

Thực hiện trên hệ thống theo các bước hướng dẫn sau:

Bước 1: Vào hệ thống chọn module Explore, chọn tập dữ liệu cần tổng hợp và tạo mới một tập dữ liệu tổng hợp từ IMPUTER

The screenshot shows the SQL Lab interface. On the left, the 'EXPLORE' button is highlighted with a red box. The main area displays a dataset named 'storagehub.hanh_chinh.quanly_t...' with 8 columns: 'abc', 'cccc', 'dia_diem_tai_nhan', 'quan_huyen', 'so_nguoi_bi_thu', 'so_nguoi_chet', 'ten', and 'thoi_gian_xay_ra'. The 'IMPUTER' button in the top right corner is also highlighted with a red box.

Bước 2: Chọn cột dữ liệu để thuật toán Imputer giúp điền các giá trị thay thế vào những vị trí thiếu.

The screenshot shows the 'Add Column' dialog box in the SQL Lab interface. The 'so_nguoi_bi_thu' column is selected for the Imputer algorithm. The 'Add offset' button is highlighted with a red box. The background shows the same dataset as in the previous screenshot.

Bước 3: Chọn thuật toán xử lý cho cột dữ liệu đã chọn.

The screenshot shows the SQL Lab interface. In the center, the 'Add Column' dialog box is open. Under the 'COLUMN' section, the 'THUẬT TOÁN' (Algorithm) dropdown menu is expanded, showing a list of algorithms: 'interpolate', 'interpolate', 'simple', 'knn', and 'bayesian'. The first 'interpolate' option is selected and highlighted. In the background, a table is visible with columns: 'anhuynh', '# so_nguoi_bi_thu', '# so_nguoi_chet', 'abc ten', and 'thoi_gian_xay_ra'. The table contains data for various locations and counts.

Trong đó:

- o Thuật toán nội suy (interpolation) được sử dụng để ước lượng giá trị tại các điểm trung gian dựa trên các giá trị đã biết của một tập dữ liệu.
- o Thuật toán Simple được dùng để xử lý các bài toán như tìm kiếm tuyến tính, sắp xếp nổi bọt, hoặc tính giai thừa.
- o Thuật toán KNN là một thuật toán dựa trên khoảng cách để phân loại hoặc dự đoán giá trị của một điểm dữ liệu mới dựa trên các điểm dữ liệu đã biết trong không gian đa chiều.
- o Thuật toán Payesian là một trong những thuật toán phân loại phổ biến nhất trong học máy, đặc biệt là đối với các bài toán phân loại văn bản và lọc thư rác.

Bước 4: “RUN” chạy test dữ liệu được tổng hợp.

SQL Lab > Tạo mới

Nhập tên ...

SOURCE:

OFFSET:

- ☒ so_nguoi_bi_thuong(interpolate)
- ☒ so_nguoi_chet(interpolate)
- [+ Add offset](#)

TEST

KẾT QUẢ Tổng cộng có 100 dữ liệu được trả về. Thời gian truy vấn 00:00:00.07

# so_nguoi_bi_thuong	# so_nguoi_chet
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0

Bước 5: Nhập tên cho tập dữ liệu mới được tổng hợp, chọn Lưu thông tin tập dữ liệu được tổng hợp bằng công cụ IMPUTER

SQL Lab > Tạo mới

LƯU

SOURCE:

OFFSET:

- ☒ so_nguoi_bi_thuong(interpolate)
- ☒ so_nguoi_chet(interpolate)
- [+ Add offset](#)

TEST

KẾT QUẢ Tổng cộng có 100 dữ liệu được trả về. Thời gian truy vấn 00:00:00.07

# so_nguoi_bi_thuong	# so_nguoi_chet
0	1
0	1
0	1
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0
1	0