**ĐỀ BÀI TẬP CHO CÁC NHÓM**

1. **Tiền xử lý dữ liệu**

Trình bày đầy đủ các bước tiền xử lý dữ liệu bằng WEKA

* Sử dụng 1 tập dữ liệu của UCI có dữ liệu thiếu, dữ liệu liên tục, dữ liệu rời rạc.
* Khuyến khích: Tìm hiểu thêm các chức năng WEKA.

1. **Xây dựng cây quyết định**

* Làm bài tập 4.1, 4.2, 4.3 trong tài liệu [1]. (các bước tính toán)
* Trình bày ứng dụng Naïve Bayes trong WEKA ( [2], trang 139) cho một tập dữ liệu đã tiền xử lý. (tập dữ liệu khác Weather.norminal.arff)
* Trình bày các bước tạo training set và testing set trong WEKA ( [2] trang 150)

1. **Luật kết hợp**

* Trình bày thuật toán Naïve (Naïve Algorithm) tìm luật kết hợp ([2],240)
* Làm bài tập 12.1, 12.2 trong tài liệu [1]
* Sử dụng WEKA hoặc R tìm luật kết hợp của bài tập 12.1, 12.2 trong tài liệu [1]
* So sánh kết quả khai phá luật kết hợp và các luật phân lớp ([2], trang 360)
* Liệt kê tên một số thuật toán khai phá luật kết hợp. (nếu được chỉ ra ý tưởng chính của các thuật toán)

1. **Phân cụm**

* Thử nghiệm phân cụm dữ liệu bằng WEKA/R cho các thuật toán:

K-means, HCA (Hierarchical Clustering Algorithms).

* Trình bày thuật toán DBSCAN ([2], trang 203).

**Tài liệu tham khảo**

[1] Data Mining, Theories, Algorithms, and Examples – Nong Ye, CRC Press

[2] Data Ming and Data Warehousing, Parteek Bhatia