Towards Unbiased Air Quality Data Analysis: A Holistic Framework for London

Ejaz Hussain [1]

University of East London
School of Architecture, Computing and Engineering
ACE Research, Impact & Innovation Conference September 2024
(Abstract & Presentation Slides)

The research thesis presents a conceptual framework designed to address and mitigate bias in air quality data analysis for London. With the increasing reliance on advanced data science techniques, particularly machine learning (ML), bias defined as any deviation that may compromise the objectivity or fairness of analytical outcome, which has emerged as a significant challenge. Such bias threatens the accuracy and reliability of air quality assessments, which are crucial for evidence-based policymaking. Despite the widespread adoption of machine learning in this field, current methodologies often overlook the potential for bias to infiltrate various stages of data processing, from collection through to model development and analysis. The research identifies a critical gap in existing literature, which has predominantly focused on other critical aspects of air quality monitoring, such as sample fairness, model transparency and data collection mechanisms, while neglecting the complex biases that can arise during data analysis. Data biases can lead to inaccurate assessments, potentially resulting in misguided policies that fail to address the root causes of air pollution effectively. To address this oversight, the thesis proposes a holistic framework that integrates multiple air quality datasets, including those from London Air and UK Local Government Monitoring sites, into a unified dataset optimised for unbiased analysis and vice versa. A key innovation of this framework is the introduction of a novel scoring methodology designed to evaluate and mitigate bias risks throughout the data analysis life cycle. This scoring method considers various factors, including the reliability of data sources, potential inaccuracies in sensor readings, and the presence of confounding variables that could unintentionally introduce bias. By systematically assessing these factors, the framework aims to minimise bias at every stage of the analysis, thereby enhancing the overall validity and reliability of the findings. The significance of this research lies in its potential to provide a robust and systematic approach to ensuring unbiased air quality data analysis. Accurate and reliable data are essential for developing effective strategies to combat air pollution, a pressing concern for London and other major urban cities. By improving the accuracy of air quality predictions and reducing the risk of bias, this framework supports the development of more targeted and impactful policies aimed at improving air quality and public health outcomes. Furthermore, the framework serves as a valuable resource for researchers and policymakers, offering a systematic process for identifying and addressing bias in complex air quality data analysis. The research also highlights the ethical implications of biased data analysis, emphasising the need for transparency and accountability in the use of advanced data science techniques in public policy. By fostering ethical and unbiased practices, this study aims to contribute to more informed and effective decision-making in the realm of environmental policy. In conclusion, this thesis advances the field of air quality analysis by proposing a novel, holistic framework

that not only addresses existing gaps in the literature but also offers practical solutions for mitigating bias in data analysis. The findings have broad implications for both the academic community and policymakers, ultimately supporting the goal of achieving cleaner air and healthier environments for urban populations.

Keywords: Air Quality, London Air Quality, Data Analysis, Predictive Air Quality Analysis, Bias in Air Quality, Bias Classifications, Types of Bias, Bias Checklist, Bias Framework, Climate Data Science, Bias Scoring, Air Quality Machine Learning.

Presentation Slides: Attached