

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**GVHD**: Nguyễn Thị Anh Thư

**Nhóm Sinh viên thực hiện:**

1. Nguyễn Ngọc Gia Khiêm MSSV: 21520287

2. Ngô Thùy Yến Nhi MSSV: 21521230

3. Mai Quốc Bảo MSSV: 21521850

4. Võ Thị Bích Ly MSSV: 21522317

5. Trần Kim Thanh MSSV: 21522605

6. Ngô Kỳ Anh MSSV: 21521825

7. Hoàng Xuân Lộc MSSV: 22520788

**BÁO CÁO ĐỒ ÁN**

**MẠNG XÃ HỘI**

**ĐỀ TÀI**

**DỰ ĐOÁN KẾT QUẢ HỌC TẬP CỦA HỌC KỲ TIẾP THEO CỦA SINH VIÊN UIT**

**TP. Hồ Chí Minh, tháng … năm ….**

**NHẬN XÉT CỦA GIẢNG VIÊN**

**…………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………**

## Bổ sung phần kết quả thực nghiệm

Sau khi báo cáo vào ngày 11/12/2024, nhóm em đã chỉnh sửa theo góp ý của cô, thực hiện tiến hành chia lại dữ liệu thực nghiệm và đề xuất kịch bản thực nghiệm tập trung vào việc phân chia dữ liệu theo thời gian, Dữ liệu học kỳ 1 được sử dụng làm đặc trưng đầu vào (features), trong khi dữ liệu học kỳ 2 được sử dụng làm nhãn mục tiêu (labels). Nhóm thực hiện mô phỏng tình huống thực tế, bằng cách sử dụng kết quả của học kỳ 1 là cơ sở để dự đoán kết quả học kỳ 2. Dữ liệu sau đó được chia thành hai phần: dữ liệu huấn luyện bao gồm các dòng có năm học nhỏ hơn 2021, và dữ liệu kiểm tra bao gồm các dòng có năm học từ 2021 trở đi.

Kết quả thực nghiệm của tập kiểm tra sau khi đã chỉnh sửa - tập trung vào hai chỉ số, accuracy – đánh gía độ chính xác của mô hình và precision - đảm bảo rằng mọi học sinh được phân loại chính xác trong nhóm của họ mà không có sự nhầm lẫn. Đây là kết quả trên tập test:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | Trước khi combine | | | Sau khi combine | | |
| Accuracy | Precision | | Accuracy | Precision | |
| XGBoost | 95% | Class 0 | 97% | 95% | Class 0 | 97% |
| Class 1 | 91% | Class 1 | 91% |
| Class 2 | 86% | Class 2 | 86% |
| Class 3 | 93% | Class 3 | 93% |
| Class 4 | 97% | Class 4 | 97% |
| RandomForest | 96% | Class 0 | 97% | 98% | Class 0 | 98% |
| Class 1 | 93% | Class 1 | 96% |
| Class 2 | 88% | Class 2 | 93% |
| Class 3 | 93% | Class 3 | 96% |
| Class 4 | 97% | Class 4 | 99% |
| AdaBoost | 93% | Class 0 | 97% | 94% | Class 0 | 96% |
| Class 1 | 89% | Class 1 | 89% |
| Class 2 | 85% | Class 2 | 85% |
| Class 3 | 93% | Class 3 | 93% |
| Class 4 | 97% | Class 4 | 97% |
| SVM | 94% | Class 0 | 97% | 86% | Class 0 | 93% |
| Class 1 | 93% | Class 1 | 77% |
| Class 2 | 87% | Class 2 | 64% |
| Class 3 | 87% | Class 3 | 78% |
| Class 4 | 95% | Class 4 | 89% |

*Nhận xét:*

+ Có thể thấy, dữ liệu sau khi được combine giúp tăng hiệu suất của hầu hết các mô hình, đặc biệt là Random Forest và XGBoost. Vấn đề mất cân bằng dữ liệu chưa được giải quyết triệt để khi Class 2 thường có Precision thấp nhất.