

2024 전기 착수 보고서

머신러닝, 자연어 처리 기반 중고차 추천 플랫폼



팀명	도미노	담당교수	유영환 교수님
	202012142		이은진
학번	201824606	이름	최희웅
	201913543		최성빈

- Table

1. Introduction.....	3
2. Project Description.....	4
2.1 Project Objectives.....	4
2.2 Expected Benefits.....	5
2.3 Development Environment and Design Diagram.....	6
3. Development Process and Technical Explanation.....	7
3.1 LLM & fine-tuning.....	7
3.2 Web Application and AI Chatbot Server Development..	10
3.3 Deployment and Operation Plan.....	11
4. Development Schedule and Role Allocation.....	12
4.1 Development Schedule.....	12
4.2 Role Allocation.....	13

1. Introduction

오늘날 머신러닝과 자연어처리 기술의 급격한 발전을 통해 검색과 사용자 추천시스템은 혁신적으로 변화하고 있다. IR(Information Retrieval) 분야에서는 기존 검색 엔진이 AI 챗봇을 통한 새로운 정보 탐색 방식에 도전하고 있으며, 마이크로소프트 사의 New Bing 3에서는 LLM(Large Language Model)을 기반으로 검색 결과를 개선하는 첫 시도를 선보였다. 이러한 LLM을 활용한 대화형 챗봇으로 사용자와의 대화를 통해 사용자의 요구에 맞는 맞춤형 대답을 생성할 수 있다.

최근 몇 년간 온라인 플랫폼 중고차 시장은 급격한 성장세를 보이고 있다. 경제적 불확실성, 신차 가격 상승 그리고 온라인 플랫폼을 통해 과거보다 편리해진 중고차 검색 및 조회는 많은 사람들이 새 차 대신 중고차를 선택하게 만드는 주요 요인이 되어왔다. 하지만 다양한 중고차 매매 플랫폼 속에서 소비자들은 방대한 선택지 중 자신에게 적합한 차량을 찾기 위해 많은 시간을 소비하고 있다. 특히나 차에 대해서 잘 모르는 사람들은 차량의 상태, 가격, 성능 등의 다양한 요소들을 종합적으로 고려하지 못한 채 자신에게 맞지 않는 차를 구매하는 결과로 이어지는 결과가 많다.

머신러닝과 자연어처리를 활용한 중고차 추천 AI 플랫폼은 이러한 문제점을 해결하려고 한다. 사용자가 AI 챗봇과 대화를 하여 자동차에 대해 잘 모르는 사람들도 중고차를 구매할 때 확인해야 하는 중요한 정보들을 확인하여 쉽고 편리하게 확인할 수 있다.

대량의 중고차 데이터를 전처리기법을 통해 정제하고 중고차 구매시 반드시 확인해야 하는 중요한 정보들을 더한 후에 사전 학습된 LLM에 추가학습(fine-tuning)시키는 방식으로 중고차 추천 챗봇을 만들 수 있다. 이를 통해 누구나 빠르고 쉽게 실시간으로 사용자의 선호도와 요구사항(예산, 선호하는 차종, 연식, 주행거리, 연료타입)을 바탕으로 개인화된 차량 추천을 받아 최소한의 예산으로 최대한의 만족을 가져갈 수 있는 차량을 선택할 수 있다.

2. Project Description

2.1 Project Objectives

중고차 추천 AI 플랫폼은 자신에게 맞는 중고차를 찾고 있거나 중고차를 살 때 중요한 정보들을 알지 못해서 정하지 못하는 사용자를 위한 서비스이다. 사용자는 플랫폼 안에 탑재된 챗봇과의 대화를 통해 자신의 선호도와 요구사항을 고려한 맞춤형 중고차를 추천 받는다.

동작과정

1. 사용자가 챗봇에게 자신이 원하는 차량의 요구사항에 대해 대화로 입력한다.
2. 대화를 통해 받은 사용자의 정보를 토대로 차량을 추천해주고 사용자가 적은 양의 정보만을 제공해 준 경우 차종(세단, SUV, 스포츠카..), 목적(차박, 출퇴근용, 의전용..), 운전성향(여유로운 주행, 스포티한 주행..), 연료타입(디젤, 가솔린, 하이브리드, 전기..), 가격, 브랜드, 연식, 주행거리, 환경, 등을 자세한 설명과 함께 추가적으로 제공해 줄 것을 요청한다.
3. 로그인기능을 추가해 사용자가 최근에 검색하거나 입력받았던 내용들을 토대로 이어서 추천을 할 수 있다.

개발목표

pretrained model로 KULLM3 모델을 선정하고 시스템 프롬프트 튜닝에 맞게 데이터를 수집한다. 수집한 데이터를 정제하여 fine tuning이 가능한 상태로 만든 후 fine tuning 과정을 거쳐서 모델을 완성시킨다.

모델을 챗봇 형태로 웹어플리케이션을 통해 이용 가능하게 만든다. 이를 통해 사용자들의 요구사항과 선호도를 입력하면 실시간으로 중고차 차량을 추천해주는 모델을 배포할 수 있다. 또한 사용자에게 필요하다고 생각되는 정보를 제공할 수 있다.

로그인 기능을 통해서 사용자들의 정보를 백엔드에서 관리하고 저장할 수 있다.

2.2 Expected Benefits

머신러닝, 자연어 처리 기반 중고차 추천 플랫폼은 다음과 같은 기대효과가 있다.

- **개인 조건별 중고차 추천**

대화를 통한 소비자의 니즈를 파악 사용자의 예산, 선호하는 차종, 연식, 주행거리, 연료타입 등 개인화된 요구사항을 반영하여 최적의 중고차를 추천한다.

- **편의성과 효율성**

웹 어플리케이션 기반 챗봇은 사용자들이 쉽게 접근할 수 있으며, 언제든지 대화를 마저 이어나갈 수 있다. 메신저와 유사하게 채팅을 통한 서비스 제공에 의해 진입장벽이 낮다. 또한 실시간으로 중고차 정보를 제공받을 수 있어 소비자는 방대한 정보를 일일이 검색하고 비교하는 시간을 획기적으로 절약할 수 있다.

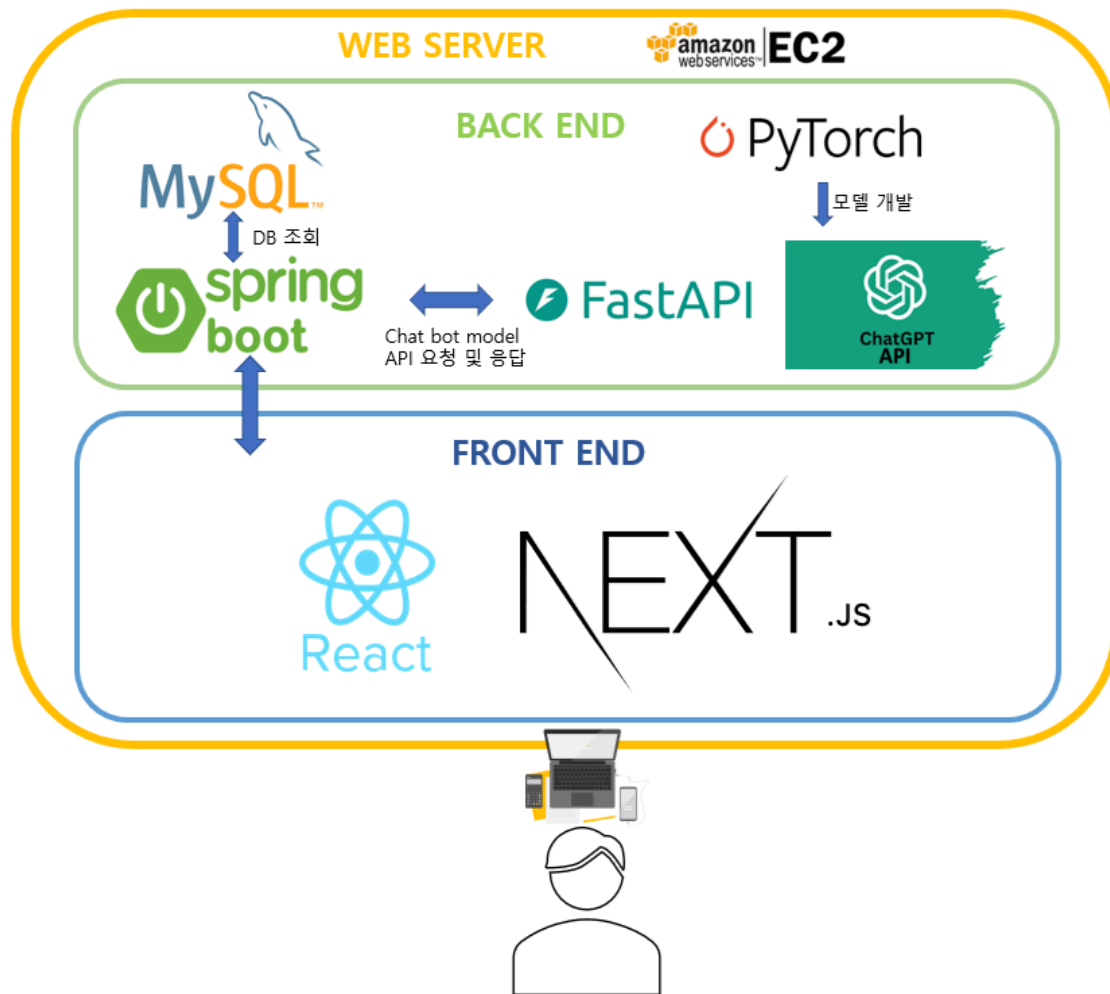
- **경제적 이익**

최소한의 예산으로 최대한의 효용을 가져갈 수 있는 차량을 선택할 수 있도록 할 수 있다.

- **정보 비대칭성 완화**

중고차 시장은 차에 대해 잘 모르는 소비자와 차에 대해 제대로 된 설명을 제공하지 않는 판매자 간의 정보 비대칭성이 크다. 그리고 이 정보 비대칭성을 이용해 판매자는 보다 더 많은 이익을 얻을 수 있다. 하지만 중고차 추천 플랫폼 챗봇을 통해 차량에 대해 잘 모르는 사람들도 쉽게 이해하고 필요한 정보를 얻을 수 있어 정보 비대칭성을 완화하므로 기존의 플랫폼보다 더욱 소비자 친화적이다.

2.3 Development Environment and Design Diagram



- 사용자 : 채팅 인터페이스를 통해 중고차 추천 요청.
- Front-end (React, Next.js): 사용자 인터페이스, RESTful API 요청 전송.
- Back-end (Spring Boot): RESTful API 서버, MySQL 데이터베이스와 통신.
- Chatbot API 서버 (FastAPI + ChatGPT): ChatGPT와 통신하여 응답 생성.
- 데이터베이스 (MySQL): 차량 데이터 저장 및 관리.
- 클라우드 서비스 (AWS EC2): 전체 시스템 호스팅.

3. Development Process and Technical Explanation

3.1 LLM & fine-tuning

1) LLM(Large Language Model)

LLM은 "Large Language Model"의 약어로, 매우 큰 규모의 매개변수를 가진 신경망 기반 언어 모델을 지칭한다. 이 모델은 대규모 텍스트 데이터를 사용하여 광범위한 언어 이해 및 생성 작업을 수행할 수 있는 딥 러닝 알고리즘이다.

이러한 모델은 수십억 개 이상의 매개변수를 포함하고 있으며, 이 매개변수는 모델이 언어의 다양한 측면을 학습하고 표현할 수 있도록 도와준다. 주로 "transformer"라는 신경망 아키텍처를 기반으로 하며, 텍스트 데이터를 처리하고 이해하는 방법을 학습한다.

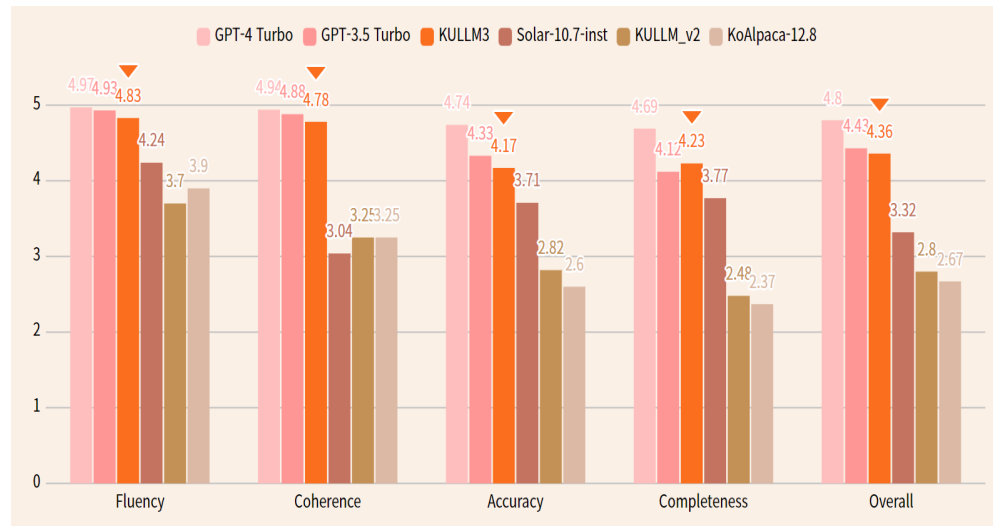
이러한 LLM 모델은 주로 "self-supervised learning"이라는 학습 방법을 사용하여 훈련된다. 이는 레이블이 지정되지 않은 대규모 텍스트 데이터를 사용하여 모델을 학습하는 기술로, 모델은 입력 데이터를 자체적으로 예측하고, 이를 통해 학습을 진행한다.

LLM은 최신 자연어 처리 기술의 핵심 부분으로서, 다양한 분야에서 활용되고 있다. 예를 들어, 검색 엔진, 자동 요약 시스템, 대화형 AI 시스템 등에서 LLM이 활용되어 사용자와의 상호작용을 향상시키고, 다양한 언어 작업을 자동화하는 데 기여하고 있다.

KULLM

고려대학교의 NLP&AI 연구실과 HIAI 연구소가 협력하여 개발한 한국어 특화 대형 언어 모델인 '구름(KULLM)'은 한국어 자연어 처리 분야에서 주목받는 최신 모델 중 하나이다. 이 모델은 대규모의 한국어 텍스트 데이터를 사용하여 학습되었으며, 고품질의 한국어 지시어 데이터셋을 활용하여 한국어 지시를 따르는 능력을 향상시켰다.

특히 KULLM3은 GPT-4 Turbo와 비슷한 성능을 보여, 기존의 한국어 모델보다 우수한 생성 능력을 보인다. 이 모델은 유창성, 응집성, 정확도, 완결성 등의 평가에서 우수한 성능을 보였으며, 한국어로 된 지시 사항을 이해하고 실행하는 능력이 뛰어나다.



2) fine-tuning

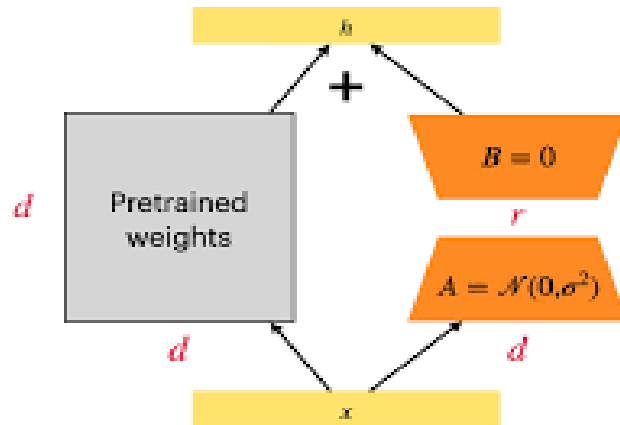
fine-tuning은 사전에 훈련된 모델에 데이터를 업데이트 하여 특정 작업 혹은 도메인에 대해 적합성을 높이는 작업이다. 해당 작업은 머신러닝 모델이 초기 학습에서 사용된 데이터의 논리와 이해력만을 반영할 수 있다는 점에서 유래한다. 따라서 모델은 새로운 데이터를 이해하는 것에 어려움을 겪을 수 있으며 특히나 대규모 모델에서 이 점이 부각된다. fine-tuning은 언어 모델 영역인 자연어 처리(NLP) 부분에서 많이 사용된다. 이 방식을 이용한 LLM모델이 점차 개발되고 배포되고 있는 중이다.

LoRA(Low-Rank Adaptation)

LoRA란 모델의 파라미터를 전체적으로 조정하는 대신 모델의 파라미터 중 일부만을 조정하여 특정 작업에 맞게 모델을 최적화하는 기술이다. GPT와 같은 대규모 언어 모델 같은 대규모 모델의 미세조정에 유용하며, 계산 비용과 학습 시간을 절약하면서도 효과적인 성능 향상을 노릴 수 있다.

Low Rank Adaptation의 원리

- Low Rank Matrix
 - 낮은 순위의 행렬을 사용하여 기존 모델의 파라미터를 근사한다.
 - 모델의 파라미터 수를 줄이면서, 원본 모델의 주요 데이터를 유지할 수 있게 해준다.
- Adaptation
 - 주로 모델의 특정 층(layer)에 적용되며, 이 층의 파라미터를 낮은 순위의 행렬로 대체하여 새로운 데이터셋에 대해 미세조정할 수 있다.
 - 모델의 나머지 부분은 그대로 유지되며, 선택된 층의 파라미터만이 최적화된다.



해당 fine-tuning 기법을 바탕으로 중고차 추천과 관련된 데이터를 학습시킬 수 있다.

3) Model Development

1. Selection of Pre-trained Model

pre trained model로 KULLM3 를 선택한다. KULLM3 는 고급 명령 수행 능력과 유창한 대화 능력을 갖춘 모델이다. 특히 GPT 3.5 turbo를 가까이 따르는 명령 수행에서 뛰어난 성능을 보인다.

2. System Prompt Tuning

KULLM3 model은 시스템 프롬프트 없이 평가 시 기대보다 낮은 성능을 보일 수 있다. 이렇게 성능을 올려주는 시스템 프롬프트 튜닝은 모델이 fine-tuning되기 전에 수행되는 기술이다. 이는 기존 모델에 fine-tuning 하게 될 데이터의 일부를 제공하여, 모델이 해당 데이터에 대해 추론하게 한다. 모델이 추론해낸 결과를 바탕으로 fine-tuning 시킬 데이터의 형태를 결정한다.

3. Data Collection

데이터의 형태가 프롬프트 튜닝을 통해 결정되었다면 국내의 중고차 유명 사이트와 국토교통부의 자료를 크롤링을 통해 수집한다. 그리고 사람들이 중고차를 구매할 때 중요하게 여기는 추가적인 요소 목적(차박, 출퇴근용, 의전용..), 운전성향(여유로운 주행, 스포티한 주행..), 중요사고유무 등을 추가하여 정리한다. 크롤링은 BeautifulSoup을 활용하여 진행한다.

4. data preprocessing

결측값을 처리하고 이상값 탐지 및 제거를 한다. 그리고 데이터를 정제하여 HTML 태그, 특수문자, 불필요한 공백 등을 제거한다. 데이터 형식을 통일하고 범주형 데이터를 모델이 이해할 수 있는 숫자 형식으로 변환한다. 이를 통해 수집한 데이터를 fine tuning이 가능한 형식으로 준비한다.

5. Model Fine-tuning

준비된 데이터를 사용해서 모델을 fine tuning 한다. 모델이 새로운 데이터셋을 학습하면서 파라미터를 조정하게 된다.

6. Model Evaluation

Fine-tuning한 모델을 평가한다. 문제점이 있다면 다시 이전 단계로 돌아가서 보완한다.

3.2 Web Application and AI Chatbot Server Development

1) Chatbot Server within the Platform

챗봇 AI 모델의 서버 부담을 줄이기 위해 이를 API 형태로 분리하고 스케일 아웃을 한다. 이를 통해 로드밸런싱을 도입할 수 있고 서버 간 트래픽을 효율적으로 분산시키고 서버의 증설 및 축소를 간단하게 할 수 있다. AI 모델은 구동에 많은 메모리를 필요로 하므로 이를 위해 별도의 고성능 서버를 마련해서 AI 챗봇 API 서버를 구축해야 한다. 이를 통해 AI 모델의 처리 속도를 향상시켜서 사용자 경험을 크게 개선시킬 수 있다.

2) Backend Server

사용자가 챗봇에 입력을 하면 백엔드 서버는 그 요청을 받아서 처리한다. 회원가입과 로그인 기능에 있어서는 사용자의 세션을 유지하고 기존에 저장된 사용자의 채팅 기록과 선호도 등에 대한 정보를 불러와서 제공한다. 사용자의 채팅 요청이 있다면 방금 서버로 도착한 채팅에다 기존 데이터베이스에 저장되어 있던 사용자의 채팅기록과 선호도 등의 정보를 포함 내용이 입력이 된다. 출력은 백엔드 서버에서 다시 한번 더 가공되어 사용자에게 전달된다.

3) User Interface

화면의 UI는 현재 나와있는 Chat gpt, claude 등의 AI 채팅 서비스 UI와 유사하게 하면서도 중고차 관련 정보를 가독성 높게 제공할 수 있도록 수정한다. React 기반의 Next.js 프레임워크를 사용하여 서버 사이드 렌더링으로 클라이언트 페이지를 개발한다.

3.3 Deployment and Operation Plan

AWS EC2 인스턴스를 설정하여 백엔드 서버와 데이터베이스를 호스팅한다. 필요한 경우 AWS RDS를 사용하여 MySQL 데이터베이스를 별도로 관리할 수 있다. 네트워크 설정 및 보안 그룹 설정을 통해 인스턴스에 대한 접근을 제어한다. 사용자가 쉽게 접근할 수 있도록 도메인을 설정하고, HTTPS 통신을 위해 SSL 인증서를 적용한다. 수정 및 문제점 건의 게시판을 만들어서 사용자의 피드백을 지속적으로 수집하고 분석하여 서비스 품질을 향상시킨다.

4. Development Schedule and Role Allocation

4.1 Development Schedule

6월	7월					8월					9월					10월		
5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3
DB 설계	UI 디자인																	
	API문서 작성																	
	데이터 수집																	
				언어 모델 fine tuning, 테스트														
							중간보고서 제출											
								백엔드 개발										
									프론트 개발									
											배포 서버 환경 구축							
													테스트/배포					
																최종발표		

4.2 Role Allocation

이름	역할
이은진	데이터 수집, 프론트엔드 개발, 백엔드 개발, AI 모델 제작, UI 디자인
최희웅	데이터 수집, 프론트엔드 개발, 백엔드 개발, AI 챗봇 api 제작(FastAPI)
최성빈	데이터 수집, 프론트엔드 개발, 백엔드 개발, AI 모델 제작, 서버 및 배포 환경 구축