* Giới thiệu chung:
  + Là thuật toán có giám sát
    - Thuật toán có giám sát (supervised learning) là một loại phương pháp trong học máy (machine learning) mà mô hình máy tính được huấn luyện thông qua dữ liệu đã được gán nhãn (labeled data). Trong quá trình huấn luyện, mô hình học từ các cặp dữ liệu đầu vào và nhãn tương ứng và cố gắng tạo ra một mô hình hoặc hàm ánh xạ từ dữ liệu đầu vào đến đầu ra mong muốn.
    - Thuật toán có giám sát được gọi là "có giám sát" bởi vì quá trình huấn luyện yêu cầu sự giám sát từ phía con người. Con người cần cung cấp dữ liệu đã được gán nhãn để mô hình có thể học từ các mẫu và điều chỉnh các tham số để dự đoán đầu ra chính xác hơn trên các dữ liệu mới
* **Tại sao chọn Bag of Words (BoW) cho bài toán xác định ngôn ngữ?**
* Bag of Words (BoW) là phương pháp lý tưởng cho bài toán xác định ngôn ngữ nhờ vào sự đơn giản và dễ triển khai. BoW biến văn bản thành các vector số bằng cách đếm tần suất xuất hiện của từ, giúp xử lý nhanh chóng và hiệu quả các tập dữ liệu lớn. Phương pháp này phù hợp cho các tình huống không yêu cầu phân tích sâu về ngữ nghĩa hoặc cấu trúc câu, đồng thời có khả năng mở rộng tốt và dễ dàng kết hợp với các kỹ thuật khác. Sự phổ biến và thành công của BoW trong các ứng dụng xử lý ngôn ngữ tự nhiên cũng chứng minh tính hiệu quả của nó.