Cho tập dữ liệu hãy thực hiện các bước phân tích sau: Xây dựng mô hình Logistic Regression nhằm phân lớp sinh viên có định hướng nghề nghiệp (HUONGNGHIEP) hay không dựa trên các điểm toán các học kì năm cấp 3, điểm thi NGONNGU, LOGIC, UNGXU và giới tính của sinh viên. Biết rằng, ta quy ước với định hướng nghề nghiệp thì No = 1 và Yes = 0

1. Tải dữ liệu diemthidaihoc.csv và hiển thị các dòng dữ liệu
2. Với các giá trị định lượng bị thiếu ta thay bằng giá trị 0
3. Tiến hành EDA dữ liệu, ví dụ: Hãy thống kê các độ đo cơ bản của dữ liệu và nhận xét
4. Chia tập dữ liệu thành tập train và test với tỉ lệ 80:20 và random state là 16
5. Tiến hành transform theo hàm f(x) = √=|x| dữ liệu input đầu vào trên tập train
6. Tiến hành EDA dữ liệu sau khi transform. Ví dụ: Hãy vẽ biểu đồ phân phối các biến định lượng và cho nhận xét
7. Tiến hành tiếp tục Polynomial Scaling dữ liệu đầu vào trên tập train với bậc là 2
8. Xây dựng mô hình Logistic Regression với thông số solvers= liblinear, max\_iter = 1000 và penalty = L2
9. Hãy đánh giá mô hình trên tập train và test. Ví dụ: so sánh kết quả accuracy trên tập train và test đưa ra nhận xét
10. Ứng dụng dự báo, với điểm toán các học kì 1,2 các năm lớp 10,11,12 và giới tính lần lượt là: 5.0, 6.5, 7.0, 7.0, 6.5, 6.0, giới tính là nam và điểm thi NGONNGU, LOGIC, UNGXU lần lượt là 5.5, 7.0, 4.5 thì sinh viên này có cần hỗ trợ định hướng nghề nghiệp hay không? (lưu ý, ta chỉ định hướng nghề nghiệp nếu giá trị dự báo là No)