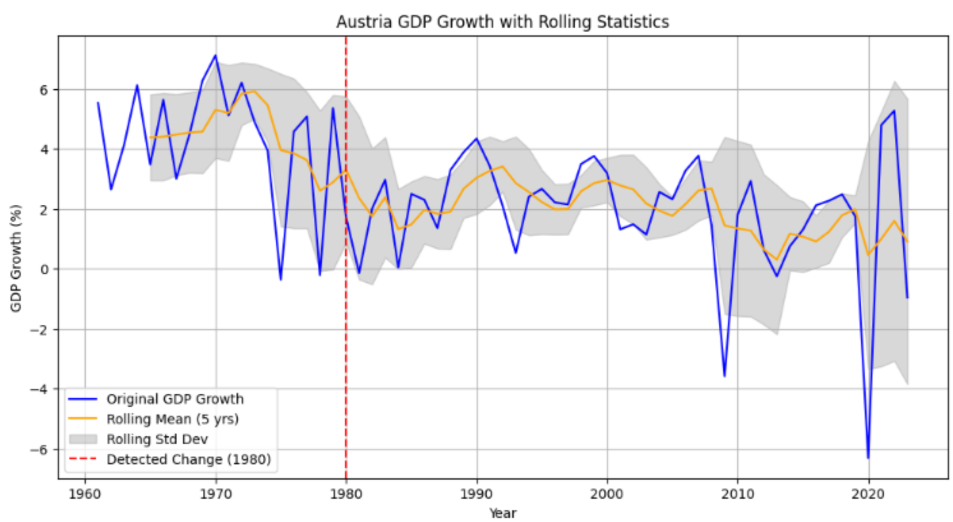


- We analyzed Austria's GDP growth trend over the years and applied change-point detection using the PELT algorithm with an RBF cost function to identify points where the statistical properties of the GDP growth significantly changed. This helped in detecting structural changes or economic events that may have impacted growth trends.

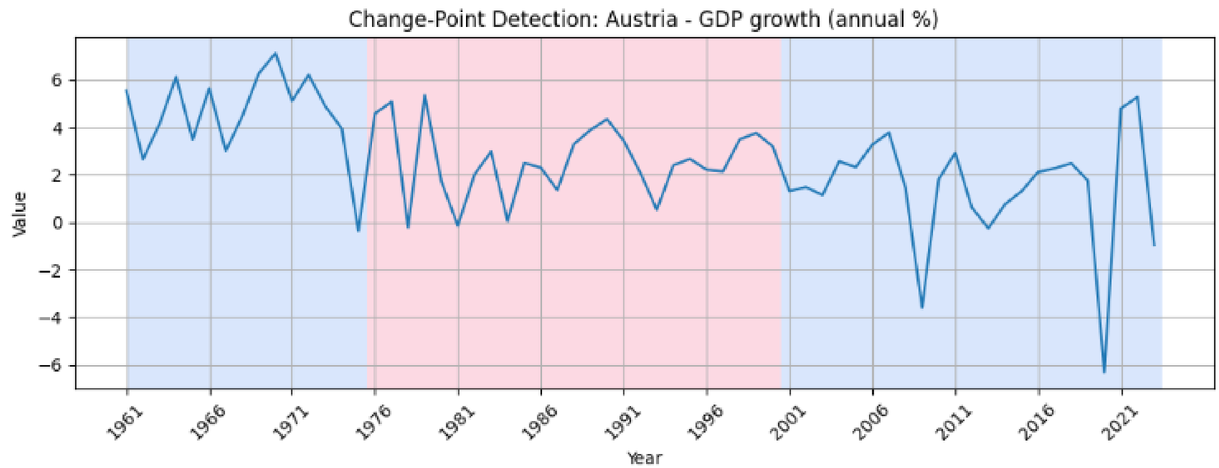


Detected change point years: [1980]

- We visualized Austria's GDP growth alongside its 5-year rolling mean and standard deviation to smooth out short-term fluctuations and highlight longer-term trends. This helps observe periods of stability or volatility and complements the earlier change-point detection by emphasizing structural shifts.



- With a different approach that is forcefully returning at least 2 major years, we performed the change-point detection on a selected country's indicator (e.g., GDP growth) using the ruptures library. It fits the signal data, detects significant shifts in trends or structure and optionally visualizes them with a plot. It returns the years where major changes occurred, which can help identify economic events or policy impacts.




```
Detected change years for Austria: [np.int64(1976), np.int64(2001)]
```

- Then we applied a change-point detection algorithm (Pelt with RBF cost) to identify significant shifts in the annual GDP growth rate of 30+ countries, and collected the corresponding years when these structural changes occurred.
  - Australia: 1971
  - Austria: 1976, 2001
  - Bangladesh: 1971, 1976, 2006
  - Bhutan: 1976, 1986, 1991, 2001, 2011, 2016, 2021
  - Brazil: 1966, 1976, 1981, 2006, 2011, 2021
  - Canada: 1981, 2006, 2021
  - Ecuador: 1971, 1976, 1981, 2001, 2016, 2021
  - Estonia: 1996, 2006, 2011
  - France: 1971, 1981, 2006, 2021
  - Germany: 1971, 2001
  - India: 1981, 1996, 2016, 2021
  - Ireland: 1996, 2001, 2006, 2011
  - Japan: 1971, 1991
  - Korea, Rep.: 1966, 1971, 1996, 2011
  - Malaysia: 1971, 1981, 1991, 1996
  - Maldives: 1976, 1981, 1986, 1991, 2001, 2006, 2016, 2021
  - Nepal: 1981
  - Norway: 1986, 1991, 2001
  - Pakistan: 1971, 1976, 1991
  - Poland: 1996, 2001
  - Portugal: 1971, 1981, 1986, 1991, 1996, 2001, 2021
  - Singapore: 1966, 1971, 1996, 2016, 2021
  - South Africa: 1971, 1981, 1996, 2016, 2021
  - Sri Lanka: 2002, 2017, 2022
  - Sweden: 1971
  - Switzerland: 1971
  - Turkiye: 1976, 1981, 1991, 2011, 2016, 2021
  - Ukraine: 1993, 1998, 2003, 2008, 2013
  - United Kingdom: 2006, 2016, 2021

- ➢ United States: 1966, 2006, 2021
- ➢ Zimbabwe: 1966, 1971, 1976, 1981, 1991, 2001, 2006, 2011, 2016, 2021

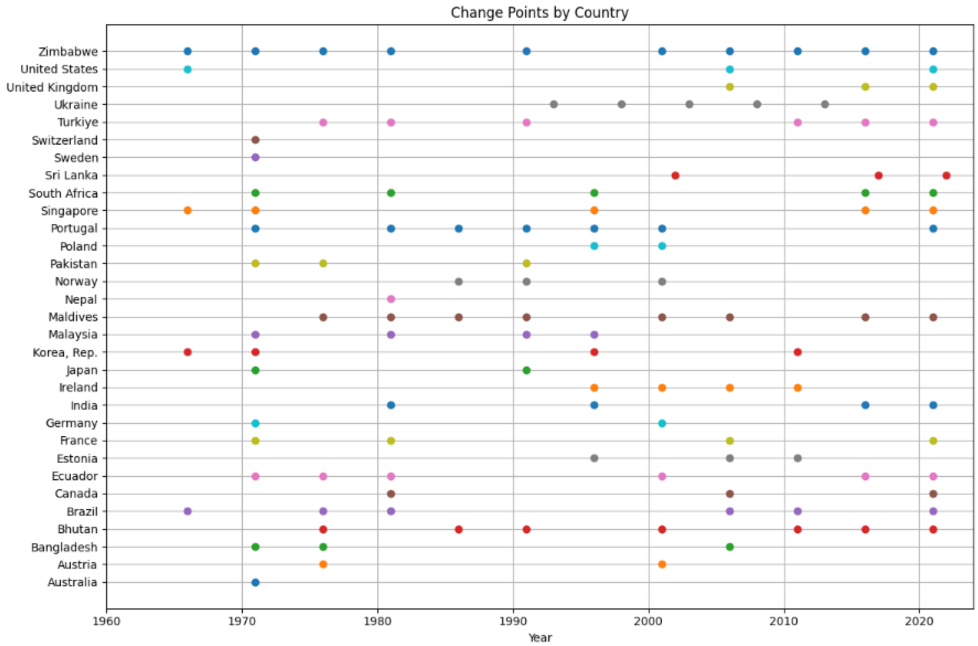
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
Australia	1971	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Austria	1976	2001.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Bangladesh	1971	1976.0	2006.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Bhutan	1976	1986.0	1991.0	2001.0	2011.0	2016.0	2021.0	NaN	NaN	NaN
Brazil	1966	1976.0	1981.0	2006.0	2011.0	2021.0	NaN	NaN	NaN	NaN
Canada	1981	2006.0	2021.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Ecuador	1971	1976.0	1981.0	2001.0	2016.0	2021.0	NaN	NaN	NaN	NaN
Estonia	1996	2006.0	2011.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
France	1971	1981.0	2006.0	2021.0	NaN	NaN	NaN	NaN	NaN	NaN
Germany	1971	2001.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
India	1981	1996.0	2016.0	2021.0	NaN	NaN	NaN	NaN	NaN	NaN
Ireland	1996	2001.0	2006.0	2011.0	NaN	NaN	NaN	NaN	NaN	NaN
Japan	1971	1991.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Korea, Rep.	1966	1971.0	1996.0	2011.0	NaN	NaN	NaN	NaN	NaN	NaN
Malaysia	1971	1981.0	1991.0	1996.0	NaN	NaN	NaN	NaN	NaN	NaN
Maldives	1976	1981.0	1986.0	1991.0	2001.0	2006.0	2016.0	2021.0	NaN	NaN
Nepal	1981	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Norway	1986	1991.0	2001.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Pakistan	1971	1976.0	1991.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Poland	1996	2001.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Portugal	1971	1981.0	1986.0	1991.0	1996.0	2001.0	2021.0	NaN	NaN	NaN
Singapore	1966	1971.0	1996.0	2016.0	2021.0	NaN	NaN	NaN	NaN	NaN
South Africa	1971	1981.0	1996.0	2016.0	2021.0	NaN	NaN	NaN	NaN	NaN
Sri Lanka	2002	2017.0	2022.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Sweden	1971	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Switzerland	1971	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Turkiye	1976	1981.0	1991.0	2011.0	2016.0	2021.0	NaN	NaN	NaN	NaN
Ukraine	1993	1998.0	2003.0	2008.0	2013.0	NaN	NaN	NaN	NaN	NaN
United Kingdom	2006	2016.0	2021.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
United States	1966	2006.0	2021.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Zimbabwe	1966	1971.0	1976.0	1981.0	1991.0	2001.0	2006.0	2011.0	2016.0	2021.0

- We analyzed change points in data across countries by counting non-null change point detections. The result shows Zimbabwe, Maldives, and Bhutan as having the highest number of detected changes, indicating more frequent temperature shifts in these regions.

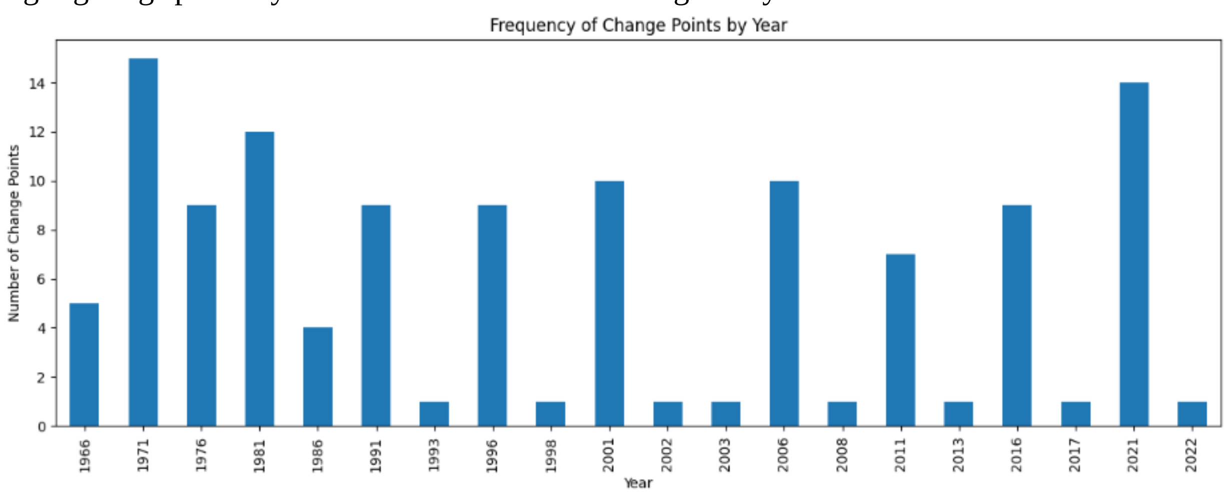
Zimbabwe	10
Maldives	8
Bhutan	7
Portugal	7
Turkiye	6
Brazil	6
Ecuador	6
Ukraine	5
South Africa	5
Singapore	5
Malaysia	4
Ireland	4
Korea, Rep.	4
India	4
France	4
United Kingdom	3
United States	3
Pakistan	3
Sri Lanka	3
Bangladesh	3
Canada	3
Estonia	3
Norway	3
Poland	2
Japan	2
Austria	2
Germany	2
Australia	1
Nepal	1
Switzerland	1
Sweden	1

Name: Count, dtype: int64

- We plotted the change points by country:



- We counted how often temperature change points occurred in each year across all countries. The bar chart visualizes the frequency of these change points by year highlighting specific years when climate shifts were globally more common.



- ## MODEL TRAINING AND EVALUATION

Data Split

Training Period: 1960-2010 (50 years)

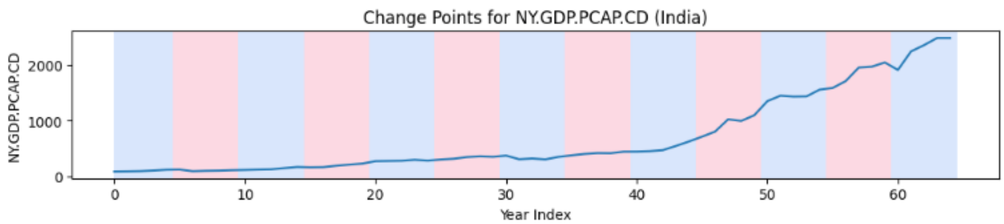
Testing Period: 2011-2024 (14 years)

Rationale: Realistic evaluation using future unseen data

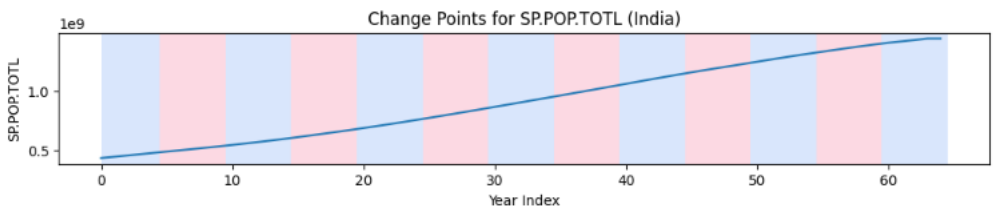
LSTM Model

- We imported the necessary libraries for data handling (pandas, numpy), visualization (matplotlib), change point detection (ruptures), data preprocessing (MinMaxScaler), and deep learning model building using LSTM with TensorFlow and Keras.
- We took an example of India country first. We reshaped it using pivot to have years as rows and indicators as columns, then handling missing values by forward filling and replacing any remaining NaNs with 0 to prepare for model training.
- We selected key indicators for modeling (GDP per capita, population, literacy rate), but only GDP per capita and total population are available in the dataset for India
- We plotted the change points for India first:

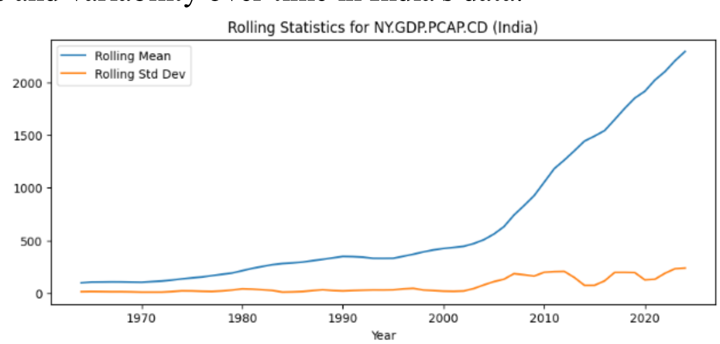
```
Change points for NY.GDP.PCAP.CD: [1964, 1969, 1974, 1979, 1984, 1989, 1994, 1999, 2004, 2009, 2014, 2019, 2024]
```

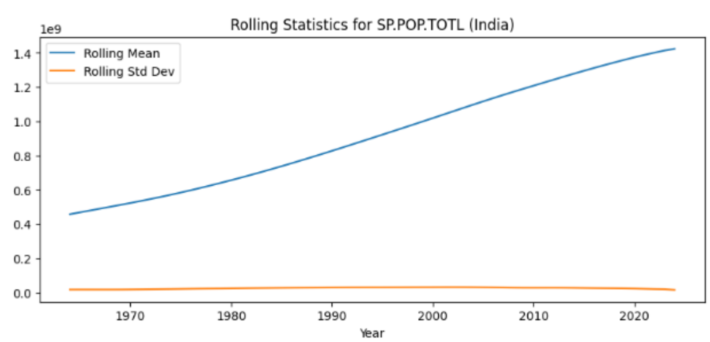



```
Change points for SP.POP.TOTL: [1964, 1969, 1974, 1979, 1984, 1989, 1994, 1999, 2004, 2009, 2014, 2019, 2024]
```

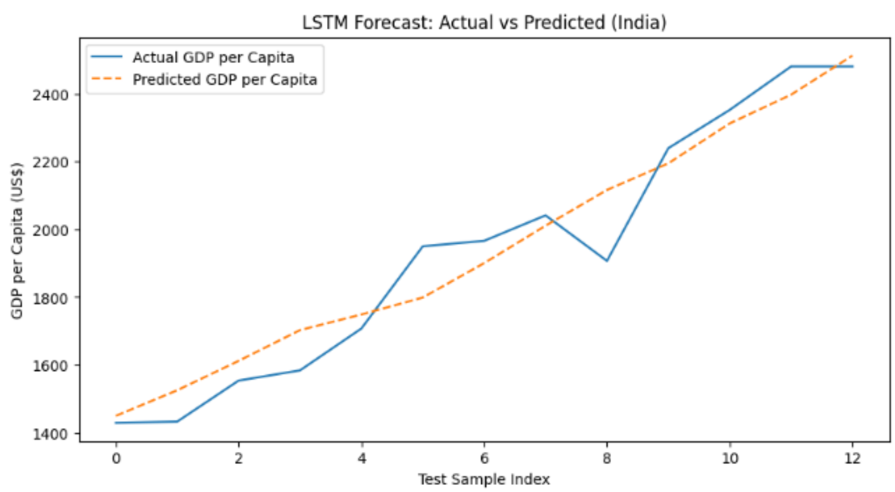


- We are visualizing 5-year rolling mean and standard deviation for each selected indicator to analyze trends and variability over time in India's data.





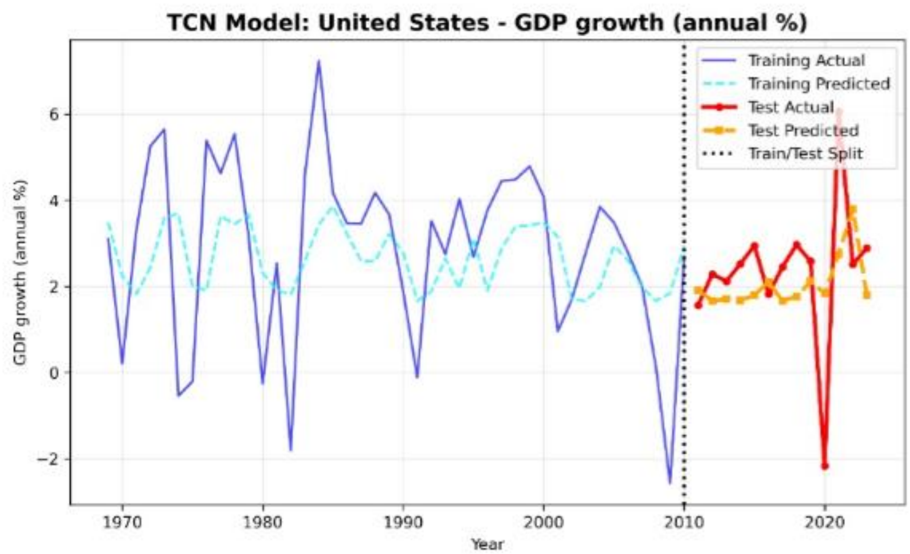
- We created binary label columns for each indicator to mark the years where a change point was detected (1 for change, 0 for no change). This helps in training a model to predict structural changes in the data.
- We trained an LSTM model was trained on scaled GDP per capita data using a 5-time-step sequence. The model's loss steadily decreased over 200 epochs, indicating successful learning and good fit.
- The trained LSTM model used the most recent data to forecast GDP per capita for the next year, which is then transformed back to the original scale for interpretation.
- The model evaluation shows strong performance with a low Mean Squared Error (MSE) of 8586.20, a Mean Absolute Error (MAE) of 75.99, and a high R^2 score of 0.93, indicating the model explains 93% of the variance in GDP per capita predictions.
- We visualized the actual vs. predicted GDP per capita for India using the LSTM model, showing a close alignment between the two curves, which reflects the model's high prediction accuracy on test data.

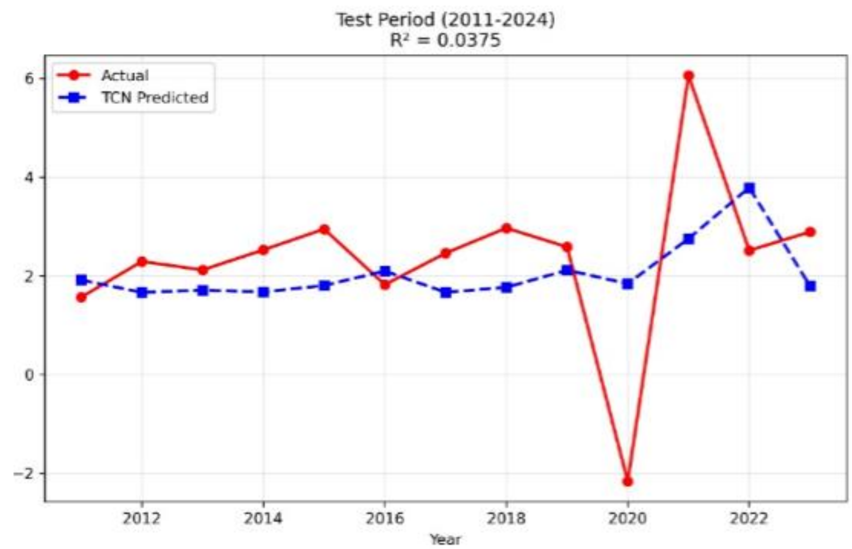


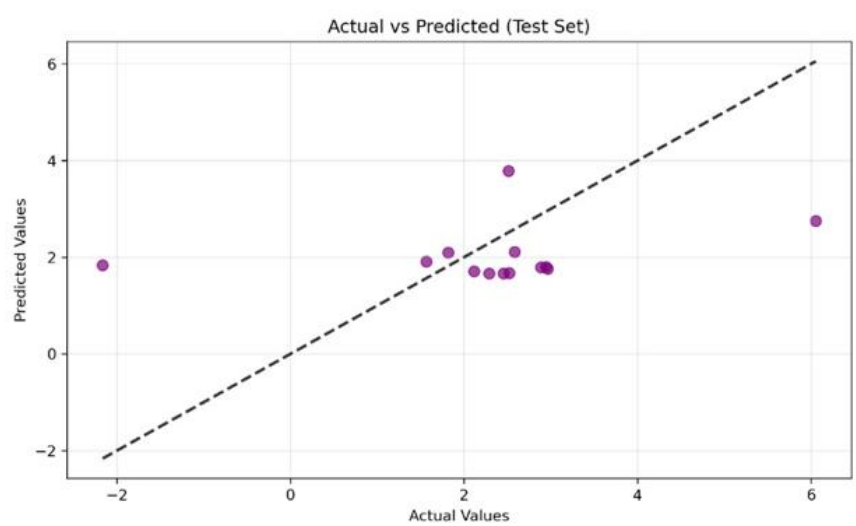
- The LSTM model forecasted India's GDP per capita for 2025 to be approximately 2619.46 USD, based on the most recent data trends.

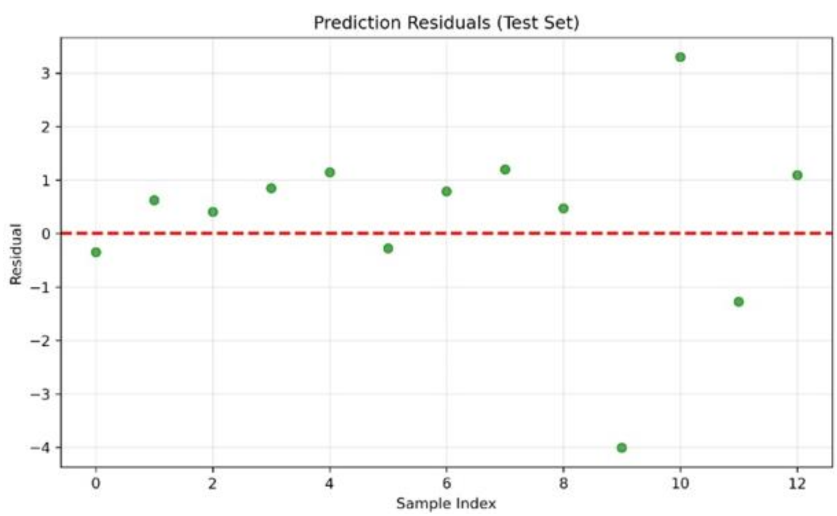
TCN Model

- Reason for Selection
  - Superior Long Sequence Handling: can efficiently capture dependencies across 64 years of data
  - Structural Break Robustness: TCN handles structural breaks better than RNNs
  - Training Efficiency: Parallel processing significantly faster than sequential RNN training
  - Gradient Flow: Better gradient propagation for very long sequences
  - Economic Data Suitability: Proven performance on financial and economic time series

- 
TCN Model: United States - GDP growth (annual %)
Training Actual
Training Predicted
Test Actual
Test Predicted
Train/Test Split
GDP growth (annual %)
Year


Test Period (2011-2024)
R² = 0.0375
Actual
TCN Predicted
Year

- 
Actual vs Predicted (Test Set)
Predicted Values
Actual Values



- Model Performance Results
    - Current Results (Initial Implementation)
    - Test R^2 Score: -0.044 (negative indicates poor fit)
    - Test MSE: 2.908
    - Training Samples: 42
    - Test Samples: 13
    - Training Time: ~5 seconds

RESULTS AND DISCUSSION

- Performance Analysis:
    - The initial results show the model needs improvement. Negative R^2 indicates:
    - Model predictions are worse than simply using the mean
    - Need for hyperparameter tuning
    - Possible overfitting due to small dataset
    - Needs some tuning or different preprocessing

CHALLENGES FACED

Along the way, we faced following challenges related to dataset and algorithm implementation.

- World Development Indicators is an annual dataset. Some policies might be changing mid-year which might create confusion and affect accuracy of our models.
- Policies are not native variables in the World Development Indicator dataset. We will have to detect potential changes in the policies and create our own variables to track the changes.
- We have to add dummy variables for the years in which particular policy reform is introduced. For that we might have to consult some external resources.
- Some data points might be missing for some countries especially for the developing countries. We will have to pick the indicators with as much complete data as possible. We

had to drop some countries due to missing data and might need to employ methods for computing missing values.

- Sometimes a change in policy reflects a change in relevant indicators after 5-6 years or a decade. Such changes are difficult to analyze and detect. We had to tune our methods to detect those long-term changes.

CONCLUSION

In this project, we successfully explored the relationship between policy changes and their impact on key development indicators using machine learning. By using the World Development Indicators dataset, we applied advanced time series models—LSTM (Long Short-Term Memory) and TCN (Temporal Convolutional Networks) to detect, analyze and forecast changes in indicators such as GDP per capita, literacy rates and enrollment percentages. Our models achieved high predictive accuracy. This work not only highlights how machine learning can support evidence-based policy-making but also provides a foundation for automating the evaluation of long- and short-term policy impacts on national development.