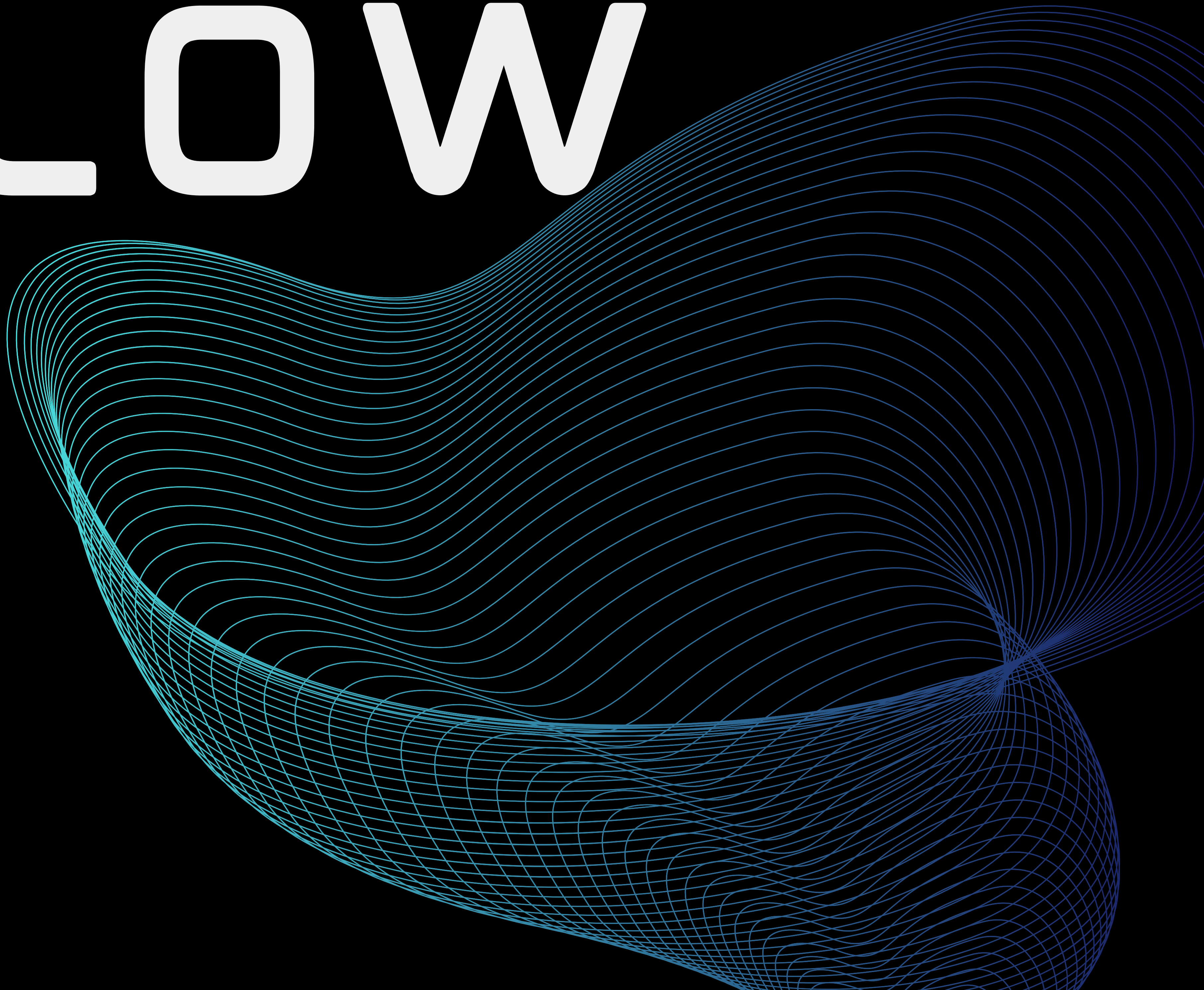
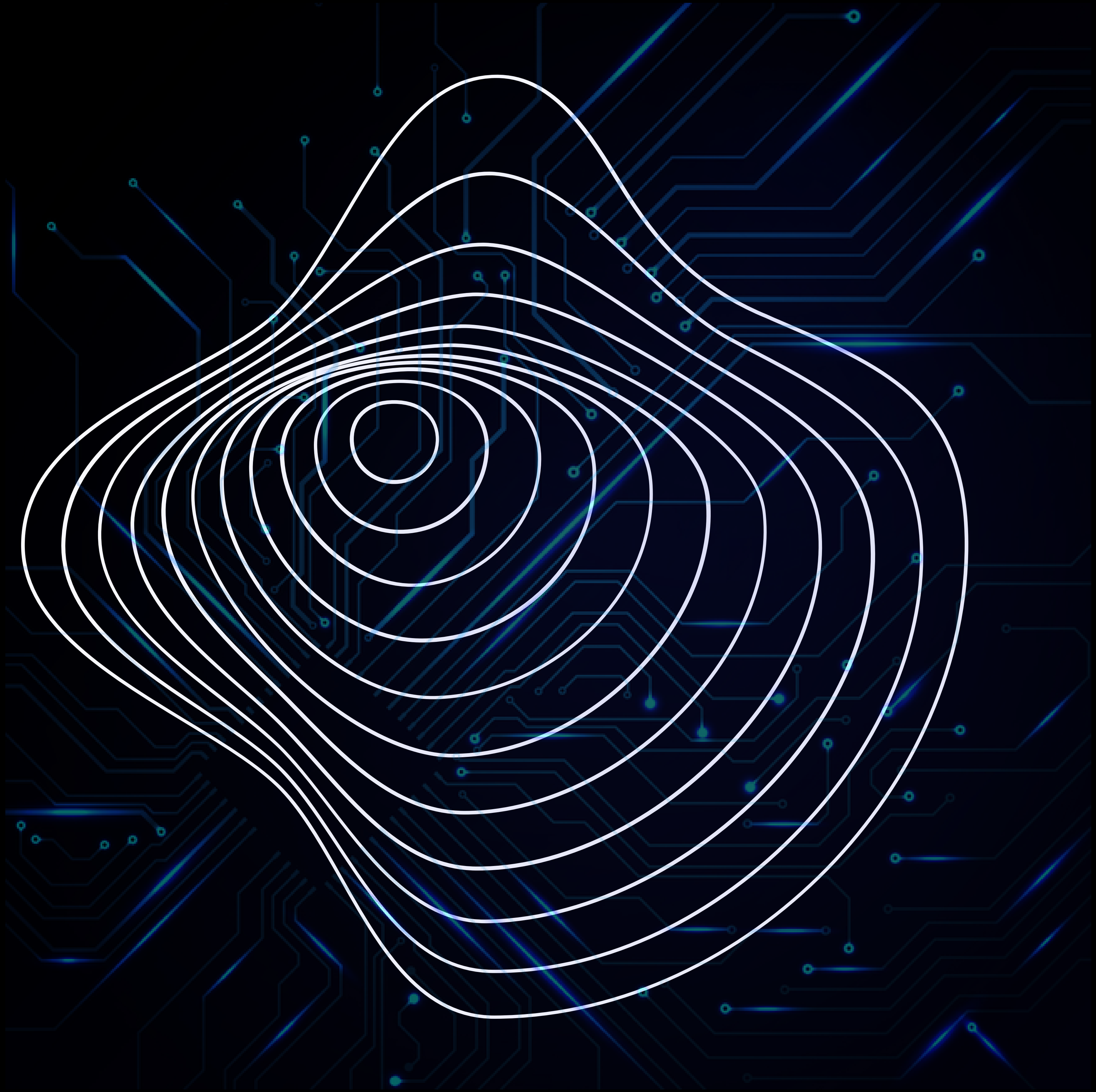
WATTFLOW

Machine Learning
for STP Energy Forecasting





Problem Statement

Literature Survey

Dataset

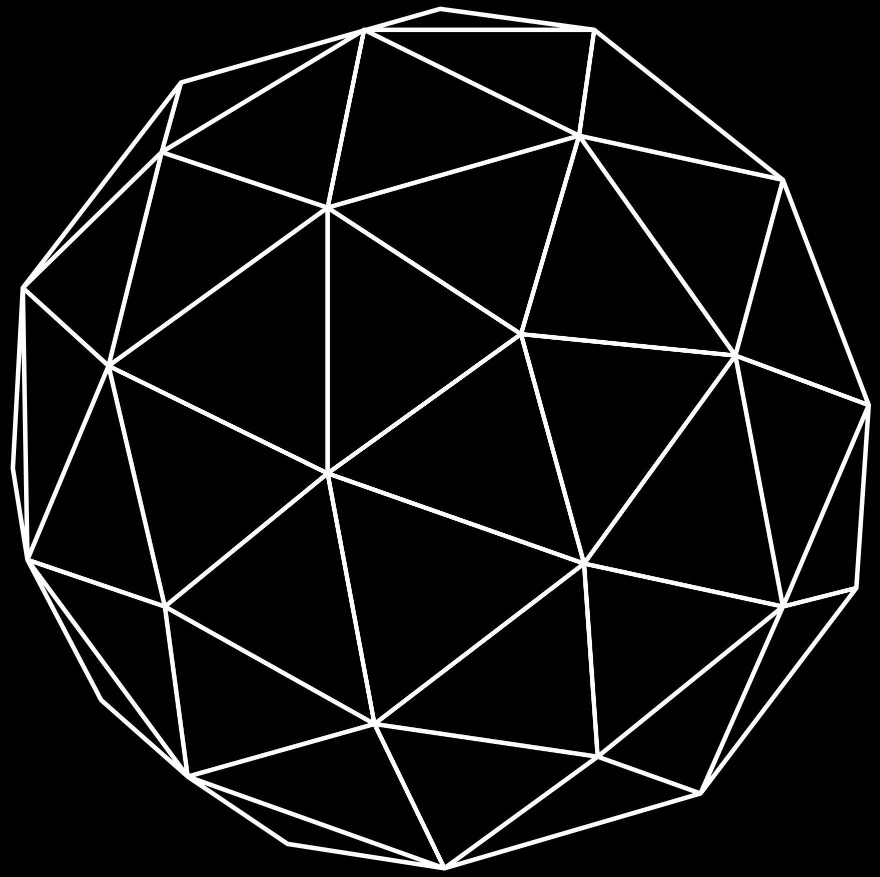
Methodology

Metrics

Which problem at Plaksha?

STP

Problem Statement



The energy consumption of the STP is a critical operational metric that directly impacts our operational costs and environmental footprint.

- Gain insights into the factors affecting energy consumption.
- Implement energy-saving strategies based on real-time predictions.
- Reduce operational costs and minimize environmental impact.

Literature Survey

Paper 1: Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach

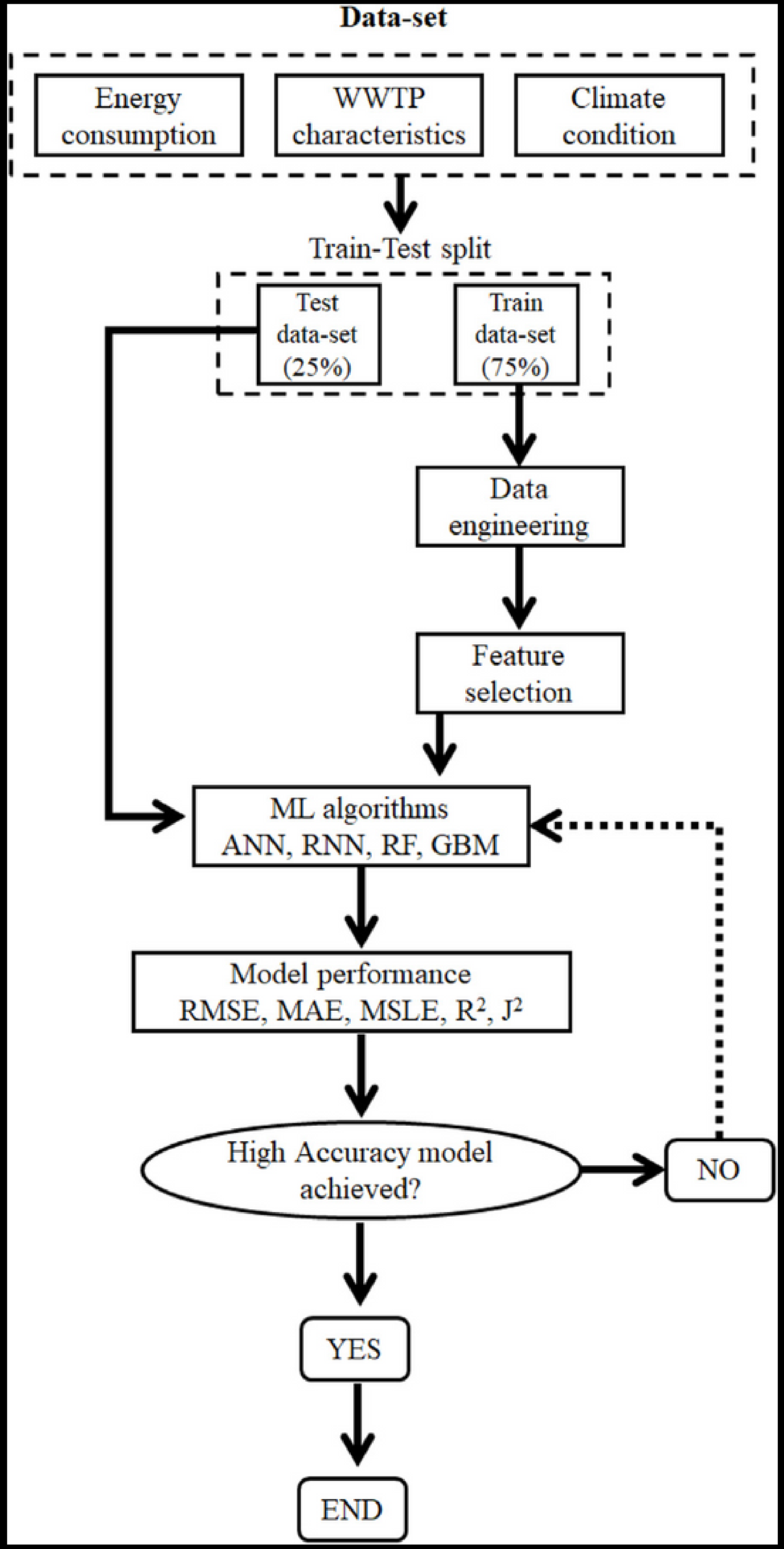


Fig: Methodology
(B. Faramarz et al.,)(

East Melbourne WWTP (2014-19)
+
Weather station data

Findings:

1. Positive correlation of energy consumption and weather data
2. TN, BOD, ammonia, daily temperature, humidity, and influent flow had the highest impact on the EC in Melbourne east WWTP.
3. GBM algorithm revealed the best performance for prediction among other algorithms showing its prediction power in nonlinear irregular patterns

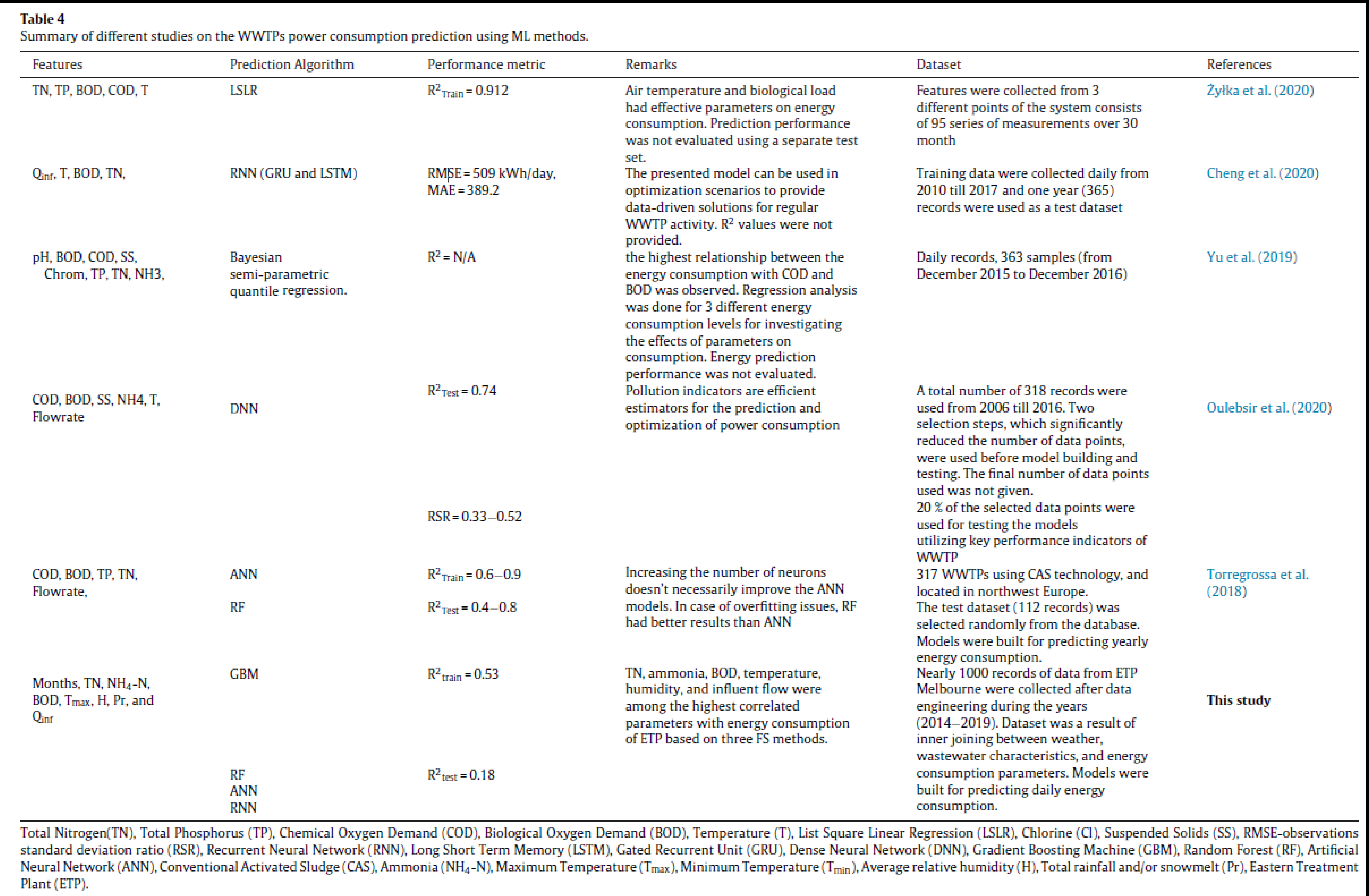


Table 4

Summary of different studies on the WWTPs power consumption prediction using ML methods.

Features	Prediction Algorithm	Performance metric	Remarks	Dataset	References
TN, TP, BOD, COD, T	LSLR	$R^2_{Train} = 0.912$	Air temperature and biological load had effective parameters on energy consumption. Prediction performance was not evaluated using a separate test set.	Features were collected from 3 different points of the system consists of 95 series of measurements over 30 month	Żyłka et al. (2020)
Q_{inf}, T, BOD, TN,	RNN (GRU and LSTM)	RMSE = 509 kWh/day, MAE = 389.2	The presented model can be used in optimization scenarios to provide data-driven solutions for regular WWTP activity. R^2 values were not provided.	Training data were collected daily from 2010 till 2017 and one year (365) records were used as a test dataset	Cheng et al. (2020)
pH, BOD, COD, SS, Chrom, TP, TN, NH3,	Bayesian semi-parametric quantile regression.	R^2 = N/A	the highest relationship between the energy consumption with COD and BOD was observed. Regression analysis was done for 3 different energy consumption levels for investigating the effects of parameters on consumption. Energy prediction performance was not evaluated.	Daily records, 363 samples (from December 2015 to December 2016)	Yu et al. (2019)
COD, BOD, SS, NH4, T, Flowrate	DNN	$R^2_{Test} = 0.74$	Pollution indicators are efficient estimators for the prediction and optimization of power consumption	A total number of 318 records were used from 2006 till 2016. Two selection steps, which significantly reduced the number of data points, were used before model building and testing. The final number of data points used was not given.	Oulebsir et al. (2020)
		RSR = 0.33–0.52		20 % of the selected data points were used for testing the models utilizing key performance indicators of WWTP	
COD, BOD, TP, TN, Flowrate,	ANN	$R^2_{Train} = 0.6–0.9$	Increasing the number of neurons doesn't necessarily improve the ANN models. In case of overfitting issues, RF had better results than ANN	317 WWTPs using CAS technology, and located in northwest Europe.	Torregrossa et al. (2018)
	RF	$R^2_{Test} = 0.4–0.8$		The test dataset (112 records) was selected randomly from the database. Models were built for predicting yearly energy consumption.	
Months, TN, NH_4-N, BOD, T_{max}, H, Pr, and Q_{inf}	GBM	$R^2_{train} = 0.53$	TN, ammonia, BOD, temperature, humidity, and influent flow were among the highest correlated parameters with energy consumption of ETP based on three FS methods.	Nearly 1000 records of data from ETP Melbourne were collected after data engineering during the years (2014–2019). Dataset was a result of inner joining between weather, wastewater characteristics, and energy consumption parameters. Models were built for predicting daily energy consumption.	**This study**
	RF ANN RNN	$R^2_{test} = 0.18$			

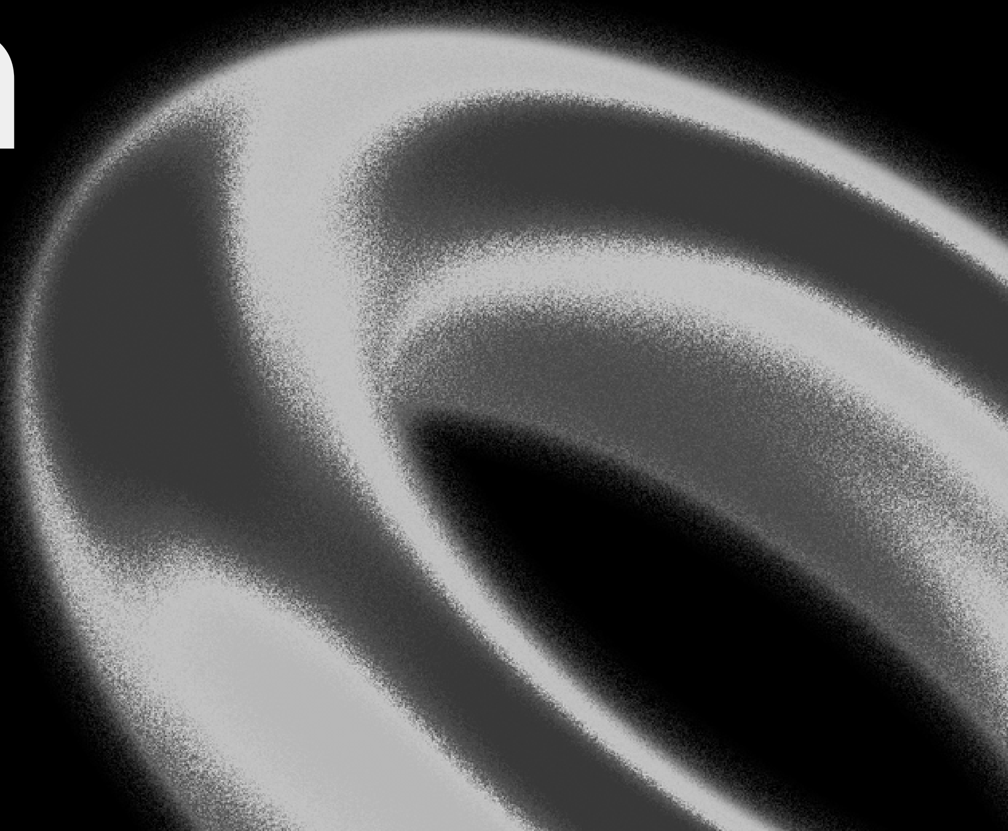
Total Nitrogen(TN), Total Phosphorus (TP), Chemical Oxygen Demand (COD), Biological Oxygen Demand (BOD), Temperature (T), List Square Linear Regression (LSLR), Chlorine (Cl), Suspended Solids (SS), RMSE-observations standard deviation ratio (RSR), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), Dense Neural Network (DNN), Gradient Boosting Machine (GBM), Random Forest (RF), Artificial Neural Network (ANN), Conventional Activated Sludge (CAS), Ammonia (NH_4-N), Maximum Temperature (T_{max}), Minimum Temperature (T_{min}), Average relative humidity (H), Total rainfall and/or snowmelt (Pr), Eastern Treatment Plant (ETP).

OTHER STUDIES

Paper 2: Review Paper on similar studies

Gap identified through literature survey:

Limited Indian Energy consumption prediction studies



Features/Dataset Preprocessing

Preprocessing

All data was present in undigtized physical registers, thus required extensive manual and image to table extractor programs

Use of an API to get relevant hour by hour weather data

Since all data was incremental, data was subjected to a difference operation with the value underneath it(on the spreadsheet)

Removal of factually wrong data

LOG SHEET Operation & Maintenance of STP/ETP Capacity : 50 KLD Location : Alaknanda NEW UNIVERSITY ECO GROUP

SHIFT I Date : 20-9-23 Operator Name : Amrinder Singh Operator Signature : Asingh

Time Hr. | Sewage Feed Pump | Air Blower | Sludge Recycling Pump | Dosing Pump | Filter Feed Pump | Filter Press Feed Pump | Inlet Flow Meter Reading | Outlet Flow Meter Reading | Energy Meter Reading | Remarks

Back wash 10:10 to 10:20 10:20 to 10:30

SHIFT II Date : 20-9-23 Operator Name : JUJHAR SINGH Operator Signature :

Back wash 3:10 to 3:20 3:20 to 3:30

SHIFT III Date : 20-9-23 Operator Name : Harpreet Singh Operator Signature : Asingh

Backwash 19:10 to 19:20 19:20 to 19:30

Total Inlet KLD : 42 KLD

Total Outlet KLD : 37 KLD

Total Daily Energy Consumption KW / Day : 59 KWh

Chemical Consumption Per Day :

Sodium Hypo Chloride : 1 litre

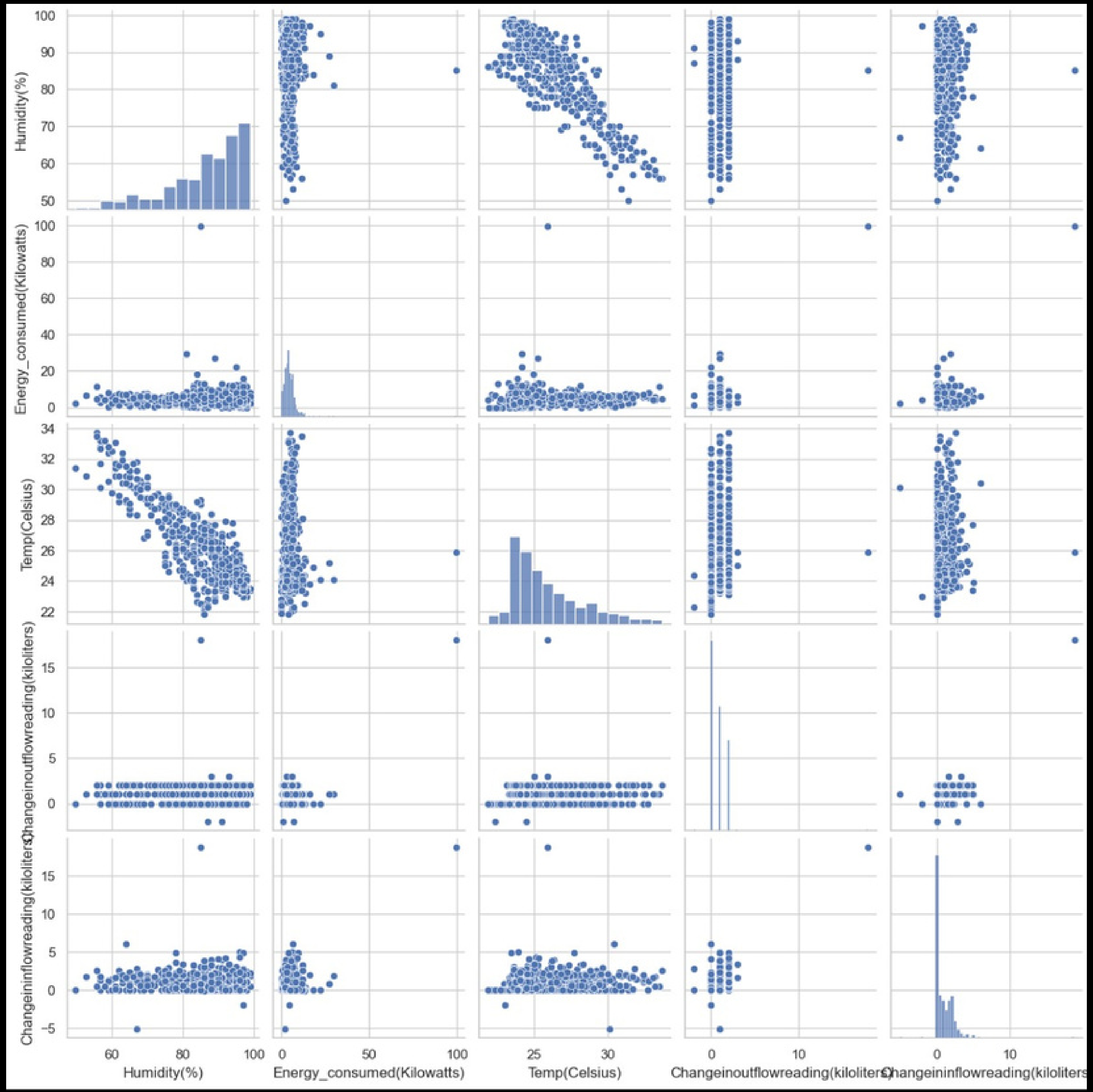
Polyelectrolyte :

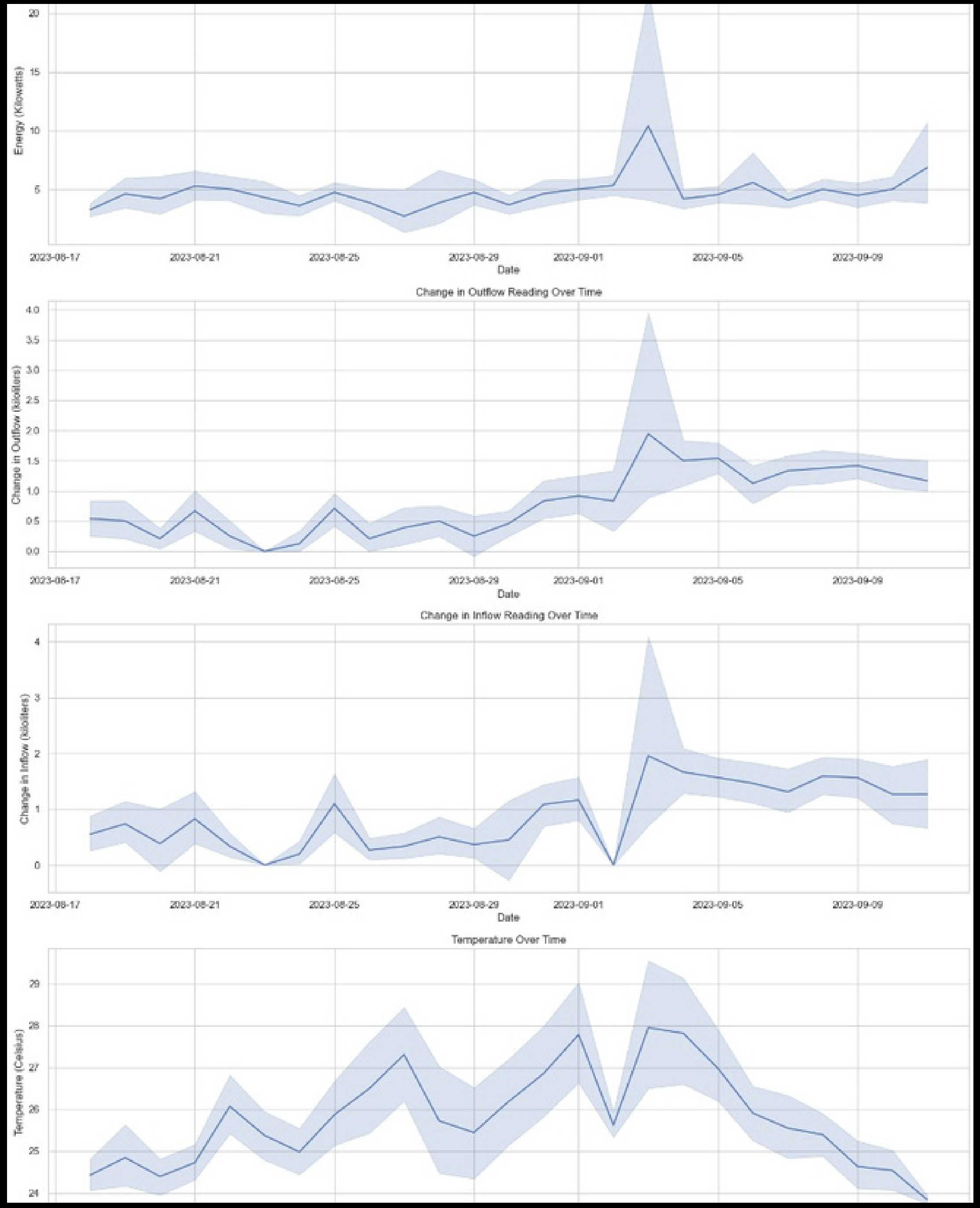
Miscll. :

(Plant in-charge)

(Customer Representative)

Preprocessing

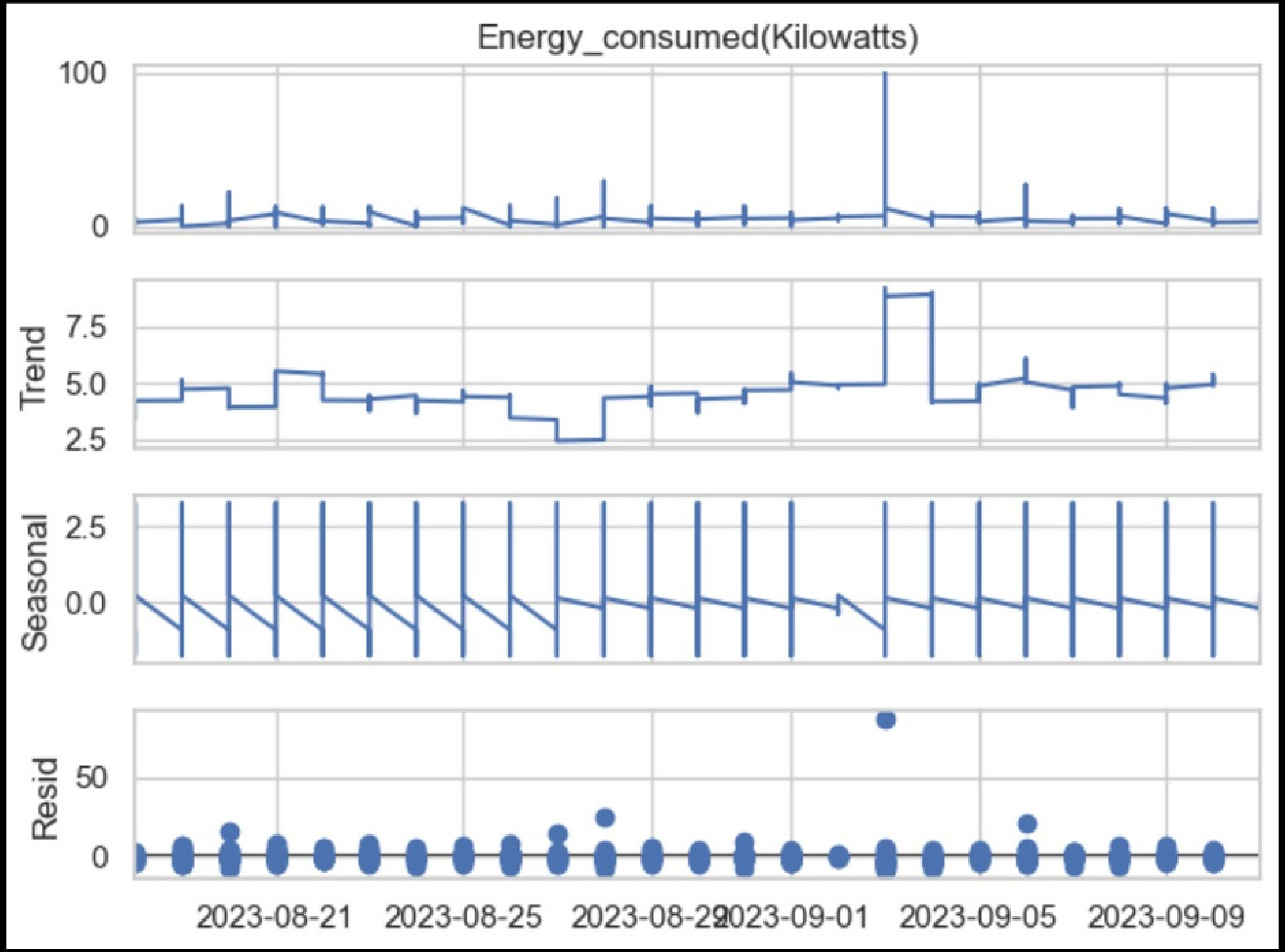


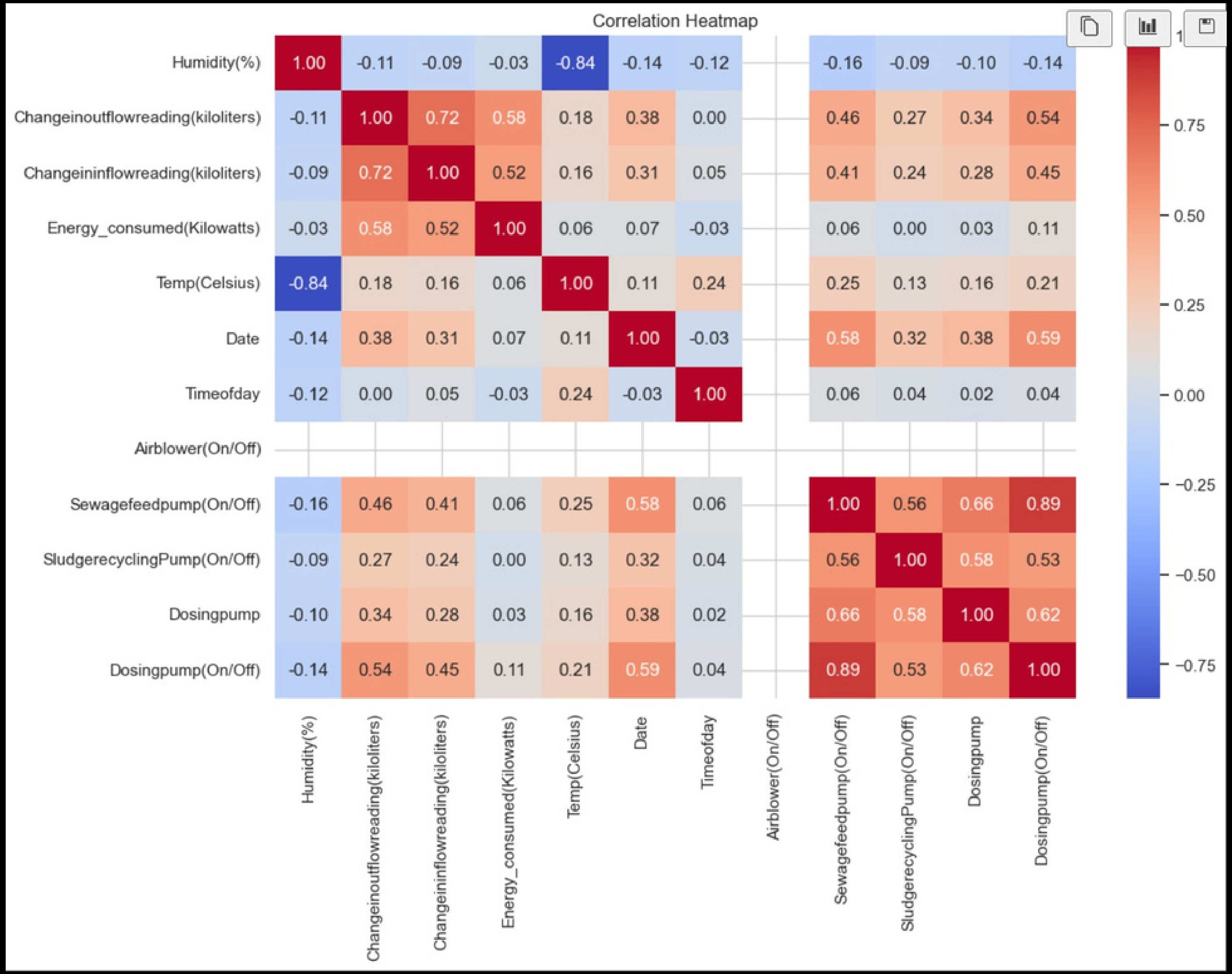

Change in Outflow Reading Over Time

Change in Inflow Reading Over Time

Temperature Over Time

Preprocessing

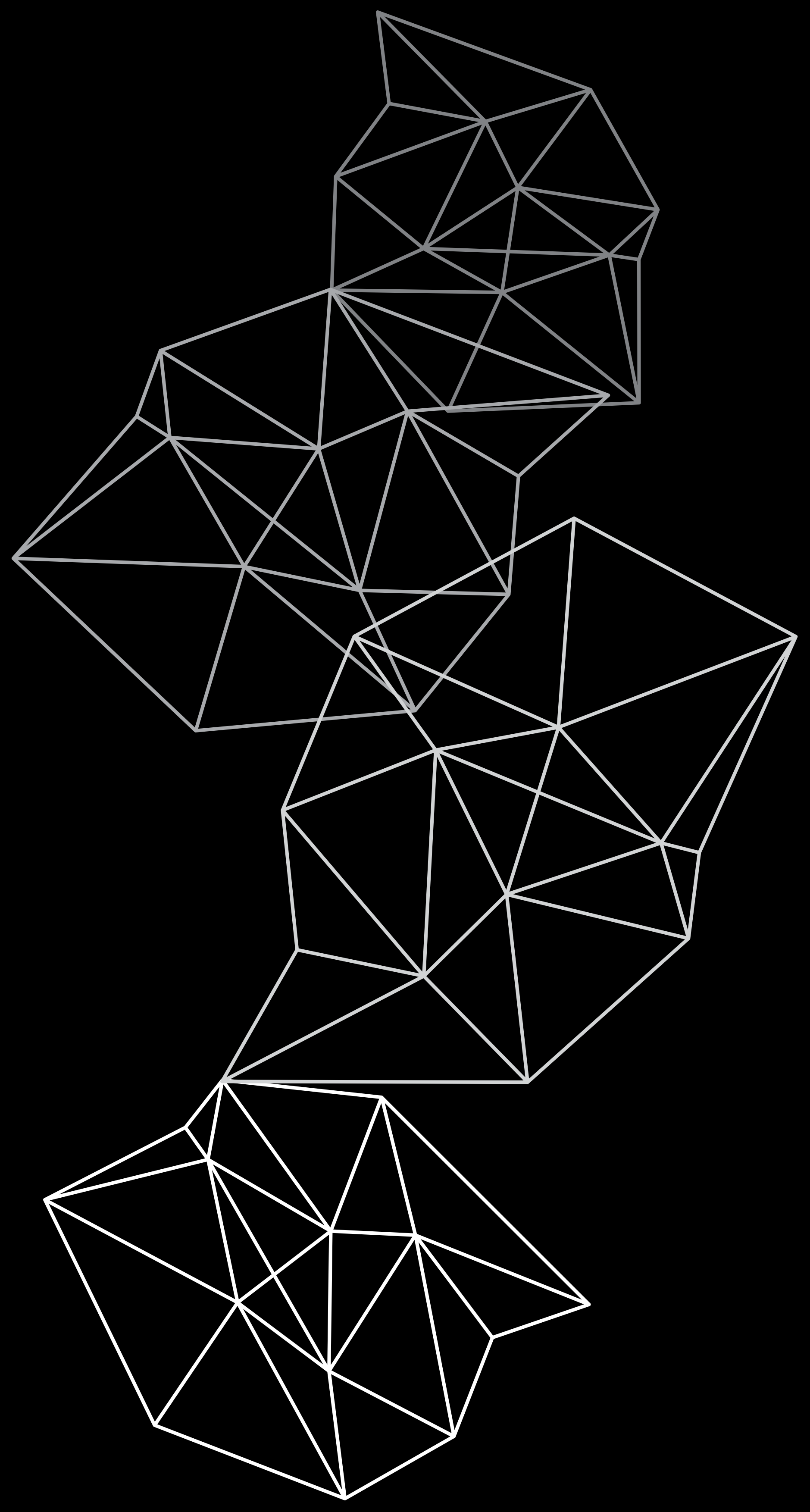




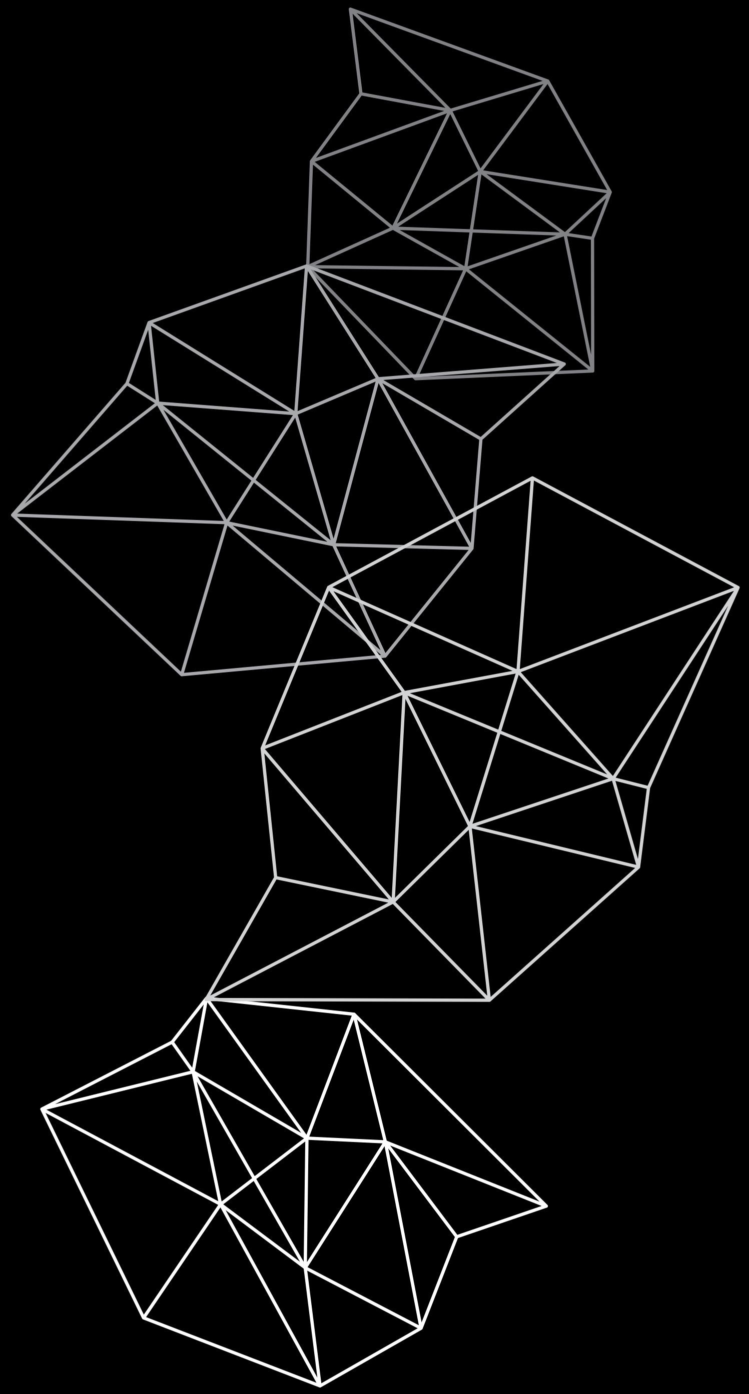
	Humidity(%)	Changeinoutflowreading(kiloliters)	Changeininflowreading(kiloliters)	Energy_consumed(Kilowatts)	Temp(Celsius)	Date	Timeofday	Airblower(On/Off)	Sewagefeedpump(On/Off)	SludgerecyclingPump(On/Off)	Dosingpump	Dosingpump(On/Off)
Humidity(%)	1.00	-0.11	-0.09	-0.03	-0.84	-0.14	-0.12		-0.16	-0.09	-0.10	-0.14
Changeinoutflowreading(kiloliters)	-0.11	1.00	0.72	0.58	0.18	0.38	0.00		0.46	0.27	0.34	0.54
Changeininflowreading(kiloliters)	-0.09	0.72	1.00	0.52	0.16	0.31	0.05		0.41	0.24	0.28	0.45
Energy_consumed(Kilowatts)	-0.03	0.58	0.52	1.00	0.06	0.07	-0.03		0.06	0.00	0.03	0.11
Temp(Celsius)	-0.84	0.18	0.16	0.06	1.00	0.11	0.24		0.25	0.13	0.16	0.21
Date	-0.14	0.38	0.31	0.07	0.11	1.00	-0.03		0.58	0.32	0.38	0.59
Timeofday	-0.12	0.00	0.05	-0.03	0.24	-0.03	1.00		0.06	0.04	0.02	0.04
Airblower(On/Off)												
Sewagefeedpump(On/Off)	-0.16	0.46	0.41	0.06	0.25	0.58	0.06		1.00	0.56	0.66	0.89
SludgerecyclingPump(On/Off)	-0.09	0.27	0.24	0.00	0.13	0.32	0.04		0.56	1.00	0.58	0.53
Dosingpump	-0.10	0.34	0.28	0.03	0.16	0.38	0.02		0.66	0.58	1.00	0.62
Dosingpump(On/Off)	-0.14	0.54	0.45	0.11	0.21	0.59	0.04		0.89	0.53	0.62	1.00

Features

- Date(index)
- Hourly time
- Filters/pumps: Sewage feed pump, Sludge recycling pump, Dosing pump, Filter feed pump, air blower
- Inlet flow meter reading
- Outlet flow meter reading
- Energy meter reading
- Derived Features- Consumed Energy, Volume of inlet/outlet flow



Features



Sewage Feed Pump: designed to move sewage from one location to another within a sewage system

Sludge Recycling Pump: used to circulate or transfer sludge—a semi-solid byproduct of wastewater treatment—within a treatment system

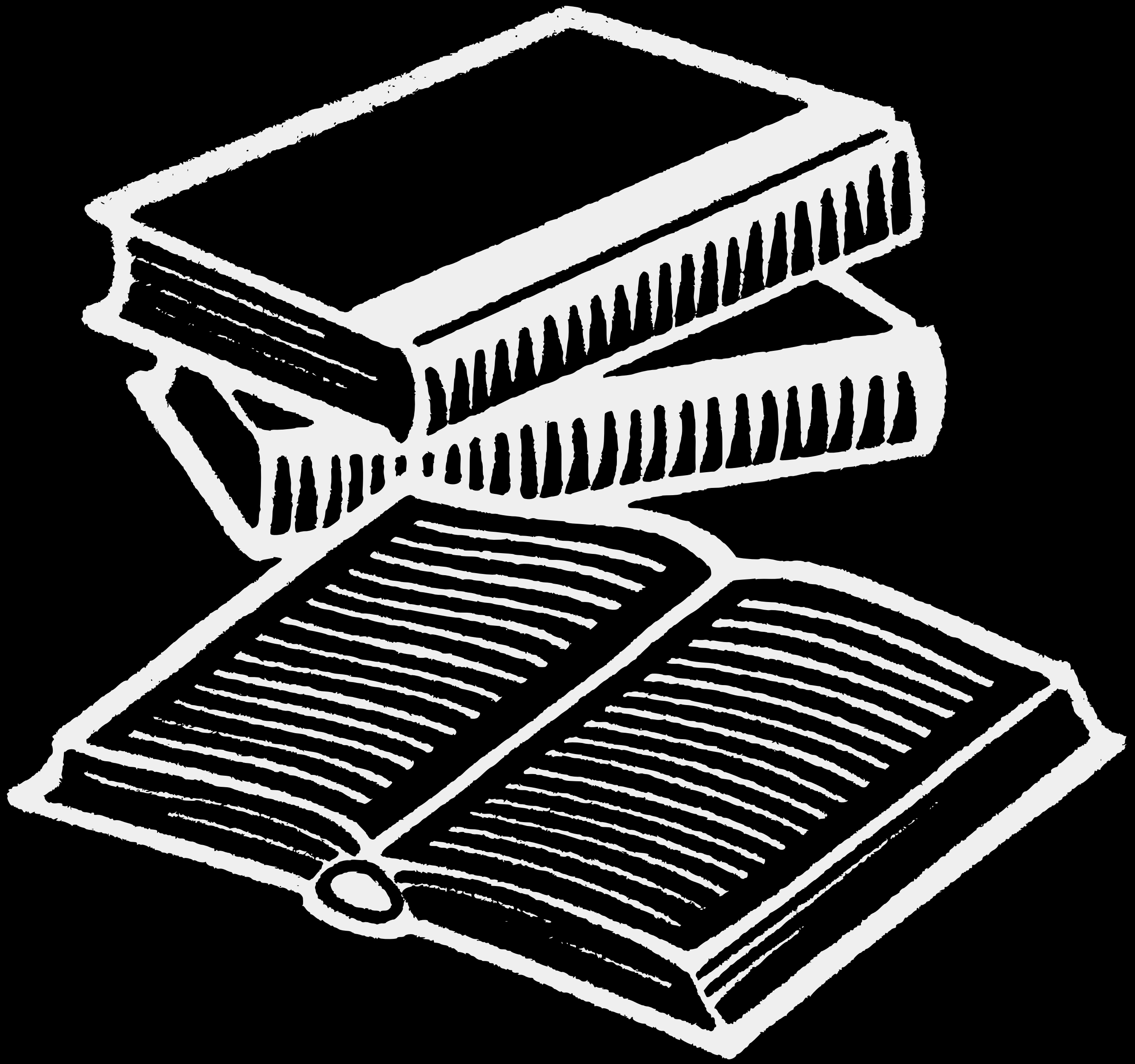
Dosing Pump: delivers precise amounts of fluids, such as chemicals or additives, into a system at specific intervals or rates.

Filter Feed Pump: supplies liquid to a filtration system, ensuring a consistent flow of fluid through the filters to separate contaminants or particles from the liquid.

Dataset

Columns/Features: 15

Datapoints/Rows: 1416



Methodology

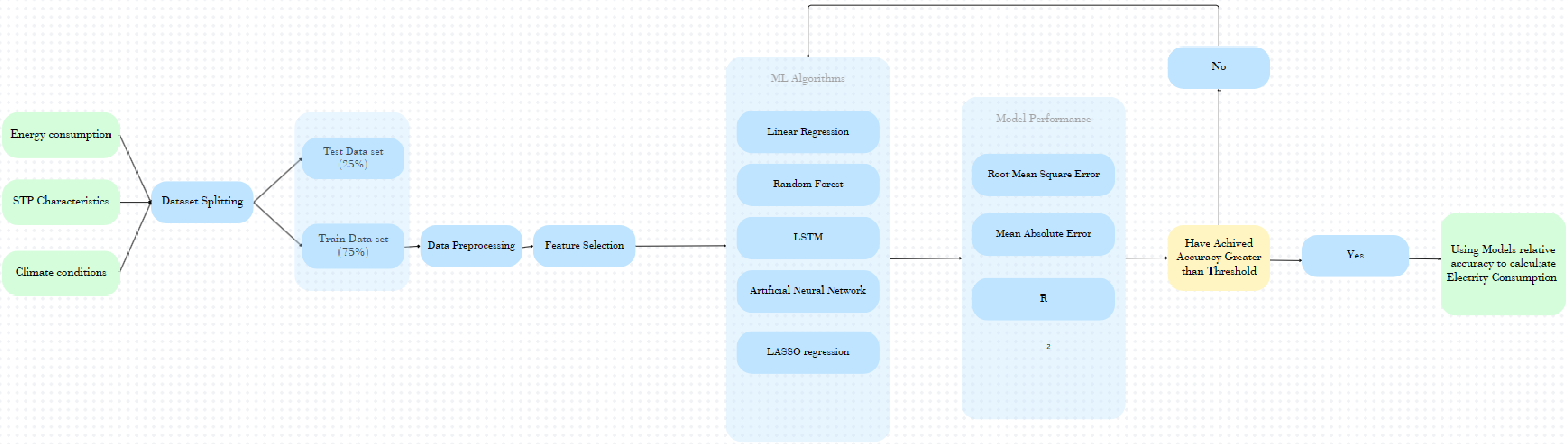


Fig: ML Methodology - different algorithms evaluated for finding algorithm with best performance

Metrics

Error Metrics

MSE - Mean Squared Error

$$\frac{1}{n}\sum(Y_i-\hat{Y}_i)^2$$

RMSE- Root Mean Squared Error

$$\sqrt{MSE}$$

R2- Coefficient of Determination:

$$1-\frac{\sum(Y_i-\hat{Y}_i)^2}{\sum(Y_i-\bar{Y})^2}$$

ANN

Artificial Neural Networks (ANNs) are a type of machine learning model inspired by the structure and function of the human brain. Composed of interconnected nodes (neurons) organized into layers, ANNs excel in learning complex patterns and relationships within data.

Regression

Linear regression is a foundational statistical method used for modeling the relationship between a dependent variable and one or more independent variables. Its primary goal is to understand and predict the behavior of the dependent variable based on the independent variables

LASSO Regression

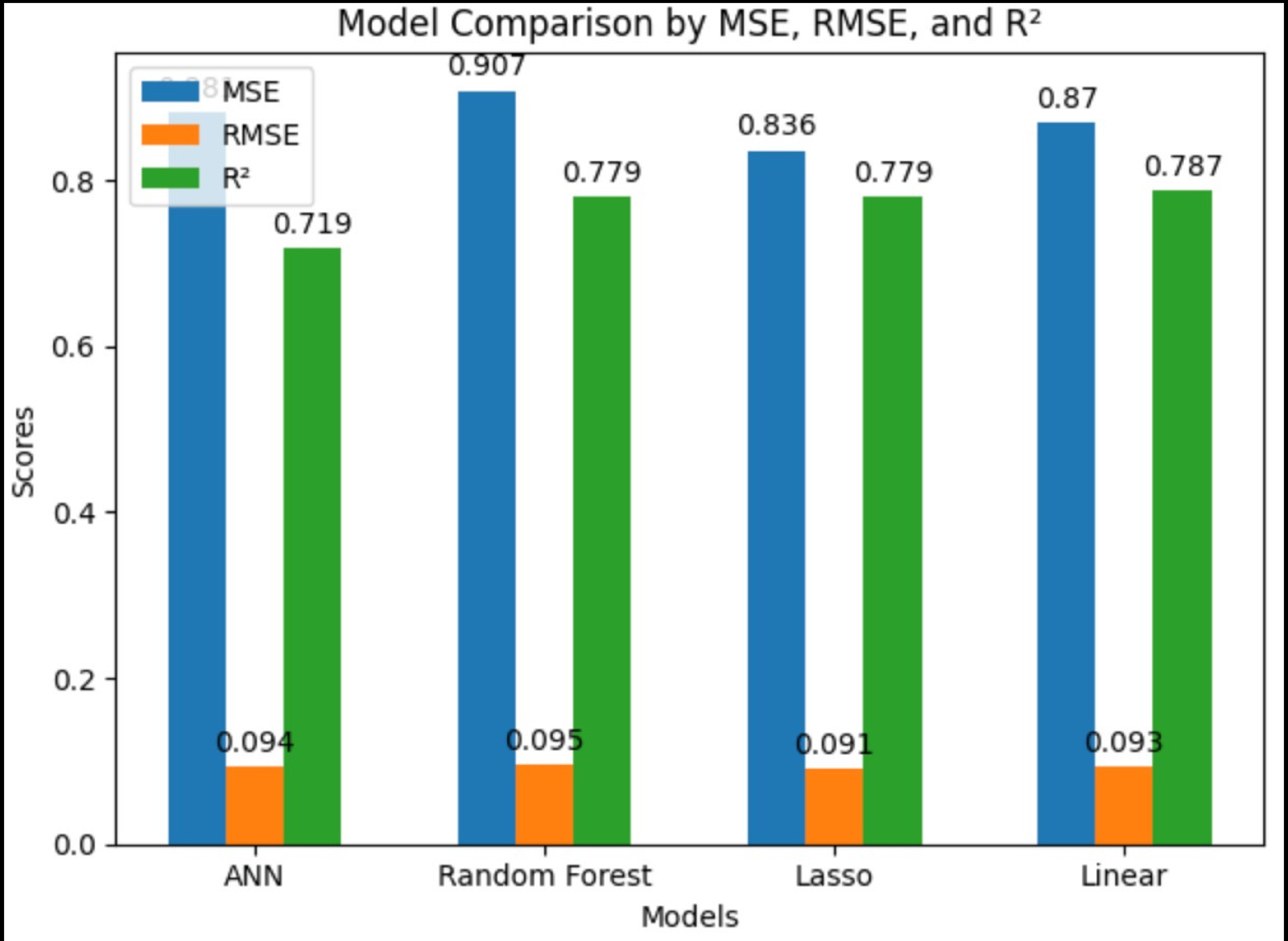
LASSO (Least Absolute Shrinkage and Selection Operator) Regression is a regularization technique used in linear regression analysis. It's employed for feature selection and regularization by penalizing the absolute size of the regression coefficients.

Random Forest

Random Forest is a popular machine learning algorithm used for both classification and regression tasks. It's an ensemble learning method based on the concept of decision trees, where multiple trees are built and aggregated to make predictions.

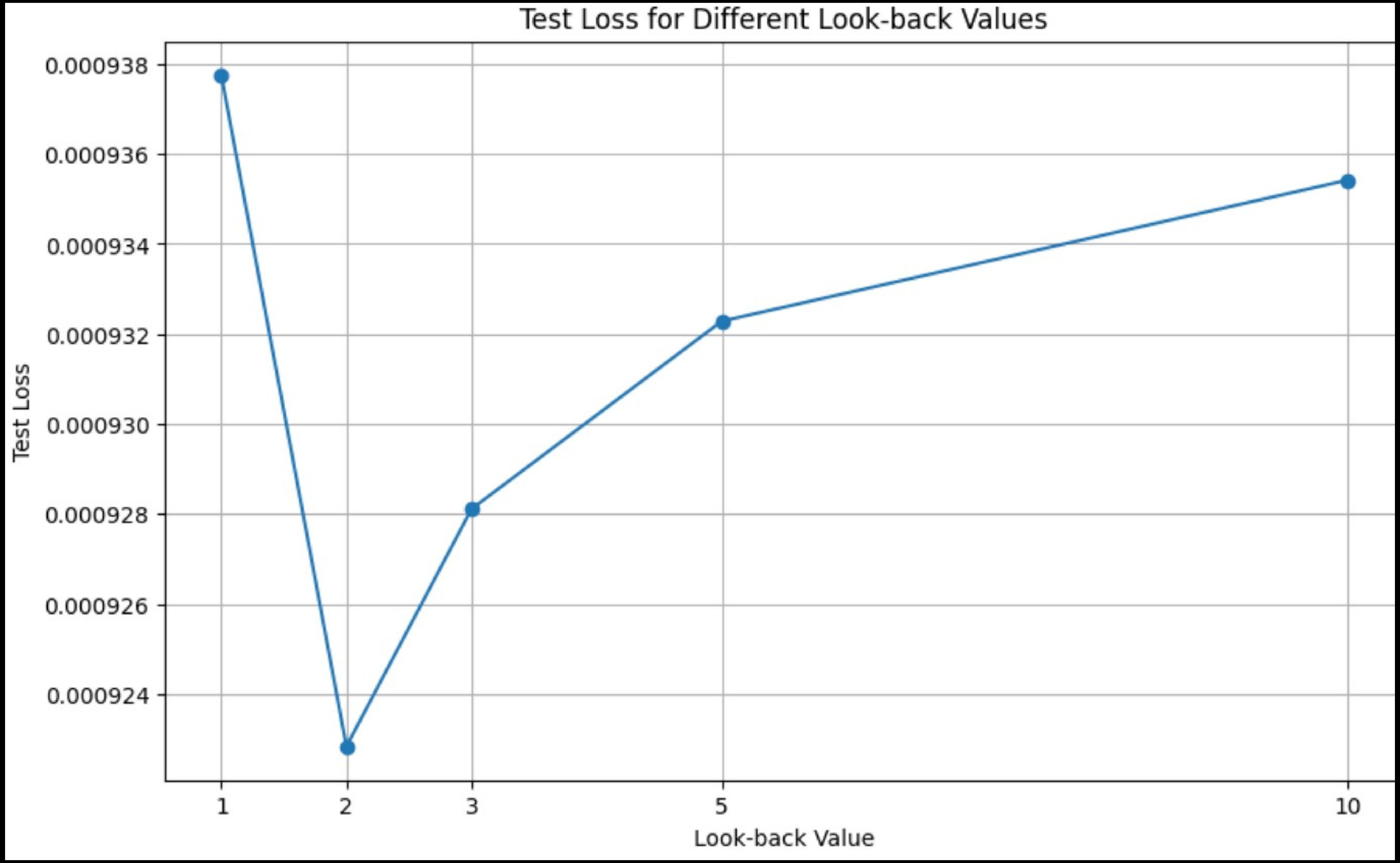
LSTM

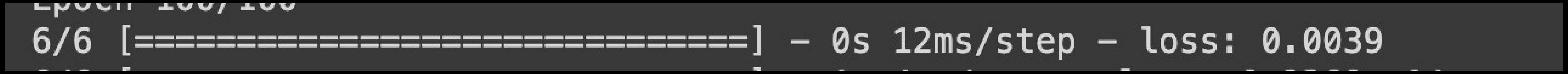
LSTM, short for Long Short-Term Memory, is a type of recurrent neural network (RNN) architecture designed to address the limitations of traditional RNNs in capturing and learning long-range dependencies in sequential data



Model Comparison by MSE, RMSE, and R²
MSE
RMSE
R²
0.907
0.836
0.87
0.719
0.779
0.779
0.787
0.094
0.095
0.091
0.093
Scores
0.0
0.2
0.4
0.6
0.8
ANN
Random Forest
Lasso
Linear
Models

LSTM





```
Epoch 100/100
6/6 [==============================] - 0s 12ms/step - loss: 0.0039
```

Deployability?

Thank You!