

Data Science

데이터 시각화

강사 양석환

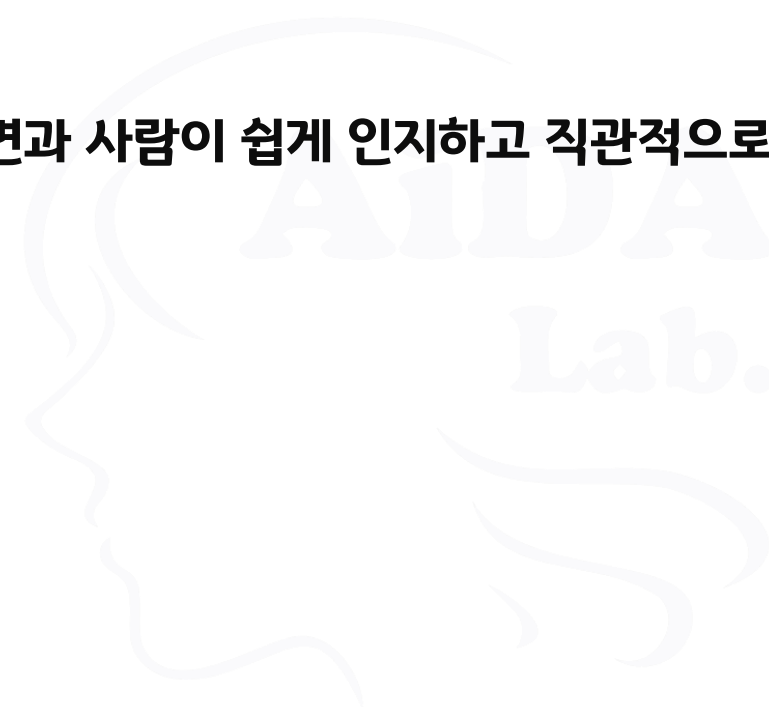


데이터 시각화



• 데이터 시각화의 정의

- 데이터의 분석 결과를 쉽게 이해할 수 있도록 시각적으로 표현하고 전달하는 과정과 기법
- 데이터 사이의 관계를 대표할 수 있는 특징을 도표나 차트 등으로 표현하여 **정보를 명확하고 효과적으로 전달하는 것을 목적으로 함**
- 시각화를 위해서는 데이터의 의미를 정확하게 표현하는 기능적인 측면과 사람이 쉽게 인지하고 직관적으로 이해할 수 있는 심미적인 측면을 모두 고려해야 함



• 데이터 시각화의 특성

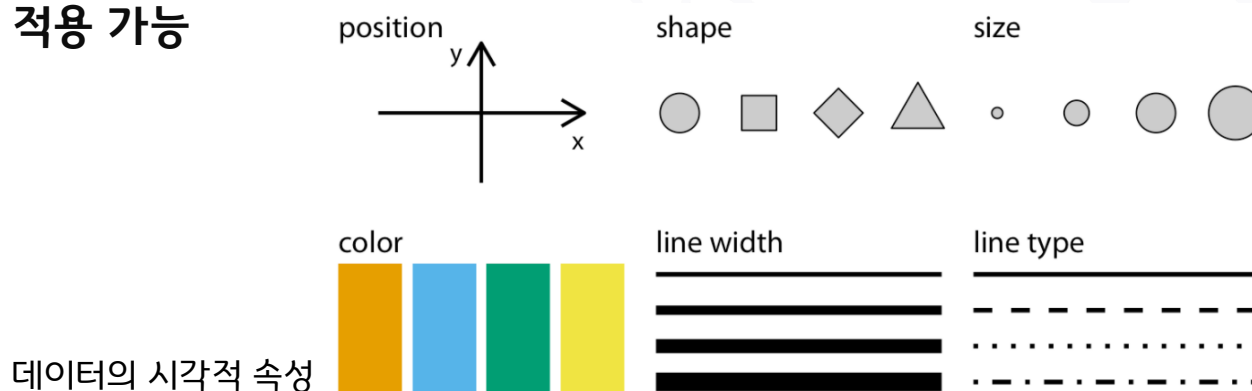
- 자료로부터 정보를 습득하는 시간을 절감하고 데이터에 대한 즉각적인 판단이 가능함
- 데이터의 특징과 패턴, 추세를 직관적으로 제공하여 데이터의 의미에 대한 이해를 효과적으로 도움
- 하나의 시각화 자료를 통해 시간, 공간, 분포 등 다양한 측면에서의 표현이 가능함
- 그래픽 도구를 이용함으로써 데이터의 의미, 관계, 차이, 분포 등을 선명하게 표현함

• 데이터 시각화의 접근 방법

접근 방법	정의
통계적 그래픽	데이터의 분포와 통계적 정보를 2차원 또는 3차원 공간에 시각적으로 표현하는 방법
주제 지도학	특정 주제의 지리적 분포와 패턴을 나타내는 지도 형태를 이용한 표현 방법

• 데이터의 시각적 속성

- 데이터 시각화는 체계적이고 논리적인 방식을 통해 데이터 값을 시각적 속성으로 변환한 다음, 그 속성들을 이용하여 최종 그래프를 만듦
- 데이터 값을 정량화 가능한 시각적 속성으로 나타내어 그래픽으로 표현
- 자주 사용되는 시각적 속성은 **위치, 형태, 크기, 색, 굵기, 선 유형** 등이 있음
 - 위치, 크기, 선 굵기, 색 속성은 연속형 데이터와 이산형 데이터에 모두 적용 가능
 - 형태, 선 유형 속성은 이산형 데이터에만 적용 가능



• 데이터의 시각화 방법

- 데이터 시각화는 목적에 따라 다양한 방법이 존재함
- 시간, 분포, 관계, 비교 및 공간을 중심으로 시각화 하는 방법이 주로 사용됨

시각화 방법	주요 도구
시간 시각화	막대 그래프, 누적 막대 그래프, 점/선 그래프
분포 시각화	히스토그램, 파이 차트, 도넛 차트, 트리 맵, 누적 연속 그래프
관계 시각화	산점도, 버블 차트, 히트 맵
비교 시각화	히트 맵, 체르노프 페이스, 스타 차트, 평행 좌표계, 다차원 척도법
공간 시각화	지도 매핑

• 데이터 시각화의 영역

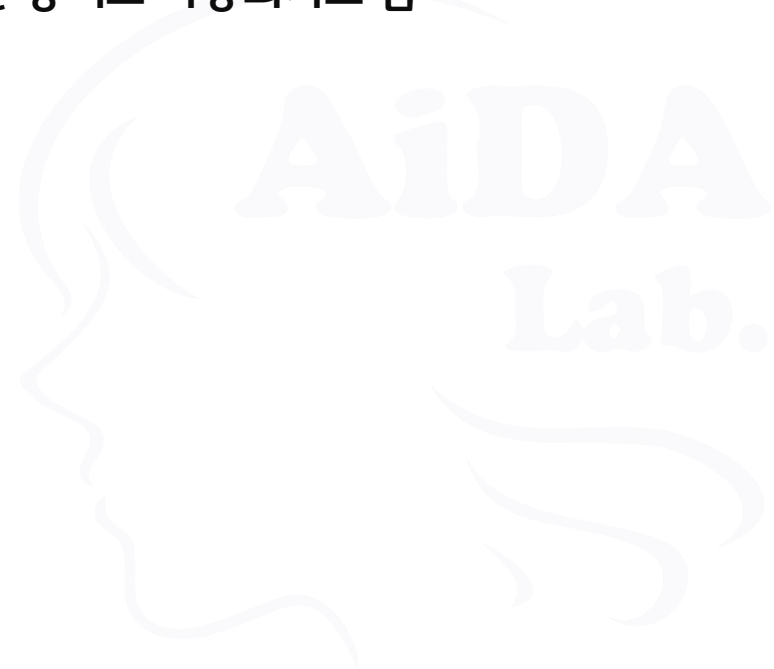
• 정보 시각화(Information Visualization)

- 방대한 양의 정보를 한 번에 사용자가 보고 이해할 수 있도록 직관적으로 표현하는 방법에 중점을 둠
- 데이터 시각화는 그래픽을 이용해 정보를 명확하게 표현하는 것을 강조하며, 정보 시각화는 큰 범위의 집합에 대한 시각적 표현 방법을 강조함
- 수치정보 뿐만 아니라 텍스트나 지형정보 같은 비수치 정보까지 포함함
- 데이터 시각화에서 한 단계 더 나아간 정보 형태의 가공 과정을 거침
- 카토 그램(Cartogram, 지도도표), 분기도(Cladogram), 개념도(Concept Map), 계통도(Dendrogram), 네트워크 다이어그램(Network Diagram), 트리 맵(Tree Map), 하이퍼볼릭 트리(Hyperbolic Tree) 등 다양한 도구를 사용함

- 데이터 시각화의 영역

- 정보 디자인(Information Design)

- 정보를 구성하여 효율적으로 사용할 수 있게 하는 디자인 기술 및 업무를 말함. 시각 디자인의 하위 영역
 - 정보를 효과적으로 전달하기 위한 수단으로서 그래픽 디자인을 강조하는 용어로 사용되기도 함
 - 데이터 시각화, 정보 시각화, 인포그래픽을 모두 포괄하는 개념



• 데이터 시각화의 영역

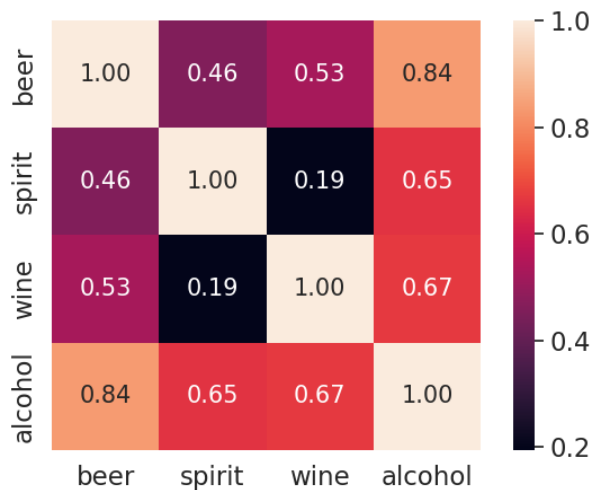
• 인포그래픽(Infographic)

- 복잡한 수치나 글로 표현되어 있는 정보와 지식을 차트, 지도, 픽토그램, 다이어그램, 일러스트레이션 등을 활용하여 한 눈에 파악할 수 있도록 시각적으로 표현하는 것
- 설득형 메시지를 전달하기 위해서 주로 사용되며, 정보를 빠르고 분명하게 표현하는 것에 중점을 둠
- 인포그래픽은 스토리를 통해 정보를 전달하려는 경향이 강함
- 데이터 시각화는 주로 전문 영역에서 활용되는 반면, 인포그래픽은 일반인을 대상으로 특정 정보와 메시지를 전달하기에 적합함

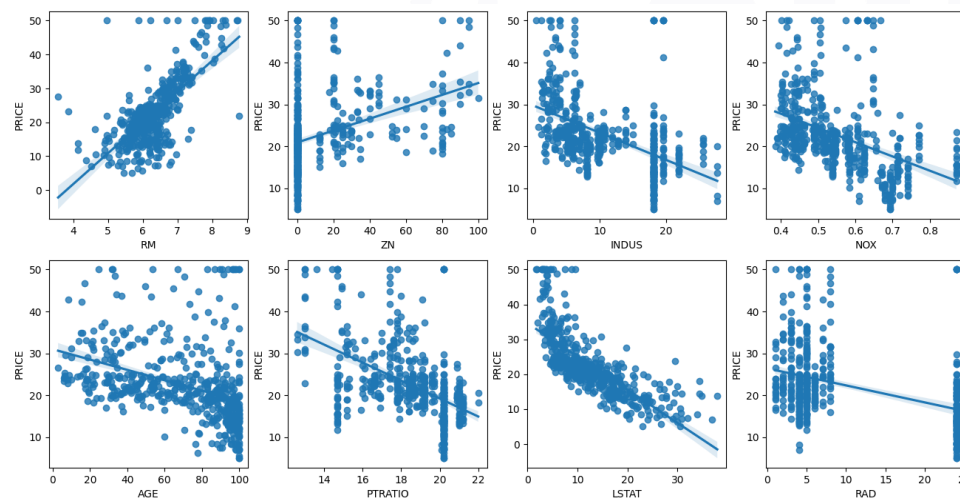


• 회귀 모델

- 회귀 모델은 여러가지 변수들을 동시에 비교하고 전체에서 식별이 되는 부분에 대한 수치, 정도를 표현하는 비교 시각화 기법으로 표현
- 또는 각기 다른 변수들과의 관계를 표현하는 관계 시각화 기법으로도 표현 가능
- 변수들 간의 관계 분석을 위해서는 주로 히트맵(Heat Map)과 산점도(Scatter Plot)를 많이 활용함



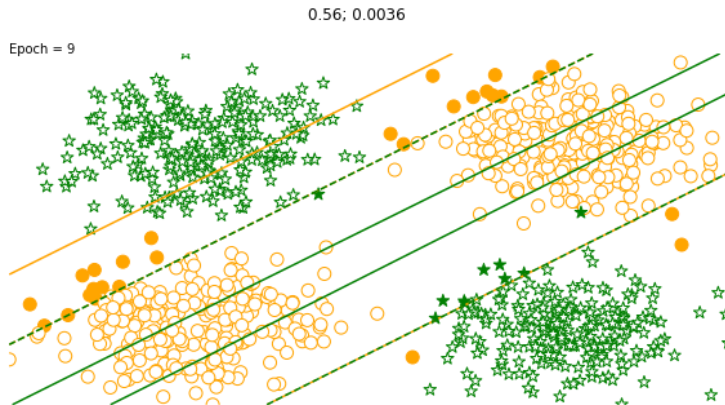
단순선형회귀분석의 히트맵 시각화 예시



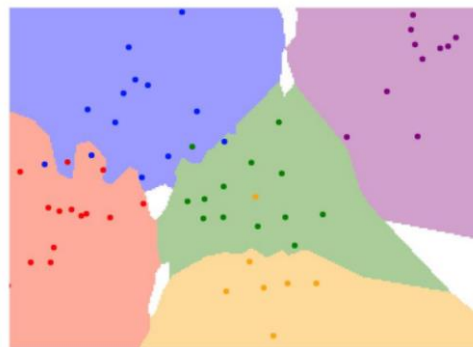
다항회귀분석의 산점도 시각화 예시

• 분류 모델

- SVM: 산점도와 구분선을 통한 비교 시각화 기법으로 활용 범위와 영역을 구분함
- KNN: 비교 시각화의 평행 좌표계로써 변수들과의 연관성 및 그룹 데이터의 경향성을 파악함
- 의사결정나무: 관계 시각화 기법의 트리 다이어그램으로 시각화

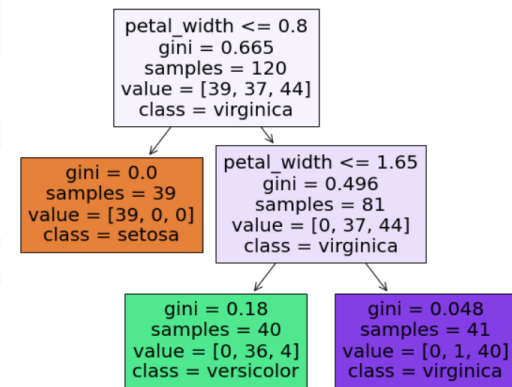


SVM 시각화 예시



K = 3

KNN 시각화 예시



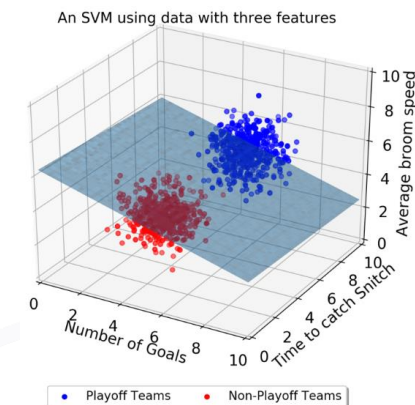
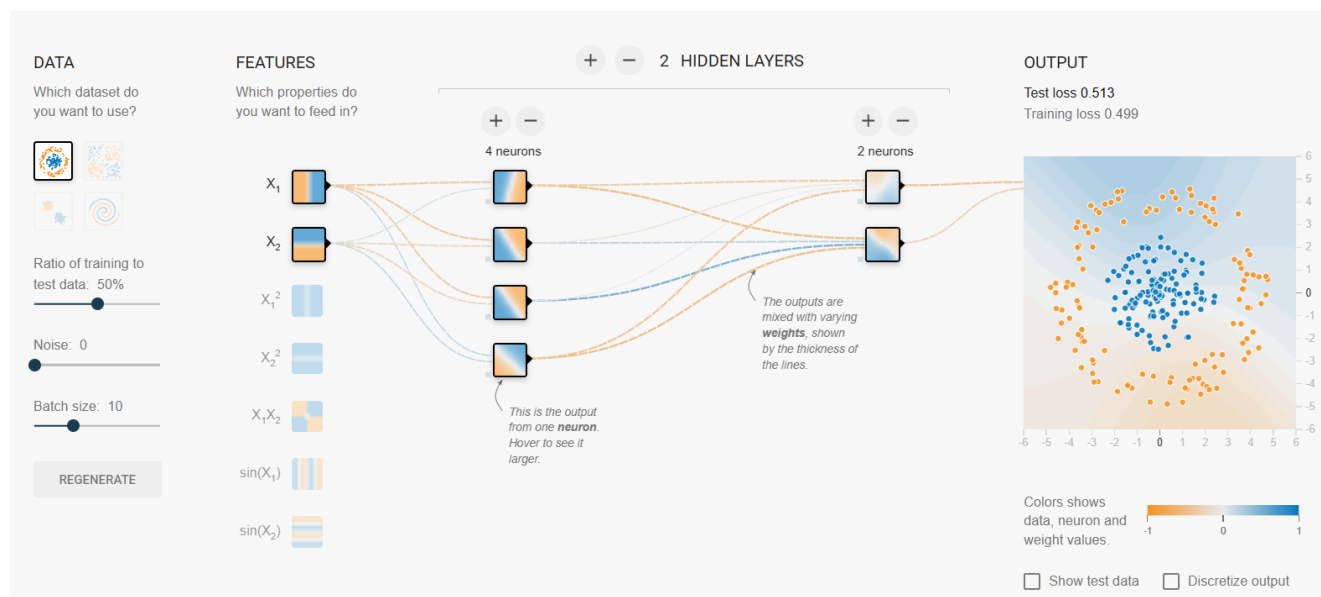
의사결정나무 시각화 예시

• 딥러닝 모델

- 딥러닝 모델은 모델 아키텍처에서 파라미터, 가중치 시각화 및 특징 차원감소를 통해 시각화를 사용
 - 네트워크 구조를 위한 노드-링크 다이어그램(Node-Link Diagrams for Network Architectures)
 - 뉴런과 연결가중치를 표현하는 노드로 시각화 수행
 - 차원 축소(Dimensional Reduction), 산포도(Scatter Plots)
 - t-SNE 또는 PCA를 이용하여 차원을 축소함으로써 데이터를 2차원으로 표현할 수 있음
 - 측정을 위한 선도표(Line Charts for Temporal Metrics)
 - 진행상황에 따른 결과를 선도표로 표현
 - 객체 기반 분석 및 탐색(Instance-based Analysis & Exploration)
 - 객체 그룹을 분석하고 분류 정확도를 확인함

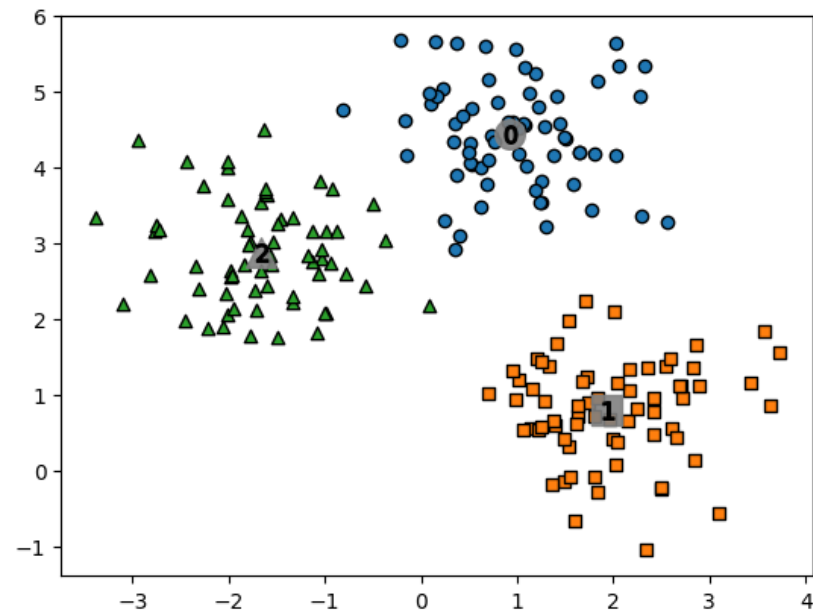
• 딥러닝 모델

• 시각화 예시



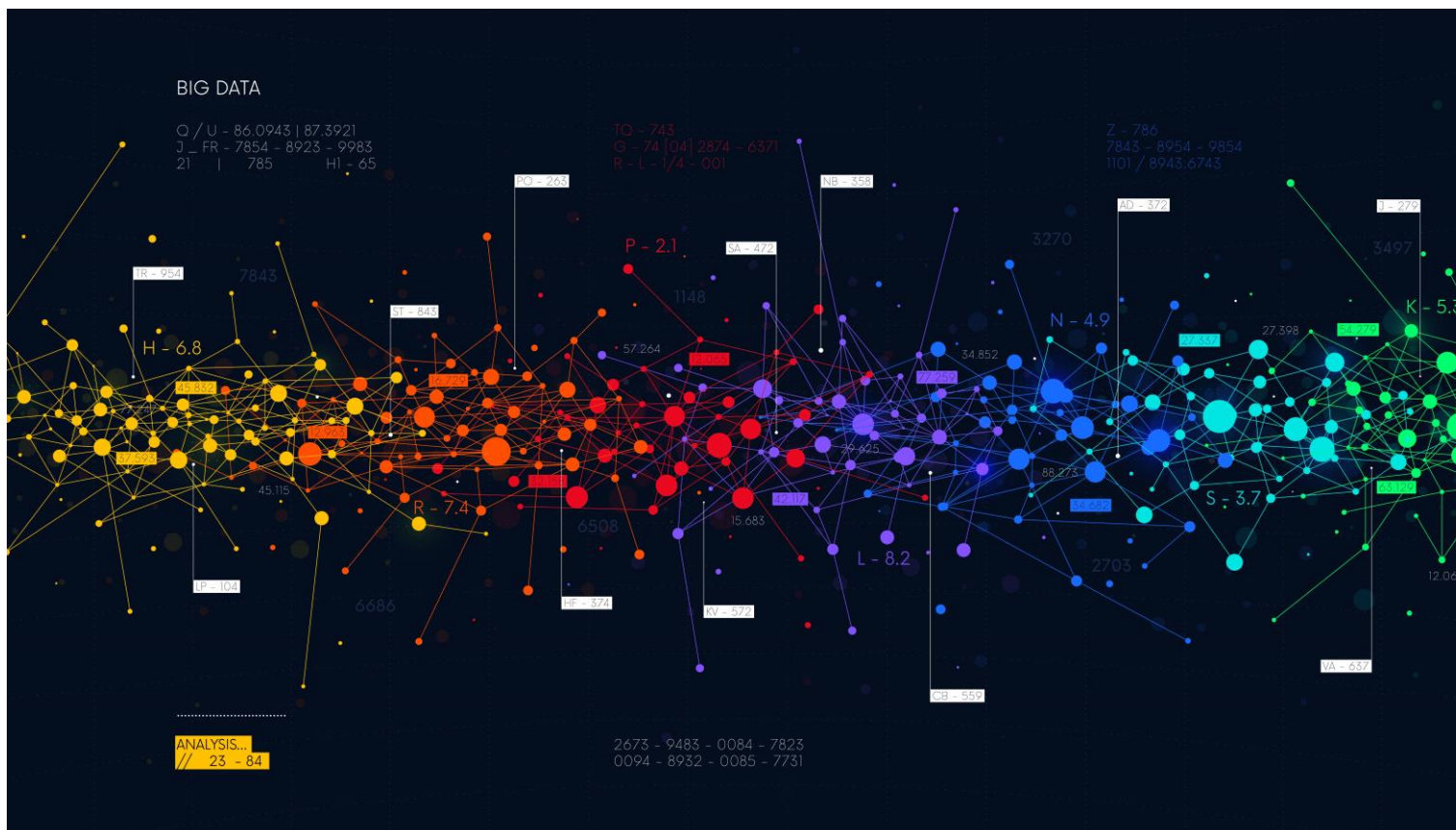
- 군집분석 모델

- 주로 그룹 클러스터 별 단위로 산점도를 이용하여 시각화



- **연관분석 모델**

- 각 연관규칙 별로 연관성 있는 항목(item)끼리 묶은 후, 관계시각화 기법인 네트워크 그래프를 활용하여 시각화를 수행



**THANK
YOU**

