جامعة محمد الخامس بالرباط
Université Mohammed V de Rabat

École Nationale Supérieure d'Informatique et d'Analyse des Systèmes
Centre d'Études Doctorales en Sciences des Technologies de l'Information et de l'Ingénieur

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Madame Ihsane GRYECH

Soutiendra publiquement sa thèse de Doctorat en Informatique

Le Samedi 21 Janvier 2023 à 10H00 au Grand amphi à l'ENSIAS

Intitulé de la thèse

MACHINE LEARNING FOR URBAN POLLUTION MONITORING AND ITS IMPACT ON INFANT HEALTH

Devant le Jury composé de :

Président :

Pr. Hanan El Bakkali, PES, Université Mohammed V, Rabat

Directeur de thèse :

Pr. Abdellatif Kobbane, PES, ENSIAS, Université Mohammed V de Rabat

Co-Directeur de thèse :

Pr. Mounir Ghogho, PES, Université Internationale de Rabat

Rapporteurs :

Pr. Juan Carlos Augusto, Professor, Middlesex University, London, United Kingdom

Pr. Desmond C. McLernon, Reader, University of Leeds, Leeds, United Kingdom

Pr. Mohammed Boulmalf, PES, Université Internationale de Rabat, Rabat

Examinateur:

Pr. Houda Benbrahim, PES, ENSIAS, Université Mohammed V de Rabat

Abstract: Air pollution has been and still is one of the major environmental issues facing our planet and contributing to its global warming. Many scientific studies have shown that air pollution is the root cause of many respiratory and cardiovascular diseases, and cancers. The respiratory health of children is particularly affected by air pollution, as their lungs have not been fully developed yet.

The number of environmental pollution-induced deaths of children under five years old in Morocco is the highest in North Africa. However, Moroccan citizens are not informed about the spatial distribution of air quality, particularly in urban areas. Indeed, capturing the spatial variability of air pollution in cities requires a dense deployment of monitoring stations. This is an issue for developing countries, such as Morocco, because the deployment and maintenance of these monitoring stations are extremely expensive.

In this thesis, we aim to improve air pollution monitoring in urban areas by means of Data Science and IoT solutions. We focus on addressing three main research axes: urban air pollution modeling, air pollution monitoring and data measuring, and spatio-temporal air pollution forecasting. We approach the first axis by analyzing the air quality trends in Paris, France and predicting the concentration of different pollutants using Machine Learning approaches. This axis also considers real-world issues such as missing data from out-of-order stations, as a way to improve the monitoring process. In the context of the air pollution monitoring research axis, we design a sensor deployment strategy based on mobile and nomadic sensors as well as on medical data collected at a children hospital. The monitoring process resulted in the "MoreAir" open dataset, a Moroccan air pollution dataset available to the research community, professionals and governmental entities interested in regulating and solving air quality issues in under-represented areas. For air pollution forecasting, we explore the previously acquired dataset to inspect the various factors contributing to a higher concentration of pollutants. The richness of the dataset allowed us to improve the effectiveness and dependability of air quality monitoring and modeling. A key finding of this thesis is outlined in the first axis of our thesis, which investigates correlations between air pollution concentrations and road traffic and meteorological data.

The non-linear impact of traffic flows and meteorology on the concentrations of pollutants in the atmosphere was better explained by SVM than other machine learning models. Additionally, it is shown that more accurate prediction models can be obtained when some pollutants' concentrations are used as predictors. This allows for inferring the concentration of some pollutants from those of other pollutants, thus reducing the number of air pollution sensors. Furthermore, our study provides evidence that predicting the data of out of order measurement stations can be accomplished by using data from other stations. The second axis of the thesis considered the involvement of the people most affected by urban pollution and looked in detail at the monitoring fields methods and applications in order to determine the necessary functionalities to collect

interesting and useful data. Moreair was introduced as a research solution that can provide insights on air quality data, as well as 52 temporal and geographical features. Another axis uncovers the most important aspects of understanding air quality distribution in Morocco, as well as the factors required to monitor the spatio-temporal distribution of air quality in urban areas. These same features are then used as predictors to forecast air quality. The findings of this thesis reveal several practical applications worthy of consideration and relevant to a wide range of contexts.

Keywords: Air Quality, Air Quality Monitoring, Geographical Information Systems, Machine Learning, Mobile sensing, IoT, Particulate matters, Pollutants' Prediction, Pollution Hot-spots identification, Pollution Forecasting, Respiratory health, Urban air pollution.