ECAS-7 2025 Conference

The 7th International Electronic Conference on Atmospheric Sciences

04–06 June 2025 | Online

MDPI

Particulate Matter $PM_{2.5}$ Concentration Prediction in Tashkent Using Machine Learning

Umida Madiyarova[1], Jaloliddin Erkinov[1], Moulay Rachid Babaa[2]

[1]Department of Software Engineering, New Uzbekistan University, Tashkent, Uzbekistan

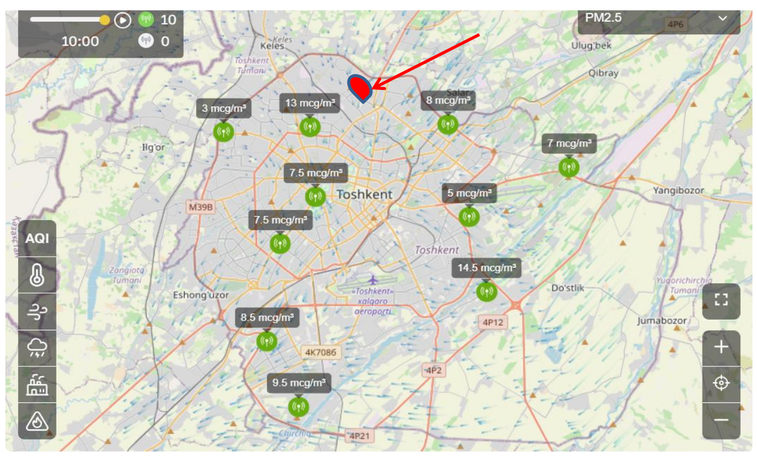
[2]Department of Chemical and Materials Engineering, New Uzbekistan University, Tashkent, Uzbekistan

INTRODUCTION & AIM

Particulate matters, specifically PM2.5, is widely recognized as a significant air pollutant due to strong scientific consensus regarding its effects on health and well-being [1]. It also has critical connections with other key air pollutants, such as sulfur dioxide and nitrogen dioxide. Central Asian cities often experience hazardous air quality due to several factors, including home heating in winter, industrial emissions, vehicle traffic, frequent dust storms, and limited environmental regulations [2]. These sources of air pollution have serious health implications for residents and contribute to the growing environmental challenges in the region. In this work, simple Machine Learning (ML) models were used to predict the hourly concentrations of PM2.5 in Uzbekistan's largest and most populous city, Tashkent considering meteorological variables . Five models including Linear regression, Ridge, LASSO and ensemble models namely Random Forest, and XGBoost were compared by taking into account, temperature, relative humidity, precipitation and timestamp for temporal patterns capture.

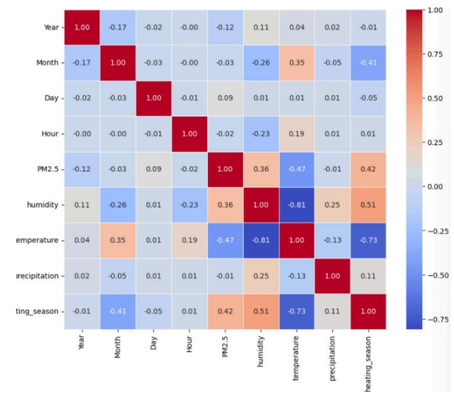
METHOD

this research utilized publicly available air quality data collected from U.S. Embassy Air Quality Monitoring Station in Tashkent (indicated by a red pin on the map). The dataset comprises hourly observations spanning from January 2023 to October 2024, providing nearly two years of continuous records. This time frame encompasses all seasonal variations, offering a good basis for capturing seasonal patterns and meteorological influences on $PM_{2.5}$ levels. The collected data includes the following variables: $PM_{2.5}$ (µg/m³),Temperature (°C), Humidity (%, Precipitation (mm) and Timestamp (used for time-series indexing). The dataset was subjected to essential preprocessing. Missing values in temperature and humidity were imputed using the mean of the two temporally closest data points. Missing values in the $PM_{2.5}$ were left unaltered to avoid introducing artificial bias.To prepare the data for machine learning, all numerical features were normalized using the Min-Max scaling technique. In addition, heating season, a binary feature indicating whether the observation occurred during the older months (typically November to March) was introduced to account for elevated emissions from residential heating. The models were evaluated based on their performance using metrics such as mean absolute error (MAE), root mean squared error (RMSE), and R^2.



RESULTS & DISCUSSION

Heat Map



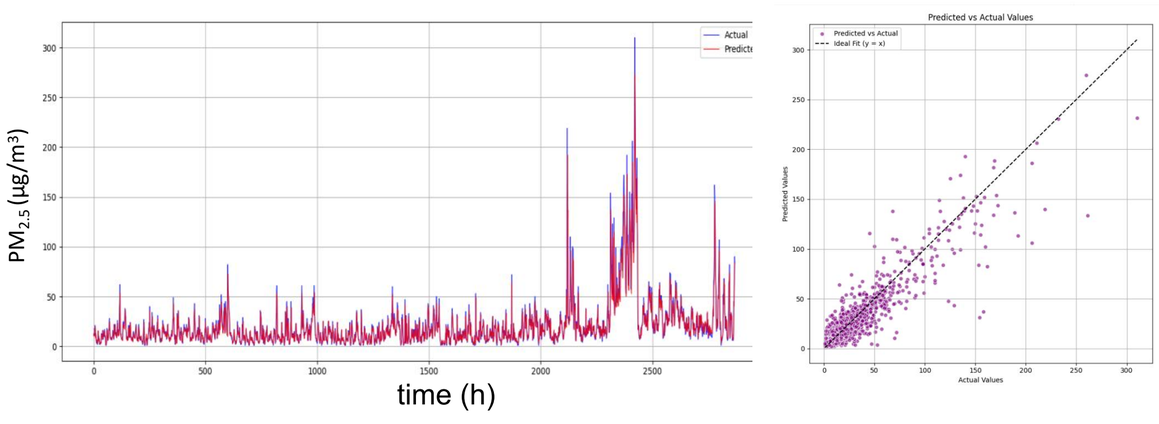
$PM_{2.5}$ concentration is negatively correlated with temperature indicating higher pollution during colder months (residential heating) while showing a moderate positive correlation with humidity.

Performance Metrics of Machine Learning Models

Model	MAE	RMSE	R^2
Linear Regression	5.6227	10.4067	0.8337
Ridge Regression	5.5788	10.3547	0.8354
Lasso Regression	**5.4968**	**10.3308**	**0.8361**
Random Forest	6.2856	12.7472	0.7505
XGBoost	6.2576	12.5441	0.7584

Simpler regularized linear models like Lasso and Ridge offered the most balanced performance between accuracy and generalization

Lasso regression predictions-comparison with actual values:



LASSO presents a good balance between accuracy and interpretability

CONCLUSION

Five ML models were succesfully developed and used for ambiant $PM_{2.5}$ concetrations prediction in Tashkent.

These results suggest that Lasso, as an interpretable ML model is the best for reducing large errors and is thus more robust in capturing variations in $PM_{2.5}$ levels.

FUTURE WORK / REFERENCES

[1]Sarawut Sangkham, Worradorn Phairuang, Samendra P. Sherchan, Nattapon Pansakun, Narongsuk Munkong, Kritsada Sarndhong, Md. Aminul Islam, Pornpun Sakunkoo, An update on adverse health effects from exposure to PM2.5, Environmental Advances, Volume 18, 2024,100603, https://doi.org/10.1016/j.envadv.2024.100603.

[2] Tursumbayeva M, Muratuly A, Baimatova N, et al. (2023). Cities of Central Asia: New hotspots of air pollution in the world. Atmospheric Environment. DOI: 10.1016/j.atmosenv.2023.119901