Paper Title

Development of Heavy Rain Damage Prediction Model Using Machine Learning Based on Big Data

Paper Link

https://www.hindawi.com/journals/amete/2018/5024930/

Summary

This research paper presents a machine learning-based model designed to predict the extent of rainfall destruction in the Seoul Capital Area, which is situated in South Korea. The model leverages an extensive dataset comprising meteorological information from the years 1994 to 2015. It employs various techniques such as decision trees, bagging, random forests, and boosting. Following a comprehensive performance evaluation, the boosting model emerged as the most effective, yielding an impressive Area Under the Curve (AUC) value of 95.87%.

1. Motivation

The impetus behind this study arises from the escalating frequency of environmental calamities, such as floods and tsunamis, which are further compounded by the effects of climate change and swift urban development. Within South Korea, torrential rainfall is responsible for approximately 65% of the cumulative destruction caused by natural disasters. The ability to forecast the magnitude and consequences of such catastrophes beforehand has the potential to greatly augment the efficacy of both disaster mitigation and prevention efforts.

1.2 Contribution

This research makes a valuable contribution to the advancement of the field by recognizing the limitations of previous studies, which primarily relied on linear regression analysis and often focused solely on meteorological factors. The current study combines machine learning techniques with extensive datasets to develop a more reliable predictive framework for significant precipitation-related problems.

1.3 Methodology

The Seoul Capital Area was chosen for the study due to its notable occurrence of intense rainfall damage. The approach encompassed the creation of response variables derived from data pertaining to heavy rain damage, as well as explanatory variables obtained from meteorological data. The analysis of the data was performed through the utilization of two algorithms: Algorithm 1, which employed weather observation data from the present day, and Algorithm 2, which made use of historical weather data.

1.4 Conclusion

According to the study, the boosting model, which includes meteorological observation data from 1 to 4 days earlier, is the most successful for predicting significant rain damage. With its high AUC value and constant performance, this model can be used in proactive disaster management systems. The research emphasizes the power of machine learning and big data in enhancing disaster prediction models.

2. Limitations

2.1 The Initial Limitation

The study admits data quantity and variety limitations. It was based mostly on 22 years of severe rain damage data and hydrometeorological data, implying that including more diverse data could improve the model's forecast performance.

2.2 Second Restrictions

Only hydrometeorological data were used in the study. Additional data on catastrophe preventive efforts, disaster recovery funds, and socioeconomic characteristics could help to refine and improve the prediction model.

Synthesis

This research marks a substantial advancement in the use of machine learning and big data to predict heavy rain damage. It highlights the promise of this approach in disaster management by picking a suitable study topic, applying sophisticated machine learning techniques, and overcoming the limitations of earlier studies. The model developed could help to reduce heavy rain damage by implementing early and effective disaster management and preventive actions.