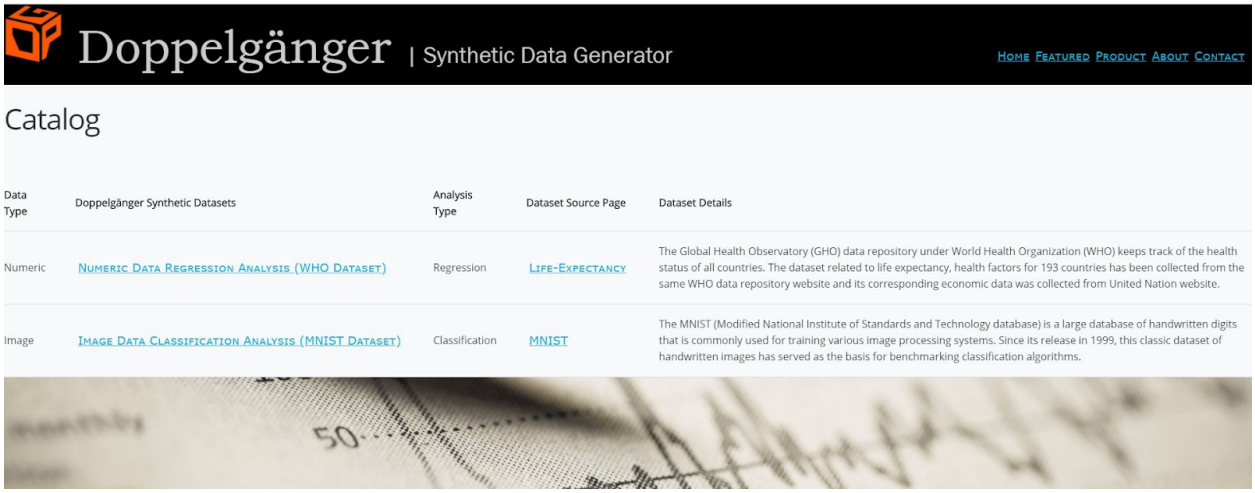
DOPPELGÄNGER
SYNTHETIC DATA GENERATOR

Application User Guide

1. Click on the Application URL - http://101.127.128.81:8080/

Below home page will appear -



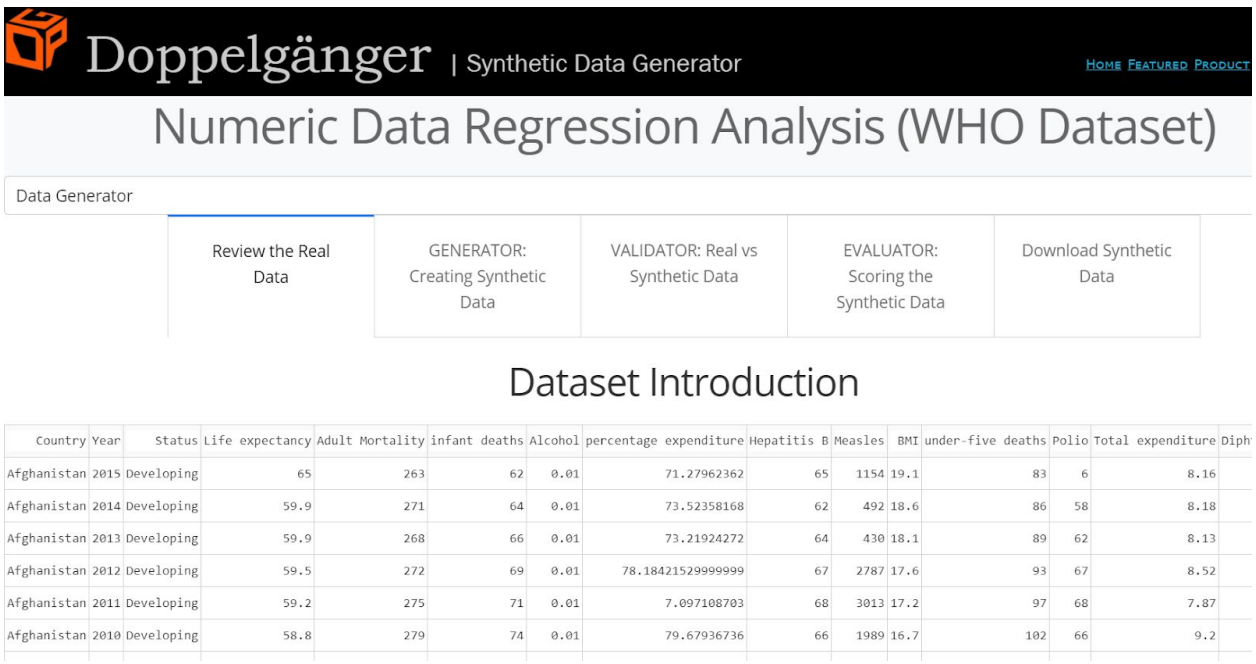
2. On **Home Page**
   - Under **Catalog** -> **Doppelgänger Synthetic Datasets**
     - Click the respective links to navigate to the **Doppelgänger Synthetic Datasets Pages**
       - **Numeric Data Regression Analysis (WHO Dataset)**
       - **Image Data Classification Analysis (MNIST Dataset)**

   - Under **Catalog** -> **Data Source Page**
     - Click on the respective links to navigate to the source datasets
       - **Life-Expectancy**
       - **MNIST**

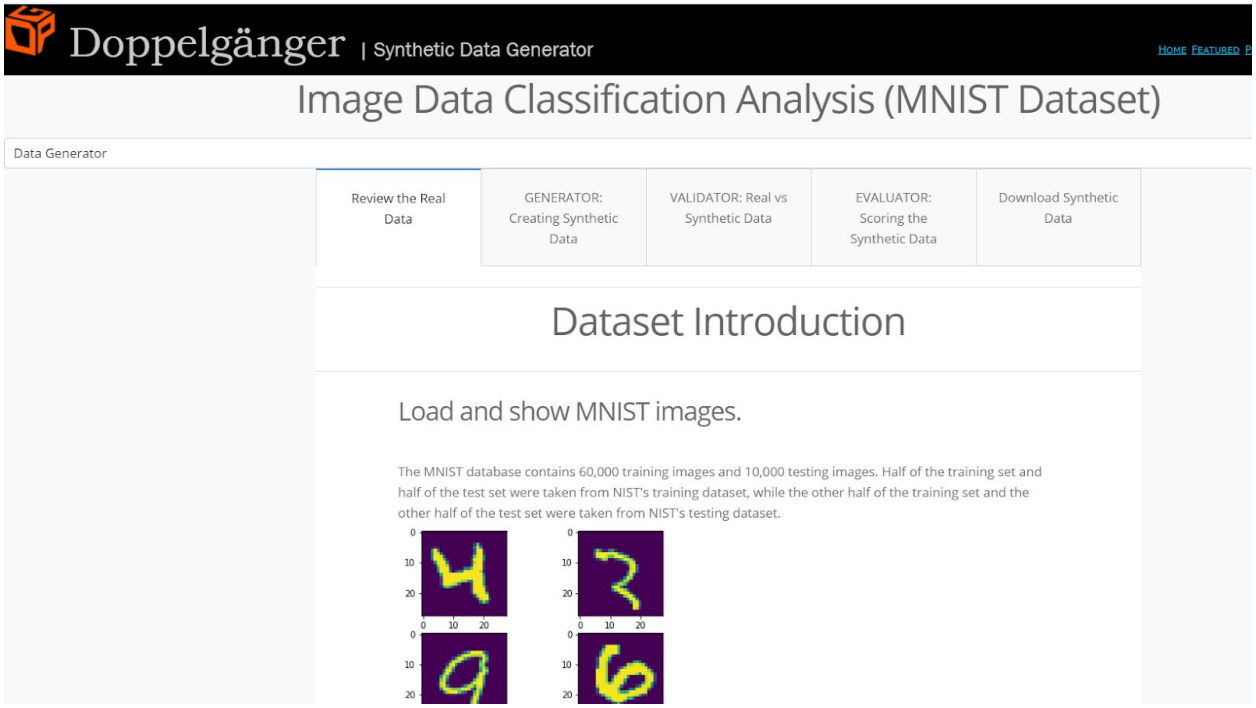
3. After navigating to any of the **Doppelgänger Synthetic Datasets Pages**, a tabbed interface will appear:

- By Default “Data Generator” drop-down will be selected as shown below -

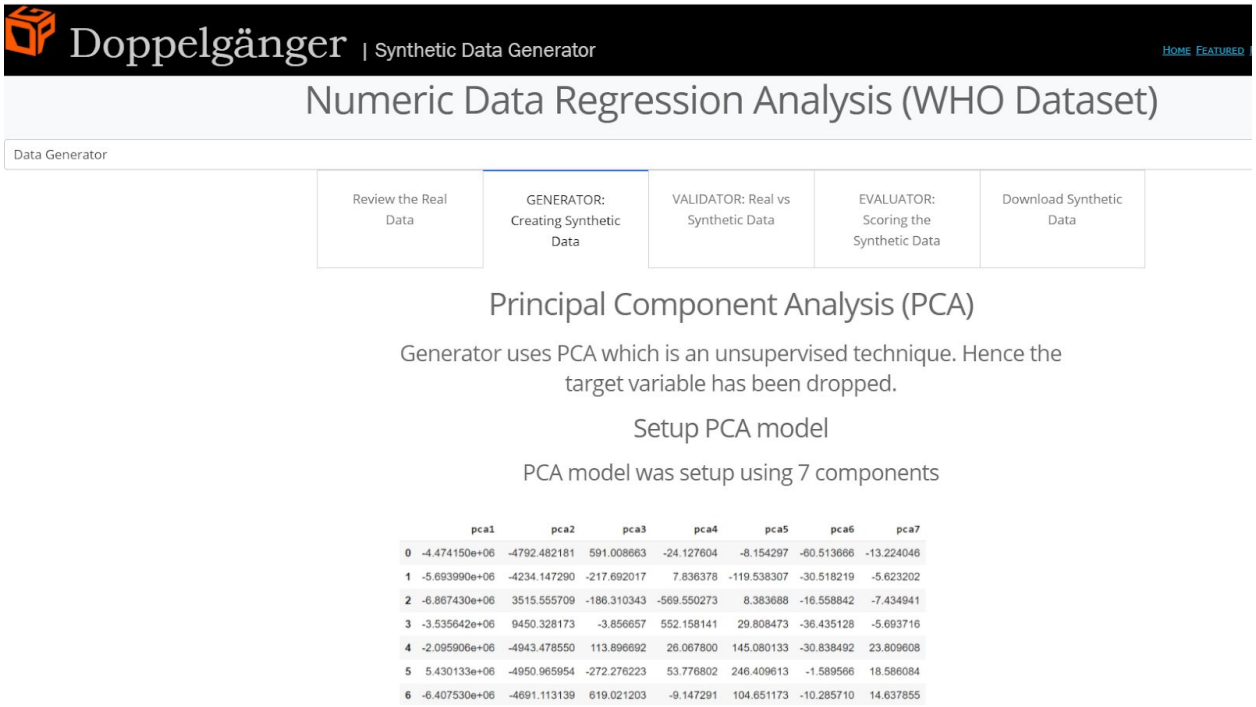
3.1 Click/View Tab **“Review the Real Data”** - To review original source dataset attributes -



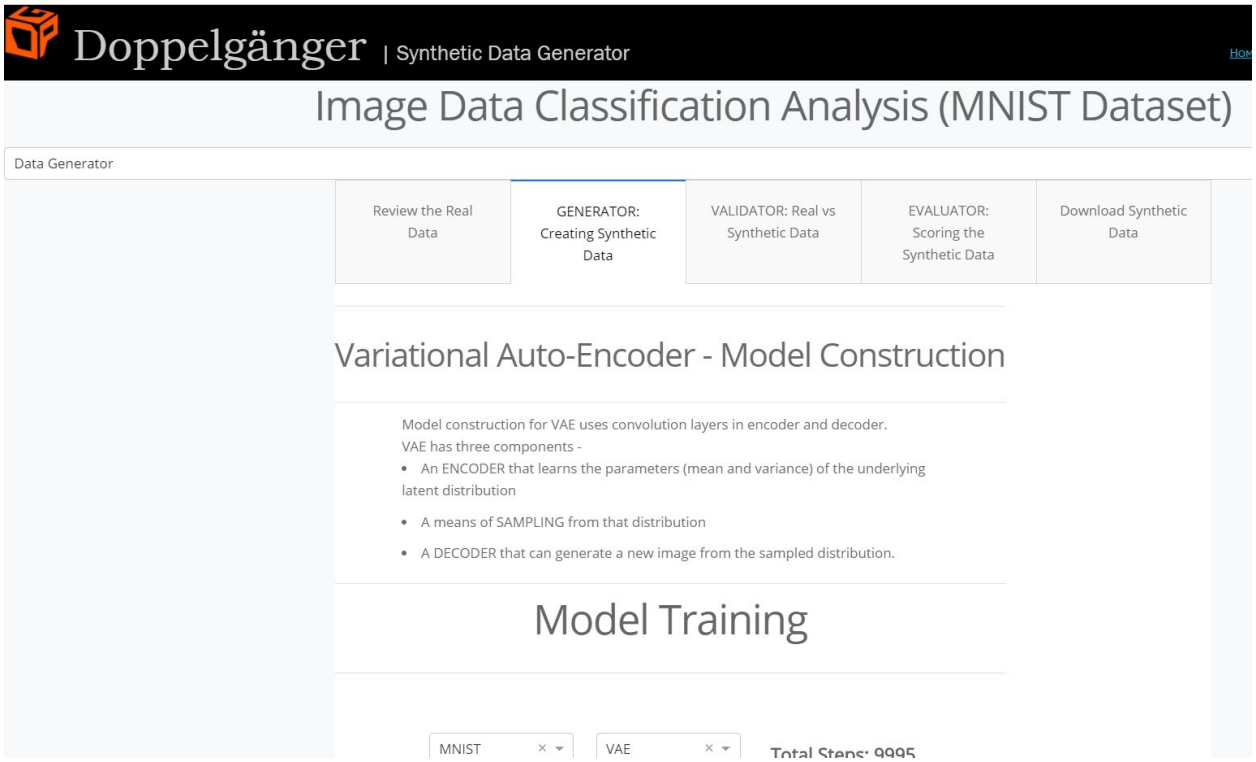
Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diph
Afghanistan	2015	Developing	65	263	62	0.01	71.27962362	65	1154	19.1	83	6	8.16	
Afghanistan	2014	Developing	59.9	271	64	0.01	73.52358168	62	492	18.6	86	58	8.18	
Afghanistan	2013	Developing	59.9	268	66	0.01	73.21924272	64	430	18.1	89	62	8.13	
Afghanistan	2012	Developing	59.5	272	69	0.01	78.18421529999999	67	2787	17.6	93	67	8.52	
Afghanistan	2011	Developing	59.2	275	71	0.01	7.097108703	68	3013	17.2	97	68	7.87	
Afghanistan	2010	Developing	58.8	279	74	0.01	79.67936736	66	1989	16.7	102	66	9.2	



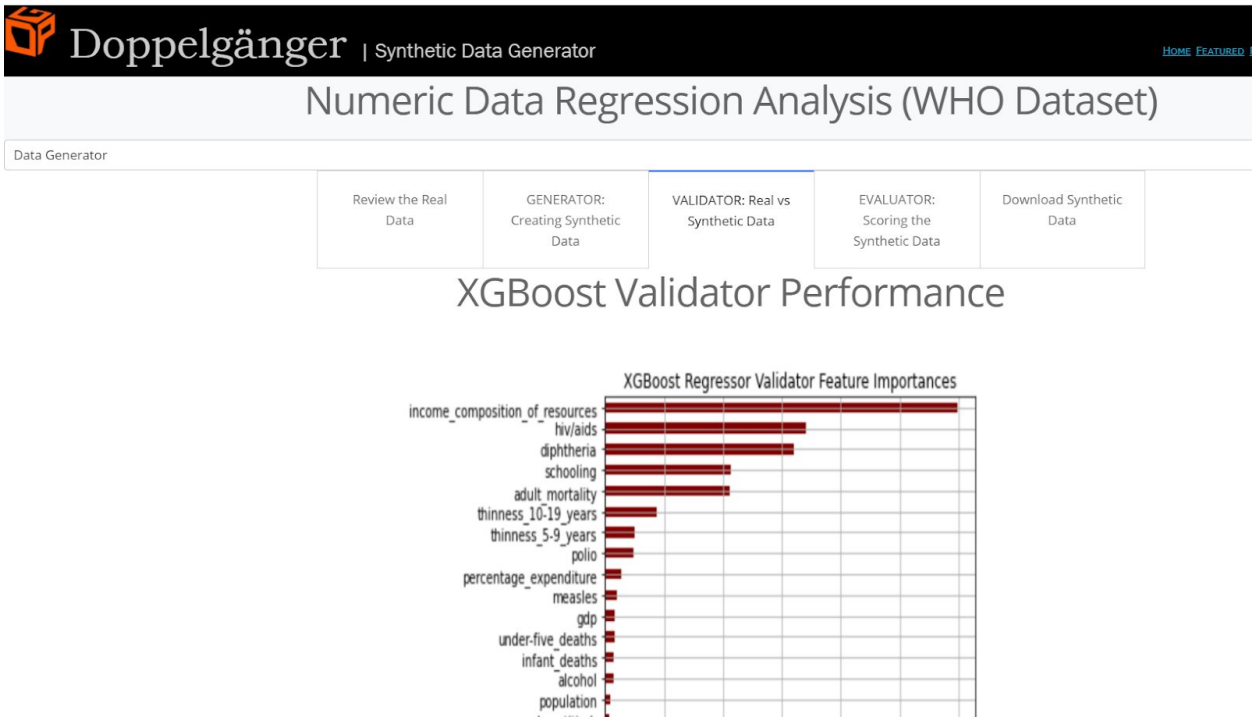
3.2 Click on Tab **"GENERATOR: Creating Synthetic Data"** - to review Synthetic Data generation strategy and model training aspects -

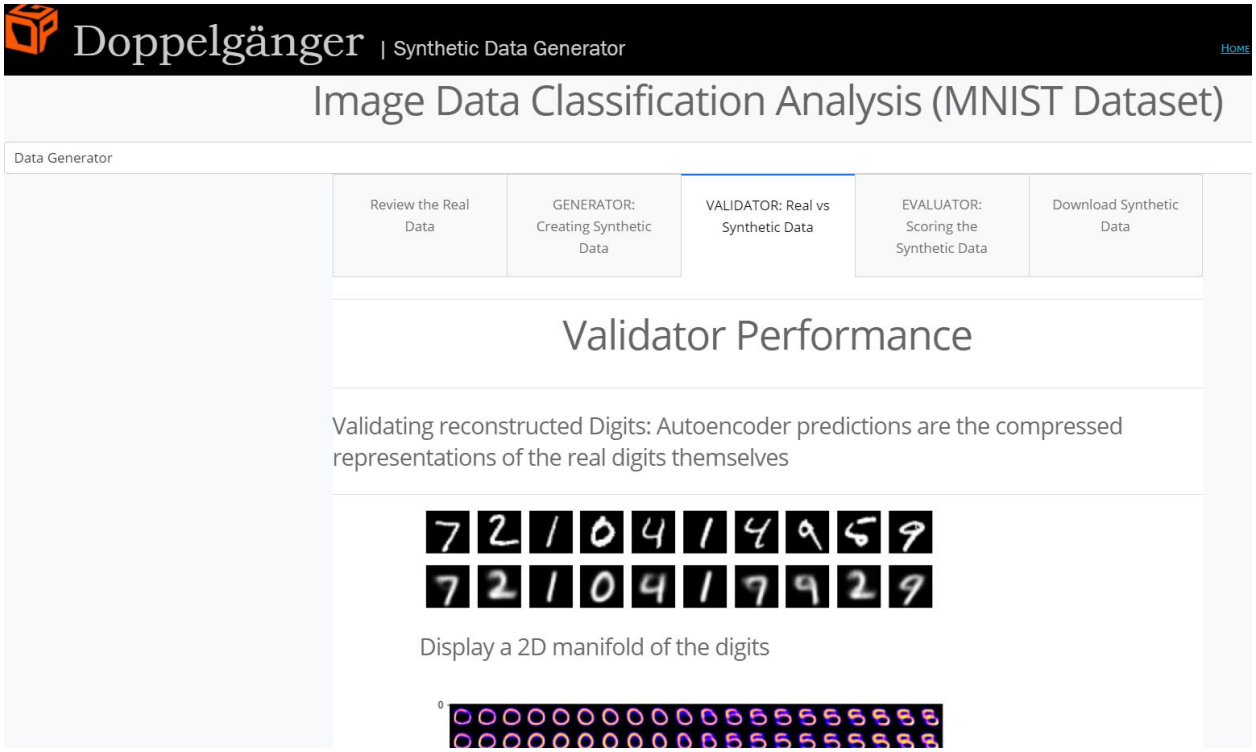


	pca1	pca2	pca3	pca4	pca5	pca6	pca7
0	-4.474150e+06	-4792.482181	591.008663	-24.127604	-8.154297	-60.513666	-13.224046
1	-5.693990e+06	-4234.147290	-217.692017	7.836378	-119.538307	-30.518219	-5.623202
2	-6.867430e+06	3515.555709	-186.310343	-569.550273	8.383688	-16.558842	-7.434941
3	-3.535642e+06	9450.328173	-3.856657	552.158141	29.808473	-36.435128	-5.693716
4	-2.095906e+06	-4943.478550	113.896692	26.067800	145.080133	-30.838492	23.809608
5	5.430133e+06	-4950.965954	-272.276223	53.776802	246.409613	-1.589566	18.586084
6	-6.407530e+06	-4691.113139	619.021203	-9.147291	104.651173	-10.285710	14.637855



3.3 Click on Tab **"VALIDATOR: Real vs Synthetic Data"** - to review validator -





3.4 Click on Tab **"EVALUATOR: Scoring the Synthetic Data"** - to review evaluator -

Doppelgänger | Synthetic Data Generator

Home

Image Data Classification Analysis (MNIST Dataset)

Data Generator

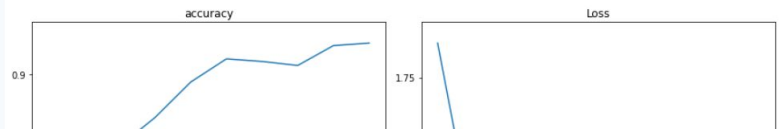
Review the Real Data | GENERATOR: Creating Synthetic Data | VALIDATOR: Real vs Synthetic Data | EVALUATOR: Scoring the Synthetic Data | Download Synthetic Data

Evaluator Performance

Comparing Real vs Synthetic Data Results

Train another CNN using the synthetic data to predict the test (real) data

CNN Evaluator - Loss and Accuracy



Doppelgänger | Synthetic Data Generator

Home Featured

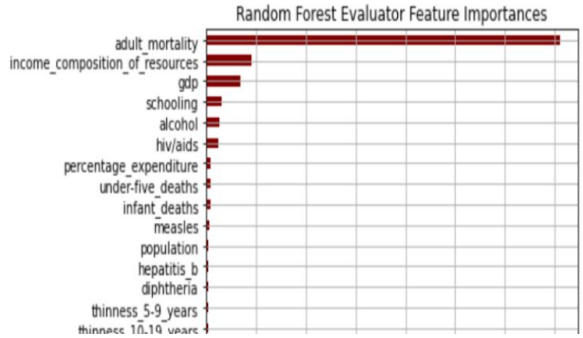
Numeric Data Regression Analysis (WHO Dataset)

Data Generator

Review the Real Data | GENERATOR: Creating Synthetic Data | VALIDATOR: Real vs Synthetic Data | EVALUATOR: Scoring the Synthetic Data | Download Synthetic Data

Random Forest Evaluator Performance

R2 Score of RandomForestRegressor : 0.88, Root Mean Squared Error Score of RandomForestRegressor : 3.17



3.5 Click on Tab **“Download Synthetic Data”** - to download the Synthetic Data and see summary of the overall workflow that was applied -