

2025년도 2학기 빅데이터응용프로그래밍

중간고사/기말고사 과제 제출 가이드

1. 개요

「빅데이터 응용프로그래밍」 과목은 학생 각자가 자신의 관심사에 기반하여 데이터 분석의 전 과정을 직접 수행함으로써, 문제 정의에서 데이터 수집·분석·해석에 이르는 과정을 자기주도적으로 체계적으로 학습하도록 하는 데 그 취지를 두며, 이에 따라 중간고사에서는 데이터 분석 계획서(PPT)를, 기말고사에서는 실행 및 결과 보고서를 제출하도록 하여 실제 데이터 분석 역량과 문제 해결 능력을 함께 향상시키는 것을 목표로 한다. 또한 이를 통해 학생들은 이론 중심의 학습을 넘어 실무에서 요구되는 데이터 활용 능력과 분석적 사고력을 함양하고, 다양한 데이터 도구 및 분석 기법을 통합적으로 적용하는 경험을 쌓을 수 있도록 한다.

2. 중간고사 (데이터 분석 계획서)

중간고사 레포트에 포함될 내용은 아래와 같다. 아래 내용은 참고의 용도이며 필요한 내용은 추가할 수 있다.

2.1 표지 및 기본정보

- 과목명 / 학번 / 이름 / 제출일
- 프로젝트 주제명

2.2 주제 선정 이유 및 배경

- 왜 이 주제를 선택했는가? (관심 분야, 사회적 이슈, 산업적 필요 등)
- 분석을 통해 해결하고자 하는 문제 정의
- 예상되는 인사이트 또는 기대효과

* 예시: “요식업 리뷰 데이터를 통해 지역별 고객 만족도 요인 분석”, “국내 주식시장 뉴스 데이터를 이용한 감성분석 기반 종목 트렌드 예측”

2.3 데이터 수집 계획

- 데이터 출처: 공공데이터, Kaggle, SNS API, 웹 사이트 크롤링, 내부 데이터 등
- 데이터 유형: 정형 / 반정형 / 비정형
- 수집 방법: Python Crawling, Open API, CSV 다운로드 등

- 예상 데이터량: 건수, 크기 등

* 주의사항: 개인정보 또는 민감정보의 경우 블라인드 처리 / 데이터 출처 표시 필수

2.4 데이터 전처리 및 분석 방법 계획

- 결측치(오류데이터) / 이상치 처리 방법
- 텍스트 데이터의 경우: 형태소 분석, 불용어 처리 등
- 사용가능한 분석 기법 활용
- 시각화 처리(matplotlib, pandas 등)

2.5 분석 도구 및 기술 스택

- Python (pandas, matplotlib, scikit-learn 등)
- R, Excel, SQL, Google Colab, Power BI 등
- 사용 이유 명시

2.6 기대 결과 및 활용 방안

- 분석을 통해 얻을 수 있는 인사이트
- 실제로 어떤 의사결정이나 서비스 개선에 활용될 수 있을지

2.7 일정 계획

- 과제를 진행하는 일정(결과 보고서까지의 일정을 포함)

2.8 참고문헌 및 출처

- 참고한 논문, 사이트, 데이터 출처 명시

3. 기말고사 (실행 및 결과 보고서)

기말고사 레포트에 포함될 내용은 아래와 같다. 아래 내용은 참고의 용도이며 필요한 내용은 추가할 수 있다.

추가로, 기말고사 결과 보고서의 발표 영상을 추가로 제출(제출 영상에는 결과 보고서와 발표자 본인의 음성이 포함되어야 하며, 발표시간은 5분~6분 이내로 한다.)

3.1 표지 및 요약

- 과목명 / 학번 / 이름 / 제출일
- 프로젝트명 / 작성자 / 요약

3.2 주제 및 목적 재정리

- 계획서 대비 주제 변경 여부 및 이유

- 목표 명확히 재기술

3.3 데이터 수집 및 전처리

- 실제 확보한 데이터 설명 (샘플 코드 및 수집 정보 스크린샷 포함)
- 데이터 크기, 주요 변수 설명
- 전처리 과정 요약 (코드 및 화면 스크린샷)

3.4 분석 과정

- 수행한 분석 방법의 단계별 설명
- 모델 학습, 성능 비교, 통계 분석 결과 등 (가능한 경우 처리)
- 시각화 결과 포함 (그래프, 차트, 워드클라우드 등)

3.5 주요 결과 및 해석

- 데이터로부터 도출된 핵심 인사이트
- 가설 검증 결과
- 예측 정확도나 유의미한 패턴 설명

3.6 결론 및 한계

- 분석 결과의 의미
- 프로젝트 수행 중 한계점 (데이터 부족, 시간 제약 등)
- 향후 보완 방향 제시

3.7 참고자료 및 코드 링크

- 데이터 출처
- GitHub / Colab 링크 (진행한 과제의 주소의 경우는 퍼블릭오픈으로 하여 주소 추가)

4. 평가기준

중간고사와 기말고사로 제출된 리포트는 아래 기준을 통해서 평가를 진행한다.

4.1 중간고사

- 주제의 독창성(흥미롭고 실질적인 주제를 선정하였는지 여부) - 20%
- 데이터 확보 적절성(신뢰가 가능한 데이터를 활용하였는지와 충분한 데이터를 확보할 수 있는지의 여부) - 20%
- 분석 설계 / 기법 활용(분석 절차가 논리적인지에 대한 여부와 기술 활용이 잘 사용되었는지에 대한 여부) - 30%
- 표현력 / 보고서 완성도(PPT 및 보고서의 구성 및 설명의 완성도가 잘 되었는지의 여부) - 20%

- 위 항목 이외에 추가 점수를 줄 수 있는 주관적인 요소(수업 시간의 활용도, 수업 시간의 태도 등) - 10%

4.2 기말고사

- 주제의 독창성(흥미롭고 실질적인 주제를 선정하였는지 여부) - 15%
- 데이터 확보 적절성(신뢰가 가능한 데이터를 활용하였는지와 충분한 데이터를 확보할 수 있는지의 여부) - 15%
- 분석 설계 / 기법 활용(분석 절차가 논리적인지에 대한 여부와 기술 활용이 잘 사용되었는지에 대한 여부) - 15%
- 결과의 타당성 및 해석(분석된 결과가 타당하게 표현되었는지의 여부와 해석 부분의 객관적인 완성도 평가) - 15%
- 표현력 / 보고서 완성도(PPT 및 보고서의 구성 및 설명의 완성도가 잘 되었는지의 여부와 발표를 통해서 본인의 주제를 잘 전달하는지 여부) - 30%
- 위 항목 이외에 추가 점수를 줄 수 있는 주관적인 요소(수업 시간의 활용도, 수업 시간의 태도 등) - 10%

5. 제출 일정

- 제출 일정은 원격수업(<https://eclass.dongyang.ac.kr/>) 사이트 수업과목 공지사항 참조

6. 기타

중간고사에는 "왜 이 데이터를 분석해야 하는가"를 설득하는 게 중요한 포인트이며, 기말고사에는 "데이터가 무엇을 말해주는가"를 증명하는 게 핵심적인 요소에 기반해서 제출(발표 영상 포함)하여야 합니다.