**[과제제출] 마케팅사례로 배우는 빅데이터 기초와 비즈니스 활용**

**1. 데이터 분석과 빅데이터 분석의 차이점**

**데이터 분석의 유형**

첫 번째 기술적 분석이다. 기술적 분석이란 점수 카드나 대시보드에 기초하여 이뤄지고 과거에 무슨 일 이 발생했는지를 밝혀주는 것이다 그 후 분석 결과를 어떻게 해석하고 어떻게 행동으로 옮길지 판단하는 것은 인간이 수행한다.

두 번째 진단적 분석이다. 진단적 분석이란 과거에 축적된 행동 데이터를 토대로 인과관계를 찾아내는 것으로서 왜 그 일이 일어났는지를 밝히는 것이다. 이 분석 결과도 기술적 분석과 마찬가지로 어떻게 해 석하고 어떻게 행동으로 옮길지 판단하는 것은 인간이 수행한다.

세 번째 예측 분석이다. 예측 분석은 통계학적 모델을 사용해 미래에 무슨 일이 어느 정도의 확률로 일어날지를 예측하는 것이다. 기존의 데이터베이스에서 과거의 데이터를 참조하는 데이터 마이닝의 확장 기능이라고 할 수 있다.

네번째 처방적 분석이다. 처방적 분석은 차세대 분석법이라고 일컬어진다 즉 예측되는 사태를 위해서 무엇을 하면 좋을지 처방하는 것이다. 처방적 분석은 앞서 소개한 과거에 일어난 것을 해석하는 기술적 분석 그것이 왜 일어났는지를 밝히는 진단적 분석 그리고 앞으로 어떤 일이 일어날지를 추측하는 예측 분석을 조합하여 이루어진다

**빅데이터 분석 8단계**

(1) 문제정의 문제 정의는 기업이 해결하고자 하는 현재 상태의 개선이 무엇이고, 왜 그것을 해 결해야 하는지, 무엇을 달성할 것인지를 명확히 하는 것입니다. 문제를 정의할 때 에는 통계적인 시각, 즉 데이터의 수집과 분석으로 해결할 수 있도록 정의해야 합니다.

(2) 분석결과 탐색 문제를 해결하기 위해서 어떤 데이터 분석 결과가 필요한지 파악해야 합니다. 문제가 잘 정의되더라도 그 문제를 해결할 수 있는 분석 기법이 분명하지 않은 경우가 많습니다. 문제 해결을 위해서 필요한 분석 결과가 어떤 것인지를 고민해야 합니다.

(3) 분석기법 결정 원하는 결과를 만들어내는 분석기법이 하나인 경우보다는 여러 개인 경우가 많습니다.

(4) 데이터 지정 데이터 지정 시 데이터 유형 및 수집주기 등을 고려해야 한다. 현업과 통계 전문가 의 밀접한 의사소통과 협력이 필요합니다.

(5) 데이터 수집 및 전처리 해당 데이터를 수집, 저장, 전처리를 하는 작업이 필요합니다. 저장관리, 품질관리, 보안관리를 고려해야 합니다.

(6) 분석기법 적용 분석기법을 적용·평가하는 과정을 거쳐서 최적 모델을 선정한 뒤 이 최적 모델로 최종적인 분석 결과를 도출해야 합니다.

(7) 분석결과 해석 도출된 결과로부터 문제 해결의 방향을 제시하는 의미 있는 인사이트를 도출합니다.

(8) 액션플랜 전개 및 평가 추출한 인사이트를 기반으로 데이터 중심의 의사결정을 내리고 실무에 적용하여 피드백을 받습니다.

**2. 데이터 전처리 기술 3가지**

1) 데이터 필터링

* 필터링이란 오류의 발견 후 보정과 삭제 또는 중복성 검사 등을 수행
* 실제 사전 테스트에서 필터링 과정을 수행하게 되고 필터링 기준을 최적화하여 활용
* 데이터 필터링 작 업을 수행할 때에는 사전 정의된 기준에 의거하여야 하며 오류에 대한 이력을 저장해야 한다.

2) 데이터 변환

데이터 변환은 데이터의 특정 변수를 정해진 규칙에 따라 변환하는것.

데이터 변환 기술은 평활화, 집계, 정규화, 일반화, 속성 생성이며,

**평활화란** 데이터로부터 잡음제거를 위해 데이터 추세에 벗어나는 값들을 변환하는 기법.

**집계란** 다양한 차원의 방법으로 데이터를 요약하는 기법이다. 복수의 속성을 하나로 줄이거나 유사한 데이터 객체를 줄이고 스케일을 변경하는 기법을 사용한다.

**정규화란** 데이터를 정해진 구간 내에 들도록 하는 기법이다 정규화를 할 때 사용하는 통계 방식은 최소최대 정규화 Z스코어 정규화 소수 스케일링이다.

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

**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

3) 데이터 정제

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