Nhóm: Data Discretization

Thời gian: 2/2/2023

Người tham dự: thầy Trung, Dương, Hưng, Dũng

Nội dung:

* Công việc đã làm: Train và test trên nhiều tập data cho thuật toán đề xuất
* Vấn đề đang gặp phải: Data sinh chưa đủ tổng quát -> khi tune threshold để tối ưu f2 trên tập train thì kết quả rất tốt (f2 hơn 0.9) nhưng kết quả trên tập test khá thấp (f2 nhỏ hầu như nhỏ hơn 0.3)
* Công việc tuần tới:
* Cải thiện thuật toán sinh data (làm mờ ngưỡng threshold khi sinh data, tìm cách sinh data hiệu quả hơn)
* Chạy lại thuật toán đề xuất với data mới và so sánh với EWB, EFB
* Bài toán mở rộng: phát hiện sự thay đổi ngầm trong data (sự thay đổi này không làm thay đổi distribution của data -> psi không phát hiện được) -> đề xuất dùng transition matrix