

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

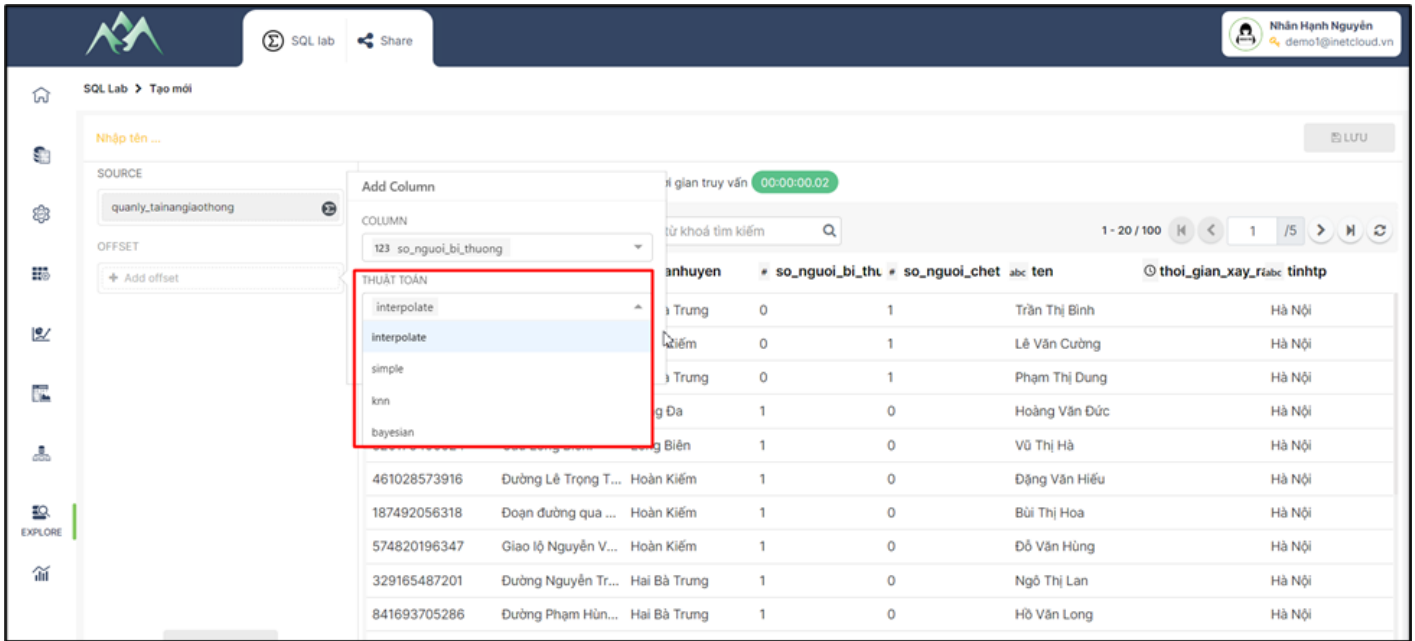
Bước 1: Vào hệ thống chọn module Explore, chọn tập dữ liệu cần tổng hợp và tạo mới một tập dữ liệu tổng hợp từ IMPUTER

The screenshot shows the SQL Lab interface. On the left, the 'EXPLORE' module is selected. 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 is highlighted in the top right corner.

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 imputation. The dialog box also shows the 'so_nguoi_chet' column. The 'EXPLORE' module is selected on the left.

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



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.

