vào file code để đọc hiểu code trước , còn các file data từ từ t giải thích cho

theo flow :

crawthegioididong -> preprocessing -> after correct preprocessing -> phoBert