**< RFM 기법과 K-means 알고리즘을 이용한 고객 분류 >**

-> k-means 알고리즘의 작동방식 이해 & 설명

**1. k-means 알고리즘 이란**

주어진 데이터를 k개의 클러스터로 묶는다. 초기 k개의 중심점을 잡고 각 데이터와 중심점 사이의 거리를 계산하여 해당 데이터에서 가장 가까운 클러스터를 찾아 데이터를 배당한다.

k-means 알고리즘에서는 최적의 k값이 존재하지 않는다. 데이터에 따라 k값이 달라지게 되며 클러스터링 결과를 통해 최적의 k값을 찾아야 한다.

**2. 클러스터링 평가방법**

1) 내부평가 -> Silhouette coefficient & Calinski Harabaz

클러스터 내의 밀집도와 클러스터 간의 분포를 통해 평가

라벨이 없는 데이터 셋에 대한 클러스터링 결과 평가 가능

1-1) Silhouette coefficient 계산식

1-2) Calinski Harabaz 계산식 -> 계산속도가 빠르다는 장점

2) 외부평가

라벨이 존재하는 데이터의 경우 라벨을 제외하고 클러스터링 알고리즘을 통해 결과를 라벨과 비교하여 알고리즘 평가

**< 의류 영상에서 컬러 영상 척도를 이용한 감성 인식 >**

**1. k-means 클러스터링**

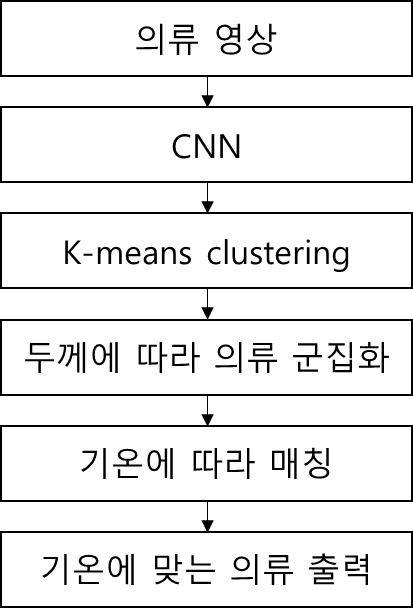
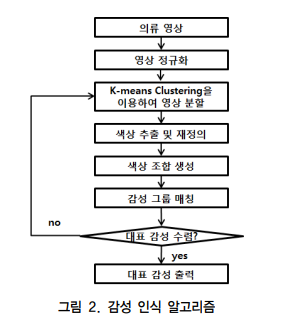
- k-means 알고리즘은 초기 중심점으로부터 클러스터의 경계선을 새로 도출하고, 그 후 새로운 클러스터의 중심점을 재계산한다. 이런한 과정을 반복하되, 만약 클러스터의 경계가 바뀌지 않거나, 중심점이 변하지 않으면 반복을 중지한다.



C(x,y): 점 (x,y)에 있는 데이터 값

: k번째 클러스터의 중심값

**2. 시스템 전체 과정(알고리즘)**

 우리 주제

**3. k-means 클러스터링 자세한 과정**

- 클러스터 개수 : 태그 개수? -> 기온에 따라