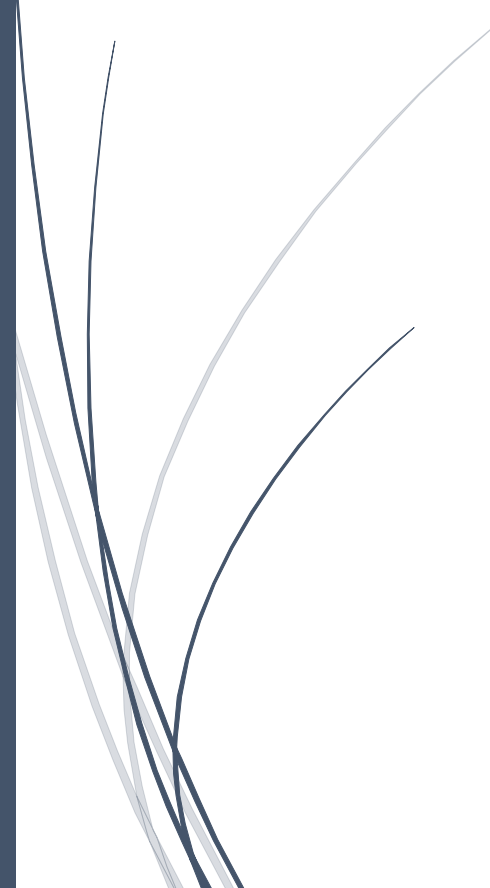


The logo for RADemics, featuring the text "RADemics" in white on a blue arrow-shaped background pointing to the right. The arrow is part of a larger blue horizontal bar that is positioned over a dark blue vertical bar on the left side of the slide.

RADemics

Predictive Analytics for Environmental Pollution Detection Using Machine Learning

A decorative graphic consisting of several thin, curved lines in shades of blue and grey, originating from the bottom left corner and extending upwards and to the right, resembling stylized grass or reeds.

Baseera A, Sharmila Joseph
VIT Bhopal University

Predictive Analytics for Environmental Pollution Detection Using Machine Learning

¹Baseera A, Associate Professor, School of Computing Science and Engineering (SCOPE), VIT Bhopal University, Kothrikalan, Sehore, Madhya Pradesh, India. baseera.a@vitbhopal.ac.in

²Sharmila Joseph, Assistant Professor, School of Computing Science and Engineering, VIT Bhopal University, Indore Road, Bhopal, Madhya Pradesh, India. Sharmila.joseph@vitbhopal.ac.in

Abstract

Environmental pollution, in its various forms, poses significant threats to human health, ecosystems, and the global climate. With urbanization and industrialization escalating, effective monitoring and management of pollution have become paramount. Predictive analytics, powered by machine learning (ML), has emerged as a transformative approach to forecasting and mitigating the impacts of environmental pollutants. This chapter explores the applications of machine learning in environmental pollution detection, focusing on key areas such as air quality monitoring, water pollution prediction, and noise control. By integrating data from diverse sources, such as IoT sensors, satellite imagery, and meteorological data, machine learning models offer enhanced accuracy in forecasting pollution levels, identifying pollution hotspots, and optimizing management strategies. Special emphasis is placed on the development of hybrid models that combine regression, classification, and clustering techniques to provide a holistic view of pollution dynamics. Additionally, the chapter examines the role of predictive analytics in addressing emerging environmental challenges such as eutrophication and algal blooms, where machine learning algorithms are deployed to predict and manage these complex phenomena. Despite challenges in data quality and model generalization, machine learning continues to demonstrate immense potential in transforming environmental monitoring systems. The integration of real-time data, advanced machine learning techniques, and adaptive systems is poised to significantly improve decision-making in pollution control and environmental sustainability.

Keywords: Environmental Pollution, Machine Learning, Predictive Analytics, Hybrid Models, Air Quality Monitoring, Noise Control.

Introduction

Environmental pollution has become one of the most pressing global challenges, affecting air, water, soil, and noise levels across the planet [1]. With the continuous growth of industrialization, urbanization, and transportation systems, the consequences of pollution are increasingly evident in deteriorating ecosystems, public health problems, and economic costs [2]. Air quality, water contamination, and excessive noise levels not only impact the immediate environment but also pose long-term threats to biodiversity, human health, and the climate [3]. Pollution has been linked to a wide range of health issues, including respiratory diseases, cardiovascular problems, and neurological disorders. In response, governments, policymakers, and environmental organizations are increasingly turning to advanced technologies to manage and mitigate pollution's impact [4]. Among the most promising innovations in this area is predictive analytics powered by machine

learning, a technology that offers the potential to forecast pollution levels, identify pollution sources, and inform more effective policy-making and public health interventions [5].

Machine learning (ML) has gained significant attention for its capacity to handle large volumes of environmental data and extract actionable insights from complex datasets [6]. Traditional pollution monitoring methods, such as manual sampling and laboratory analysis, are labor-intensive and time-consuming, limiting their ability to provide real-time information [7]. In contrast, machine learning models can continuously analyze data from a wide range of sources sensors, satellite imagery, weather stations, and social media platforms to generate timely, real-time predictions of pollution levels [8]. By processing vast amounts of real-time environmental data, machine learning can quickly detect changes in pollution patterns and provide early warnings of potential risks [9]. These capabilities make machine learning an essential tool in environmental pollution detection, offering more accurate, faster, and scalable solutions compared to conventional methods [10].

One of the most important applications of machine learning in environmental pollution detection is air quality monitoring [11]. Air pollution, particularly in urban areas, has become a major health concern, with fine particulate matter (PM_{2.5}), nitrogen oxides (NO_x), sulfur dioxide (SO₂), and ozone contributing to poor air quality and health risks [12]. Machine learning algorithms can predict air quality by analyzing data from fixed air quality sensors, meteorological conditions, traffic patterns, and industrial emissions [13]. These predictive models are capable of forecasting air pollution levels hours or days in advance, enabling governments and public health agencies to issue warnings and take preventive actions [14]. machine learning models can identify the sources of pollution, helping policymakers target interventions more effectively. By integrating machine learning into air quality monitoring systems, cities can better manage air pollution, reduce exposure to harmful contaminants, and improve public health outcomes [15].