

Nghiên cứu về làm sạch dữ liệu

TRƯƠNG NHÃ BÌNH *

Trường Đại học Kinh tế - Tài chính TP.HCM
Nhận bài: 20/09/2022 - Duyệt đăng: 15/04/2023
(* Liên hệ: binhntn@uef.edu.vn - Tel: 0939466277

Tóm tắt:

Trong giai đoạn chuyển đổi công nghệ số, dữ liệu đóng vai trò vô cùng quan trọng. Các doanh nghiệp muốn duy trì và phát triển phải linh hoạt chuyển đổi phương thức kinh doanh giữa hoạt động trực tiếp và hoạt động trực tuyến. Từ đó lượng dữ liệu thu được ngày càng nhiều. Tuy nhiên dữ liệu chỉ có giá trị khi đã được làm sạch. Nếu dữ liệu không được xử lý kỹ sẽ dẫn đến hàng loạt các vấn đề như lỗi liên kết, mô hình sai đặc điểm kỹ thuật, lỗi trong ước lượng tham số và phân tích sai dẫn đến rút ra kết luận sai. Bài báo tập trung vào giới thiệu tầm quan trọng, những lợi ích của việc làm sạch dữ liệu, các lỗi dữ liệu, cũng như các phương pháp, kỹ thuật làm sạch phổ biến hiện nay khi xử lý dữ liệu.

Từ khóa: làm sạch dữ liệu, phương pháp làm sạch, kỹ thuật làm sạch dữ liệu, dữ liệu lớn.

Abstract:

In the period of digital technology transformation, data truly plays an important role. Businesses that want to maintain and grow must flexibly change their business methods between offline and online. So the amount of data is increasing significantly. However, the data is only valuable when it has been cleaned. If data is not handled carefully, it will lead to problems such as link errors, model misspecification, errors in parameter estimation and wrong analysis leads to wrong conclusion. This article focuses on introducing the importances, benefits of data cleaning, data errors, as well as methods and techniques when processing data.

Keywords: data cleaning, data cleaning methods, data cleaning techniques, big data.