USING GENERATIVE AI TO FORECAST ELECTRICITY DEMAND AND SUPPLY GAPS: SUB-SAHARAN AFRICA VS. OTHER REGIONS

Project Presenter

Bharat Kumar Jhawar
Adeniyi Olaolu

Hamoye Spring' 24
Team Biggan

Content

- Introduction
  - Problem Statement
  - Problem Objective
  - Process Overflow
- Data Collection
- Data Cleaning and Preprocessing
- Exploratory Data Analysis & Feature Selection
- Model Training and Evaluation
- Model Deployment
- Conclusion
- References

Introduction

- **Problem Statement**

  Building a generative AI model capable of accurately forecasting electrification demand, supply, and the resulting gap to guide effective and sustainable energy access interventions in SSA.

Introduction

Objectives

01 02 03 04 05

Comparative Analysis

Electrification landscape: SSA among other world regions / Countries within SSA region

Guiding Interventions

Provide data-driven insights to guide targeted and effective interventions for improving sustainable energy access in SSA, considering both demand and supply gap factors

Dynamic Adaptation

Ensure the AI model can dynamically adapt to new data inputs to continuously refine predictions and intervention strategies over time.

Develop a Generative AI Model

Robust Generative AI model to accurately forecast the electricity demand, supply, and resulting gap for SSA

Predict Demand/Supply Gap

Utilize machine learning techniques to forecast the electricity demand and supply gap for Sub-Saharan Africa

Introduction

1. Data Collection
2. Data Cleaning and Preprocessing
3. Exploratory Data Analysis & Feature Selection
4. Model Training and Evaluation
5. Model Deployment

Process Overflow

Data Collection

Global Electrification Database by Harvard Dataverse

The World Bank data

World Energy Statistics and Balances IAE 50

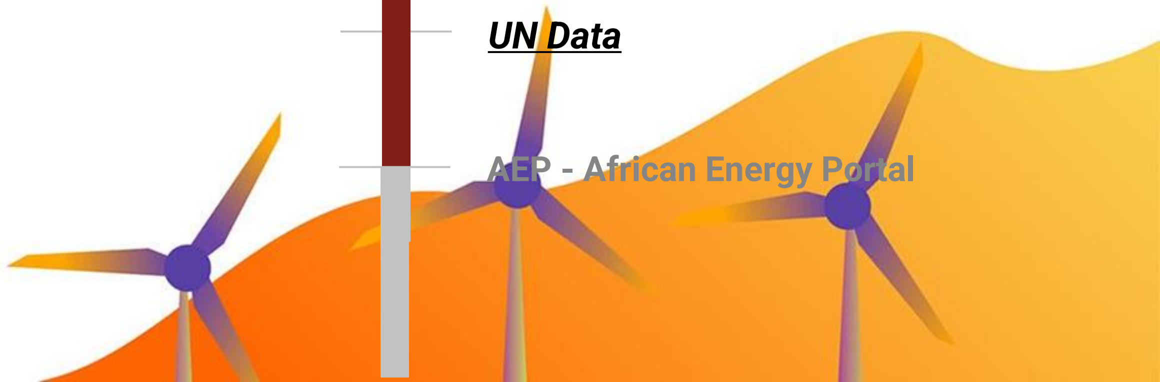
Tracking SDG 7

Our World in Data

Ember

UN Data

AEP - African Energy Portal



Data Cleaning and Preprocessing

Steps of Data Cleaning are as follows:

- Data types were corrected
- Units of measurement were standardized for uniformity,
- Categorical variable were encoded into numerical variable.
  For example - 'income_group' to 'income_group_num'
- Column names were standardized.

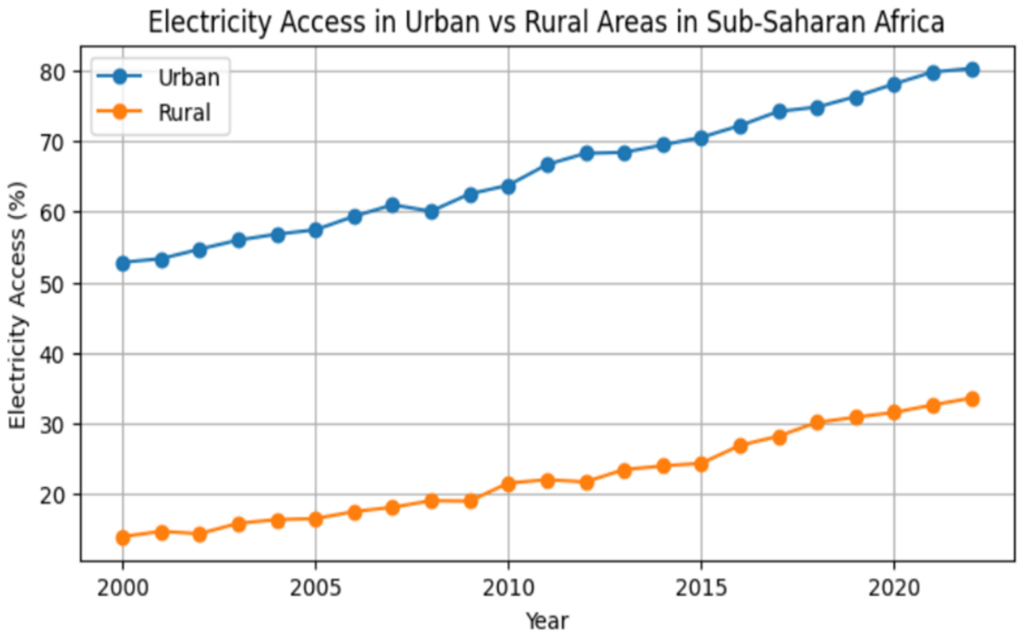
Data Cleaning and Preprocessing

- Calculated 'net_imports' values where 'el_exports' and 'el_imports' were present by combining data from two sources; replaced missing values with 0 where necessary.
- Resulted in the df_all dataset, covering the years 1990-2023 for 49 SSA countries and world regions (2166 rows, 39 columns).
- Imputed missing values using four methods, selecting the best method (Prophet) for modeling.
- Excluded South Sudan, removed some similar and less important metrics, and dropped years with the most missing data.
- Final data subset for modeling covered the years 2000-2022 for 48 SSA countries (1104 rows, 31 columns).

Exploratory Data Analysis

Comparative Analysis

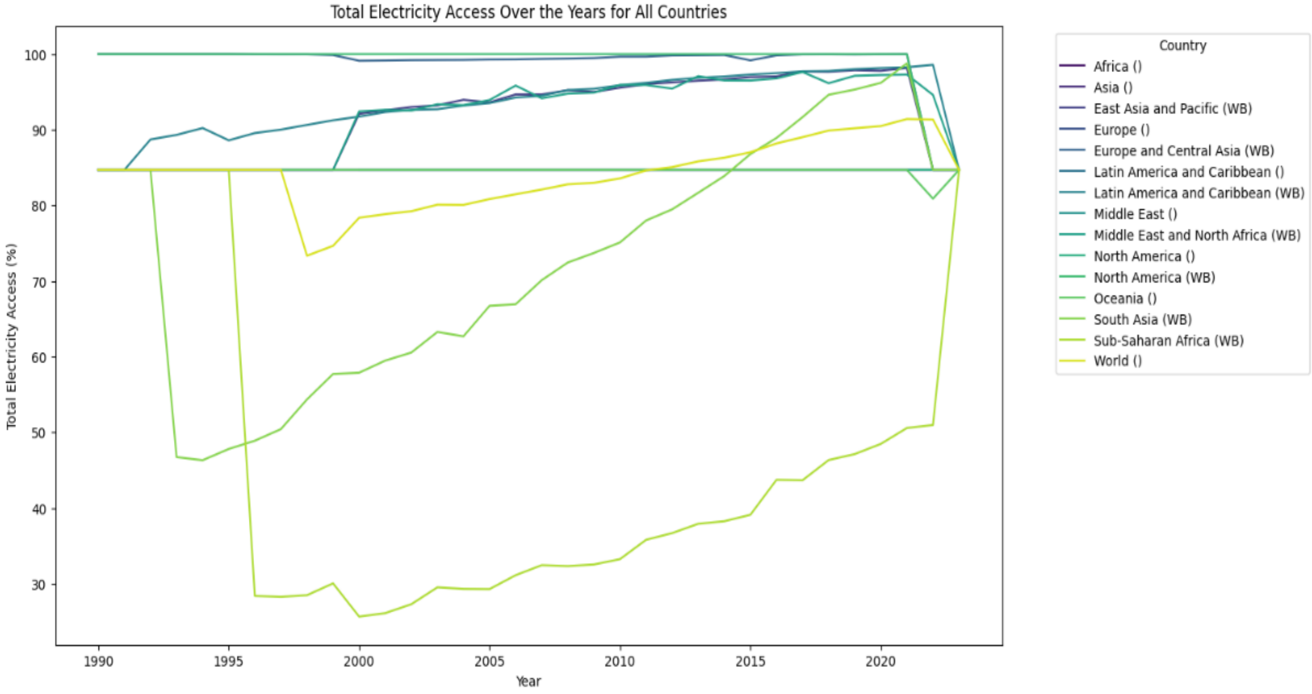
Electricity access within SSA regions between urban and rural areas.



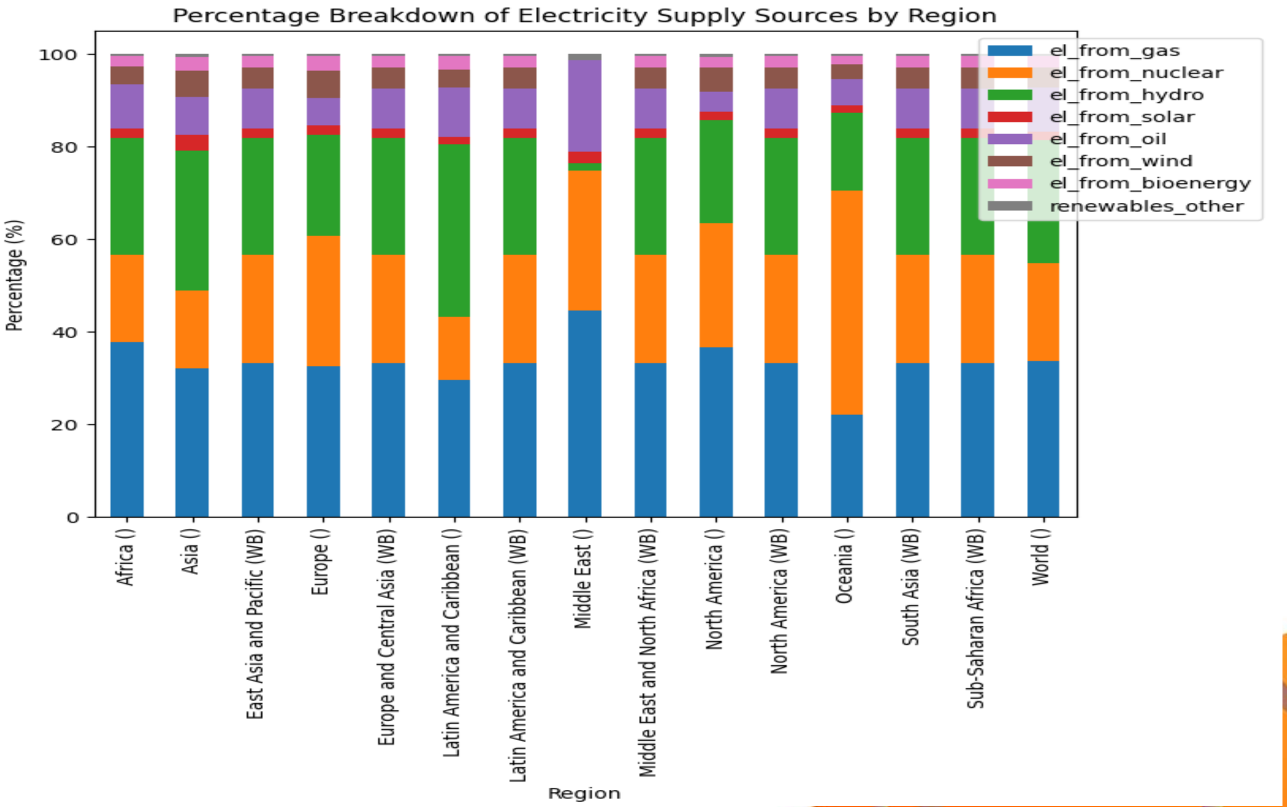
Exploratory Data Analysis

Comparative Analysis

Total Electricity access over all the countries comparing with Sub Saharan Africa..

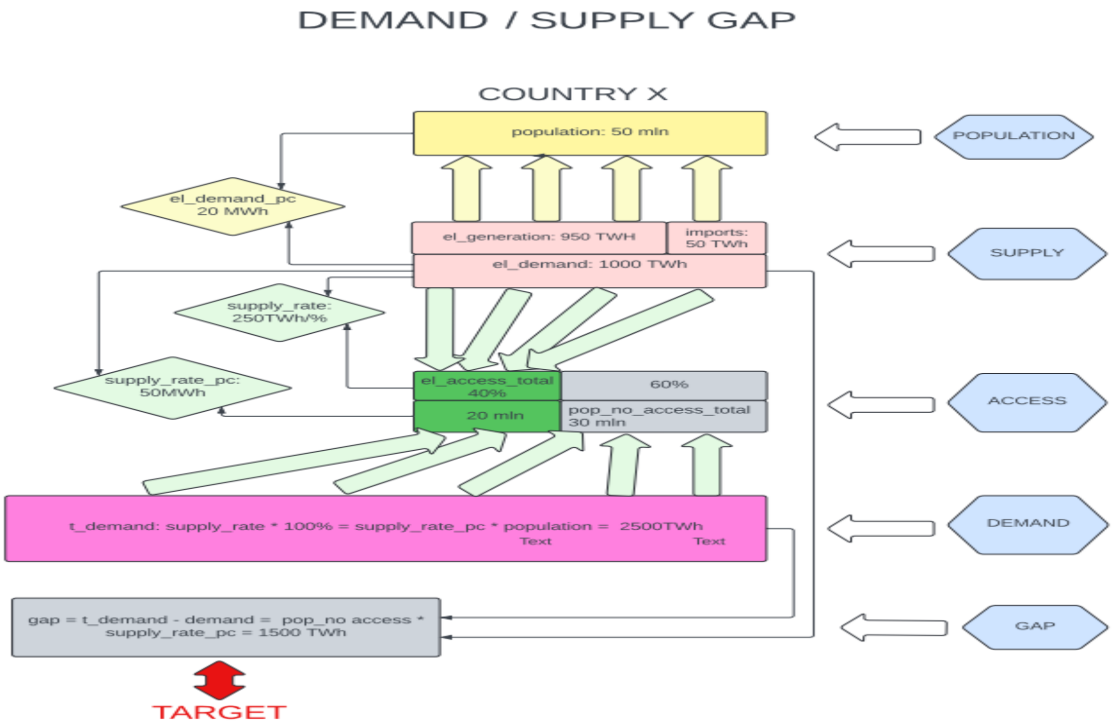


Exploratory Data Analysis



Composition of electricity generation sources in Sub-Saharan Africa compared to other regions.

Exploratory Data Analysis



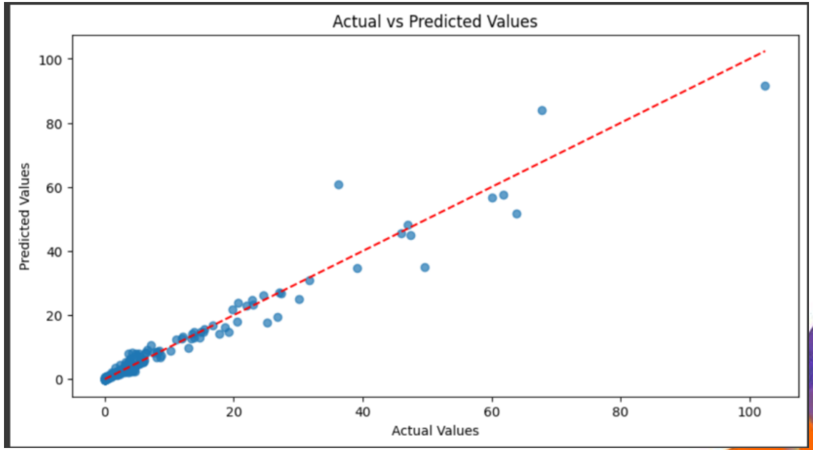
Exploratory Data Analysis

Category	Definition	Metric
SUPPLY	Actual amount of electricity delivered within a country, including generated and imported electricity (TWh)	**‘el_demand’**
ACCESS	Percentage of the total population with access to electricity (%)	**‘el_access_total’**
SUPPLY RATE	Amount of electricity delivered to 1% of the population with access to electricity (TWh/%)	`el_demand / el_access_total`
TARGET DEMAND	Amount of electricity needed to provide access to 100% of the population (TWh)	`100% * el_demand / el_access_total`
GAP	Difference between the amount of electricity needed and what is delivered (TWh)	`TARGET DEMAND - SUPPLY`

Model Training and Evaluation

1. Gradient Boosting

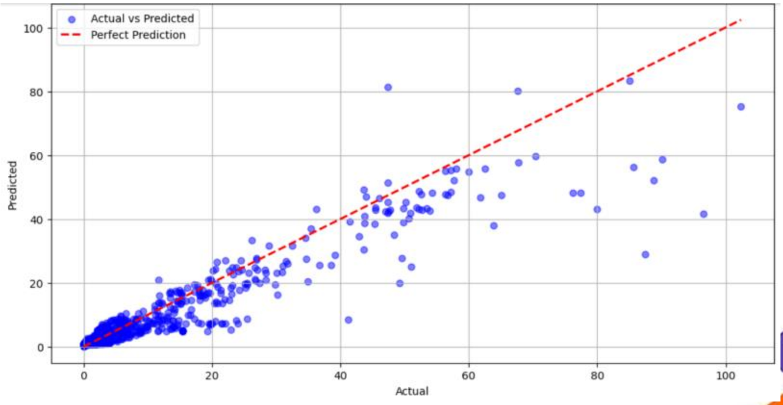
Mean Absolute Error (MAE): 1.2387692483234058
Root Mean Squared Error (RMSE): 2.925424694714788



Model Training and Evaluation

2. Sequential Model from Keras with Dense Layer Type

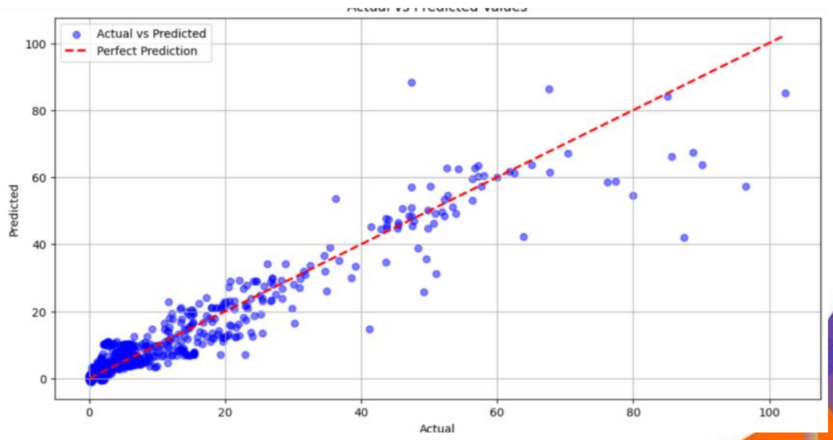
Lowest Train Loss: 32.5914
Lowest Test Loss: 36.7255
Combined RMSE: 5.1829



Model Training and Evaluation

3. Sequential Model with LSTM Layer Type

Train Loss: 20.6781
Test Loss: 14.7151
Combined RMSE: 4.4142



Model Training and Evaluation

Model Back-Testing

Model Training and Evaluation

Three main standard things that will affect the results are :-

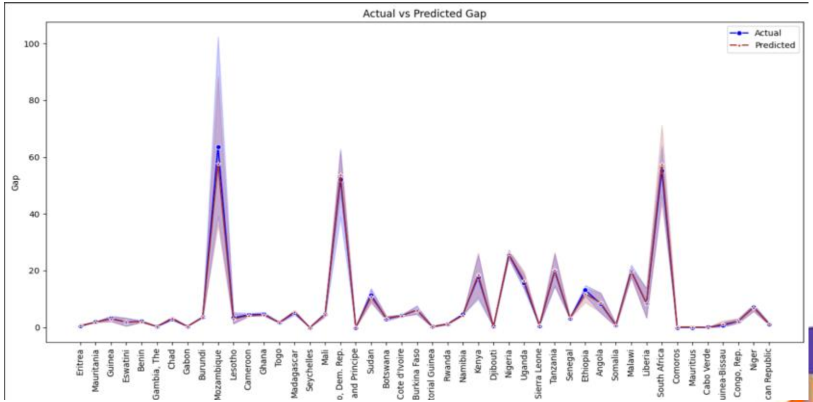
1. Feature Scaling
2. Outlier Handling
3. Feature Selection

This refinement is crucial for developing a robust forecasting tool for the electricity demand and supply gap in Sub-Saharan African countries. so we added **'t_demand'** and "**supply_rate**"

Model Training and Evaluation

1. Gradient Boosting

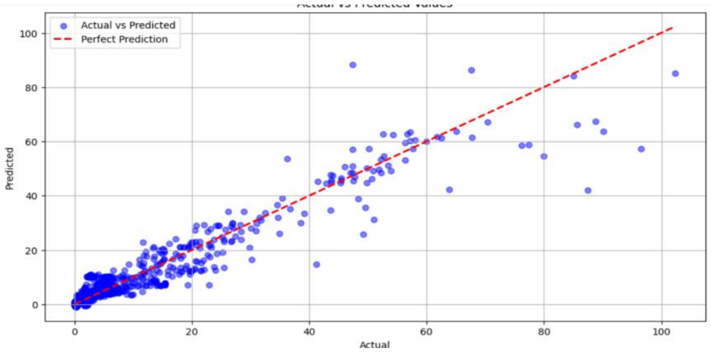
Mean Absolute Error (MAE): 0.5792657362366925
Mean Squared Error (MSE): 2.972355404579559
Root Mean Squared Error (RMSE): 1.724052030705442

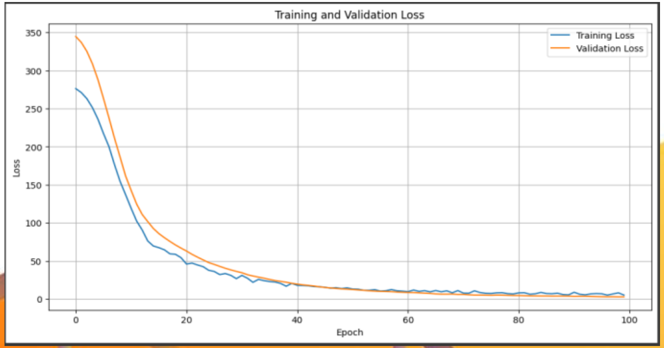


Model Training and Evaluation

2. Sequential Model with LSTM Layer Type

Train Loss: 2.2201149463653564
Test Loss: 2.0234854221343994
Combined RMSE: 1.4767493285573132
Combined MAE: 0.9081776610234112



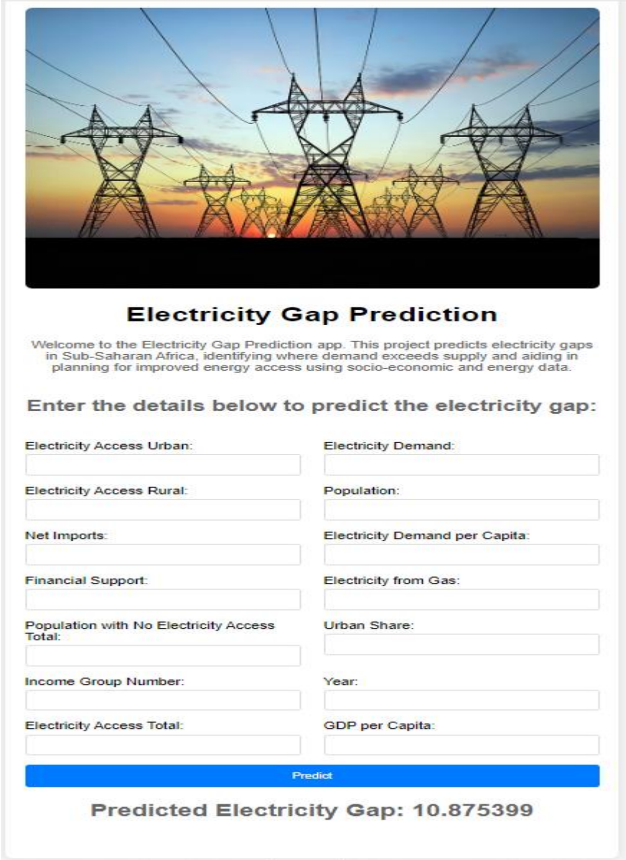


The LSTM model achieved the lowest train and test loss, with a combined RMSE of approximately 1.48 and a combined MAE of approximately 0.91.

Model Deployment

- Deploying Our Machine Learning Model with **Flask** and **Render**
- **Project Structure**:
  - **Key Components**:
    - **Flask Application**: Acts as the engine that powers our backend, handling requests, processing data, and returning results.
    - **HTML Frontend**: The interface through which our users interact with the application.
    - **Machine Learning Model**: The core component that makes predictions based on the input data.
- **Deployment with Render:**
  - **Steps**:
    - **Set Up Render**: Linked our project repository to Render.
    - **Configure Service**: Pointed Render to our Flask application.
    - **Deploy**: Render handles the deployment process, making our app live.

Model Deployment



A sneak peak at the website

Conclusion

1. Re-evaluation process significantly enhanced model performance.
2. Optimized LSTM model now delivers accurate predictions for electricity demand and supply gap in Sub-Saharan African countries.
3. Insights derived from the model are crucial for informed energy planning and policy-making.
4. Improved predictions aid in enhancing energy access in the region.

References

The following are the Github link for project proposal, deliveries such as EDA and Modeling notebook:-

- https://github.com/zaratti/BigGAN-Capstone-Project-Deliverables

References

The following are the dataset that we have used to create a dataset:-

- Tracking SDG 7 - The Energy Progress Report - https://trackingsdg7.esmap.org/results

- Our World in Data - https://ourworldindata.org/grapher/share-with-access-to-electricity-vs-per-capita-energy-consumption

- Ember - https://ember-climate.org/app/uploads/2022/07/yearly_full_release_long_format.csv

- UN Data - Unated Nations Data Retrieval System - https://data.un.org/Data.aspx?d=EDATA&f=cmID%3AEC

- AEP - Africa Energy Portal - https://africa-energy-portal.org/database

Thank You

Any Questions?