

TÓM TẮT

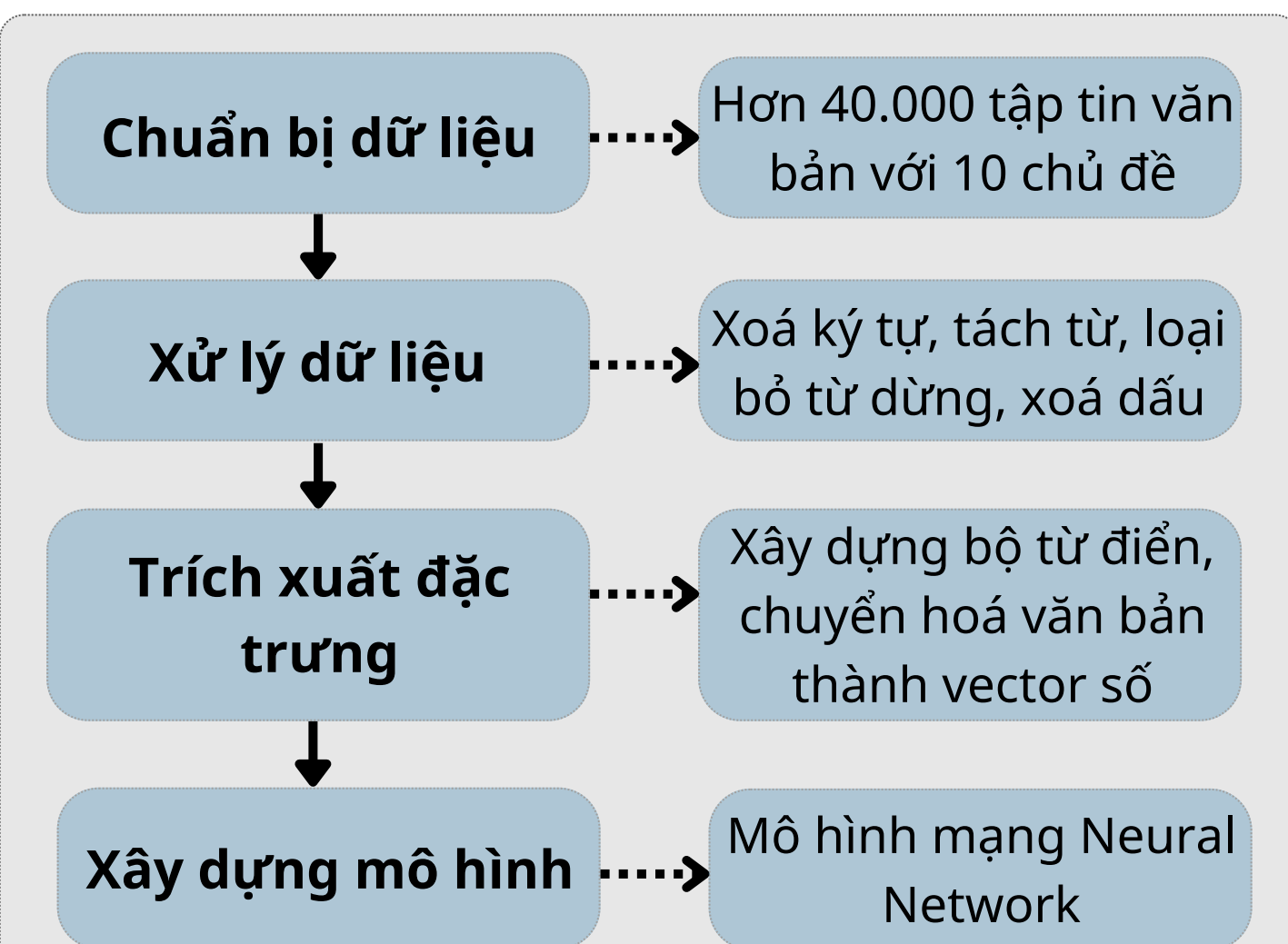
Mô hình nghiên cứu là sản phẩm cuối kỳ của môn học Trí Tuệ Nhận Tạo với sự hướng dẫn của PGS.TS Nguyễn Trường Thịnh. Nghiên cứu trình bày về phương pháp xử lý văn bản tiếng Việt và xây dựng mô hình phân loại văn bản tiếng Việt bằng mạng Neural Network.

TÍNH CẤP THIẾT

Do sự gia tăng của số lượng tin tức hàng ngày trên mạng Internet và nhu cầu tìm kiếm văn bản của người dùng cũng tăng nên việc phân loại các bài báo điện tử là điều cần thiết, nó có thể giúp con người quản lý lượng tin tức lớn và tìm kiếm thông tin một cách nhanh chóng hơn.

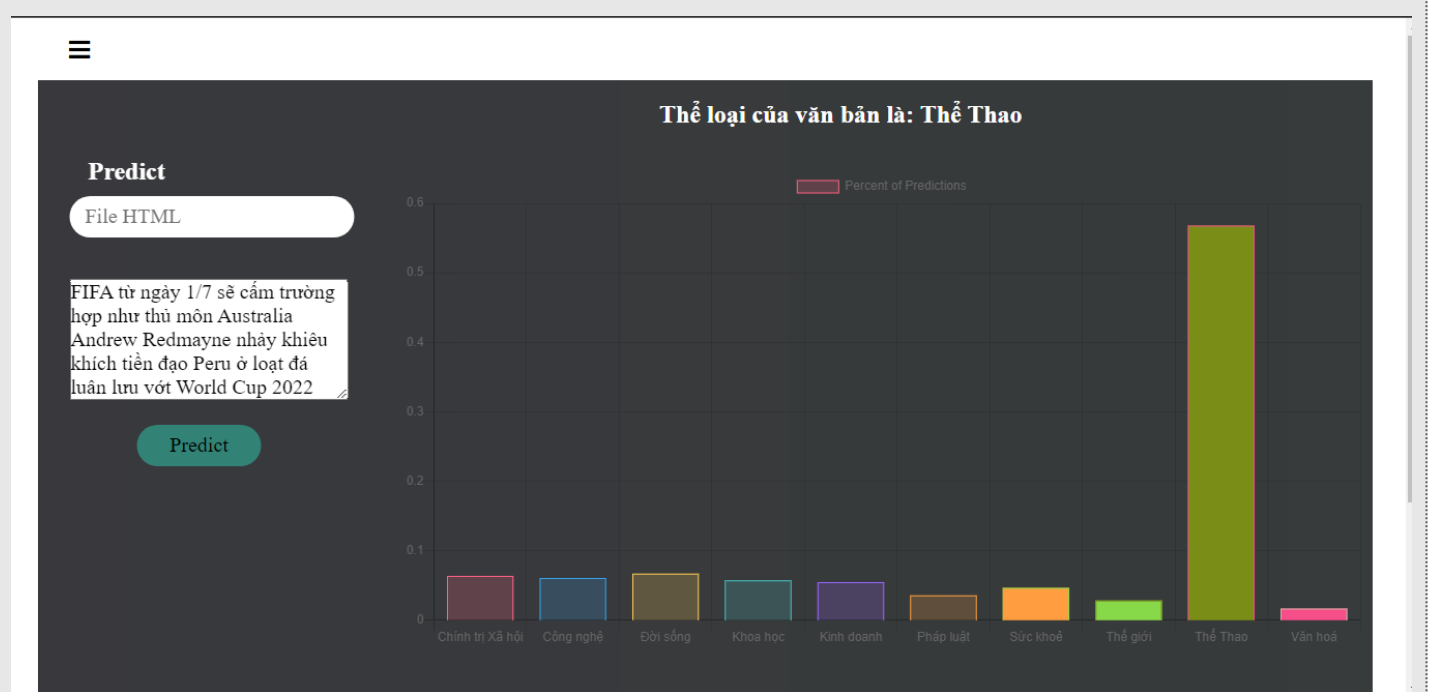
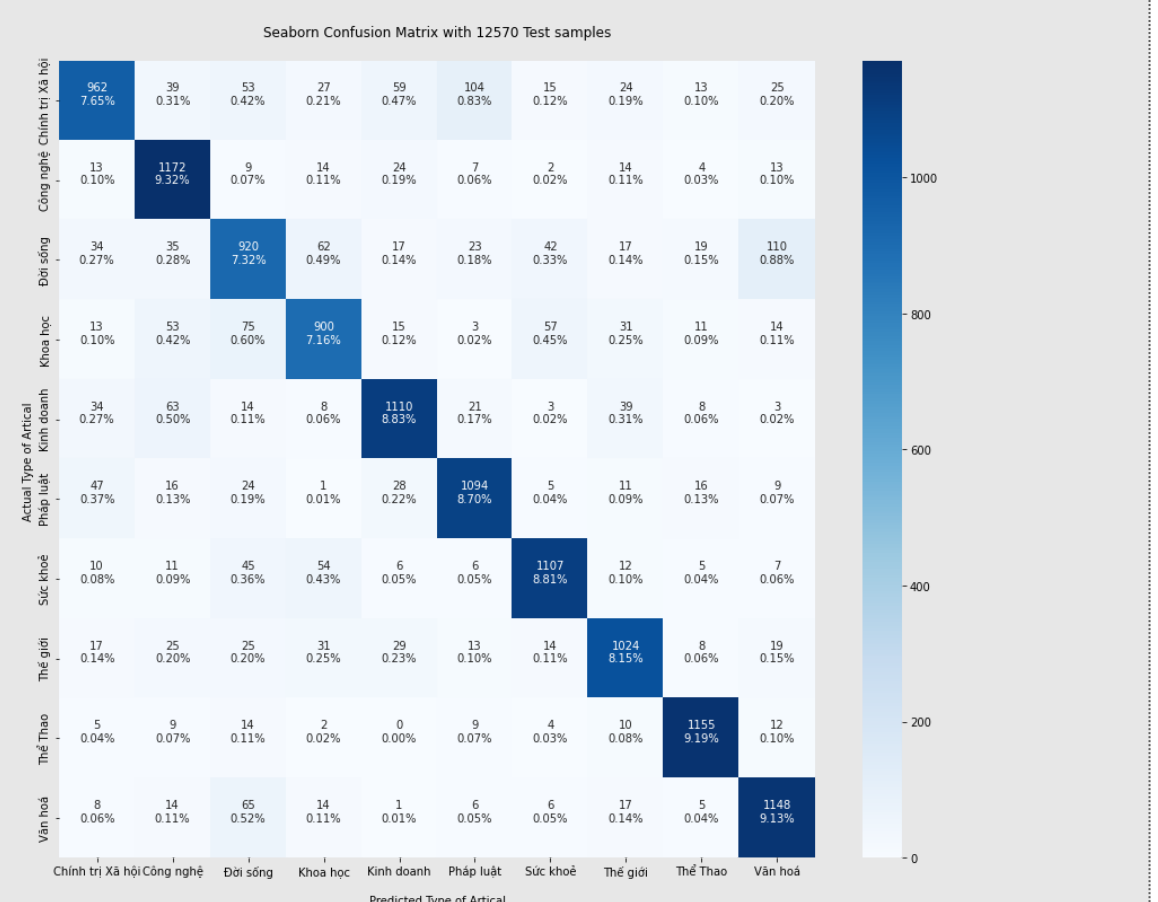
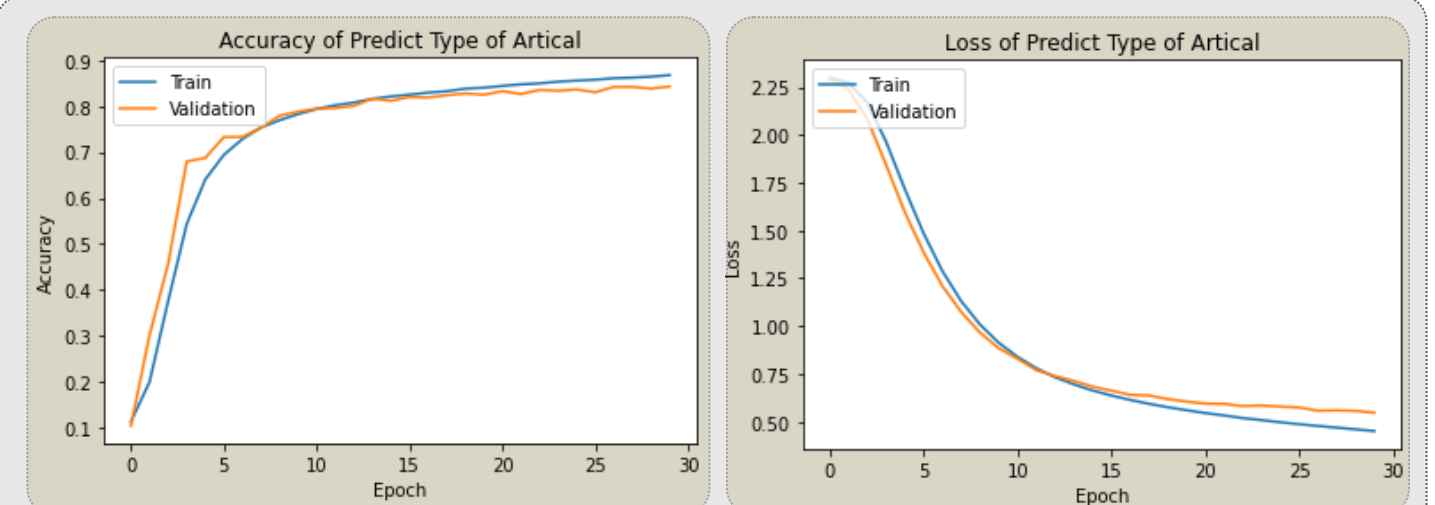


PHƯƠNG PHÁP VÀ MÔ HÌNH



- Phương pháp tách từ: Sử dụng từ điển tiếng Việt kết hợp phương pháp So khớp dài nhất
- Trích xuất đặc trưng: Gán các từ trong văn bản bằng các trọng số tương ứng của từ trong từ điển
- Xây dựng mô hình: Mạng Neural Network qua các lớp Embedding, GlobalAveragePooling1D và Dropout.

KẾT QUẢ



Giao diện phân loại báo điện tử

KẾT LUẬN

- Mô hình phân loại có độ chính xác tương đối khả quan 84.2%
- Thiết kế giao diện phân loại báo điện tử thời gian thực
- Có thể tiếp tục phát triển bằng cách thay đổi phương pháp tách từ và mô hình để độ chính xác cao hơn. Thêm cơ sở dữ liệu để lưu lại các văn bản đã phân loại theo thời gian thực