Project Proposal

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

Descriptive Epidemiological and Predictive Modeling Study Of Air Pollution Impact On Urban Public Health In Indonesia

Faradhysa Camila (MCS231034)

Outline

- Introduction
- Literature Review
- Research Methodology
- Initial Result
- Conclusion

Introduction

Problem Background

- Indonesia's rapid urban development results in high emissions from transport, industries, and buildings, causing significant air pollution, particularly in major cities.
- High pollution levels in these urban areas lead to severe health impacts, including increased occurrences of respiratory diseases, heart attacks, and strokes, especially among vulnerable groups like children, the elderly, and those with pre-existing health conditions.
- Addressing the pollution problem is challenging due to inadequate statistical analysis and research focused on urban settings in Indonesia, resulting inefficient solutions.
- Most studies are built on partial information, necessitating the accumulation and systematization of extensive data on pollution concentrations and health indicators.
- The study aims to create a solid theoretical framework and develop predictive models to better understand the extent of air pollution and its immediate effects on public health in Indonesian cities.

Problem Statement

Environmental and public health concerns such as air pollution remains to be a significant threat affecting urban areas in Indonesia. The problem concerned by this study is pollution, particularly air pollution, as it affects the health of the people in urban centres of Indonesia. The goal of this research is to estimate the impact on human health caused by air pollution, as well as estimate the future consequences of pollution and construct the corresponding models.

Research Objective

1. To assess the current levels of air pollution in major Indonesian urban areas.
2. To identify the most prevalent health outcomes related with air pollution in urban areas.
3. To develop predictive models that can forecast future air pollution levels and related health risks.

Scope

1. The research focuses on collecting and analysing air quality data of urban areas in Indonesia which are in Jakarta and Surabaya.
2. The research focuses on collecting and analysing health data on respiratory and cardiovascular diseases.
3. The target of the research is to integrate air quality and health data to identify correlations and trends.
4. The descriptive analysis is to identify approaches to mitigate the health risks caused by air pollution.
5. The predictive model is to forecast future air pollution levels and health risks.

Literature Review

Air Pollution in Indonesia

- Indonesia faces significant air pollution challenges due to rapid urbanization, industrial activities, and car emissions, particularly in cities like Jakarta and Surabaya. High levels of pollutants (PM2.5, PM10, NOx, SO2, VOCs) have led to an increase in respiratory and cardiovascular diseases among urban residents. More extensive research combining air quality and medical data is needed to fully understand these health effects.
- Policy responses in Indonesia have been varied and often hindered by limited resources. Measures like stricter emission limits and promoting public transportation have shown some improvement, but more focused strategies are needed to mitigate the health impacts of urban air pollution. This study aims to explore the health effects of air pollution in major urban areas and support the development of effective policies.

Public Health in Indonesia Urban Area

- In Indonesia's urban areas, the accessibility of public health is influenced by environmental conditions, social factors, and healthcare infrastructure. While urbanization has improved healthcare access, it has also led to new health challenges related to environmental pollution and lifestyle changes. Urban residents face increased rates of pulmonary diseases, such as asthma and chronic obstructive pulmonary disease (COPD), as well as cancer and cardiovascular conditions linked to pollutants like PM2.5 and NOx.
- Despite the evident health impacts of air pollution, environmental health is often overlooked in health planning in Indonesia. This study aims to emphasize the importance of air pollution as a significant health predictor and advocate for integrating environmental health considerations into public health practices to better address these issues.

Descriptive Epidemiology

Descriptive epidemiology will be used to analyze trends in air pollution and its health effects in urban Indonesia. This approach helps identify geographical and temporal patterns, assess current pollution levels, and determine vulnerable populations and high-risk areas. The goal is to develop effective policies to address air pollution and its health impacts.

Predictive Modeling

Predictive modeling will be used to forecast future air quality and health impacts in Indonesian cities, employing statistical methods and machine learning techniques. The study will use data on air quality and health to develop and test models that inform policies for mitigating pollution and its effects.

Integrated Approach

This study will use a multi-disciplinary approach to analyze air pollution and its health risks by integrating various data sources, including air quality data and health records. Acknowledging the complex nature of air pollution and its impacts, the research will employ descriptive epidemiology and predictive model to provide a comprehensive understanding of how air pollution affects health in Indonesian cities and to develop effective prevention measures.

Research Methodology

Data Science Project Life Cycle

Problem Definition	Identify and understand the research issue, air pollution's impact on health in Indonesian cities, defining pollutants, health endpoints, and regions.
Data Collection	Collect standardized and accurate data from government sources, websites, and hospital records to link air pollution with health outcomes.
Data PreProcessing	Clean and standardize data to address missing values, remove unnecessary information, and ensure data quality before analysis.
Data Modelling	Use mathematical and statistical methods to identify patterns and relationships, employing descriptive models for insights and predictive models for forecasts.
Evaluation	Assess model accuracy and relevance by testing with data and comparing results to actual outcomes to ensure the conclusions are valid.

Data Source and Collection Methods

Air Quality Data

Purpose: Gather data to determine pollution rates in specific urban region.

Data Source:

- IQAir Real-time Data
- Badan Pusat Statistik (BPS)

Health Data

Purpose: Determine the link between air pollution and health, focusing on respiratory and cardiovascular diseases.

Data Source:

- Badan Pusat Statistik (BPS)
- Satu Sehat (Indonesia Ministry of Health)
- Local Hospitals

Data Pre-Processing

Data Cleaning

Remove noise and prepare data for analysis.

Steps:

- Transfer Data
- Remove Duplicates
- Handle Missing Values

Data Transformation

Make data suitable for analysis by standardizing formats.

Steps:

- Standardization
- Create Additional Columns
- Data Formatting

Data Integration

Combine air quality and health data into a single dataset.

Steps:

- Combine Datasets
- Validate Data
- Final validation

Analytical Methods

1. **Descriptive Analysis**
   Descriptive Analysis uses basic statistics and visualizations (e.g., charts, heat maps) in Power BI and Python to summarize data properties and reveal patterns in air pollution and health impacts.

2. **Predictive Analysis**
   Predictive Analysis uses historical data and tools like Power BI's forecasting algorithms and Python's ARIMA and Prophet models to estimate future air pollution levels and health risks.

3. **Statistical Analysis**
   Statistical Analysis involves using tools like Power BI and Python libraries to test hypotheses and determine relationships between air pollution and health outcomes through correlation and regression analyses.

4. **Machine Learning**
   Machine Learning employs Python's Scikit-learn to build predictive models (e.g., decision trees, random forests) that forecast future air pollution and health impacts, capturing complex, nonlinear relationships.

Initial Result

Exploratory Data Analysis

	Primary Data	Secondary Data
Visualization	Use Power BI to create line charts, bar charts, and scatter plots from directly collected air quality and health data.	Use Power BI and Python to create line charts, histograms, and heat maps from pre-existing sources like IQAir and Ministry of Health records.
Descriptive Statistic	Calculate mean, median, mode, standard deviation, and range using Power BI for direct pollutant concentrations and health metrics.	Calculate averages, standard deviations, and interquartile ranges using Power BI and Python for existing pollutant and health data.
Initial Insight	Derive trends such as higher PM2.5 levels in dry seasons linked to increased respiratory disease admissions.	Identify patterns like climate changes affecting air quality and respiratory diseases, corroborating or contrasting with primary data.
Feature Engineering	Use Python to create new features like moving averages and lagged variables from original data.	Use Python to create new features like abnormal pollutant levels and interaction terms from existing data.

Machine Learning

The initial machine learning models are developed using Python's Scikit-learn to predict air pollution levels and potential health impacts. Techniques such as decision trees, random forests, and support vector machines (SVM) are used. For instance, a random forest model might predict the likelihood of high pollution days, while an SVM model could estimate the risk of hospitalization due to respiratory diseases. These models are evaluated on a specific dataset to establish a performance baseline, with feature importance scores helping to identify key predictors and hyperparameters being adjusted for optimization. Cross-validation techniques are applied to avoid overfitting and ensure the models are generalizable.

Conclusion

Conclusion

This research examines the impact of air pollution on health in Indonesian cities using descriptive epidemiology and predictive analysis, leveraging both primary and secondary data on air quality and health outcomes. Exploratory data analysis (EDA) identified relationships between pollutants like PM2.5 and PM10 and respiratory and cardiovascular diseases, providing insights into pollutant cyclicity and spatial distribution. Machine learning models demonstrated potential for predicting future air pollution levels and related health risks. Descriptive analysis highlighted areas with high pollution and associated health effects, enhancing the understanding of the current state.

The predictive analysis and machine learning aimed to foresee future pollution and health concerns, aiding in better management and prevention of adverse effects. Statistical analysis was conducted to ensure the validity and reliability of the results obtained. In this project, the use of Microsoft Power BI together with Python was instrumental in managing diverse data and enabling deep analysis of the complex dataset.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

univteknologimalaysia utm_my utmofficial

Thank You

www.utm.my

innovative • entrepreneurial • global