Power consumption of Tetouan city prediction

DAWID PAŁKA

JAN JURASZ

Dataset

Data was collected by a public service operator, Supervisory Control and Data Acquisition system (SCADA) and has information on power consumption and environmental factors every 10 minutes for the period between 2017-01-01 and 2017- 12-31.

Dataset doesn’t contain any missing values.

DateTime	Temperature	Humidity	Wind Speed	general diffuse flows	diffuse flows	Zone 1 Power Consumption	Zone 2 Power Consumption	Zone 3 Power Consumption
2017-01-01 00:00:00	6.559	73.8	0.083	0.051	0.119	34055.69620	16128.87538	20240.96386
2017-01-01 00:10:00	6.414	74.5	0.083	0.070	0.085	29814.68354	19375.07599	20131.08434
2017-01-01 00:20:00	6.313	74.5	0.080	0.062	0.100	29128.10127	19006.68693	19668.43373
2017-01-01 00:30:00	6.121	75.0	0.083	0.091	0.096	28228.86076	18361.09422	18899.27711
2017-01-01 00:40:00	5.921	75.7	0.081	0.048	0.085	27335.69620	17872.34043	18442.40964

Additional features

Under the suspicion there may emerge regularities (people sleeping,working etc.) in data, we created additional categorical features

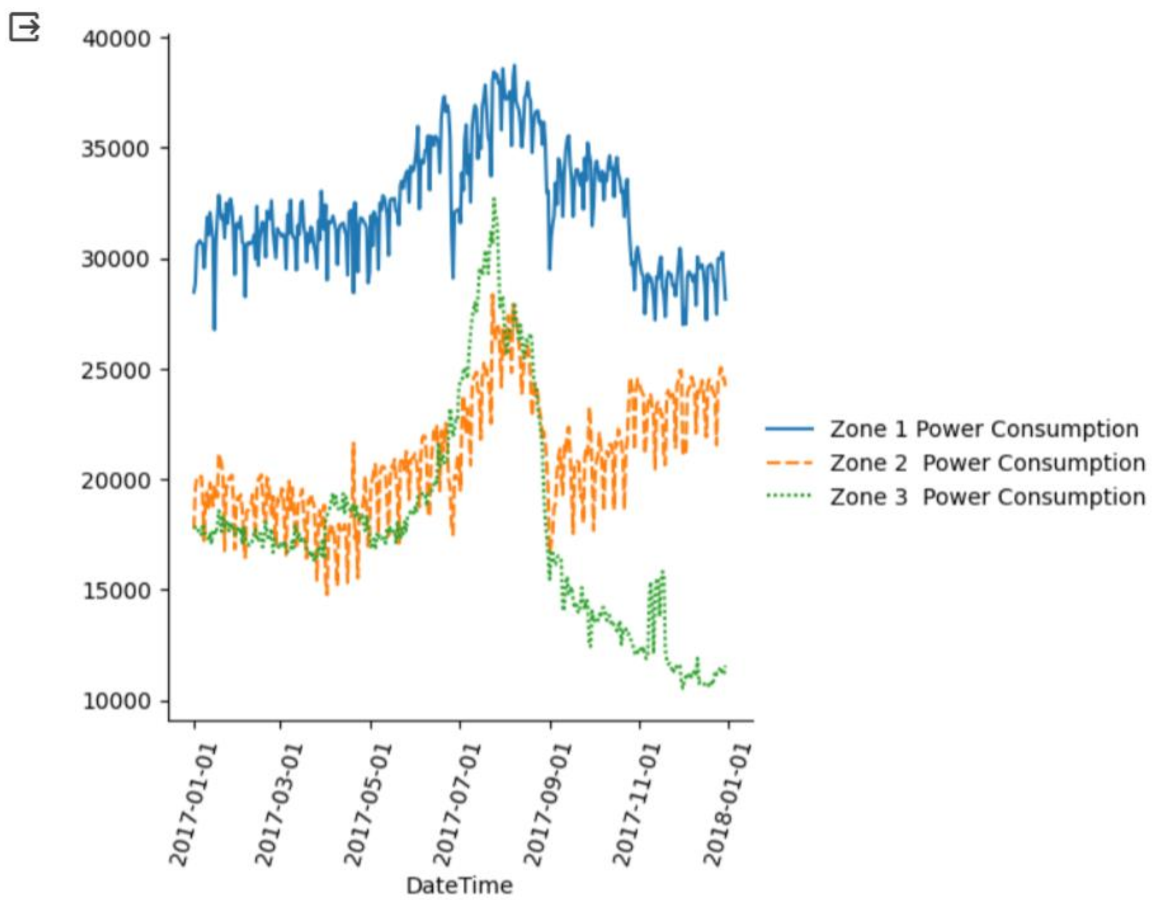
- Time of the Day
- Day of the week

We used OHE for these features.

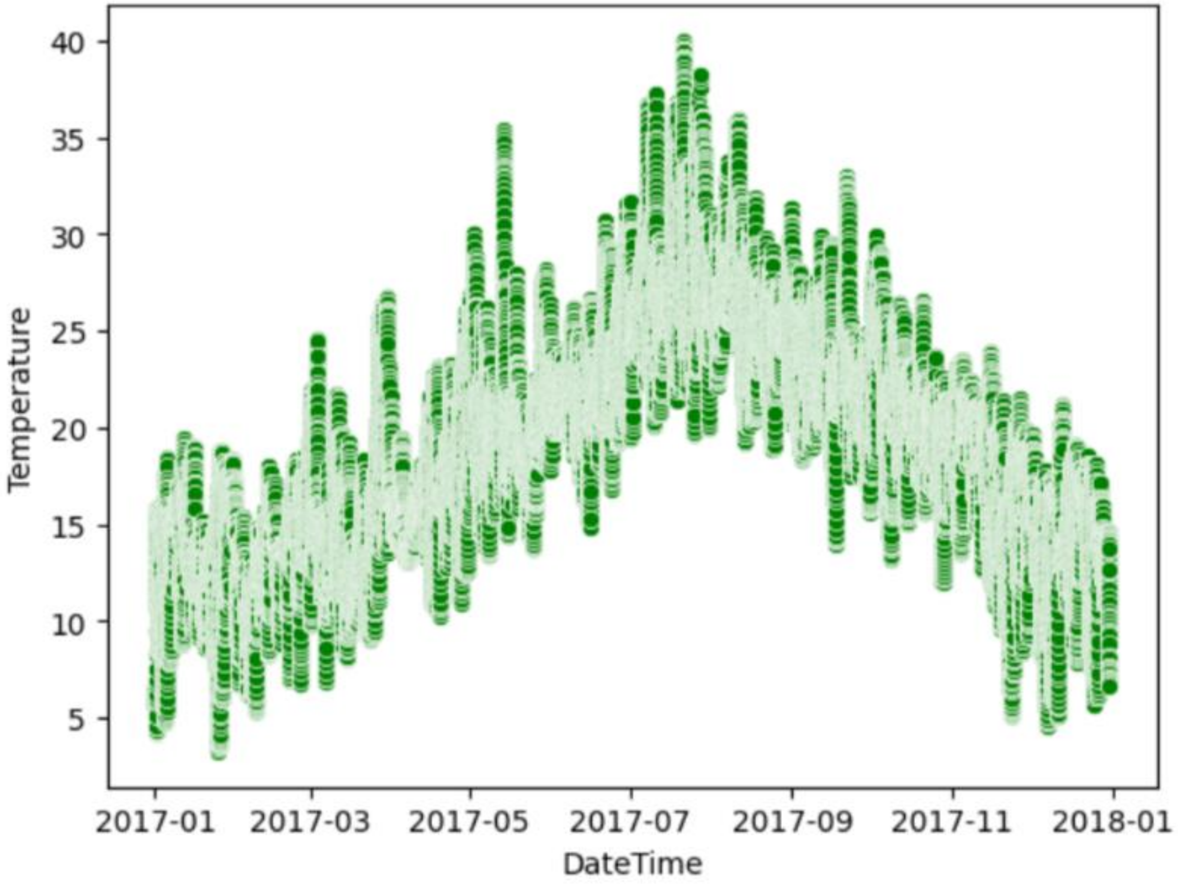
```
def feature_date(df):
    df['TimeOfDay'] = pd.cut(df['DateTime'].dt.hour,bins=[0,6,12,18,24],
                            labels=['morning','midday','evening','night'],include_lowest=True)
    df['DayOfWeek'] = df['DateTime'].dt.dayofweek
feature_date(df)
```

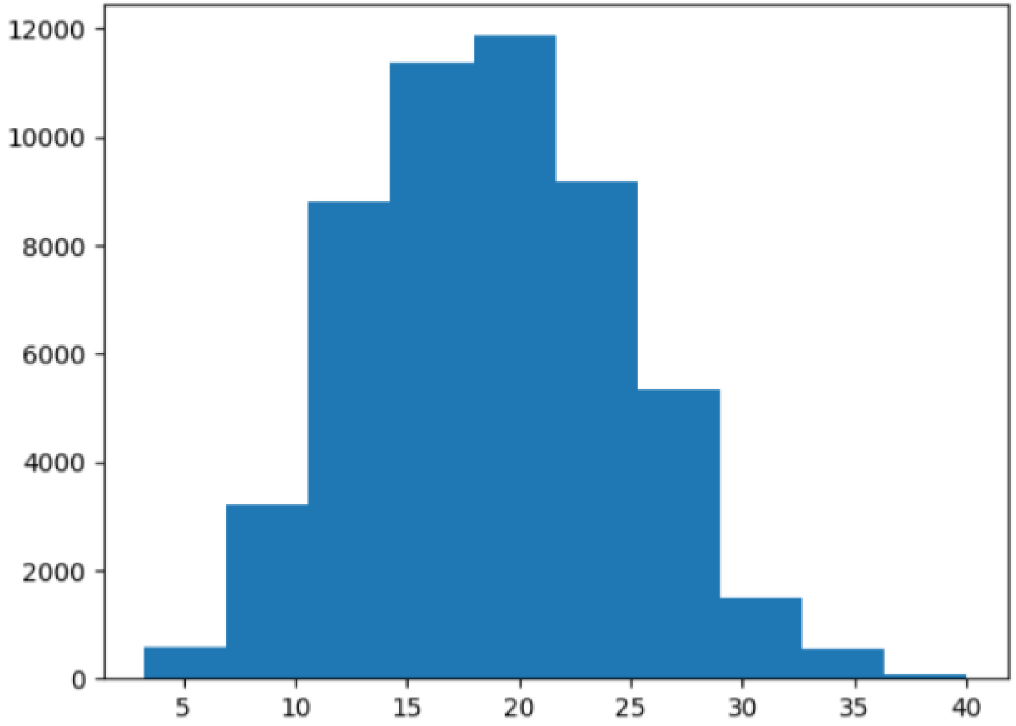
Power consumption in Tetouan

Distribution network is powered by 3 zone stations: Quads, Boussafou and Smir.

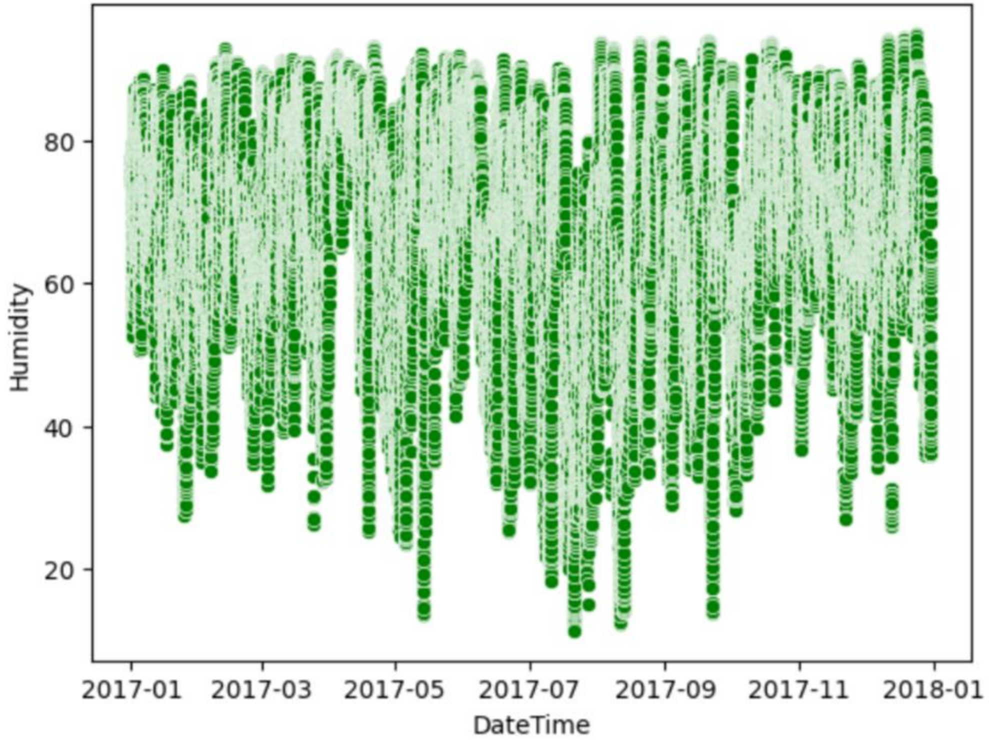


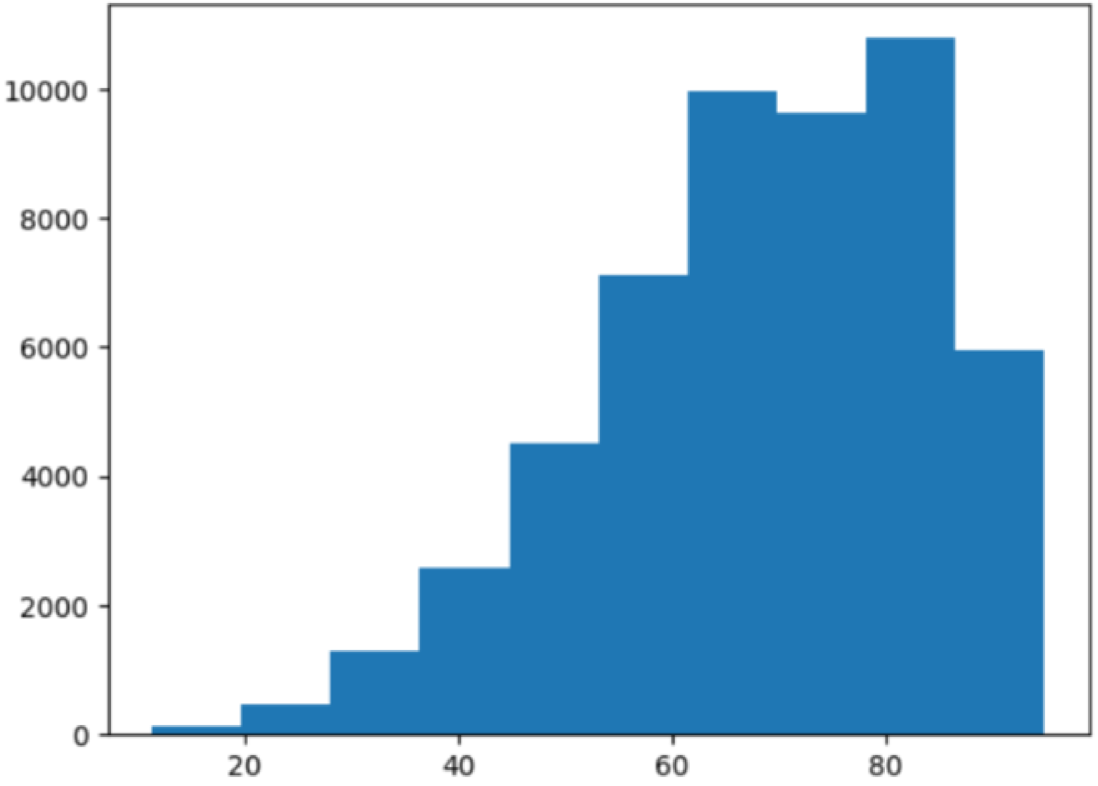
Temperature in Tetouan



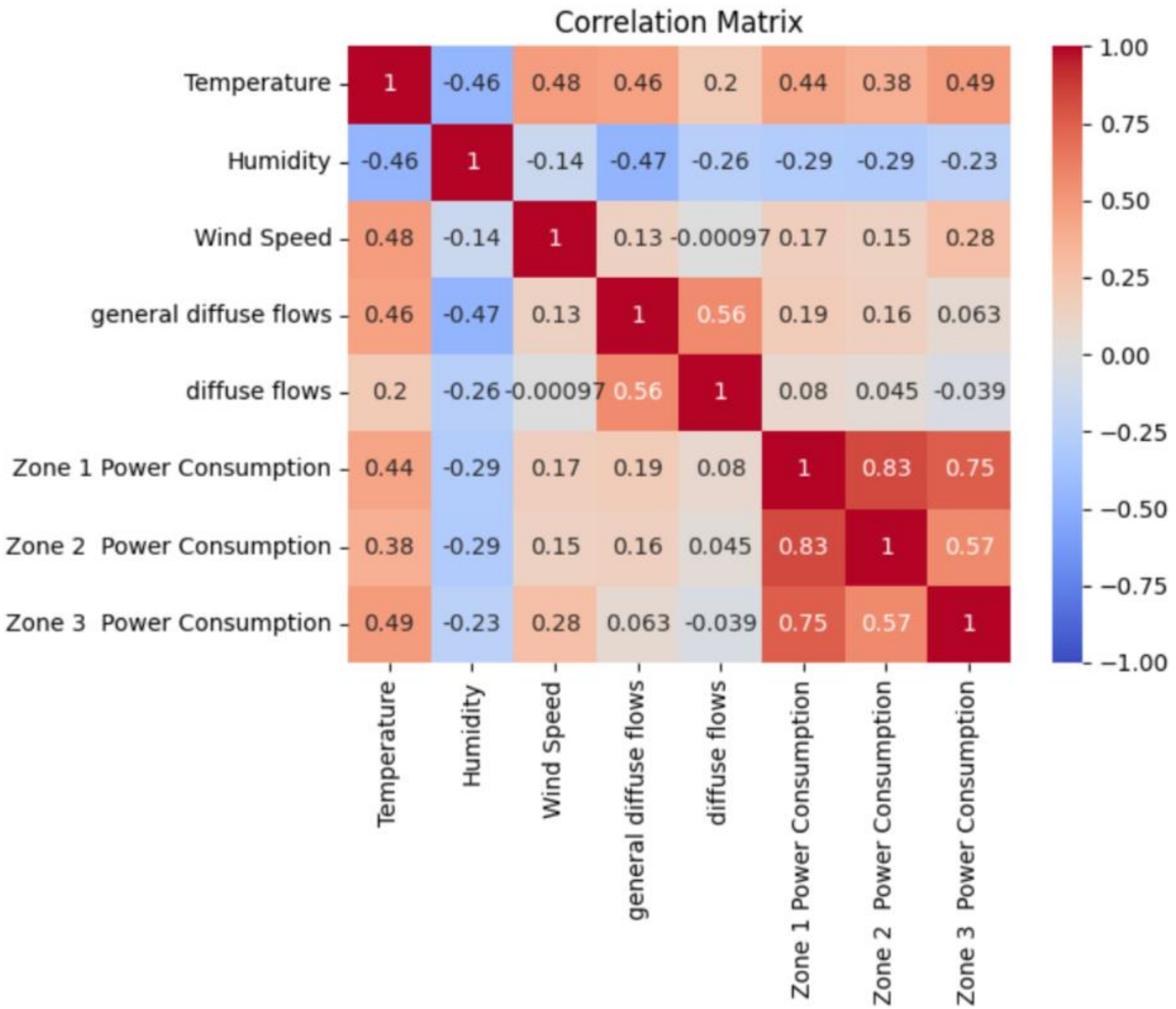


Humidity in Tetouan

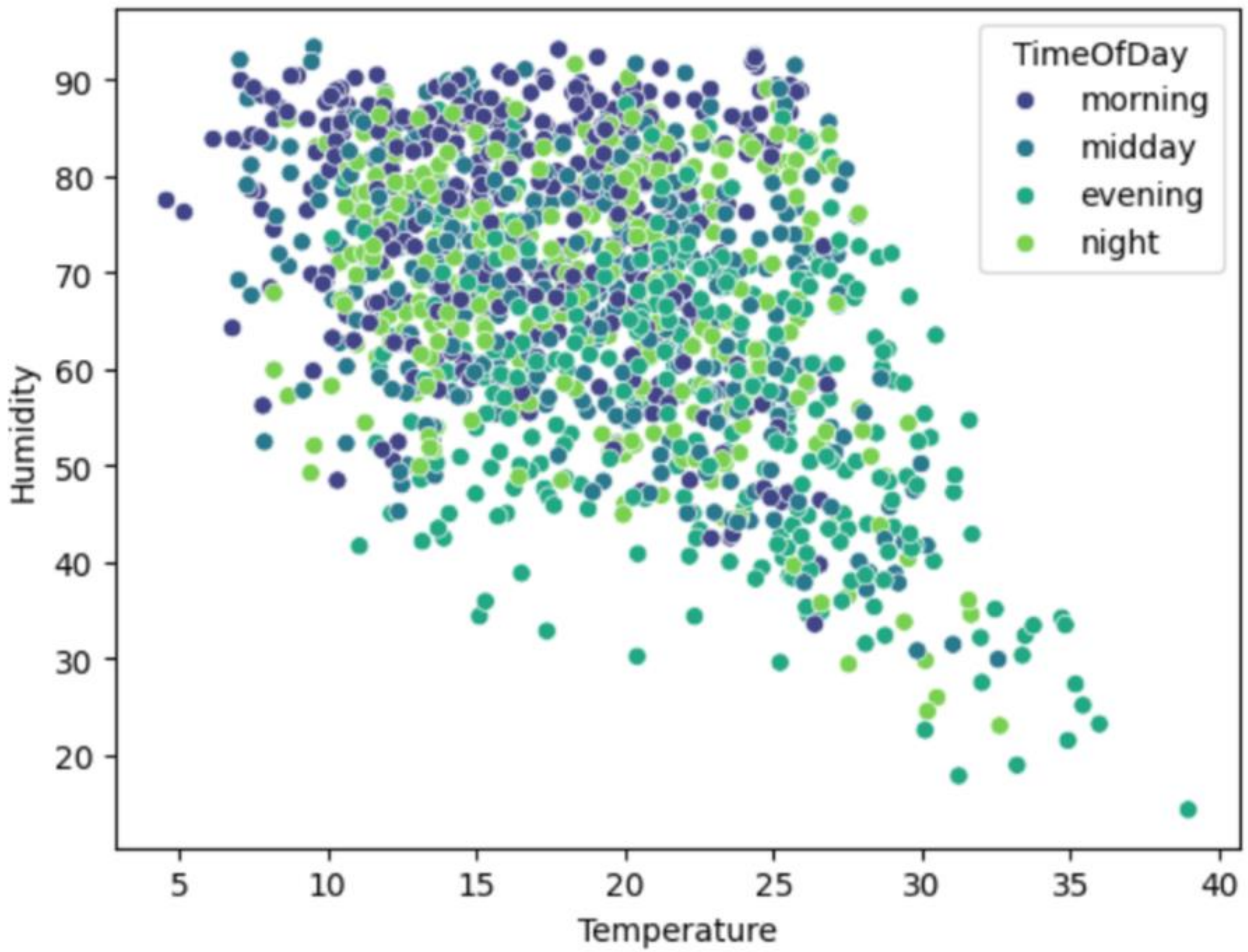




Correlation between features



Humidity, Temperature Correlation



Findings

Power consumption in each zone varies throughout the year, and each location should be considered independently.

Energy consumption tends to increase with higher temperature levels.

Positive and negative correlations exist between humidity, temperature, and diffuse flows.

Energy usage typically increases in hours 8-16 and 17-22.

Over the week it also typically decreases on Sundays.

Pre-model preperations

A key think was to split the date column to separated columns representing the minute , the hour ,the day of the month ,the month the and year .

```
dr_first_zone_x['day_of_month'] = pd.to_datetime(dr_first_zone_x['DateTime']).dt.day
dr_first_zone_x['year'] = pd.to_datetime(dr_first_zone_x['DateTime']).dt.year
dr_first_zone_x['month'] = pd.to_datetime(dr_first_zone_x['DateTime']).dt.month
dr_first_zone_x['hour'] = pd.to_datetime(dr_first_zone_x['DateTime']).dt.hour
dr_first_zone_x['minute'] = pd.to_datetime(dr_first_zone_x['DateTime']).dt.minute
```

It was essential because every model that we have used needed a numerical value, not categorical one .

```
columns_to_normalize = ['DateTime_abs_normalized', 'Temperature', 'Humidity']

for col in columns_to_normalize:
    dr_first_zone_x[col] = scaler.fit_transform(dr_first_zone_x[col].values.reshape(-1, 1)).flatten()
```

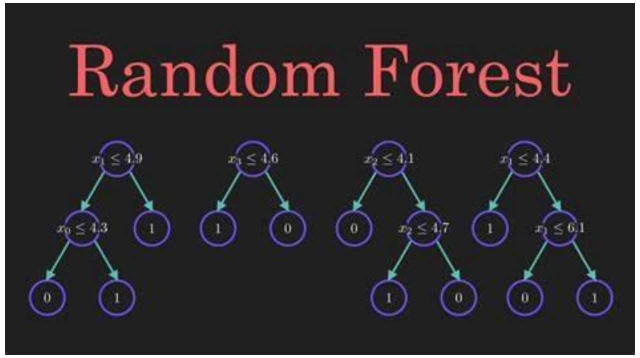
Also to make outputs more clear we normalized y value .

Model candidates

Taking into consideration previous research we found three models :

- XGBoost

- RandomForest



-Tensorflow

XGBoost

Zone 1:

MSE : 0.000611

MAE: 0.017678

MAP: 4.44%

Zone 2:

MSE :0.000679

MAE: 0.019102

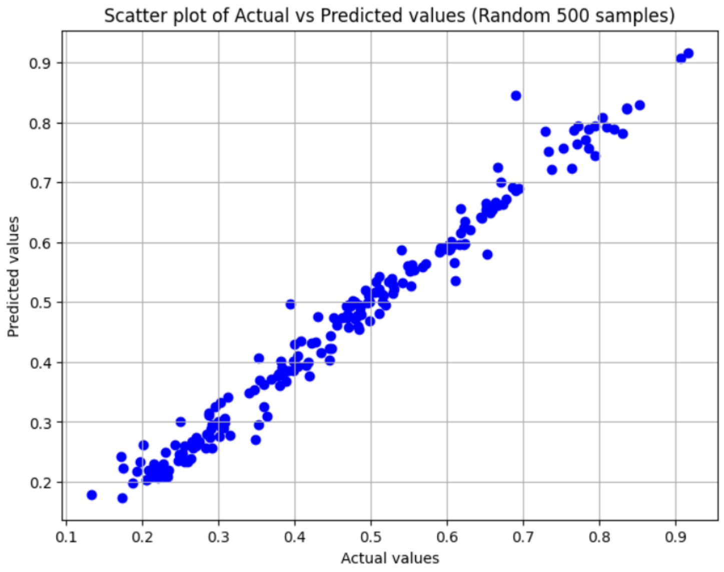
MAP: 5.93%

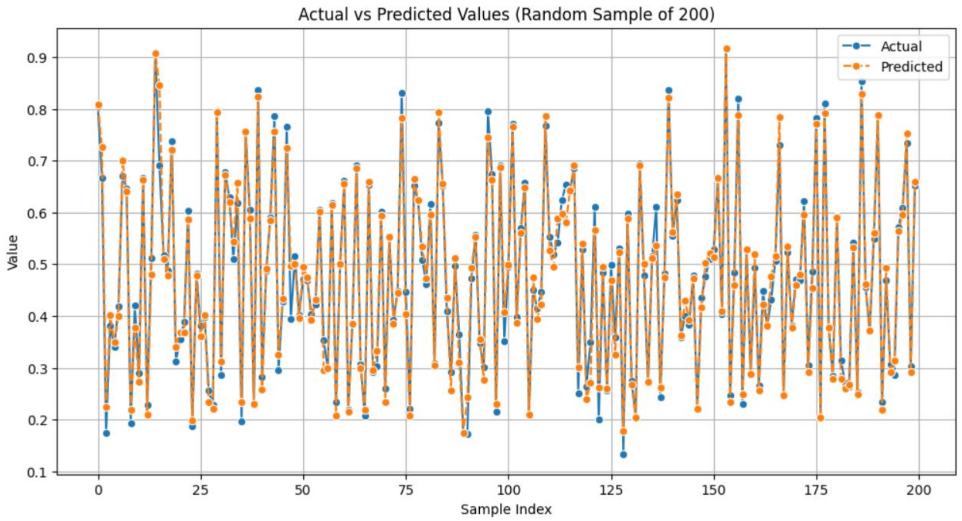
Zone 3:

MSE : 0.000188

MAE: 0.009692

MAP: 4.62%





RandomForest

Zone 1:

MSE : 0.000473

MAE: 0.013868

MAP: 3.5%

Zone 2:

MSE : 0.000469

MAE: 0.013826

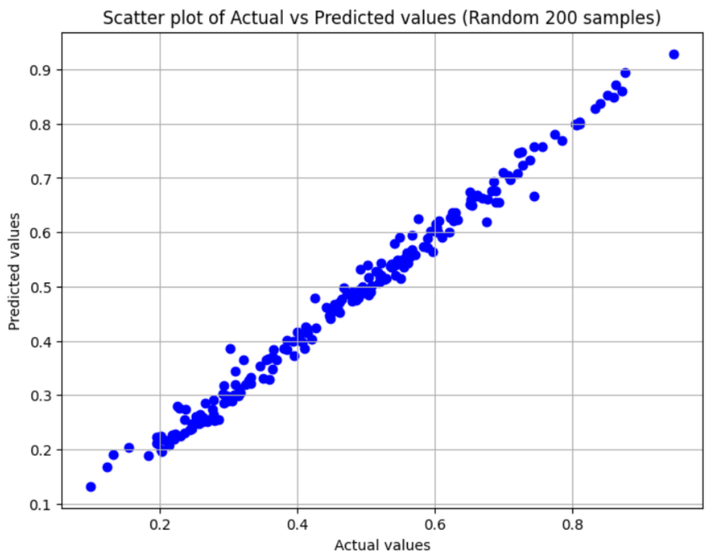
MAP: 4.3%

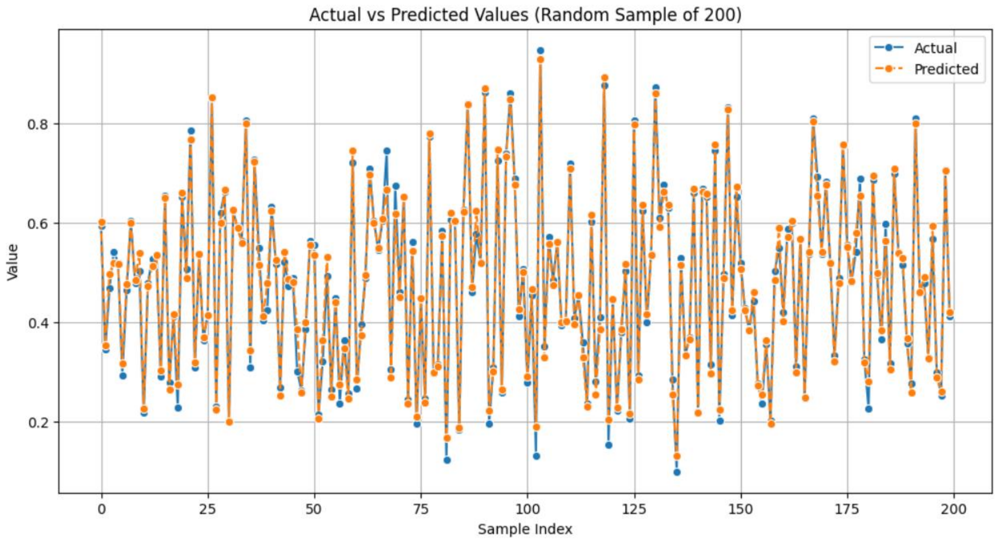
Zone 3:

MSE : 0.000119

MAE: 0.006957

MAP: 3.28%





Tensorflow

Zone 1:

MSE : 0.001768

MAE: 0.031303

MAP: 7.9%

Zone 2:

MSE : 0.002373

MAE: 0.037234

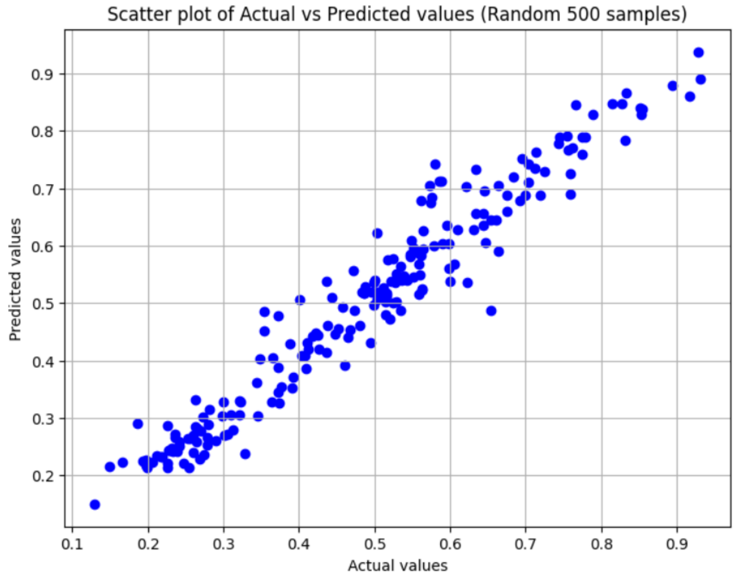
MAP: 11%

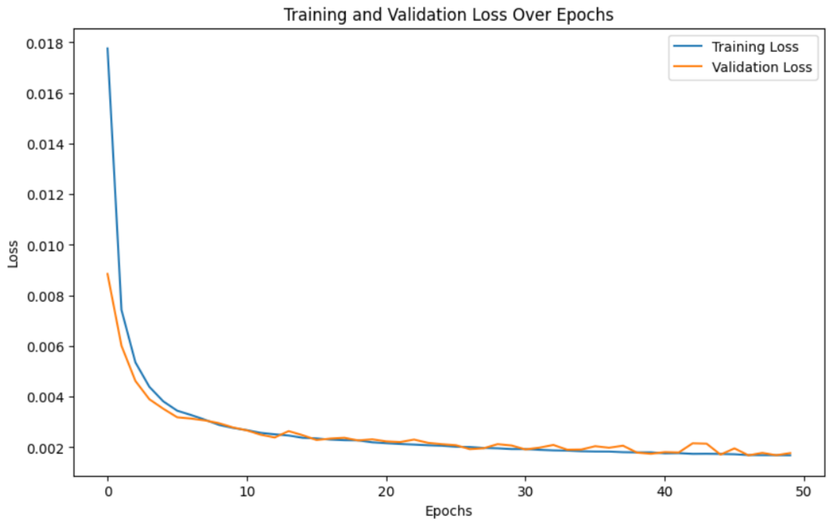
Zone 3:

MSE : 0.00574

MAE: 0.017973

MAP: 9.74%





Total performance

Mesures	XGBoost	RandomForest	Tensorflow
Mean MSE:	0.000492	0.000354	0.003290
Mean MAE:	0.015491	0.111550	0.028837
Mean MAP:	4.99%	3.69%	9.55%

Final resoults on test data

Mesures	ZONE 1	ZONE 2	ZONE 3
MSE	0.000469	0.000502	0.000125
MAE	0.014027	0.014062	0.006953
MAP	3.45%	4.58%	3.12%

Thank you for your time