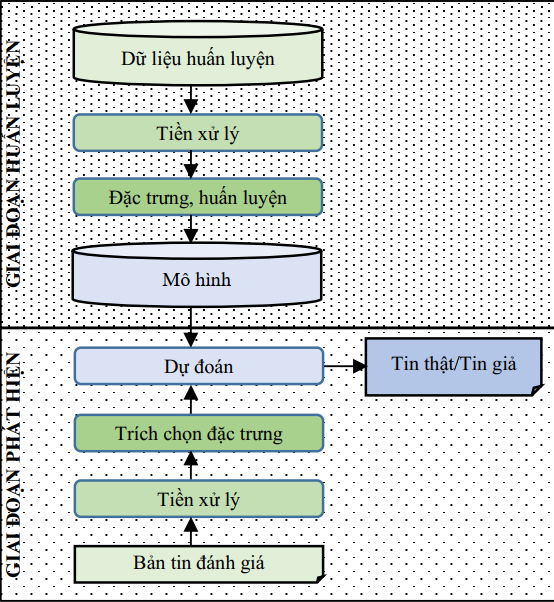
PHẦN 3. CƠ SỞ LÝ THUYẾT

1. Giới thiệu chung về phương pháp học máy
   1. *Phương pháp*

Đa số các nghiên cứu đều theo hướng sử dụng phương pháp học máy giám sát hoặc học máy bán giám sát để huấn luyện tạo mô hình nhằm mục đích phân loại tệp dữ liệu và dự đoán.

Mô hình tổng quát của hướng tiếp cận này:



Bước đầu tiên trong mô hình này là giai đoạn thu thập tập dữ liệu để xây dựng cơ sở dữ liệu huấn luyện. Trong cơ sở dữ liệu này bao gồm các bản tin đã được gán nhãn là tin giả hoặc tin thật. Trong trường hợp học máy giám sát, tất cả các dữ liệu dùng để huấn luyện đều phải được gán nhãn, trong trường hợp học bán giám sát thì bao gồm cả dữ liệu đã gán nhãn và chưa gán nhãn.

Giai đoạn tiền xử lý cho phép sử dụng các kỹ thuật xử lý ngôn ngữ tự nhiên để làm sạch dữ liệu, loại bỏ các thông tin không có ích và biểu diễn lại dữ liệu.

Giai đoạn trích chọn đặc trưng