RESEARCH ON SOIL ENVIRONMENTAL
AND HEALTH RISK ANALYSIS BASED
ON MACHINE LEARNING.

ZHAO ZHIHAN

UNIVERSITI TEKNOLOGI MALAYSIA

5.0 Discussion

This research applies machine learning techniques to accurately and reliably assess the human health risks posed by soil pollution. Data including soil pollutant concentrations, pH values, industrial and agricultural activities are collected for analysis, revealing the internal relationships and connections between soil pollution and human health risks.

The research first preprocesses the data, then conducts feature engineering, and finally develops models.

The final results show that in areas with frequent industrial and agricultural activities, there are significant and specific phenomena of soil pollutants, which have an impact on human health.

5.1 Interpretation of the Results

This study demonstrates that machine learning can uncover the potential links between soil pollution and human health. The Random Forest model achieved an accuracy rate of 90%, and the XGBoost model is slightly superior to the Random Forest model in terms of accuracy. These two models showcase the non - linear relationships between environmental pollution and health outcomes. The area under the ROC curve reaching 0.95 indicates the high precision of the models.

Compared with traditional methods, machine - learning models have a greater ability to integrate and mine data. They can effectively handle multivariate data and achieve higher accuracy.

5.2 Discussion of the Implications of Findings

This research holds great significance for environmental management and related policy - making. In terms of environmental management, relevant regulations can be imposed on the high - risk industrial and agricultural pollutants identified by the model. By restricting pollutant emissions, the health of nearby residents can be

improved. Regarding risk supervision, it can provide support for the development of a soil health risk supervision system. This contributes to the management of pollutants and the prevention of pollution, thereby reducing the harm of soil pollution to public health.

5.3 Future Work

Based on the current research gaps, future research can be carried out in the following directions: Strengthen the integration of multivariate data to uncover the internal relationships among different data. On this basis, enhance the interpretability of the model, so as to more efficiently demonstrate the relationship between soil pollution and health hazards. Subsequently, develop a system capable of real - time monitoring of soil pollution, enabling real - time surveillance of soil pollution and improving public health. Finally, strengthen cross - regional validation to generalize the model to more diverse geographical environments, so that relevant policies can be formulated according to local conditions.