

2023.08.07~08.25

# GCP기반 머신러닝

## 프로젝트

## 1. 프로젝트 개요

- 프로젝트명: 예) GCP 기반 머신러닝 프로젝트
- 목적: GCP 서비스를 활용하여 효율적으로 데이터 전처리, 모델 개발, 훈련 및 배포를 수행하여 머신러닝 모델을 구축하고 실제 환경에서 활용하는 것
- 프로젝트 기간: 14일(112Hr)

## 2.프로젝트 목표

- 예측 모델 개발: GCP를 활용하여 고급 머신러닝 알고리즘을 사용하여 예측 모델을 개발. 이 모델은 데이터의 패턴을 학습하여 미래 값을 예측하거나 분류 작업을 수행할 수 있다.
- 대규모 데이터 처리 및 분석: GCP의 대용량 데이터 처리 기능을 활용하여 대규모 데이터를 효율적으로 처리하고 분석. 이를 통해 데이터의 패턴과 동향을 발견하거나 복잡한 계산을 수행할 수 있다.

- 이미지/음성/텍스트 처리: GCP의 머신러닝 API를 활용하여 이미지, 음성 또는 텍스트 데이터를 처리하고 분석하는 솔루션을 개발. 이를 통해 이미지 분류, 음성 감지, 자연어 처리 등 다양한 작업을 수행할 수 있다.
- 실시간 예측 서비스 제공: 훈련된 머신러닝 모델을 GCP의 Vertex AI에 배포하여 실시간 예측 서비스를 제공. 이를 통해 실시간으로 사용자의 요청에 대해 모델의 예측 결과를 제공할 수 있다.

- 비즈니스 인텔리전스 구축: GCP의 데이터 분석 및 시각화 도구를 활용하여 비즈니스 데이터를 분석하고 시각화. 이를 통해 의사 결정에 필요한 정보를 제공하고 인사이트를 도출한다
- 자동화된 프로세스 최적화: GCP의 자동화 및 프로세스 최적화 기능을 활용하여 비즈니스 프로세스를 자동화하고 최적화. 이를 통해 효율성을 향상시키고 비용을 절감할 수 있다.

### 3. 역할 분담 : 인원 4 명

#### (1) 프로젝트 리더(팀장):

- 프로젝트의 계획, 일정, 목표 설정 등을 관리한다
- 팀 구성원 간의 협력을 조율하고 커뮤니케이션을 관리
- 리소스 할당 및 우선순위 결정을 수행
- 프로젝트 진행 상황을 모니터링하고 보고서 작성 등을 담당

## (2) 데이터 엔지니어:

- 데이터 수집 및 저장을 위한 파이프라인을 구축
- 데이터의 전처리, 변환, 정제 작업을 수행.
- 데이터의 품질을 향상시키기 위한 작업을 담당.
- GCP의 데이터 관련 서비스 (BigQuery, Cloud Storage 등)를 활용합니다.

### (3) 머신러닝 엔지니어/과학자:

- 머신러닝 모델을 개발하고 훈련시킨다.
- 피처 엔지니어링 및 피처 선택을 수행한다.
- 모델의 하이퍼파라미터 튜닝 및 성능 개선 작업을 수행한다.
- 모델의 평가 및 결과 해석을 수행한다.



#### (4) 시스템 아키텍트:

- 프로젝트의 아키텍처를 설계하고 리소스를 관리한다
- GCP의 다양한 서비스를 조합하여 프로젝트의 구조를 정의한다.
- 시스템의 확장성, 안정성, 보안성을 고려하여 아키텍처를 최적화한다.

## (5) 개발자/프로그래머:

- 모델 개발 및 훈련에 필요한 코드를 작성.
- 모델 배포를 위한 API 엔드포인트를 개발하고 관리한다.
- 데이터 파이프라인, 자동화된 프로세스 등을 구현한다.

## (6) 비즈니스 분석가/전문가:

- 비즈니스 요구사항을 이해하고 모델의 목표 및 성과 지표를 설정한다.
- 프로젝트 결과물이 비즈니스 목표에 부합하는지 평가하고 의사 결정에 활용한다.

## (7) 문서 작성 및 보고 담당자:

- 프로젝트의 진행 상황, 결정사항, 보고서 등을 문서화한다. 팀 구성원 간의 정보 공유를 위한 문서 작성 및 관리를 담당한다.

## 4. 프로젝트 개발 단계

### (1) 계획 및 환경 설정 단계

- 프로젝트 목표 및 주제와 범위 정의
- GCP 계정 생성 및 프로젝트 생성
- 프로젝트 리소스 관리 및 보안 설정
- GCP 서비스 및 API 활성화

## (2) 데이터 수집 및 전처리 단계

- 데이터 수집 계획 수립 및 데이터 소스 확보
- 데이터 전처리 파이프라인 구축 (GCP Dataflow, Dataproc 등)
- 데이터 품질 향상을 위한 클리닝 및 변환 작업

### (3) 데이터 저장 및 관리 단계

- 데이터 저장 옵션 검토 및 선택 (BigQuery, Cloud Storage 등)
- 데이터베이스 스키마 설계 및 구축 (BigQuery 등)
- 데이터 저장 및 관리를 위한 접근 권한 및 보안 설정

## (4) 머신러닝 모델 개발 단계

- GCP Vertex AI 사용을 위한 모델 개발 환경 설정
- 피쳐 엔지니어링 및 데이터 분할 (Train, Validation, Test)
- 머신러닝 알고리즘 선택 및 모델 개발 (TensorFlow, PyTorch 등)
- 하이퍼파라미터 튜닝 및 교차 검증 수행

## (5) 모델 훈련 및 평가 단계

- 분산 훈련을 위한 GCP 리소스 활용 (Vertex AI Training 등)
- 훈련된 모델의 성능 평가 및 모델 선택
- 모델의 예측 성능 향상을 위한 실험 및 수정 작업



## (6) 모델 배포 및 관리 단계

- 훈련된 모델을 GCP Vertex AI에 배포
- 모델 엔드포인트 설정 및 API 제공
- 모델의 실시간 예측 요청 처리 및 모니터링 설정

## (7) 문서화 및 마무리 단계

- 프로젝트 관련 문서 작성 (프로젝트 보고서, 코드 설명 등)
- 프로젝트 결과물을 공유할 수 있는 형식으로 정리
- 시각화

## 5. 프로젝트 결과 시연 예시:

- 웹 애플리케이션 또는 스크립트를 작성하여 사용자가 텍스트를 입력하면 해당 텍스트의 감정을 분석하는 서비스를 제공한다.
- 사용자가 영화 리뷰를 입력하면 API를 호출하여 모델의 예측 결과(긍정 또는 부정)를 반환 받는다.
- 시연 결과를 확인하여 입력한 영화 리뷰의 감정 분석이 올바르게 수행되는지 확인합니다. 모델의 성능 및 예측 정확도를 평가한다.

## 6. 프로젝트 일정

- 1주차 : 프로젝트 계획 수립 및 역할 분담, GCP 서비스 구성, 데이터 수집
- 2주차 : 머신 러닝 구현, 모델 학습 및 배포 및 평가, PPT발표 자료 작성
- 3주차 : 프로젝트 발표 및 시연, 프로젝트 보완

## 7. 발표자료 작성시 주의

- 프로젝트에 사용된 각 **GCP** 서비스들에 대한 설명을 반드시 포함시킨다
- 사용된 소스 코드나 스크립트는 되도록 확대하여 시연 시 보일 수 있도록 하고 지나치게 세세한 소스코드는 생략한다
- 팀원 모두 발표할 수 있도록 각각 별도로 발표자료를 준비한다(공통부분은 서로 공유)

## 프로젝트 진행시 출석 및 시간 엄수

- 정상 수업과 동일하게 50분 수업 10분 휴식으로 진행
- 출석 및 수업 시간 엄수
- 프로젝트 중에도 필요할 경우 약간의 강의를 진행 될 수도 있음
- 프로젝트 진행 시 발생하는 문제점 도움 요청

### \*\* 1일 수업 시간표 \*\*

[오전]	09:00~09:50	10:00~10:50	11:00~11:30
[오후]	01:00~01:50	02:00~02:50	03:00~03:50
	04:00~04:50	05:00~05:50	

감사합니다