

KỸ THUẬT HỌC MÁY PHÂN LỚP VỚI DỰ BÁO DỊCH TẢ

MACHINE LEARNING TECHNIQUE - CLASSIFICATION TO PREDICT CHOLERA

Tác giả: [Lê Thị Ngọc Anh*](#), Nguyễn Thị Thanh Xuân, Hoàng Xuân Dậu, Bùi Trung Dũng

Tóm tắt bằng tiếng Việt:

Bệnh tả là một loại dịch bệnh truyền nhiễm nguy hiểm đối với con người. Nó không chỉ gây ảnh hưởng đến sức khỏe cộng đồng, mà còn tác động xấu đến phát triển kinh tế- xã hội. Bài báo này đề xuất mô hình dự báo về tỷ lệ dịch tả liên quan đến yếu tố khí hậu trong một khoảng thời gian định sẵn dựa trên kỹ thuật học máy thống kê rời rạc sử dụng bộ phân lớp rừng ngẫu nhiên thông qua số liệu hồi cứu về khí hậu, môi trường và số ca bệnh Tả trong giai đoạn 2007-2010 tại Hà Nội. Kết quả nghiên cứu cho thấy tính khả quan của hướng tiếp cận mô hình này để dự báo dịch bệnh là phù hợp và có khả năng áp dụng vào thực tế. Với phương pháp đề xuất, chúng tôi mong muốn sẽ xây dựng được một mô hình dự báo và áp dụng mô hình này cho việc dự báo bệnh dịch tả trong thực tế tại khu vực Hà Nội

Từ khóa: học máy; dịch tả; dự báo; rừng ngẫu nhiên; thống kê rời rạc.

Tóm tắt bằng tiếng Anh:

Cholera is an infectious disease dangerous to humans. It not only affects public health but also has negative impact on economic and social development. This paper proposes classification model using random forests to predict cholera cases in a predetermined time period with the retrospective data about the climate, the environment and the number of cases of cholera in Hanoi from 2007-2010. The results show that the positive approach of this model to predict cholera is consistent and can be put into practice. In addition, the approach can be used as an important input in the decision making process of the preventive healthcare. With the proposed approach, we hope to develop a forecasting model of cholera in Hanoi.

Key words: machine learning; cholera; forecast; random forest; bald statistic.