**1. Giới thiệu chung về nơi thực tập**

**1.1 Thông tin công ty Goline**

Công ty Cổ phần Goline (Goline Corporation) được thành lập vào tháng 02 năm 2010, là doanh nghiệp công nghệ chuyên cung cấp giải pháp phần mềm Fintech cho thị trường tài chính và chứng khoán tại Việt Nam và khu vực. Sau hơn 15 năm hoạt động và phát triển, Goline đã trở thành một trong những đơn vị uy tín hàng đầu trong lĩnh vực này.

**Thông tin cơ bản về công ty Goline:**

* **Tên công ty:** Công ty Cổ phần Goline (Goline Corporation)
* **Năm thành lập:** 02/2010
* **Vốn điều lệ:** 15.000.000.000 VNĐ
* **Chủ tịch/CEO:** MBA. Lê Ngọc Tuấn
* **Trụ sở chính:** Tầng 8, Toà nhà Kim Ánh, 78 Duy Tân, Cầu Giấy, Hà Nội
* **Chi nhánh:**
  + Đà Nẵng: Tầng 5, 15 Quang Trung, Hải Châu
  + Hồ Chí Minh: Tầng 3, 20/13 Nguyễn Trường Tộ, Q.4
  + Nhật Bản: Tầng 7, Suishin Building, Yokohama, Kanagawa
* **Website:** <https://goline.vn>
* **Email:** contact@goline.vn
* **Hotline:** 0903 304 888

**1.2 Cơ cấu tổ chức và lĩnh vực hoạt động**

Goline hoạt động theo mô hình tổ chức chuyên sâu với nhiều phòng ban và nhóm kỹ thuật, được phân chia rõ ràng theo từng chức năng và dự án. Cơ cấu tổ chức tiêu biểu bao gồm:

* Ban Giám đốc điều hành
* Bộ phận phát triển sản phẩm (Core Systems, Trading Platform, AI, Data Engineering,...)
* Bộ phận kiểm thử và chất lượng phần mềm (QA/QC)
* Bộ phận tư vấn nghiệp vụ tài chính – chứng khoán
* Bộ phận hỗ trợ kỹ thuật và khách hàng
* Bộ phận đào tạo và tuyển dụng
* Các nhóm dự án chuyên trách theo từng khách hàng hoặc giải pháp

**Lĩnh vực hoạt động chính:**

* **Tư vấn và phát triển phần mềm Fintech:**  
  Goline phát triển hệ thống giao dịch chứng khoán bao gồm cả thị trường cơ sở, OTC, phái sinh và trái phiếu, tích hợp đầy đủ các phân hệ từ Front Office đến Back Office.
* **Sản phẩm AI và hệ thống hỗ trợ đầu tư:**  
  Một số hệ thống tiêu biểu như **Gaia Advisor** (nền tảng tư vấn đầu tư bằng AI), **Robot Trading**, **Phân tích hành vi người dùng**, và các nền tảng **Copy Trade**, **Algo Trading**.
* **Tư vấn và triển khai giải pháp IT cho tổ chức tài chính:**  
  Goline xây dựng và triển khai giải pháp cho hơn 25 công ty chứng khoán tại Việt Nam, kết nối đa dạng với các hệ thống bên ngoài như ngân hàng, sở giao dịch, trung tâm lưu ký...
* **Dịch vụ gia công phần mềm (Outsource):**  
  Hợp tác cùng các đối tác trong và ngoài nước để phát triển giải pháp phần mềm chuyên biệt cho thị trường tài chính.

**1.3 Mô tả nhóm dự án thực tập**

Trong kỳ thực tập, tôi được phân công vào **nhóm phát triển ứng dụng Fin Chatbot** – một dự án trí tuệ nhân tạo (AI) với mục tiêu xây dựng một hệ thống chatbot tài chính thông minh, có khả năng hỗ trợ người dùng tra cứu thông tin, phân tích cảm xúc thị trường và trả lời các câu hỏi liên quan đến chứng khoán – tài chính.

**Nhóm dự án bao gồm:**

* **Mentor:** Trực tiếp hướng dẫn kỹ thuật và giám sát tiến độ
* **Leader:** Quản lý nhóm, phân chia nhiệm vụ và định hướng dự án
* **Các thành viên:** Gồm các sinh viên thực tập (trong đó có tôi), đảm nhận các vai trò như: thu thập dữ liệu, xử lý ngôn ngữ tự nhiên, triển khai mô hình AI, xây dựng backend API, kiểm thử hệ thống,...

**Các công việc nhóm đang triển khai:**

* Cào dữ liệu từ các trang web tài chính lớn như **CafeF**, **Vietstock**
* Tiền xử lý dữ liệu và lưu trữ vào cơ sở dữ liệu
* Phân tích cảm xúc các bài viết tài chính (sentiment analysis)
* Triển khai mô hình truy vấn văn bản thông minh (LLM hoặc RAG)
* Xây dựng hệ thống chatbot có khả năng trả lời theo ngữ cảnh người dùng

Nhóm sử dụng các công nghệ hiện đại như: **Python, FastAPI, PostgreSQL, Qdrant, Langchain, HuggingFace Transformers, Docker**, và các mô hình pretrained như **PhoBERT, FinBERT, T5,...**

**2. Nội dung công việc trong kỳ thực tập**

**2.1 Mô tả công việc chính**

Trong kỳ thực tập tại Công ty Cổ phần Goline, tôi được tham gia vào dự án xây dựng **Fin Chatbot** – một hệ thống trợ lý ảo ứng dụng trí tuệ nhân tạo nhằm hỗ trợ người dùng trong lĩnh vực tài chính – chứng khoán. Mục tiêu của hệ thống là trả lời các câu hỏi của người dùng một cách chính xác, nhanh chóng và có khả năng hiểu ngữ cảnh cũng như phân tích nội dung tài chính.

Các công việc chính tôi đảm nhiệm trong quá trình thực tập bao gồm:

* **Thu thập dữ liệu (Web Crawling):**  
  Xây dựng chương trình cào dữ liệu tự động từ các trang web tài chính lớn như **CafeF** và **Vietstock**, bao gồm tiêu đề, nội dung, thời gian, danh mục và các chỉ số liên quan.
* **Tiền xử lý và làm sạch dữ liệu:**  
  Thực hiện các bước làm sạch văn bản, loại bỏ ký tự thừa, chuẩn hóa tiếng Việt, phân tách câu, và lưu trữ vào hệ quản trị cơ sở dữ liệu **PostgreSQL**.
* **Phân tích cảm xúc tài chính (Sentiment Analysis):**  
  Áp dụng mô hình học sâu (deep learning) để gán nhãn cảm xúc (tích cực – tiêu cực – trung lập) cho các bài viết tài chính, phục vụ cho việc đánh giá xu hướng thị trường.
* **Xây dựng hệ thống truy vấn thông minh:**  
  Tham gia thiết kế kiến trúc RAG (Retrieval-Augmented Generation) để hệ thống có khả năng truy xuất thông tin theo ngữ nghĩa và trả lời dựa trên các văn bản đã thu thập.
* **Kiểm thử và đánh giá mô hình:**  
  Đánh giá độ chính xác của mô hình phân tích cảm xúc và khả năng phản hồi đúng câu hỏi của chatbot.

**2.2 Công nghệ và công cụ sử dụng**

Trong quá trình thực tập, tôi đã tiếp cận và làm việc với nhiều công nghệ hiện đại thuộc lĩnh vực AI và lập trình ứng dụng. Dưới đây là các công nghệ và công cụ tiêu biểu:

| **Nhóm công nghệ** | **Công cụ / Thư viện cụ thể** | **Mục đích sử dụng** |
| --- | --- | --- |
| **Ngôn ngữ lập trình** | Python, SQL | Phát triển pipeline xử lý dữ liệu và backend |
| **Xử lý dữ liệu** | Pandas, BeautifulSoup, Requests, LXML | Cào dữ liệu và tiền xử lý văn bản |
| **Cơ sở dữ liệu** | PostgreSQL | Lưu trữ dữ liệu văn bản đã xử lý |
| **Vector database** | Qdrant | Lưu trữ vector embedding cho hệ thống RAG |
| **Xử lý ngôn ngữ (NLP)** | Huggingface Transformers (PhoBERT, T5, FinBERT), underthesea | Phân tích cảm xúc, tóm tắt và hiểu văn bản |
| **Triển khai backend** | FastAPI, Uvicorn | Xây dựng API cho hệ thống Chatbot |
| **Truy vấn ngữ nghĩa** | Langchain | Kết nối LLM với dữ liệu tài chính |
| **Công cụ hỗ trợ** | Git, Docker, VS Code | Quản lý mã nguồn và môi trường phát triển |

**2.3 Quy trình phát triển hệ thống Chatbot**

Nhóm dự án triển khai hệ thống Fin Chatbot theo quy trình phát triển phần mềm linh hoạt (Agile) với các giai đoạn cụ thể như sau:

1. **Thu thập yêu cầu và phân tích nghiệp vụ:**  
   Xác định mục tiêu của chatbot, các loại câu hỏi cần hỗ trợ, các nguồn dữ liệu chính và luồng sử dụng của người dùng.
2. **Cào và xử lý dữ liệu tài chính:**  
   Thu thập dữ liệu từ các nguồn uy tín (CafeF, Vietstock), thực hiện làm sạch và chuẩn hóa trước khi đưa vào hệ thống.
3. **Vector hóa dữ liệu:**  
   Sử dụng các mô hình pretrained (PhoBERT, T5) để chuyển các đoạn văn bản thành vector embedding, lưu trữ vào **Qdrant** để hỗ trợ tìm kiếm ngữ nghĩa.
4. **Phân tích cảm xúc và gán nhãn dữ liệu:**  
   Xây dựng mô hình classification hoặc regression để đánh giá mức độ tích cực/tiêu cực của từng bài viết, phục vụ cho các truy vấn mang tính phân tích.
5. **Xây dựng chatbot bằng RAG:**  
   Triển khai pipeline RAG gồm: truy vấn → lọc top-k tài liệu → sinh câu trả lời bằng mô hình LLM (T5, GPT,...).
6. **Triển khai API và giao diện thử nghiệm:**  
   Xây dựng các endpoint trên **FastAPI** để kết nối hệ thống với frontend hoặc người dùng thử nghiệm chatbot.
7. **Đánh giá và cải tiến mô hình:**  
   Thu thập phản hồi từ người dùng, đánh giá độ chính xác và mức độ hài lòng để tinh chỉnh mô hình và hệ thống.

**3. Giới thiệu ứng dụng AI: Fin Chatbot**

**3.1 Mục tiêu và chức năng chính của Chatbot**

**Fin Chatbot** là một ứng dụng trí tuệ nhân tạo (AI) được thiết kế nhằm hỗ trợ người dùng trong lĩnh vực tài chính – chứng khoán bằng cách cung cấp thông tin nhanh chóng, chính xác và theo ngữ cảnh.

**Mục tiêu chính:**

* Hỗ trợ người dùng **tra cứu dữ liệu tài chính** như: diễn biến thị trường, thông tin doanh nghiệp, giá cổ phiếu,...
* Trả lời các câu hỏi tài chính một cách **tự nhiên và thông minh**, sử dụng mô hình ngôn ngữ lớn (LLM).
* Cung cấp **phân tích cảm xúc** về thị trường thông qua việc đánh giá tích cực, tiêu cực của các bài viết tài chính.
* Nâng cao trải nghiệm người dùng nhờ vào **khả năng hiểu ngôn ngữ tự nhiên tiếng Việt**, phản hồi linh hoạt và có tính ngữ cảnh.

**Một số chức năng chính:**

* Trả lời câu hỏi người dùng dựa trên thông tin cào từ CafeF, Vietstock (ví dụ: *"VNM có tin gì mới không?"*)
* Phân tích cảm xúc các bài viết về mã cổ phiếu cụ thể (ví dụ: *"Tâm lý thị trường về HPG hôm nay như thế nào?"*)
* Tóm tắt nội dung bài viết tài chính
* Gợi ý mã cổ phiếu đáng chú ý dựa trên dữ liệu tin tức gần đây

**3.2 Pipeline hệ thống: Thu thập → Xử lý → Phân tích → Trả lời**

Hệ thống Fin Chatbot được thiết kế theo một **pipeline xử lý dữ liệu và phản hồi thông minh**, bao gồm 4 giai đoạn chính:

**1. Thu thập dữ liệu**

* Cào dữ liệu từ các trang web tài chính uy tín như **CafeF** và **Vietstock**
* Dữ liệu bao gồm: tiêu đề, nội dung bài viết, thời gian đăng, mã cổ phiếu liên quan,...

**2. Xử lý dữ liệu**

* Làm sạch dữ liệu: loại bỏ HTML, ký tự thừa, chuẩn hóa tiếng Việt
* Phân tách văn bản, xử lý token, tách đoạn
* Vector hóa văn bản: sử dụng mô hình như **PhoBERT** hoặc **T5** để sinh embedding

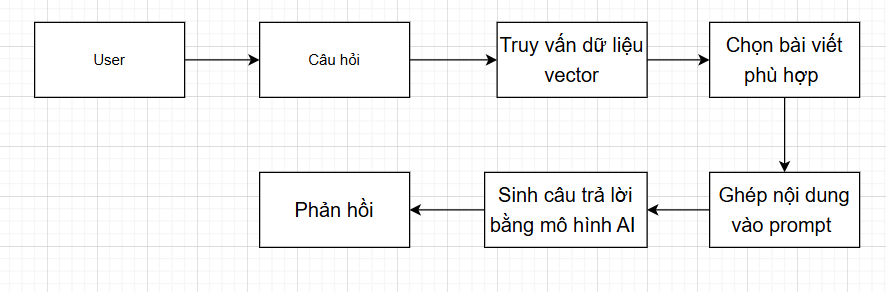
**3. Phân tích dữ liệu**

* **Phân tích cảm xúc**: sử dụng mô hình học sâu để đánh giá mức độ tích cực – tiêu cực của các bài viết
* **Tìm kiếm ngữ nghĩa**: khi người dùng đặt câu hỏi, hệ thống truy xuất các đoạn văn liên quan từ vector database (**Qdrant**)

**4. Sinh câu trả lời**

* Kết hợp tài liệu liên quan + câu hỏi của người dùng
* Sinh câu trả lời tự nhiên bằng mô hình LLM (ví dụ: **T5**, **GPT**, hoặc mô hình fine-tune)
* Trả kết quả lại qua API hoặc giao diện người dùng

Tổng quan quy trình:

 **3.3 Vai trò của AI trong từng giai đoạn**

| **Giai đoạn** | **Công nghệ AI sử dụng** | **Vai trò cụ thể** |
| --- | --- | --- |
| **Xử lý ngôn ngữ** | Tokenization, POS tagging, sentence splitting (Underthesea, PhoBERT) | Làm sạch và phân tích cấu trúc văn bản |
| **Phân tích cảm xúc** | Mô hình phân loại cảm xúc (ViBERT, FinBERT, T5) | Gán nhãn bài viết: tích cực, tiêu cực, trung lập |
| **Vector hóa văn bản** | PhoBERT, T5 encoder | Chuyển văn bản thành vector phục vụ tìm kiếm ngữ nghĩa |
| **Tìm kiếm ngữ nghĩa** | Qdrant, cosine similarity | Tìm đoạn văn phù hợp nhất với câu hỏi người dùng |
| **Sinh câu trả lời** | LLM (T5, GPT, Mistral, Qwen...) | Tạo ra phản hồi tự nhiên, có ngữ cảnh và chính xác |
| **Tóm tắt nội dung** | Sequence-to-sequence models | Rút gọn bài viết tài chính thành đoạn mô tả ngắn gọn |