CARBON EMISSION PREDICTION

APAI PROJECT WORK A.Y. 2023/24
MASTER DEGREE IN ARTIFICIAL INTELIGENCE, UNIVERSITY OF BOLOGNA

FATEMEH BOZORGI
FATEMEH.BOZORGI@STUDIO.UNIBO.IT

OUTLINE

- Introduction
  - Overview of Green Computing
  - Importance of reducing carbon emissions in computing
  - Objectives of the project
- Carbon Emission Predictive Model
  - Data exploration and key findings from data analysis
  - Introduction to the predictive model and its purpose
  - Explanation of the architecture: LSTM model implemented in PyTorch
- Model Approaches
  - First approach: Predicting the next 6 hours based on the previous 24 hours
  - Second approach:Predicting the next month based on the last month.
  - Third approach:Predicting the next month based on the last year's data
- Future Work

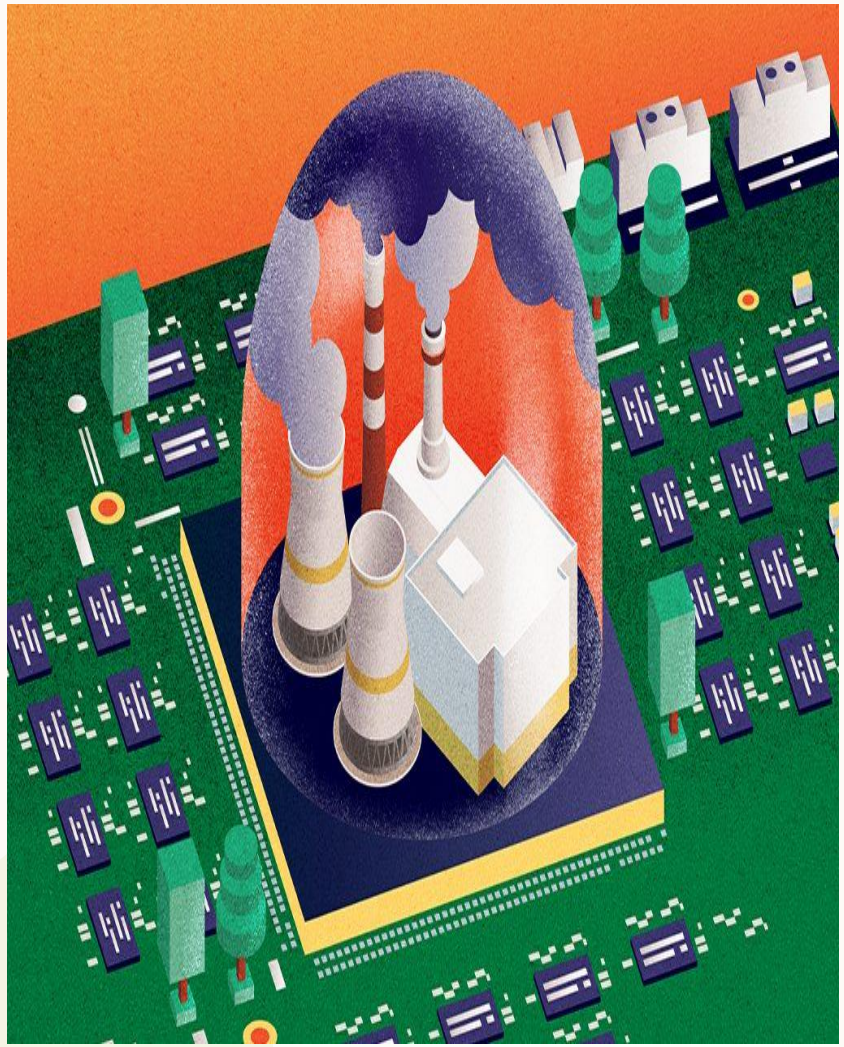
GREEN COMPUTING

- **Green computing, also called sustainable computing, is the process of developing and optimizing computer chips, systems, networks, and software in such a manner that can maximize efficiency by utilizing energy more efficiently and minimizing the negative environmental influence on the surrounding.**

S. G. Paul *et al.*, "A Comprehensive Review of Green Computing: Past, Present, and Future Research," in *IEEE Access*, vol. 11, 2023

- **Computing systems impact the environment at various stages of their life cycle. This includes the production phase, where silicon is processed, and the packaging and manufacturing of hardware occur. During operation, energy consumption contributes to environmental strain, and after their useful life, these systems generate electronic waste, posing significant recycling challenges.**

- **One of the most significant environmental impacts of large computing facilities is the emission of carbon dioxide (CO_2) and other greenhouse gases into the atmosphere.**



COMPARING CARBON FOOTPRINT IN HPC CENTERS AND OTHER SECTORS

- **Data centers generate the same amount of carbon emissions as global airlines.(1)**
- **Recent studies estimate that computing and communication technology sectors are responsible for 2% of global carbon emissions.(2)**
- **Transport accounts for around one-fifth of global carbon dioxide (CO_2) emissions — 24% if we only consider CO_2 emissions from energy.**
- **Aviation — while it often gets the most attention in discussions on action against climate change — accounts for 11.6% of transport emissions. It emits just under one billion tonnes of CO_2 each year — around 2.5% of total global emissions.(2)**

(1) Khosravi, Atefeh, Saurabh Kumar Garg, and Rajkumar Buyya. "Energy and carbon-efficient placement of virtual machines in distributed cloud data centers." *Euro-Par 2013 Parallel Processing: 19th International Conference, Aachen, Germany, August 26-30, 2013. Proceedings 19*. Springer Berlin Heidelberg, 2013.
(2) https://ourworldindata.org/co2-emissions-from-transport

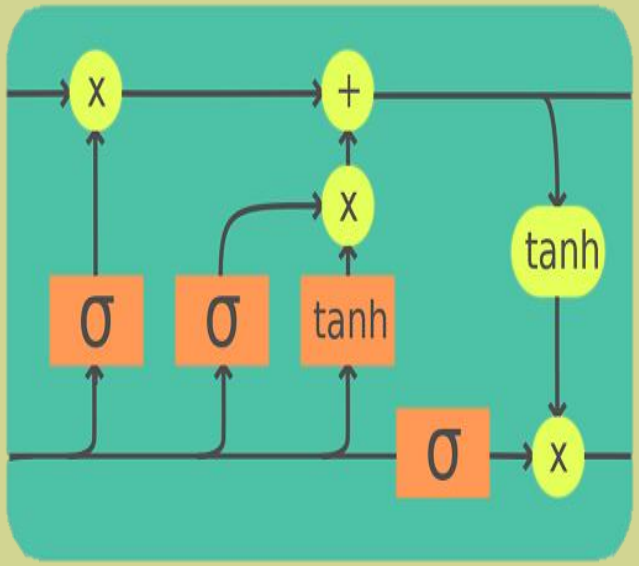
ENVIRONMENTAL EFFECT OF CO2

- **Carbon emissions contribute to global warming by trapping heat in the atmosphere, leading to climate change and its associated impacts such as extreme weather, sea level rise, and disruptions to ecosystems.**
- **Due to the serious problems that carbon emission has to the environment, we have to reduce the carbon footprint in all the involved sectors included data centers.**

The science is clear: sustainable development and climate action are inseparable (nature.com)

IDEA

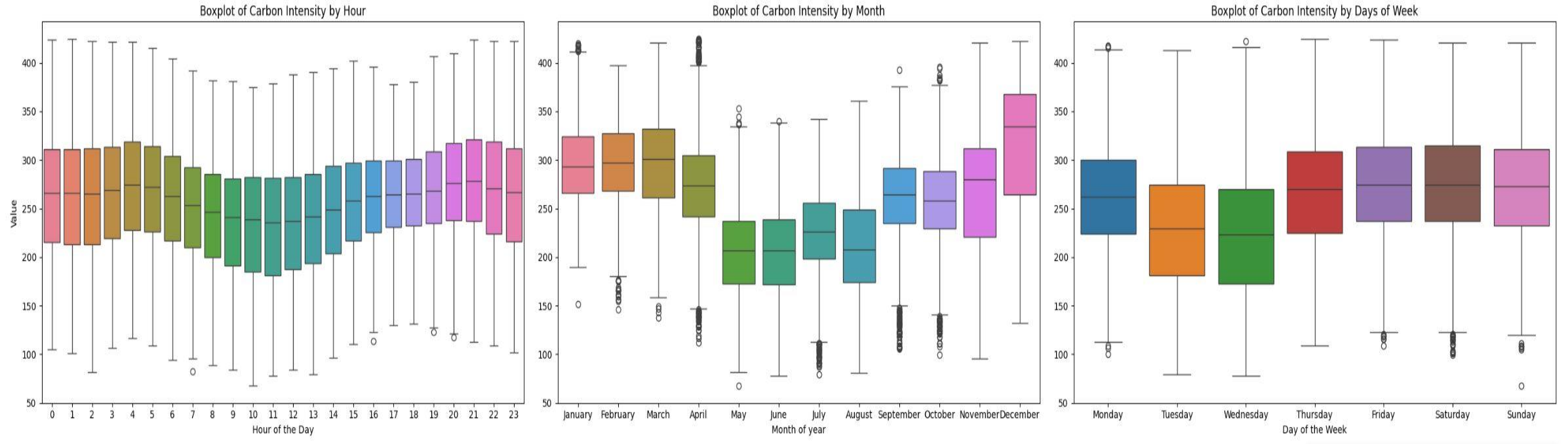
- **The purpose of this project is to build a predictor that when a task is received for doing something, the system decide to do it now or postpone it to another time to reduce the carbon.**
- **The architecture that is used as predictor is LSTM (Long Short-Term Memory) and it is implemented in pytorch.**
- **LSTM networks are effective in modeling time series data by capturing long-term dependencies, overcoming the vanishing gradient problem in traditional RNNs (Hochreiter & Schmidhuber, 1997).**



Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

DATA EXPLORATION

- After some data processing, we found some patterns in carbon emission during these three years.
- In some particular hours of day like at 11 or on Thursday and Wednesday or on May and June the carbon emission are lower than other times.
- So we can do some predictions to reduce the carbon emission based on the time that the operation will be done.



DATA

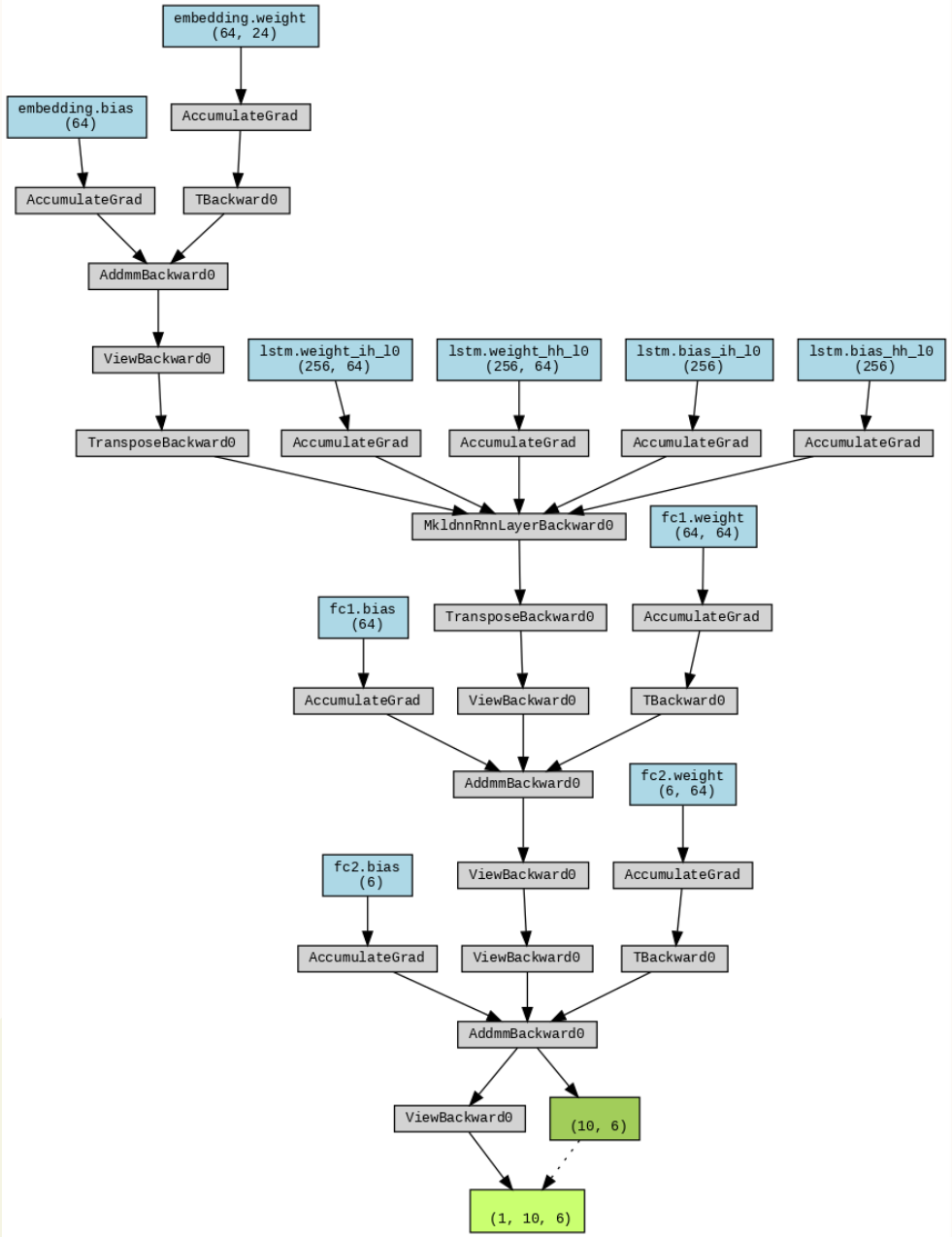
- **From 2021 to 2023, in a data center, each hours, the amount of data has been recorded and data set is created as shown in the picture.**
- **For this project we consider only one column of "Carbon Intensity gCO_2eq/kWh (direct)".**

	Datetime (UTC)	Country	Zone Name	Zone Id	Carbon Intensity gCO_2eq/kWh (direct)	Carbon Intensity gCO_2eq/kWh (LCA)	Low Carbon Percentage	Renewable Percentage	Data Source	Data Estimated	Data Estimation Method
0	2021010100	Italy	North Italy	IT-NO	303.29	389.74	35.12	30.82	entsoe.eu	False	NaN
1	2021010101	Italy	North Italy	IT-NO	303.45	387.76	36.15	33.12	entsoe.eu	False	NaN
2	2021010102	Italy	North Italy	IT-NO	295.04	377.86	37.89	33.65	entsoe.eu	False	NaN
3	2021010103	Italy	North Italy	IT-NO	295.56	380.36	37.14	32.05	entsoe.eu	False	NaN
4	2021010104	Italy	North Italy	IT-NO	308.06	394.66	34.50	31.71	entsoe.eu	False	NaN

- **This neural network consists of an embedding layer that first transforms the input into a dense representation.**
- **Following this, a LSTM layer captures temporal dependencies in both forward and reverse directions.**

- **The network then uses two fully connected layers to further process the information and produce the final output.**

- **Additionally, it employs batch normalization to ensure stable and efficient training. The model is flexible, allowing for adjustments in layer sizes and the use of dropout for regularization.**



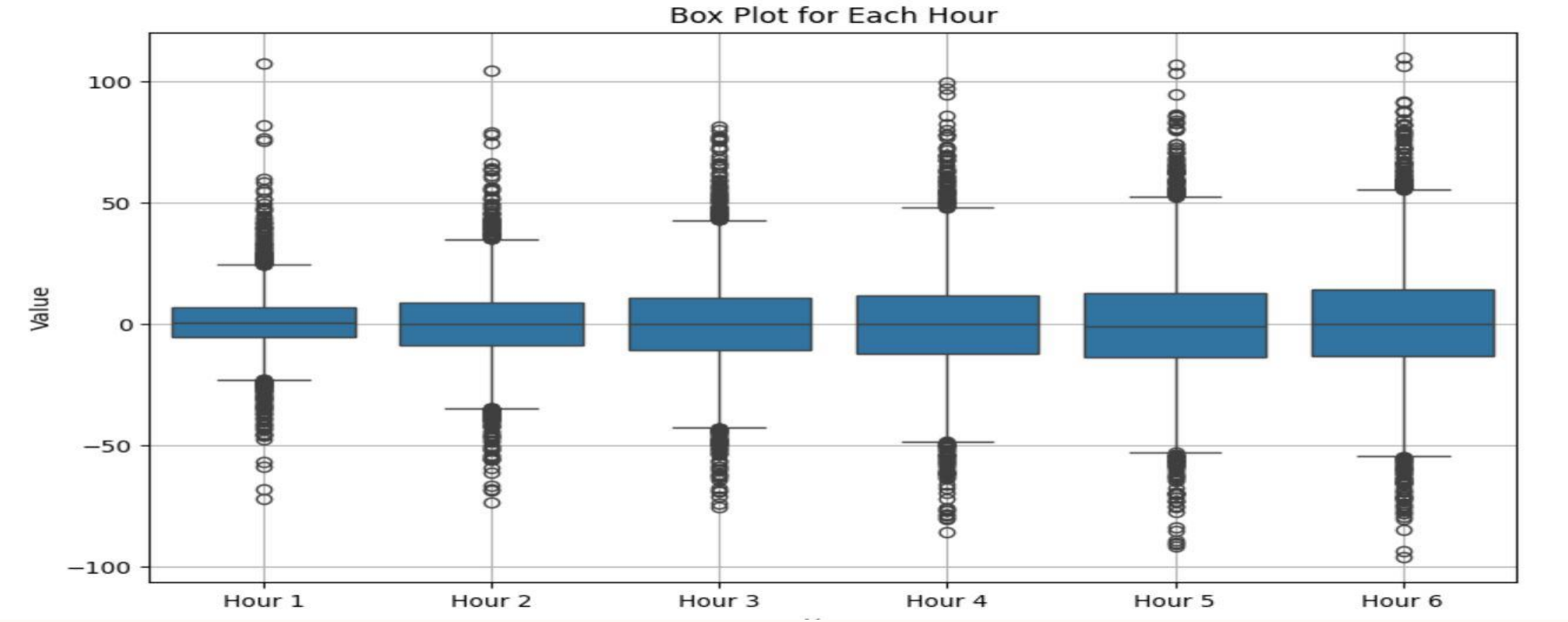
THREE DIFFERENT APPROACHES

For the architecture we consider three different set up:

- **First one is to make the data set in a way that the model receive the 24 hours and then predict the next 6 hours**
- **The other approach is the model predict the next month based on the last month.**
- **The last approach is the model take the last year data and predict the next month**

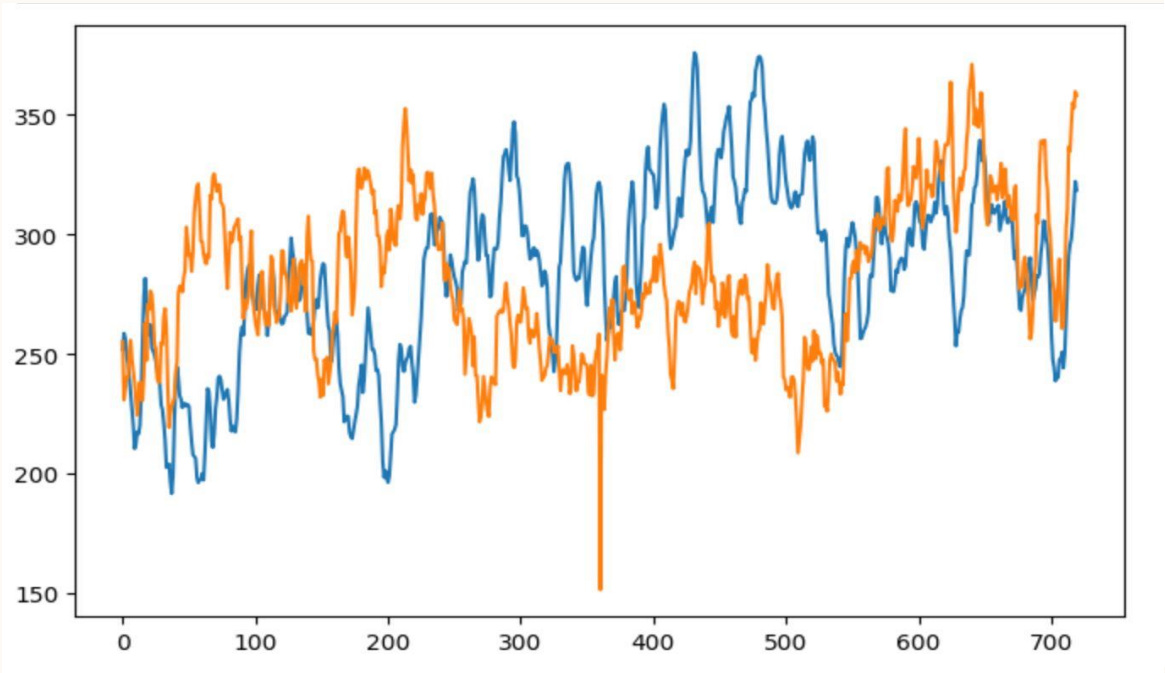
FIRST APPROACH

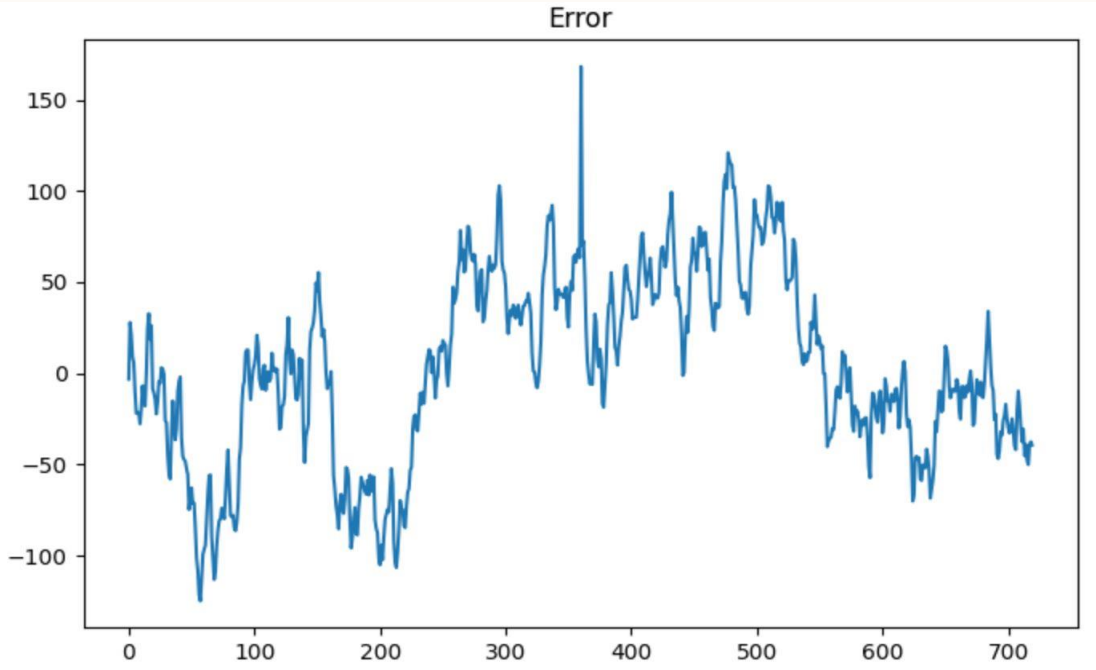
- **As it can be expected the model predicts the first hour better than the later ones and here is the errors for each hour calculated on the test set:**



SECOND APPROACH

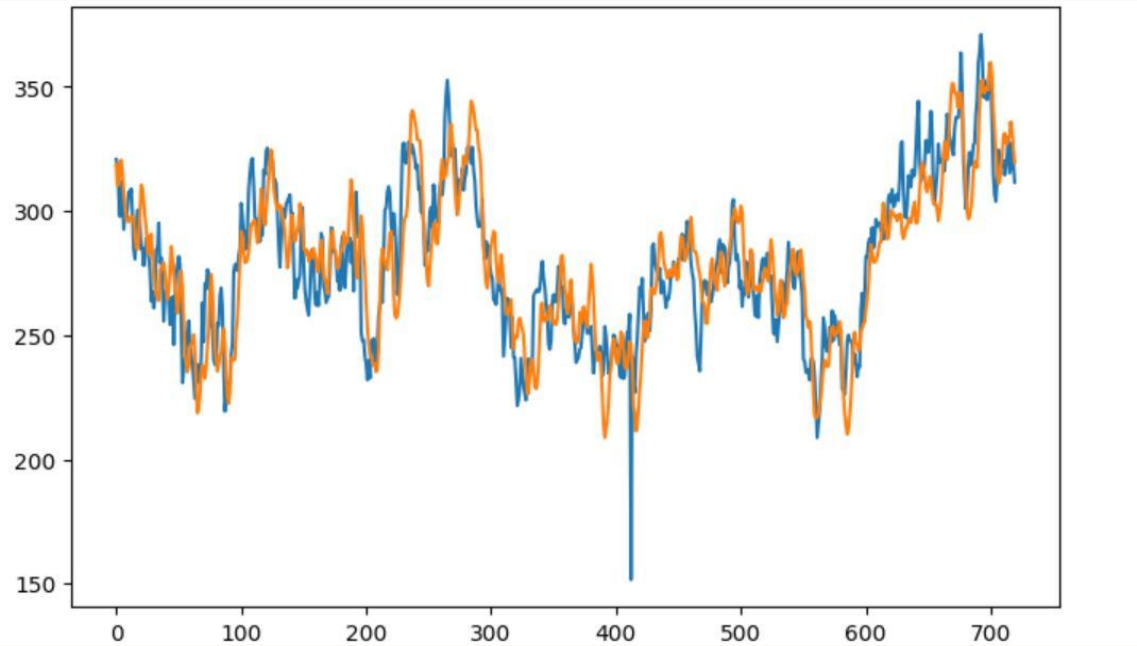
- **The second approach that take a month and predict the next month is less accurate than the previous model and as you can see the plot below show the real values(blue line) and predicted one(orange) for a month in test set.**

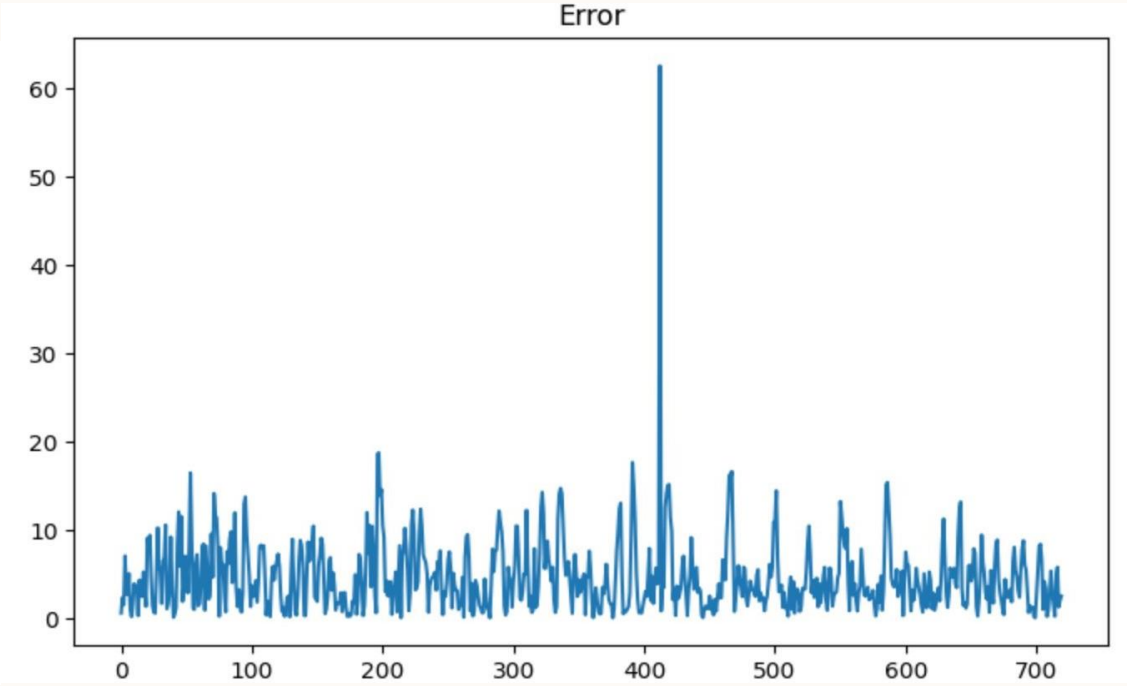




THIRD APPROACH

- **The third approach that take a year and predict the next month is more accurate than the previous models and as you can see the plot below show the real values(blue line) and predicted one(orange) for a month in test set.**





FUTURE WORKS

- Expansion of the dataset to include more variables such as temperature, humidity, and renewable energy availability.
- Integration of real-time data from multiple data centers to improve prediction accuracy.
- Implementation of adaptive scheduling algorithms to further optimize the timing of computational tasks based on predicted carbon intensity.
- Exploration of other machine learning models, such as transformers, to enhance prediction capabilities.

THANK YOU