Jio Institute

BREATHE EASY: AQI Prediction & Routing

PGP in Artificial Intelligence & Data Science

Rishita Merchant
Sanskriti Dhar
Sneha Shinde

I. Introduction

In today's fast-growing cities, air pollution has become a major concern for public health. This is especially true for people who spend a lot of time outdoors—like cyclists—who are constantly exposed to the air around them while traveling through different parts of the city. Breathing in polluted air can lead to serious health issues, especially related to the lungs and heart. This project is designed to address that problem by building a smart and reliable system that not only predicts how clean or polluted the air will be in various areas but also uses this information to help cyclists choose routes that expose them to the least amount of pollution.

Usually, when people use GPS or navigation apps, the routes suggested are based on the shortest distance or the fastest time to reach their destination. But in cities with bad air quality, these fast routes might pass through areas with high levels of harmful pollutants. As a result, cyclists might end up breathing in more toxic air, which can affect their health over time. Our system changes that approach. Instead of focusing only on speed or distance, it considers the current and predicted air quality in different locations while choosing the best route with lowest AQI , as per location of the Cyclist.

The goal of this project is to make cycling in cities healthier and safer. By using accurate air quality predictions, cyclists can make smarter choices and avoid areas with poor air conditions. This system combines powerful machine learning techniques that predict future pollution levels with smart mapping tools that calculate the safest routes in terms of air quality. All of this is presented through an easy-to-use, interactive map interface, allowing users to plan their journeys with both health and convenience in mind.

Beyond just helping individuals stay healthy, this project also supports cleaner and smarter city living. By giving people tools to avoid polluted areas and choose better routes, it encourages safer travel habits and spreads awareness about air quality. This can lead to more people choosing bicycles over cars, which helps reduce traffic, cut down on harmful emissions, and make cities more eco-friendly. In the long run, this system doesn't just protect cyclists—it also promotes a cleaner environment and supports healthier, more sustainable urban life for everyone.

II. Dataset overview

The predictive prowess of any computational model is intrinsically linked to the caliber and breadth of its training data. For this endeavor, the AQI prediction module's core relies on a meticulously compiled historical dataset, encompassing both ambient air pollutant concentrations and pertinent meteorological variables. While the specific data origin was not explicitly detailed in the Colab notebook, a typical implementation would draw from publicly accessible environmental monitoring stations (e.g., governmental environmental protection agencies, open data portals, or municipal monitoring networks) or simulated data.

2.1 Dataset

a) BandraMumbaiMPCB

	From Date	To Date	PM2.5	PM10	NO	NO2	NOx	SO2	CO	Ozone	Benzene	Toluene	Temp	RH	WS	WD	SR	BP	VWS	RF	TOT-RF
0	01-01-2021 00:00	01-01-2021 00:15	35.99	54.3	4.88	0.93	5.81	16.15	2.01	10.85	1.65		25.5	81.7	0.1	19.3	15	768.38	0.9	0	0
1	01-01-2021 00:15	01-01-2021 00:30	17.3	60.96	4.84	0.96	5.8	17.11	2.03	10.78	1.53		25.42	82.22	0.09	24.4	15	768.32	0.87	0	0
2	01-01-2021 00:30	01-01-2021 00:45	10.66	57.51	4.87	0.92	5.78	15.24	2.05	10.87	1.61		25	85.1	0.1	53.3	15	768.07	0.9	0	0
3	01-01-2021 00:45	01-01-2021 01:00	5.69	54.66	4.89	0.72	5.62	16.45	2.06	10.6	1.61		25	85.1	0.1	53.3	15	768.07	0.9	0	0
4	01-01-2021 01:00	01-01-2021 01:15	9.84	53.59	4.87	0.93	5.81	15.58	2.07	10.74	1.88		25	85.1	0.1	53.3	15	768.07	0.9	0	0
5	01-01-2021 01:15	01-01-2021 01:30	7.1	51.68	4.88	0.9	5.79	15.63	2.09	10.78	1.79		24.95	85.5	0.06	57.45	15	768.02	0.62	0	0
6	01-01-2021 01:30	01-01-2021 01:45	12.81	45.6	4.86	0.92	5.79	15.61	2.09	10.56	1.71		24.7	87.7	0.1	214.3	15	767.78	0.9	0	0
7	01-01-2021 01:45	01-01-2021 02:00	24.09	35.79	4.83	0.93	5.77	15.98	2.09	10.79	1.49		24.7	87.7	0.1	214.3	15	767.78	0.9	0	0
8	01-01-2021 02:00	01-01-2021 02:15	23.59	41.31	4.86	1.03	5.89	15.78	2.09	10.91	1.69		24.7	87.7	0.1	214.3	15	767.78	0.9	0	0
9	01-01-2021 02:15	01-01-2021 02:30	11.7	54.54	4.9	0.79	5.71	15.6	2.09	10.82	1.87		24.65	87.8	0.06	213.39	15	767.72	0.62	0	0
10	01-01-2021 02:30	01-01-2021 02:45	3.39	65.38	4.85	0.97	5.83	15.94	2.08	10.85	2.01		24.4	88.4	0.1	38.3	15	767.47	0.9	0	0
11	01-01-2021 02:45	01-01-2021 03:00	12.23	60.32	4.84	0.9	5.75	16.09	2.07	10.71	1.85		24.4	88.4	0.1	38.3	15	767.47	0.9	0	0
12	01-01-2021 03:00	01-01-2021 03:15	20.13	56.16	4.88	0.84	5.74	15.85	2.07	10.78	1.76		24.4	88.4	0.1	38.3	15	767.47	0.9	0	0
13	01-01-2021 03:15	01-01-2021 03:30	15.08	38.87	4.84	1.04	5.89	16.17	2.07	10.82	1.88		24.38	88.49	0.09	32.37	15	767.44	0.87	0	0
14	01-01-2021 03:30	01-01-2021 03:45	17.72	27.8	4.8	1.08	5.89	16.01	2.07	10.7	1.99		24.3	89	0.1	358.3	15	767.25	0.9	0	0
15	01-01-2021 03:45	01-01-2021 04:00	15	35.86	4.89	0.93	5.83	15.88	2.07	10.86	1.93		24.3	89	0.1	358.3	15	767.25	0.9	0	0
16	01-01-2021 04:00	01-01-2021 04:15	8.28	56.28	4.88	0.83	5.71	15.95	2.06	10.83	2.06		24.3	89	0.1	358.3	15	767.25	0.9	0	0
17	01-01-2021 04:15	01-01-2021 04:30	1.31	63.59	4.86	0.98	5.84	15.85	2.06	10.75	2.19		24.26	89.15	0.09	5.12	15	767.21	0.86	0	0
18	01-01-2021 04:30	01-01-2021 04:45	9.27	59.4	4.85	0.87	5.74	15.81	2.06	10.85	2.1		24.1	90	0.1	45.3	15	767.03	0.9	0	0
19	01-01-2021 04:45	01-01-2021 05:00	14.66	57.51	4.8	1.05	5.85	15.67	2.07	10.85	2.27		24.1	90	0.1	45.3	15	767.03	0.9	0	0
20	01-01-2021 05:00	01-01-2021 05:15	12.64	56.18	4.86	0.95	5.82	15.9	2.07	10.83	2.12		24.1	90	0.1	45.3	15	767.03	0.9	0	0
21	01-01-2021 05:15	01-01-2021 05:30	6.33	55.33	4.83	0.98	5.81	16.05	2.08	10.89	2.09		24.11	90	0.09	45.92	15	767.07	0.89	0	0
22	01-01-2021 05:30	01-01-2021 05:45	9.76	52.95	4.8	1.07	5.88	16.67	2.09	11.02	2.16		24.2	90	0.1	49.3	15	767.33	0.9	0	0
23	01-01-2021 05:45	01-01-2021 06:00	10.29	50.43	4.84	0.85	5.68	16.28	2.1	10.83	2.44		24.2	90	0.1	49.3	15	767.33	0.9	0	0
24	01-01-2021 06:00	01-01-2021 06:15	9.26	47.93	4.89	0.9	5.8	16.13	2.11	10.82	2.37		24.2	90	0.1	49.3	15	767.33	0.9	0	0
25	01-01-2021 06:15	01-01-2021 06:30	17.56	46.93	4.9	0.85	5.75	16.63	2.12	10.9	2.68		24.13	89.53	0.09	49.14	15	767.4	0.89	0	0
26	01-01-2021 06:30	01-01-2021 06:45	15.94	46.47	4.84	0.93	5.79	16.23	2.13	10.8	2.62		23.8	87	0.1	48.3	15	767.85	0.9	0	0
27	01-01-2021 06:45	01-01-2021 07:00	8.83	44.48	4.83	0.99	5.81	16.27	2.15	10.94	2.39		23.8	87	0.1	48.3	15	767.85	0.9	0	0
28	01-01-2021 07:00	01-01-2021 07:15		35.25	4.88	0.81	5.68	16.01	2.16	10.77	2.18		23.8	87	0.1	48.3	15	767.85	0.9	0	0
29	01-01-2021 07:15	01-01-2021 07:30	1.27	37.58	4.89	0.88	5.78	16.54	2.18	10.72	2.17		23.94	85.95	0.07	58.56	15.93	767.94	0.71	0	0
30	01-01-2021 07:30	01-01-2021 07:45	8.72	47.78	4.9	0.83	5.74	16.29	2.19	10.62	2.31		24.7	80.3	0.1	163.3	21	768.45	0.9	0	0

b) Kandivali East MumbaiMPCB

	From Date	To Date	PM2.5	PM10	NO	NO2	NOx	NH3	SO2	CO	Ozone	Benzene	Toluene	Eth-Benz	MP-Xylen	RH	WS	WD	SR	BP	AT	RF	TOT-RF
0	01-01-2021 00:00	01-01-2021 00:15	128.23	212.93	100.04	104.42	204.49	12.23	14.07	1.41	1.88	1.16		25.9	8.29	79.65	1.22	81.5			22		0
1	01-01-2021 00:15	01-01-2021 00:30	129	209	41.63	97.18	138.79	12.3	12.33	1.19	2.02	0		3.31	8.4	79.3	0.74	95.02			21.84		0
2	01-01-2021 00:30	01-01-2021 00:45	129	209	23.17	86.31	109.48	13.31	11.82	0.99	6.97	0		19.16	14.58	79.07	0.35	120.3			21.61		0
3	01-01-2021 00:45	01-01-2021 01:00	129	209	15.23	79.61	94.85	12.65	11.56	0.92	11.1	0		3.54	42.41	79.2	0.36	208.56			21.52		0
4	01-01-2021 01:00	01-01-2021 01:15	114.11	179.42	11	60.94	71.94	12.35	11.13	0.93	21.72	0		14.42	30.01	79.95	0.39	141.17			21.46		0
5	01-01-2021 01:15	01-01-2021 01:30	109	174	9.09	47.08	56.17	12.03	10.68	0.91	20.07	0		47.16	25.77	80.32	0.29	105.99			21.41		0
6	01-01-2021 01:30	01-01-2021 01:45	109	174	9.26	45.73	54.97	12.1	10.67	0.97	18.85	0		25.12	13.3	81.03	0.24	119.94			21.32		0
7	01-01-2021 01:45	01-01-2021 02:00	109	174	9.61	50.6	60.2	9.77	11.18	1.05	15.01	0		28.87	29.09	81.75	0.23	133.68			21.25		0
8	01-01-2021 02:00	01-01-2021 02:15	116.26	168.17	10.09	53.74	63.83	9.98	11.68	1.04	9.42	0		20.52	20.61	82.74	0.23	156.63			21.12		0
9	01-01-2021 02:15	01-01-2021 02:30	119	167	11.04	53.94	64.98	11.15	12.72	1.13	7.98	0		1.78	1.65	83.68	0.27	161.36			21.08		0
10	01-01-2021 02:30	01-01-2021 02:45	119	167	12.56	56.76	69.33	12.51	12.53	1.1	5.44	0		31.44	16.53	84.15	0.27	162.35			20.99		0
11	01-01-2021 02:45	01-01-2021 03:00	119	167	12.44	52.32	64.79	13.94	12.09	1.11	5.24	0		61.04	3.03	84.36	0.31	207.31			20.88		0
12	01-01-2021 03:00	01-01-2021 03:15	124.17	182.21	10.86	44.57	55.43	15.56	11.79	1.01	10.47	0		14.16	19.08	84.03	0.3	128.4			20.8		0
13	01-01-2021 03:15	01-01-2021 03:30	126	185	9.15	36.64	45.76	16.89	11.79	1	11.14	0		27.79	30.22	83.87	0.31	85.89			20.74		0
14	01-01-2021 03:30	01-01-2021 03:45	126	185	8.73	29.66	38.39	18.02	12.02	0.84	11.72	7.94		20.89	6.37	83.93	0.29	58.3			20.64		0
15	01-01-2021 03:45	01-01-2021 04:00	126	185	8.73	28.17	36.9	18.11	11.31	0.88	9.67	8.11		1.81	13.2	84.41	0.31	92.83			20.55		0
16	01-01-2021 04:00	01-01-2021 04:15	109.76	146.96	37.14	39.12	76.26	18.59	11.77	0.9	9.26	0.63		19.3	1.33	84.38	0.22	174.02			20.54		0
17	01-01-2021 04:15	01-01-2021 04:30	104	140	15.26	38.76	54.04	18.69	13.05	0.97	7.89	7.29		41.55	14.22	84.75	0.27	90.16			20.53		0
18	01-01-2021 04:30	01-01-2021 04:45	104	140	10.15	35.12	45.29	18.65	13.06	0.89	7.95	0.62		10.75	4.27	84.52	0.22	86.92			20.51		0
19	01-01-2021 04:45	01-01-2021 05:00	104		9.77	32.35	42.11	18.78															0
20	01-01-2021 05:00	01-01-2021 05:15	105.48	146.76	9.85	32.55	42.4	18.78	10.61	0.9	7.66	0.19		16.85	27.41	84.47	0.28	114.21			20.81		0
21	01-01-2021 05:15	01-01-2021 05:30	106	148	9.68	31.23	40.92	17.93	10.9	0.93	6.14	0		44.24	16.72	84.3	0.38	79.86			20.9		0
22	01-01-2021 05:30	01-01-2021 05:45	106	148	10.76	33.65	44.41	16.99	10.77	0.89	7.18	0		4.77	50.4	84.08	0.34	85.96			20.81		0
23	01-01-2021 05:45	01-01-2021 06:00	106	148	11.32	34.3	45.62	19.78	11.13	0.85	7.32	0		14.52	19.98	84.14	0.31	76.95			20.63		0
24	01-01-2021 06:00	01-01-2021 06:15	103.79	141.24	22.55	39.31	62.03	18.6	10.49	0.79	9.27	0		6.84	13.75	84.26	0.25	94.72			20.56		0
25	01-01-2021 06:15	01-01-2021 06:30	103	140	84.98	53.71	139.24	12.46	12.02	0.94	2.16	4.18		7.83	7.05	84.86	0.26	121.62			20.59		0
26	01-01-2021 06:30	01-01-2021 06:45	103	140	79.81	49.68	129.41	11.98	13.72	0.97	2.01	0.3		6.2	2.15	85.42	0.19	140.01			20.59		0
27	01-01-2021 06:45	01-01-2021 07:00	103	140	57.65	54.58	112.16	12.19	13.45	1.23	1.26	13.72		56.1	24.62	85.38	0.29	110.14			20.67		0
28	01-01-2021 07:00	01-01-2021 07:15	111.93	207.62	46.63	52.06	98.67	12.46	14.34	1.28	1.9	1.32		36.86	35	85.74	0.24	124.67			20.65		0
29	01-01-2021 07:15	01-01-2021 07:30	115	220	38.84	45.93	84.76	12.7	14.25	1.32	1.62	5.82		11.52	3.27	86.25	0.29	141.99			20.71		0
30	01-01-2021 07:30	01-01-2021 07:45	115	220	102.99	48.82	151.57	12.15	15.96	1.73	1.98	0.49		8.95	29.13	86.2	0.52	94.95	3.48		20.87		0
31	01-01-2021 07:45	01-01-2021 08:00	115	220	70.05	47.98	117.88	11.44	15.42	1.55	1.7	0		20.81	32.98	85.4	0.54	60.04	16.48		21.1		0
32	01-01-2021 08:00	01-01-2021 08:15	129.36	283.39	60.26	42.07	102.32	10.4	13.19	1.44	1.56	15.03		27.13	9.33	84.04	0.47	59.46	32.2		21.52		0
33	01-01-2021 08:15	01-01-2021 08:30	134	295	140	47.54	187.52	8.76	14.29	1.74	1.96	1.47		6.91	2.06	82.39	0.74	138.21	47.75		22.16		0
34	01-01-2021 08:30	01-01-2021 08:45	134	295	142.17	47.24	189.39	9.45	13.88	1.65	1.79	3.95		6.04	23.1	80.38	0.43	163.74	68.57		23.11		0
35	01-01-2021 08:45	01-01-2021 09:00	134	295	188.54	55.59	243.97	7.62	15.42	2.11	1.3	0.33		12.48	13.54	77.79	0.58	221.97	102.33		24.25		0
36	01-01-2021 09:00	01-01-2021 09:15	122.67	252.74	116.93	81.8	198.74	16	14.03	1.91	1.92	4.09		9.9	5.89	73.89	0.59	242.26	139.5		25.22		0
37	01-01-2021 09:15	01-01-2021 09:30	119	245	104.55	86.22	190.76	11.53	15.05	1.87	2.27	7.52		20.15	19.24	69.8	0.53	238.04	175.69		26.41		0

c) Malad West Mumbai IITM

	From Date	To Date	PM2.5	PM10	NO	NO2	NOx	NH3	SO2	CO	Ozone	Benzene	Toluene	Eth-Benzer	MP-Xylene	RH	WS	WD	BP	AT	RF	TOT-RF
0	01-01-2021 00:00	01-01-2021 00:15	157.29	243.37	11.45	27.96	35.41	14.94	2.37	1.41	34.57	0.89		0.46	1.24	82.53		328.12	964.8	21.78	0	0
1	01-01-2021 00:15	01-01-2021 00:30	140.74	221.92	8.32	19.6	25.02	14.87	6.98	1.29	42.07	0.91		0.44	1.15	83.67		330.11	987.5	21.57	0	0
2	01-01-2021 00:30	01-01-2021 00:45	141.14	222.06	8.96	19.18	25.01	14.7	7.91	1.29	44.2	0.81		0.36	0.82	84.19		333.54	985.9	21.43	0	0
3	01-01-2021 00:45	01-01-2021 01:00	139.85	220.33	7.36	19.91	24.7	14.65	6.64	1.3	34.8	0.75		0.31	0.65	84.41		332.51		21.37	0	0
4	01-01-2021 01:00	01-01-2021 01:15	88.88	133.88	8.83	21.97	27.71	13.6	3.44	1.27	30.58	0.73		0.31	0.67	84.36		315.39		21.35	0	0
5	01-01-2021 01:15	01-01-2021 01:30	42.5	54.25	7.93	21.64	26.8	13.81	8.92	1.3	27.79	0.81		0.34	0.76	84.73		340.79	981	21.25	0	0
6	01-01-2021 01:30	01-01-2021 01:45	55.75	69.54	9.94	20.17	26.63	15.61	9.31	1.34	24.96	0.98		0.38	0.88	84.54		348.29		21.27	0	0
7	01-01-2021 01:45	01-01-2021 02:00	69.66	87.73	6.68	19.78	24.09	15.15	9.64	1.34	23.06	1.06		0.45	1.02	84.91		341.29		21.25	0	0
8	01-01-2021 02:00	01-01-2021 02:15	91.32	118.53	11.77	20.7	28.36	13.83	3.69	1.4	22.05	1.04		0.46	1.02	85.21		335.16		21.21	0	0
9	01-01-2021 02:15	01-01-2021 02:30	124.08	165.75	9.68	16.85	23.11	14.34	8.87	1.3	25.52	1.06		0.45	1.05	86.1		13.39	934.3	21.06	0	0
10	01-01-2021 02:30	01-01-2021 02:45	131.14	174.45	8.17	15.2	19.99	14.19	9.43	1.25	28.07	1.01		0.4	0.91	87.82		12.7	929.9	20.84	0	0
11	01-01-2021 02:45	01-01-2021 03:00	444.16	517.01	5.94	15.72	19.29	15.43	9.48	1.3	26.98	0.9		0.34	0.68	87.89		332.32		20.86	0	0
12	01-01-2021 03:00	01-01-2021 03:15	655.03	751.95	6.63	15.17	19.48	14.37	2.25	1.2	27.91	0.81		0.33	0.61	87.45		316.83		20.91	0	0
13	01-01-2021 03:15	01-01-2021 03:30	688.43	774.96	6.23	14.39	18.08	15.7	7.22	1.23	22.42	0.76		0.32	0.63	87.43		336.34	927.1	20.77	0	0
14	01-01-2021 03:30	01-01-2021 03:45	747.51	836.87	4.56	16.12	19.06	15.89	7.86	1.31	18.53	0.77		0.34	0.67	87.31		331.03		20.75	0	0
15	01-01-2021 03:45	01-01-2021 04:00	744.58	832.93	11.34	22.99	30.36	13.54	7.29	1.38	13.31	0.84		0.43	0.96	86.92		319.54		20.81	0	0
16	01-01-2021 04:00	01-01-2021 04:15	275.67	314.25	10.52	25.93	32.78	14.27	2.48	1.37	9.25	0.93		0.48	1.07	86.93		305.35		20.85	0	0
17	01-01-2021 04:15	01-01-2021 04:30	163.07	184.28	10.18	26.07	32.69	13.9	6.94	1.42	7.34	1.05		0.53	1.06	87.25		289.92		20.82	0	0
18	01-01-2021 04:30	01-01-2021 04:45	112.05	125.81	12.2	20.59	28.53	15.1	7.42	1.47	5.37	1.17		0.62	1.29	87.76		316.93		20.86	0	0
19	01-01-2021 04:45	01-01-2021 05:00	57.34	70.66	14.46	19.27	28.52	14.31	6.37	1.45	6.85	1.25		0.67	1.63	87.64		335.19		20.95	0	0
20	01-01-2021 05:00	01-01-2021 05:15	74.79	97.75	11.73	18.58	26.21	14.96	1.86	1.37	9.13	1.25		0.65	1.64	86.91		337.93		21.06	0	0
21	01-01-2021 05:15	01-01-2021 05:30	134.32	170.6	9.88	18.78	25.2	13.64	5.59	1.36	7.06	1.2		0.6	1.45	87.33		331.25		20.98	0	0
22	01-01-2021 05:30	01-01-2021 05:45	173.53	220.16	11.46	20.62	28.07	13.1	5.76	1.41	7.99	1.15		0.56	1.35	88.18		336.34		20.86	0	0
23	01-01-2021 05:45	01-01-2021 06:00	203.16	255.32	14.03	21.9	31.03	13.43	4.78	1.46	5.98	1.14		0.55	1.29	89.12		337.21		20.73	0	0
24	01-01-2021 06:00	01-01-2021 06:15	198.77	245.67	25.02	22.21	38.48	15.94	3.57	1.62	3.71	1.14		0.54	1.22	89.89		337.63		20.67	0	0
25	01-01-2021 06:15	01-01-2021 06:30	376.88	454.22	34.08	23.28	45.44	14.13	9.7	1.63	[illegible]	[illegible]		0.57	1.36	90.33		325.29		20.6	0	0

2.2 Feature Description

- **From Date / To Date:**Represent the time window during which each data point is recorded.The dataset captures data at **15-minute intervals**, providing a fine-grained temporal resolution.

- **PM2.5 (Particulate Matter ≤ 2.5 μm):** Extremely fine particles that can deeply penetrate into the lungs and even enter the bloodstream. A key indicator of urban air pollution and a major component of smog.

- **PM10 (Particulate Matter ≤ 10 µm):** Coarser particles originating from dust, vehicle emissions, construction activities, etc. Can cause respiratory issues, especially in sensitive groups.

- **NO2 (Nitrogen Dioxide):** Formed by oxidation of NO; a reddish-brown toxic gas. Participates in photochemical reactions that produce ozone.

- **SO2 (Sulphur Dioxide):** Released from burning fossil fuels like coal and oil. Irritates the respiratory system and contributes to acid rain formation.

- **CO (Carbon Monoxide):** A colorless, odorless, and highly toxic gas. Produced by incomplete combustion; harmful even at low concentrations over long exposure.

- **Ozone (O3):** A secondary pollutant formed when sunlight reacts with NOx and VOCs. Harmful at ground level (despite being beneficial in the stratosphere); causes respiratory problems.

- **Temp (Temperature):** Recorded in °C; affects the rate of chemical reactions in the atmosphere. Influences pollutant formation, especially ozone.

This rich set of features allows for a nuanced understanding of the factors influencing AQI at a micro-temporal and micro-geographic level. Not only does this allow us to predict AQI accurately, but it also enables intelligent recommendations based on pollutant concentrations and environmental conditions in real time.

III. Methodology

To create a robust, real-time AQI prediction and route recommendation system, the methodology follows a structured and scientific flow. The first step involves data cleaning, which is critical for time-series forecasting. Time gaps are handled through forward filling and interpolation techniques. Unreliable or extreme values (especially for PM2.5 and ozone) are identified and removed using statistical methods like IQR (Interquartile Range). Timestamp fields are converted into standard datetime format to ensure proper resampling and time-based operations.

After cleaning, the dataset is enriched using feature engineering. This includes the calculation of AQI from raw pollutant data based on CPCB India's official formula. Time-based features (hour, day, weekend, etc.) are extracted to model human and traffic behavior more effectively. Lagged variables and rolling means are introduced to capture trends and temporal dependencies, which are critical in time-series modeling. All features are normalized or encoded where necessary to ensure model compatibility.

The next step involves splitting the dataset into training and testing sets, respecting temporal order. This avoids lookahead bias, which is common in naive train-test splits on time-series data. Once the data is prepared, the project employs XGBoost, a powerful ensemble machine learning model, for regression tasks. The model is trained on past intervals to forecast the AQI of future time steps. A recursive multi-step forecasting strategy is used to make forward predictions (t+1 to t+6), making the approach practical for real-world usage. Evaluation metrics like RMSE, MAE, and R^2 score are used to quantify model performance.

3.1 Data Cleaning

a) **Code Reference:** load_and_clean_data(file_path)

b) **Objective**: Convert messy sensor readings into clean, structured, time-indexed data suitable for time-series modeling.

c) **Column Cleaning:** Strips white spaces, converts column names to lowercase, replaces spaces with underscores. Renames critical columns for consistency (ozone → o3, pm2.5 → pm25, from_date → timestamp).

d) **Datetime Parsing:** Converts the timestamp to pandas datetime format using error coercion to handle corrupted entries.

e) **Filtering & Resampling:** Keeps only relevant pollutants (pm25, pm10, no2, o3). Resamples data to hourly frequency, taking the average within each hour (ensures regular time intervals).

f) **Row-wise AQI Calculation:** For every timestamp, the CPCB AQI formula is applied pollutant-wise. The final AQI value for a row is the maximum AQI among PM2.5, PM10, NO2, and O3, based on CPCB breakpoints.

3.2 Feature Engineering

a) **Code Reference:** add_features(df)

b) **Objective:** Enhance raw pollutant values with derived features that capture trends, seasonality, and volatility — improving model prediction power.

c) **Rolling Mean (3 & 7 hours):** Smooths pollutant data to capture short-term (3 hours) and slightly longer-term (7 hours) trends. E.g., pm25_roll3 indicates average PM2.5 in the past 3 hours.

d) **Rolling Standard Deviation:** Captures volatility — how much the pollutant concentration fluctuates locally. E.g., pm25_std3 flags unstable conditions.

e) Temporal Features:
- dayofweek: Identifies weekly pollution cycles (e.g., higher traffic on weekdays).
- month and day: Detect seasonal and monthly trends.

3.3 AQI Calculation

a) **Code Reference:** calculate_cpcb_aqi(pollutant, concentration)

b) **Objective:** To convert raw pollutant concentration values into standardized Air Quality Index (AQI) scores using government-defined thresholds, specifically following the Central Pollution Control Board (CPCB) guidelines.

Each pollutant (such as PM2.5, PM10, NO_2, and O_3) has specific concentration ranges, known as *breakpoints*, which are mapped to AQI values. These breakpoints help translate raw pollutant measurements into a consistent AQI scale that reflects potential health impacts.

The formula used to compute the AQI for any pollutant within a given breakpoint range is:

$$I = \left(\frac{I_{high} - I_{low}}{C_{high} - C_{low}} \right) \times (C - C_{low}) + I_{low}$$

Where:

- C = Measured concentration of the pollutant
- I = Calculated AQI value
- C_{low}, C_{high} = The lower and upper limits of the pollutant concentration range
- I_{low}, I_{high} = The corresponding AQI values for that range

After calculating AQI values for each pollutant, the final AQI for a given time and location is determined by taking the maximum AQI across all pollutants, representing the most critical pollutant affecting air quality at that moment.

3.4 Model Training

a) **Code Reference:** train_aqi_model(df, flag)

b) **Objective:** Train a powerful regression model to learn complex relationships in pollutant data and temporal features.

c) Model Chosen: XGBRegressor

d) Inputs (X): All features except the AQI column.

e) Output (y): The computed AQI (calculated_aqi).

f) Metrics Used:

- MAE (Mean Absolute Error): Measures average magnitude of error.
- RMSE (Root Mean Square Error): Penalizes larger errors more.
- R^2 Score: Proportion of variance explained by the model.

IV. Results

4.1 Model Performance Across Locations

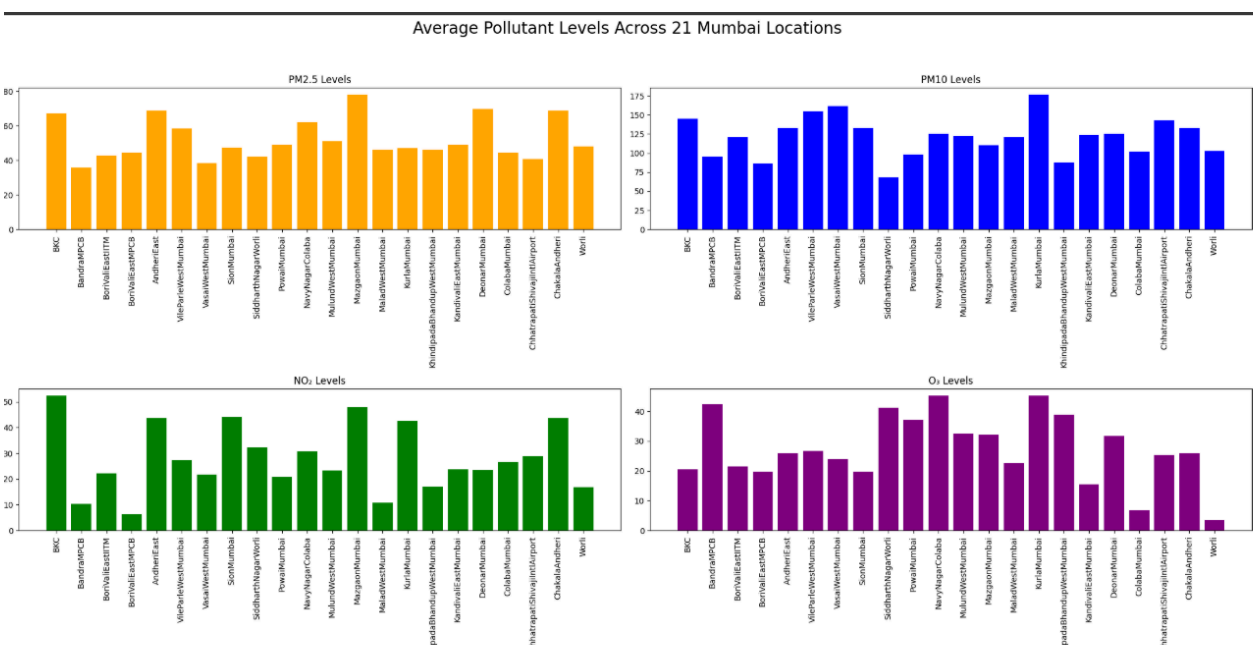
Location	MAE	RMSE	R^2
BKC	12.5851	23.1	0.9482
BandraMPCB	3.3716	8.92	0.9637
BoriValiEastIITM	4.4051	10.45	0.978
BoriValiEastMPCB	3.6721	10.7	0.9739
AndheriEast	7.7617	22.09	0.945
VileParleWestMumbai	8.7239	33.96	0.8477
VasaiWestMumbai	3.5126	6.06	0.9764
SionMumbai	6.8324	16.7	0.972
SiddharthNagarWorli	4.5236	10.5	0.9762
PowaiMumbai	3.6729	8.48	0.9888
NavyNagarColaba	5.017	14.09	0.9818
MulundWestMumbai	3.22	8.62	0.9844
MazgaonMumbai	14.2606	35.63	0.9189
MaladWestMumbai	3.7074	9.5	0.9797
KurlaMumbai	5.5374	18.33	0.9521
KhindipadaBhandupWestMumbai	5.1842	17.5	0.9682
KandivaliEastMumbai	3.1613	9.33	0.9812
DeonarMumbai	4.0476	9.88	0.9916
ColabaMumbai	3.1845	9.4	0.9839
ChhatrapatiShivajiIntlAirport	2.9099	12.38	0.9715
ChakalaAndheri	7.7617	22.09	0.945
Worli	3.0515	8.62	0.9859

These results suggest that the model performs especially well in areas like Worli, Powai, Siddharth Nagar, and Colaba, with R^2 values nearing 0.99, indicating almost perfect correlation. Areas like Mazgaon and Vile Parle, which show slightly higher error rates, may be impacted by data inconsistencies or localized pollutant sources not captured in the dataset.

4.2 Visualizations

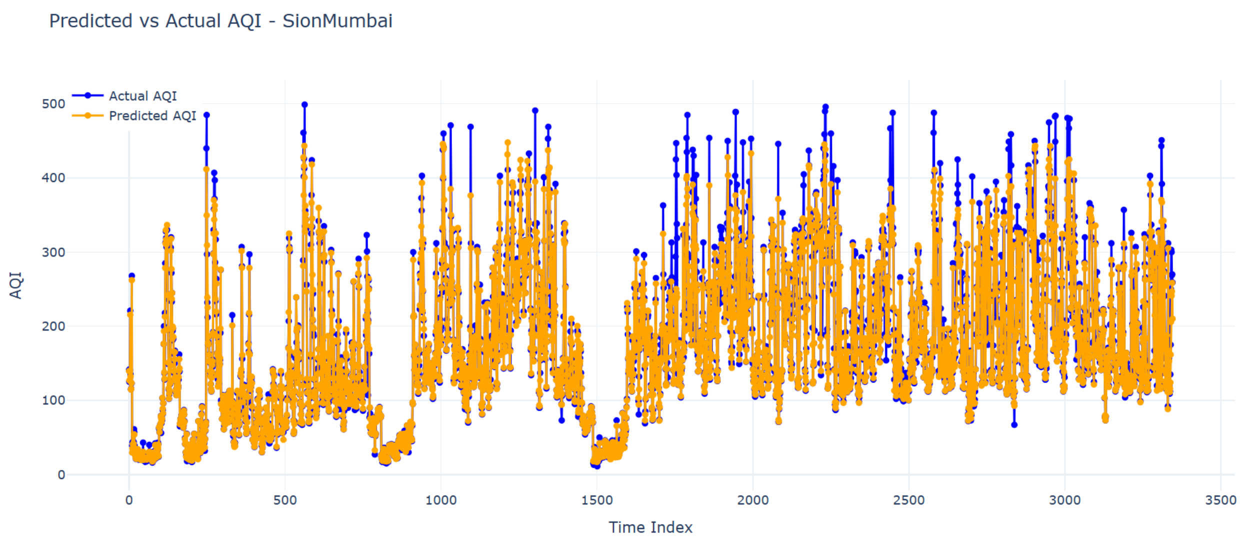
4.2.1 Pollutant Graph across each location

Pollutant trend graphs illustrate the temporal variation of major pollutants such as PM2.5, PM10, NO_2, and O_3 across different time periods. These plots help in identifying cyclical behaviors, sudden spikes, and pollutant-specific patterns, such as morning or seasonal peaks. Observing these trends provides crucial insight into which pollutants contribute most heavily to AQI fluctuations at different times, helping fine-tune both predictions and policy recommendations.



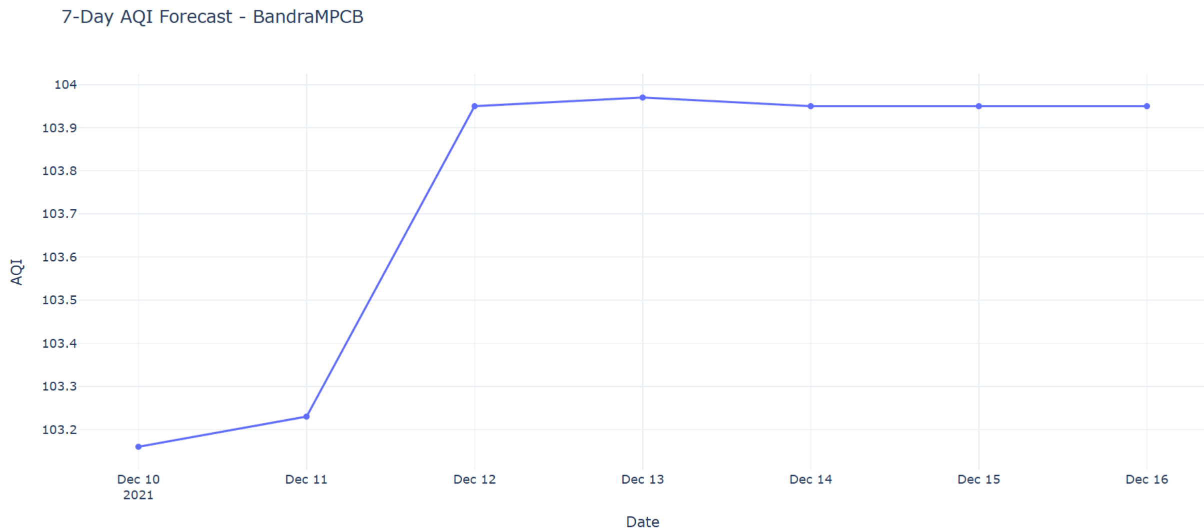
4.2.2 Predicted v/s Actual Graph

The Predicted vs Actual AQI plot serves as a key evaluation metric for the model's performance. It visually compares the AQI values forecasted by the model against the actual recorded values over time. In the graph, the predicted AQI (in orange) closely follows the trend of the actual AQI (in blue), indicating strong alignment and low deviation. This consistency across fluctuating pollution levels showcases the model's ability to capture both short-term spikes and long-term patterns effectively.



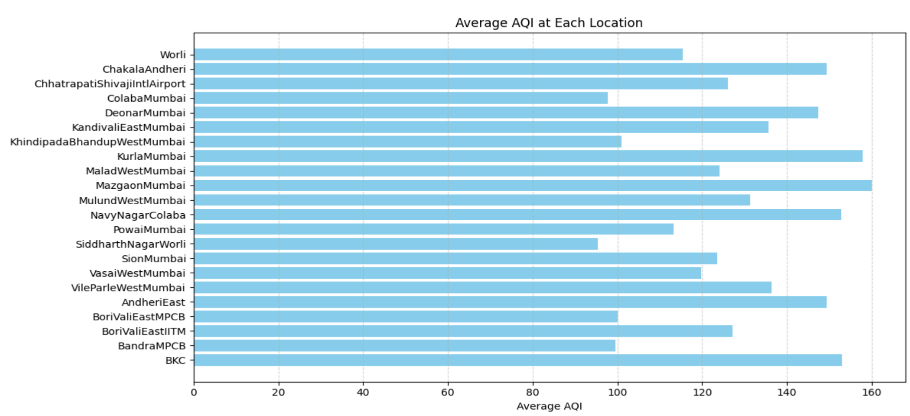
4.2.3 Forecast Graph

The forecast graph highlights the model's ability to predict future AQI values based on historical pollutant and weather data. By extending AQI projections beyond the test dataset, the graph demonstrates how the model performs in unseen conditions. The predicted line follows a realistic trajectory consistent with previous data trends, reinforcing the model's utility for real-time air quality forecasting and proactive public health measures.



4.2.4 Average AQI at Each Location

The average AQI per location chart provides a comparative overview of air quality across multiple monitoring stations in Mumbai. By aggregating data over time, it reveals consistently high or low pollution zones, helping pinpoint pollution hotspots like Mazgaon and Vile Parle. Such a visualization is useful not only for validating model results but also for urban planning, awareness campaigns, and recommending safer, low-AQI routes for cyclists and pedestrians.



V. Streamlit Integration

5.1 AQI Forecasting at Specific Coordinates

The application forecasts AQI based on user-provided latitude and longitude by referencing the nearest monitoring station or location-based data. Using machine learning models trained on historical pollutant data, it predicts future air quality for that specific region. This helps users understand not just the current AQI, but also what to expect in upcoming hours or days, enabling more informed decisions around outdoor activities and route planning.

5.2 Map Rendering with Route Visualization

Using the given coordinates, the application dynamically renders an interactive map that displays AQI levels and routes. It highlights paths between source and destination, selecting the one with the lowest predicted pollution exposure. Color-coded markers indicate air quality levels at different points, allowing users to visually assess safer routes. This spatial visualization enhances usability and helps translate data into clear, actionable insights.

VI. Conclusion

This project successfully demonstrates the power of data-driven decision making in addressing real-world environmental and health challenges. By combining AQI forecasting with geospatial optimization, it creates a dual-layered system that is both reactive and proactive. The use of robust machine learning models like XGBoost ensures accurate predictions, while Streamlit enables effective communication of results to the end user.

The broader impact of this solution lies in its applicability across cities and use cases — from urban cyclists and delivery workers to municipal planners and environmental watchdogs. By scaling the approach to more locations and integrating real-time data from IoT air sensors, the system could evolve into a real-time pollution-aware navigation engine. The incorporation of route safety, noise levels, and time optimization would further enhance its utility.