Time Series Prediction Of AQI using Facebook Prophet and YOLOV5

Omer Qureshi[1,*,†], Muhammad Osaid[1,†], Muhammad Atif Tahir[1,†] and Muhammad Nouman Durrani[1,†]

[1]National University of Computer and Emerging Sciences, Karachi, Pakistan

Abstract

This paper describes the method proposed by team VSL_FAST_NUCES for the task of predicting: Air Quality Index using Sensors and Cameras Data at MediaEval, 2022. This task aims to predict AQI levels of air by using sensors data and images from different cameras. Our proposed approach is that we collect air pollutants data from different sensors and then used Facebook Prophet to predict pollutants value and AQI. For the images data we used YOLOv5 to extract objects in images and then applied FB Prophet prediction on them to predict AQI level from images. The Best RMSE is on Sensor 7 which is 5.77%

1. Introduction

With economic development, growth in population, and other factors, air pollution has become a threat to the lives of individuals.

The MediaEval 2022 Air Quality Index challenge [1] consisted of two tasks: Task 1 and Task 2. In task 1, it is required to predict the air quality pollutants such as PM2.5, PM10, CO, NO2 & O3 by using the environmental data. On the other hand, Task 2 required predicting the air quality by using the traffic data. This paper explores the new technique to predict AQI values of sensors data and cameras images using FBProphet [2] and YOLOv5 algorithm.

2. Related Work

In the research [3], the authors estimated PM2.5 at the current time and predicted it over the short to medium term. The proposed approach utilizes the relationship between urban nature, urban traffic, and air pollution to devise an estimation model. The authors use fuzzy logic to create temporal records of the environment, transportation and urban nature.

There has been a number of methodologies to predict the air quality index. In [4] the authors propose to combine the visual features and AQI lifelogs. The features of the visual data were extracted through CNN and the AQI was calculated via the standard methodology. The purpose of the model was to establish a relationship between the visual features and AQI through a multi-layer perceptron(MLP) model.

In [5] the authors estimate AQI through regression by considering variables such as timestamp, location and weather. The algorithms include linear regression, support vector machines etc.

MediaEval'22: Multimedia Evaluation Workshop, January 13–15, 2023, Bergen, Norway and Online

*Corresponding author.

†These authors contributed equally.

omerqureshi.vsl@nu.edu.pk (O. Qureshi); mosaid.vsl@nu.edu.pk (M. Osaid); atif.tahir@nu.edu.pk (M. A. Tahir); muhammad.nouman@nu.edu.pk (M. N. Durrani)

The model evaluation was done through accuracy and the F1 score. It can be seen that the XGBoost algorithms gave the best results.

3. Approach

Two different approaches were adopted for Task 1 and Task 2 respectively and they are as follows:

3.1. Task 1

The first approach was a time series problem where the requirement was to predict the value of the respective pollutants over a period of time. For this approach, the Facebook Prophet (FbProphet) algorithm was used to predict the values of all the pollutants on an hourly basis.

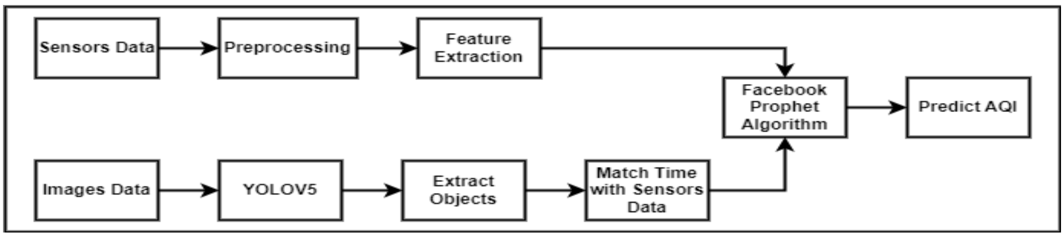
3.2. Task 2

In task 2, the pictures from camera 2 were used to extract the objects through the YOLOv5 algorithm. The main objects identified were cars, people, bicycles, trucks & motorcycles. The frequency of these objects was determined and the respective AQI at the given moment of time was considered to be the closest hour. After the propagation of the value of objects and AQI, a simple regression was run to determine the relationship between the objects and AQI.

There was a positive relationship between AQI PM 2.5, AQI PM 10, AQI CO and the presence of a person and car in the image.

The expected values of the objects over a period of time were determined through the FbProphet algorithm. These forecasted values of the objects were subsequently used with the values of the linear regression to give us an expected AQI value.

Figure 1: Working Methodology of Proposed Approach.



4. Results and Analysis

The challenge in predicting the pollutants was the unavailability of regular time series data. There were numerous instances where the data was missing for a number of days and interpolation was the only solution. However, even with interpolation, the results could vary significantly over the months as the magnitude of the values varied significantly from month to month.

For Task1, sensors 4,7 and 10 were considered. These sensors were selected for prediction as they were one of the "live" sensors.

For Task 2, we used YOLOv5 on the pictures captured from camera2 and used the values from sensor 2. The reason for selecting camera 2 and sensor 2 was the fact that they had similar longitude and latitude which means that they were in close proximity to each other. The appended consists of the results for Task1 and Task 2.

4.1. Task 1 Results

In this task, we extract AQI values for each pollutant and simple values for pollutants using FB Prophet Algorithm.

As we can see for sensor 4, date 11/25/2022 time 5 PM - 6 PM, PM2.5 48.73 value is unhealthy and their AQI is also lies in unhealthy range which is 132.76 . For CO, date 11/25/2022 time 5 PM - 6 PM, value is 3.36 which is in good range and their predicted AQI is also good 38.36. For sensor 7, date 11/23/2022 time 11 PM - 12 PM, PM10 value is is 5.95 which is good and their predicted AQI is 12.87 which is also good. For sensor 10, date 11/25/2022 time 5 PM - 6 PM, CO value is 39.89 which is hazardous and their predicted AQI is 382.88 which is also in hazardous range.

Table 1
Values for respective pollutants for sensor 4, 7 & 10.

Sensor #	Date	Time	pm2.5	pm10	co	so2	no2	o3
Sensor 4	11/18/2022	8 AM - 9 PM	36.73	42.34	3.04	1.05	0.94	0.94
		11 AM - 12 PM	37.78	43.54	2.84	1.04	1.92	1.92
		5 PM - 6 PM	46.00	51.56	3.28	1.04	0.55	0.55
	11/23/2022	8 AM - 9 PM	37.30	42.97	2.71	1.06	0.98	0.98
		11 AM - 12 PM	39.00	44.92	2.54	1.05	1.97	1.97
		5 PM - 6 PM	48.68	54.65	3.11	1.04	0.48	0.48
	11/25/2022	8 AM - 9 PM	39.46	45.56	3.13	1.05	0.87	0.87
		11 AM - 12 PM	40.51	46.75	2.93	1.05	1.86	1.86
		5 PM - 6 PM	48.73	54.78	3.36	1.04	0.62	0.62
Sensor 7	11/18/2022	8 AM - 9 PM	0.46	1.56	1.46	1.11	3.54	3.54
		11 AM - 12 PM	1.53	0.60	1.48	1.13	3.20	3.20
		5 PM - 6 PM	10.89	11.25	1.65	0.99	2.84	2.84
	11/23/2022	8 AM - 9 PM	4.92	6.91	1.31	1.05	4.50	4.50
		11 AM - 12 PM	4.07	5.95	1.33	1.07	4.29	4.29
		5 PM - 6 PM	2.87	2.06	1.50	0.93	4.08	4.08
	11/25/2022	8 AM - 9 PM	10.81	14.19	1.56	1.11	2.97	2.97
		11 AM - 12 PM	8.84	12.03	1.58	1.13	2.63	2.63
		5 PM - 6 PM	0.51	1.39	1.75	0.99	2.27	2.27
Sensor 10	11/18/2022	8 AM - 9 PM	2.80	3.35	39.47	1.03	0.15	0.15
		11 AM - 12 PM	3.22	3.94	39.18	1.04	0.16	0.16
		5 PM - 6 PM	14.23	16.14	39.40	1.03	0.15	0.15
	11/23/2022	8 AM - 9 PM	5.02	6.28	40.84	1.06	0.15	0.15
		11 AM - 12 PM	5.83	7.20	40.70	1.06	0.15	0.15
		5 PM - 6 PM	17.17	19.57	41.19	1.05	0.14	0.14
	11/25/2022	8 AM - 9 PM	0.66	1.20	39.97	1.04	0.15	0.15
		11 AM - 12 PM	1.08	1.79	39.68	1.05	0.16	0.16
		5 PM - 6 PM	12.09	14.00	39.89	1.04	0.15	0.15

4.2. Task 2 Results

The results indicate that 5 PM - 6 PM is the time slot which has the most amount of activity for all the days ; which leads to an increase in all pollutants besides O3. It is also interesting to note that the 8 AM to 9 AM slot appears to have more pollutants than 11 AM - 12 PM; this may be attributed to some morning activities by the population.

5. Discussion and Outlook

5.1. Future Work

For future research work, manual interpolating should be carried out on the missing values dataset to determine whether there is a significant change in the results. Furthermore additional factors such as the weather and proximity to coal - fired power plants can be considered.

Table 2
AQI for respective pollutants for sensor 4, 7 & 10.

Sensor #	Date	Time	PM2.5_AQI	PM10_AQI	CO_AQI	NO2_AQI	O3_AQI
Sensor 4	11/18/2022	8 AM - 9 PM	103.06	38.48	34.63	0.88	0.88
		11 AM - 12 PM	106.12	39.30	32.32	1.82	1.82
		5 PM - 6 PM	125.97	45.83	37.40	0.52	0.52
	11/23/2022	8 AM - 9 PM	103.92	38.90	30.90	0.92	0.92
		11 AM - 12 PM	108.96	40.48	28.89	1.86	1.86
		5 PM - 6 PM	133.40	48.80	35.56	0.45	0.45
	11/25/2022	8 AM - 9 PM	109.85	41.39	35.59	0.82	0.82
		11 AM - 12 PM	112.92	42.21	33.28	1.75	1.75
		5 PM - 6 PM	132.76	48.74	38.36	0.58	0.58
Sensor 7	11/18/2022	8 AM - 9 PM	26.16	8.41	15.62	4.54	4.54
		11 AM - 12 PM	28.50	9.31	15.74	4.32	4.32
		5 PM - 6 PM	52.41	18.79	16.87	3.75	3.75
	11/23/2022	8 AM - 9 PM	35.18	12.40	15.42	4.67	4.67
		11 AM - 12 PM	36.70	12.87	15.58	4.45	4.45
		5 PM - 6 PM	57.80	21.05	16.73	3.92	3.92
	11/25/2022	8 AM - 9 PM	22.60	7.62	16.29	4.30	4.30
		11 AM - 12 PM	24.94	8.52	16.41	4.07	4.07
		5 PM - 6 PM	48.86	18.00	17.54	3.51	3.51
Sensor 10	11/18/2022	8 AM - 9 PM	14.39	3.29	383.10	0.14	0.14
		11 AM - 12 PM	16.30	3.98	380.00	0.15	0.15
		5 PM - 6 PM	44.14	14.39	381.69	0.14	0.14
	11/23/2022	8 AM - 9 PM	18.48	5.68	393.83	0.14	0.14
		11 AM - 12 PM	20.65	6.50	391.95	0.14	0.14
		5 PM - 6 PM	48.05	16.71	395.69	0.13	0.13
	11/25/2022	8 AM - 9 PM	7.56	1.23	384.28	0.14	0.14
		11 AM - 12 PM	9.47	1.92	381.18	0.15	0.15
		5 PM - 6 PM	37.31	12.33	382.88	0.14	0.14

Table 3
Camera 2 / Sensor 2 - AQI Prediction

Date	Time	PM2.5_AQI	PM10_AQI	CO_AQI	NO2_AQI	O3_AQI
11/18/2022	8 AM - 9 AM	80.65	27.80	26.02	0.27	28.60
	11 AM -12 PM	77.60	26.24	25.66	0.26	30.65
	5 PM - 6 PM	82.36	28.70	26.32	0.28	25.90
11/23/2022	8 AM - 9 AM	79.86	27.40	25.95	0.27	29.19
	11 AM - 12 PM	76.92	25.92	25.62	0.26	31.18
	5 PM - 6 PM	81.92	28.53	26.32	0.27	26.27
11/25/2022	8 AM - 9 AM	80.69	27.82	26.01	0.27	28.55
	11 AM - 12 PM	77.63	26.26	25.65	0.27	30.60
	5 PM - 6 PM	82.40	28.73	26.31	0.28	25.85

5.2. Evaluation

The evaluation is carried out by considering the actual values of the sensor data. The results were compared on an hourly basis by taking the mean of the first and last values in a given hour. The RMSE for sensors were Sensor 4 (22.14%), Sensor 7 (5.77%) and Sensor 10 (6.99%).

Table 4
RMSE of Values

Sensors	RMSE
Sensor#4	22.14%
Sensor#7	5.77%
Sensor#10	6.99%

References

[1] T.-L. N.-T. T.-B. N. Minh-Son Dao, Thanh-Hai Dang, D.-T. Dang-Nguyen, Overview of mediaeval 2022 urban air: Urban life and air pollution, mediaeval, bergen, norway, 13-15 January 2022.

[2] S. J. Taylor, B. Letham, Forecasting at scale, The American Statistician 72 (2018) 37–45.

[3] T.-V. La, M.-S. Dao, K. Tejima, R. U. Kiran, K. Zettsu, Improving the awareness of sustainable smart cities by analyzing lifelog images and iot air pollution data, in: 2021 IEEE International Conference on Big Data (Big Data), IEEE, 2021, pp. 3589–3594.

[4] P.-B. Vo, T.-D. Phan, M.-S. Dao, K. Zettsu, Association model between visual feature and aqi rank using lifelog data, in: 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 4197–4200.

[5] D. Duong, Q. Le, T.-L. Nguyen-Tai, H. Nguyen, M. Dao, B. Nguyen, An Effective AQI Estimation Using Sensor Data and Stacking Mechanism, 2021. doi:10.3233/FAIA210040.