**Báo cáo Bài tập thực hành Tiền Xử Lý Dữ liệu - Bài tập 2: Xử lý dữ liệu**

**Nhóm thực hiện: Nhóm 2**

**Thông tin các thành viên:**

Lê Việt Thắng – 18520356 (Nhóm trưởng )

Trần Đăng Khoa – 18520936

Nguyễn Thế Mạnh – 18521084

Lê Quang Nhật – 18521190

Hoàng Đình Quang – 18521294

Võ Đình Tứ - 18521589

**Source code chương trình** [**source.py**](source.py)

**Thông tin mô tả source code:** [**Readme.txt**](Readme.txt.txt)

**Các bộ dữ liệu dùng để test:**

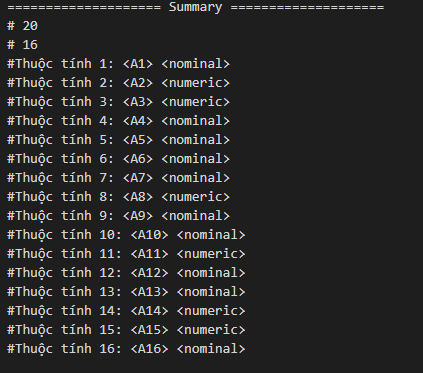
[**Test\_1.txt**](Test_1.txt)

[**Test\_2.txt**](Test_2.txt)

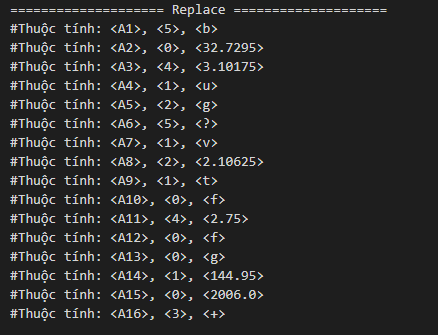
[**Test\_3.txt**](Test_3.txt)

**Kết quả thử nghiệm chương trình trên tập dữ liệu Test\_1 thể hiện qua file log:**

**1.1 Tóm tắt dữ liệu:**

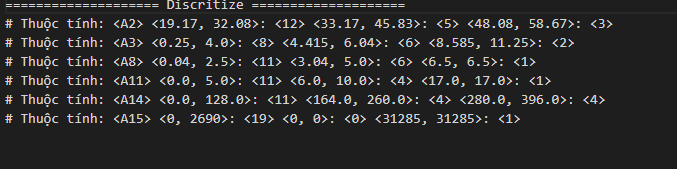
****

**1.2 Điền giá trị bị thiếu cho tất cả các thuộc tính:**

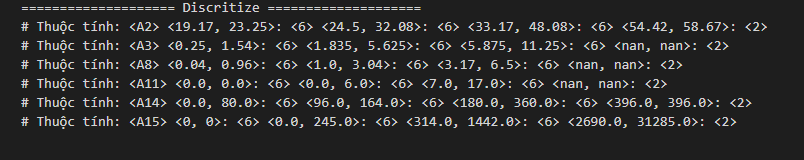
****

**1.3 Chia giỏ một hoặc nhiều thuộc tính numeric:**

**1.3.1 Chia giỏ theo độ rộng ( n = 3 ):**

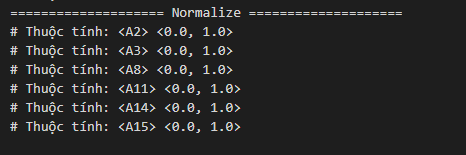


1.3.2 Chia giỏ theo độ sâu (n = 6):

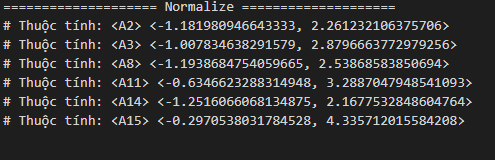


**1.4 Chuẩn hóa các thuộc tính có kiểu numeric:**

**1.4.1 Chuẩn hóa Min-max:**



**1.4.2 Chuẩn hóa Z – Score:**



**2. Kết quả nhận xét chương trình sau khi test qua với tập dữ liệu Test\_1:**

Chương trình cho qua kết quả tương đối đúng với thực tế tính toán, xuất hiện sai số trong quá trình làm tròn chữ số thập phân, tuy nhiên không đáng kể. Các tập dữ liệu đã được chuẩn hóa, thay thế dữ liệu bị mất, làm trơn bằng phương pháp chia giỏ đều đạt độ chính xác mong muốn.