

주제	세부내용
데이터 엔지니어링을 위한 파이썬 기초	<ul style="list-style-type: none">파이썬 기본 문법: 입출력 ~ 자료형 및 변수파이썬 기본 문법: 리스트 활용 ~ 조건문파이썬 기본 문법: 반복문 ~ 복합자료형파이썬 기본 문법: 함수 ~ 클래스파이썬 기본 문법: 모듈/패키지/라이브러리 ~ 파일 I/O와 예외처리
바이브 코딩	<ul style="list-style-type: none">바이브 코딩: 개념과 프롬프트 엔지니어링, 환경 구축바이브 코딩: Python Tkinter를 활용한 매크로 프로젝트
데이터 수집	<ul style="list-style-type: none">데이터 분석 프로세스, Beautiful Soup를 활용한 웹 크롤링Open API를 활용한 데이터 수집
데이터 전처리	<ul style="list-style-type: none">Pandas 라이브러리 기초: 데이터 탐색, 요약Pandas 라이브러리 활용: 데이터 추출, 정제
EDA와 시각화	<ul style="list-style-type: none">Matplotlib과 Seaborn 라이브러리 기초와 그래프다양한 그래프와 EDA 과정
머신러닝	<ul style="list-style-type: none">머신러닝 개요, Feature Engineering 및 스케일링과 인코딩머신러닝 지도 학습: Regression, Decision Tree, KNN머신러닝 비지도 학습: 유사도와 K-Means머신러닝 모델 평가 지표: MAPE, Confusion Matrix, ROC머신러닝 모델 평가 및 튜닝: Hold-out, K-Fold, GridSearch, RandomSearch
MLFlow	<ul style="list-style-type: none">MLOps, MLFlow 개념 및 구조와 설치MLFlow Tracking을 활용한 실험 관리 전략미니 프로젝트: 데이터 수집 + 분석 + ML
딥러닝	<ul style="list-style-type: none">인공신경망의 기본 구조 및 학습 과정 이해CNN을 활용한 이미지 분류 실습 (MNIST, CIFAR 등)전이학습(Transfer Learning)을 통한 모델 성능 향상RNN/LSTM 기반의 자연어 처리 이해 및 구현딥러닝 모델 학습 시 과적합 방지와 튜닝 기법 실습
생성형 AI	<ul style="list-style-type: none">Foundation Model(기반 모델)의 구조와 학습 메커니즘 이해대규모 언어모델(LLM)의 구조, 파이프라인 및 토큰화 실습프롬프트 엔지니어링 개념 이해 및 실제 적용 실습파인튜닝(Fine-tuning) vs. RAG(Retrieval-Augmented Generation) 비교질의응답(Q&A)-챗봇 시스템 설계 및 구현 실습벡터 임베딩과 RAG 기반 문서 검색·응답 처리 실습LLM을 활용한 문맥 유지형 요약 및 대화형 챗봇 서비스 구성Streamlit/Gradio 기반의 인터랙티브 챗봇 서비스 인터페이스 구현
1차 프로젝트	<ul style="list-style-type: none">프로젝트 주제 선정 및 서비스 기획서 작성데이터셋 수집LLM 파인튜닝 및 프롬프트 설계 실습질의응답(Q&A)-요약형 챗봇 기능 구현모델 성능 평가 및 실험 결과 분석모델 성능 평가 및 실험 결과 분석 및 최적화된 챗봇 모델 배포최적화된 챗봇 모델 배포

주제	세부내용
데이터 엔지니어링	<ul style="list-style-type: none"> 데이터 엔지니어링의 역할과 중요성 이해 ETL 파이프라인 구조 설계 및 구성 원리 실습 Docker, 리눅스 기반의 데이터 엔지니어링 환경 구축
실시간 데이터처리	<ul style="list-style-type: none"> Flink의 기본 개념 및 동작 원리 이해 데이터 스트림 처리(Streaming) 구조 설계 실습 Flink를 활용한 실시간 데이터 수집 및 변환 처리 구현 Kafka 연동을 통한 실시간 데이터 파이프라인 구성 실시간 모니터링 및 데이터 지연 최소화 실습 실시간 대시보드 구성 및 결과 시각화
배치 데이터 처리	<ul style="list-style-type: none"> Spark의 구조와 실행 메커니즘 이해 Spark SQL을 활용한 대규모 데이터 처리 실습 RDD의 개념과 활용 배치 처리와 스트리밍 처리의 비교 및 통합 응용 최적화 기법을 통한 Spark Job 성능 향상 Spark 클러스터 환경에서의 데이터 파이프라인 구축
데이터 구축	<ul style="list-style-type: none"> 분산 데이터 처리 프레임워크 Hadoop의 개념 및 구조 이해 하둡 에코시스템(HDFS, MapReduce, YARN 등) 구성요소 학습 하둡 시스템 기반 데이터 저장 및 처리 실습 하둡 클러스터 환경 설정 및 Job 실행 실습 HDFS 기반 데이터 수집·적재·관리 워크플로우 구축
데이터 서치	<ul style="list-style-type: none"> Elasticsearch의 개념 및 인덱스 구조 이해 데이터 색인(Indexing) 및 검색(Query) 동작 원리 실습 Elasticsearch를 활용한 검색 시스템 구축 실습 Kibana를 이용한 데이터 시각화 및 대시보드 구성 검색 효율 향상을 위한 쿼리 튜닝 및 필터링 기법 적용 Elasticsearch 클러스터 운영 및 성능 모니터링 실습
2차 프로젝트	<ul style="list-style-type: none"> 프로젝트 주제 선정 및 데이터 엔지니어링 요구사항 정의 데이터 수집 및 정제 Kafka-Flink 기반 실시간 데이터 스트리밍 처리 구현 Spark 기반 배치 데이터 처리 및 통합 저장 구조 설계 데이터 처리 결과 시각화 및 대시보드 구성 (Kibana/Streamlit) 파이프라인 성능 최적화 및 장애 대응 로깅 시스템 구축 통합 데이터 엔지니어링 시스템 시연 및 결과 보고서 작성
데이터 시스템 관리	<ul style="list-style-type: none"> 워크플로우 관리 도구 Airflow의 개념 및 구성 요소 이해 DAG(Directed Acyclic Graph) 구조를 활용한 워크플로우 설계 Airflow를 이용한 데이터 파이프라인 구축 및 스케줄링 실습 태스크 간 의존성 관리 및 자동 재시작 Airflow를 활용한 데이터 엔지니어링 시스템 자동화 운영 실습 로그 모니터링 및 실패 태스크 복구 전략 적용 운영 환경에서의 Airflow 최적화 및 확장성 관리
프로젝트 고도화	<ul style="list-style-type: none"> 프로젝트 고도화 목표 설정 실시간 SNS 데이터 수집 및 스트리밍 파이프라인 개선 (API-Kafka 활용) 수집 데이터의 정제·전처리 자동화 및 품질 관리 체계 구축 LLM을 활용한 실시간 질의응답(Q&A) 및 요약 기능 구현 데이터 엔지니어링 파이프라인과 LLM 모델 연동 통합 실시간 대시보드 및 모니터링 시스템 개발 (Kibana/Streamlit) 서비스 배포 환경 구성 및 운영 자동화 (Docker, Airflow 등) 결과 보고서 작성 및 시연 발표 준비