**문항1. 데이터 분석과 빅데이터 분석의 차이점에 4가지에 대해서 설명하시오.**

첫번째, 불확실성을 보유한 미래사회에서 빅데이터는 통찰력을 제공한다

두번째, 리스크를 보유한 미래사회에서 빅데이터는 대응력을 보유한다.

세번째, 스마트한 미래사회에서 빅데이터는 경쟁력이 된다.

네번째, 융합을 기반으로 하는 미래사회에서 빅데이터는 창조력을 보유하고 있다.

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

## 데이터 필터링

## 필터링이란 오류의 발견 후 보정과 삭제 또는 중복성 검사 등을 수행하는 것이다 실제 사전 테스트에서 필터링 과정을 수행하게 되고 필터링 기준을 최적화하여 활용하는 것이 일반적이다 데이터 필터링 작 업을 수행할 때에는 사전 정의된 기준에 의거하여야 하며 오류에 대한 이력을 저장해야 한다

## 데이터 변환

## 데이터 변환은 데이터의 특정 변수를 정해진 규칙에 따라 변환하는 것이다. 데이터 변환 기술은 평활화 집계 정규화 일반화 속성 생성이다.

## 첫번째 평활화란 데이터로부터 잡음제거를 위해 데이터 추세에 벗어나는 값들을 변환하는 기법이다.

## 두번째 집계란 다양한 차원의 방법으로 데이터를 요약하는 기법 이다 복수의 속성을 하나로 줄이거나 유사한 데이터 객체를 줄이고 스케일을 변경하는 기법을 사용한 다.

## 세번째 정규화란 데이터를 정해진 구간 내에 들도록 하는 기법이다 정규화를 할 때 사용하는 통계 방식은 최소 최대 정규화 Z 스코어 정규화 소수 스케일링이다

## 네번째 일반화란 특정 구간에 분포하 는 값으로 스케일을 변화시키는 기법이다 일반화가 잘 되어있다면 이상값이나 노이즈가 들어오더라도 크게 흔들리지 않는다

## 다섯 번째 속성 생성이란 데이터 통합을 위해 새로운 속성이나 특징을 만드는 기법이다 이는 주어진 여러 데이터 분포를 대표할 수 있는 새로운 속성이나 특징으로 대체하여 데이터를 변경 처리하는 것이 다

## 데이터 정제

## 데이터 정제는 수집된 데이터의 불일치성을 교정하기 위한 것이며 데이터 정제 기술에는 결측치의 처 리와 잡음 처리가 있다 결측치를 처리하는 방법은 해당 레코드 무시하기 자동채우기 담당자 수작업하 기가 있다 잡음을 처리하는 방법은 구간화 회기값 적용 군집화이다