Bản báo cáo mã nguồn:

1. Ý tưởng:

Lộc đã làm mã nguồn trong file model\_training.py như sau:

1. Ý tưởng là dùng mô hình LSTM (mạng RNN) để cho mô hình học và tìm lỗi đánh giá lỗi chính tả, lỗi cấu trúc câu tiếng anh.
2. Sử dụng thư viện spaCy để làm việc.

2. Quá trình thực hiện:

1. Lộc đã làm được việc tách văn bản thành các tokens, xử lý được lỗi chính tả mức cơ bản nhất (sử dụng từ điển của NLTK và NLP để xử lý sơ bộ
2. Đảm bảo tính nhất quán dữ liệu. Và hiện mô hình xử lý được các từ thuộc thì hiện tại đơn. Đang hướng đến xử lý thì hiện tại tiếp diễn và thị hiện tại hoàn thành.
3. Lưu trữ dữ liệu hơn 300 động từ bất quy tắc vào json
4. Chưa thể tiến hành việc xây dựng mô hình LSTM được vì chưa giải quyết được nhiều vấn đề về tiền xử lý dữ liệu

3. Mục tiêu sắp tới:

1. hoàn thiện việc xử lý các cấu trúc của các thì Simple.

2. hoàn thiện việc xử lý các thì Continuous

3. hoàn thiện các thì Perfect

4. hoàn thiện các thì Perfect + simple

5. hoàn thiện các thì Perfect + countinuous

6. sau khi hoàn thiện các thì tiếng anh thì sẽ dùng mô hình LLM (BERT) để phân tích chuyên sâu nội dung của văn bản.