

## ĐỀ THI THAM KHẢO THỰC HÀNH HỌC MÁY ỨNG DỤNG

Sinh viên được phép tham khảo tài liệu trong máy và không được sử dụng internet trong quá trình kiểm tra. Ngoài ra không trao đổi trong thời gian làm bài. Sinh viên có thời gian 2.5 tiếng để hoàn thành bài kiểm tra. Good luck for you 😊

### Naive Bayes

**Câu 1:** (1.5 điểm) Cho tập dữ liệu Iris từ `sklearn.datasets`. Hãy chia tập dữ liệu thành tập huấn luyện và tập kiểm tra với tỷ lệ 80:20. Sử dụng thuật toán Naive Bayes để huấn luyện mô hình và dự đoán trên tập kiểm tra và in ra độ chính xác của mô hình.

**Câu 2:** (0.5 điểm) Vẽ ma trận nhầm lẫn (confusion matrix) và nhận xét về kết quả dự đoán.

### KNN

**Câu 3:** (1.25 điểm) Tải tập dữ liệu Wine từ `sklearn.datasets` và chia tập dữ liệu theo tỷ lệ 70:30. Xây dựng mô hình KNN để phân loại dữ liệu. Sử dụng  $k = 5$ . Tính toán và in ra độ chính xác, độ nhạy (recall), và độ chính xác (precision) của mô hình.

**Câu 4:** (0.75 điểm) Thử nghiệm với các giá trị  $k$  khác nhau (1, 3, 7, 9). Vẽ đồ thị thể hiện mối quan hệ giữa  $k$  và độ chính xác của mô hình. Nhận xét về kết quả.

### Cây quyết định

**Câu 5:** (2 điểm) Cho tập dữ liệu Breast Cancer từ `sklearn.datasets`. Huấn luyện mô hình cây quyết định và đánh giá độ chính xác của mô hình trên tập kiểm tra (chia 75:25). Xuất ra hình ảnh của cây quyết định.

### SVM

**Câu 6:** (1 điểm) Dùng tập dữ liệu Digits từ `sklearn.datasets`. Huấn luyện mô hình SVM với kernel tuyến tính (linear). Đánh giá độ chính xác trên tập kiểm tra.

**Câu 7:** (1 điểm) Thử nghiệm với các kernel khác (rbf, poly). So sánh độ chính xác và thời gian huấn luyện giữa các kernel này. Đưa ra nhận xét về kernel nào hiệu quả nhất cho tập dữ liệu này.

### Multilayer Perceptron

**Câu 8:** (2 điểm) Dùng tập dữ liệu MNIST. Huấn luyện một MLP với hai tầng ẩn (hidden layers) để phân loại ảnh số từ 0-9. In ra độ chính xác trên tập kiểm tra.