**문항1. 데이터 분석과 빅데이터 분석의 차이점에 4가지에 대해서 설명하시오. (48점)**

데이터 분석은 기술적 분석, 진단적 분석, 예측 분석, 처방적 분석의 유형이 있습니다.

기술적 분석은 점수카드, 대시보드에 기초하여 과거에 무슨 일이 발생했는지를 밝혀주는 것을 의미합니다. 또한 그 결과를 어떻게 해석하느냐 하는 해석 결과에 따라 어떻게 행동할 지 인간이 결정하게 됩니다. 진단적 분석은 과거의 축적된 행동 데이터를 토대로 인과관계를 찾아내는 것을 의미합니다. 예측 분석은 통계학적 모델을 사용해 미래에 무슨 일이 어느 정도의 확률로 일어날지 예측하는 것을 의미합니다. 처방적 분석은 예측하는 사태를 위해서 무엇을 하면 좋을지 처방하는 것으로, 보통 행동의 결과는 컴퓨터가 고찰합니다.

반면, 빅데이터 분석은 데이터 분석가, 분석 기반 기술, 알고리즘과 분석 모델 이해, 데이터 통찰력 등의 특성이 필요합니다.

**문항2. 데이터 전처리 기술 3가지가 무엇인지 명칭을 작성하고, 그 개념에 대해 설명하시오.(필요시, 사용되는 세부적인 기술을 설명하시오.) (52점)**

데이터 전처리 기술은 데이터 필터링, 데이터 변환, 데이터 정제의 3가지 종류로 이루어져 있습니다.

데이터 필터링은 오류를 발견하고, 보정하고, 삭제 및 중복성 검사를 수행하는 것을 의미합니다. 실제 사전 테스트에서 필터링 과정을 수행하게 되고, 필터링 기준을 최적화하여 활용합니다. 데이터 필터링을 처리할 때 고려해야 할 기술 사항은 1. 생성된 파일의 중복성을 확인할 수 있도록 파일명, 확장자 등의 필터링 기능을 제공해야 합니다. 2. 데이터 필터링시, 사전 정의된 기준에 의거해야 하며, 오류에 대한 이력을 저장해야 합니다.

데이터 변환은 다양한 형식으로 수집된 데이터를 분석하기 쉽도록 일관성있는 형식으로 바꿔주는 것을 의미합니다. 데이터 변환 기술은 5가지로 세부화할 수 있습니다. 데이터 추세에 벗어나는 값을 변환하는 평활화, 다양한 차원의 방법으로 데이터를 요약하는 집계, 통계 기법을 활용하여 데이터를 정해진 구간 내에 집어넣는 정규화, 범용 데이터에 적합한 모델을 만드는 일반화, 데이터 통합을 위해 새로운 특성을 만드는 속성 생성이 있습니다.

데이터 정제는 수집된 데이터의 불일치성을 교정하기 위한 것으로, 2가지로 세분화됩니다. 결측치 처리방법과 잡음 처리방식이 있습니다.