



ISSN: 3022-5388

JKAI website: <https://accesson.kr/jkaia>DOI: <http://doi.org/10.24225/jkaia.2024.2.1.1>

# 클러스터링을 이용한 스마트폰 사용자 추천 시스템 만들기

## Creating a Smartphone User Recommendation System Using Clustering

Jin Hyoung AN<sup>1</sup>

Received: November 7, 2023. Revised: November 13, 2023. Accepted: November 14, 2023.

### Abstract

In this paper, we develop an AI-based recommendation system that matches the specifications of smartphones from company 'S'. The system aims to simplify the complex decision-making process of consumers and guide them to choose the smartphone that best suits their daily needs. The recommendation system analyzes five specifications of smartphones (price, battery capacity, weight, camera quality, capacity) to help users make informed decisions without searching for extensive information. This approach not only saves time but also improves user satisfaction by ensuring that the selected smartphone closely matches the user's lifestyle and needs. The system utilizes unsupervised learning, i.e. clustering (K-MEANS, DBSCAN, Hierarchical Clustering), and provides personalized recommendations by evaluating them with silhouette scores, ensuring accurate and reliable grouping of similar smartphone models. By leveraging advanced data analysis techniques, the system can identify subtle patterns and preferences that might not be immediately apparent to consumers, enhancing the overall user experience. The ultimate goal of this AI recommendation system is to simplify the smartphone selection process, making it more accessible and user-friendly for all consumers. This paper discusses the data collection, preprocessing, development, implementation, and potential impact of the system using Pandas, crawling, scikit-learn, etc., and highlights the benefits of helping consumers explore the various options available and confidently choose the smartphone that best suits their daily lives.

**Keywords :** Unsupervised Learning, Clustering, K-MEANS, DBSCAN, Hierarchical Clustering

**Major Classification Code:** Artificial Intelligence, etc

## 1. Introduction

오늘날 시장에서 참여하고 있는 소비자들은 자신이 구매하고자 하는 제품에 관하여, 그 제품이 가지고 있는 특성과 성능, 그리고 그와 관련된 다양한 정보를 종합적으로 고려하여 깊이 있는 평가와 신중한 선택을 해야 한다는 것이 일반적인 인식으로 되어있다 (Yang, 2020 & Seo,2020).

이러한 복잡한 의사결정 과정은, 인터넷의 혁신적인 발달이 정보 접근성을 향상하게 시키면서도, 동시에 소비자들이 처리해야 할 정보의 양을 대폭 늘려 소비자의 부담을 가중했다 (Ha, 2007). 제품들 사이의 품질 차이가 점차 줄어들고, 유사한 기능을 지닌 제품들이 시장에 넘쳐나는 현상은 소비자들이 어떠한 제품을 최종적으로 선택해야 할지에 대한 결정을 내리는 데 더욱 어려움을 겪는다 (Lee,

<sup>1</sup> First Author. Undergraduate Student, Medical It, Bio-convergence, Eulji University, South Korea. Email: [wsgud2075@g.eulji.ac.kr](mailto:wsgud2075@g.eulji.ac.kr)

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

2023 & Lim, 2023). 이러한 상황에서 인공지능(AI) 기반의 추천 시스템은 그 중요성이 부각되고 있다. AI 추천 시스템은 방대한 데이터를 신속하게 처리하고, 소비자의 행동, 선호도, 구매 이력 등에서 복잡한 패턴을 인식하여 이를 바탕으로 개인화된 추천을 제공함으로써 소비자에게 더 만족스러운 제품을 발견할 수 있다. 또한 AI는 지속적인 학습을 통해 변화하는 소비자의 선호와 시장 동향을 반영한 추천을 할 수 있으며, 이는 소비자가 정보를 검색하고 평가하는 데 드는 시간을 줄여 주어 기회비용을 절감하게 만든다. 특히 스마트폰 시장은 이러한 AI 기반 추천 시스템의 필요성을 더욱 뚜렷하게 드러낸다. 스마트폰은 그 사용성이 다양하고 기술적 사양이 빠르게 변화하며, 사용자의 개인적인 취향과 생활 방식을 크게 반영하는 제품이다. 'S'사는 이러한 시장의 특성을 반영하여 다양한 성능, 디자인, 가격을 갖춘 스마트폰을 제공하고 있다. 각 모델은 독특한 스펙을 가지고 있어 사용자의 세분화된 요구를 충족시킬 잠재력이 크지만, 이로 인해 소비자는 어떤 모델이 자신의 필요와 예산에 가장 적합한지를 결정하기 어려울 수 있다. 이에 본 논문은 'S'사의 스마트폰 제품 스펙을 기반으로 한 AI 추천 시스템을 개발하여, 소비자가 이러한 복잡한 결정 과정을 보다 용이하게 넘어갈 수 있도록 지원하고자 한다. 이 시스템은 소비자가 일상생활에서 가장 적합한 스마트폰을 선택하는 데 도움을 주어, 정보 검색에 소비되는 시간을 절약하고 선택의 만족도를 극대화하는 것을 목표로 한다.

## 2. Related Researches

추천 시스템에서 비지도 학습의 사용은 사용자의 명시적인 피드백이나 레이블이 지정된 학습 데이터가 제한적일 때 특히 유용하다. 스마트폰 사용자와 같이 다양한 선호도를 가진 대규모의 사용자 집단을 다룰 때, 비지도 학습 방식은 레이블이 없는 사용자 데이터에서 숨겨진 패턴과 구조를 발견하는 데 강점을 가지고 있다. 이는 사용자의 기존 행동 데이터만을 바탕으로 사용자의 선호도를 추론하고, 그들이 아직 인지하지 못한 새로운 스마트폰 모델이나 기능에 대한 관심을 예측할 수 있게 한다. 클러스터링은 이러한 비지도 학습의 한 형태로, 새로운 사용자 데이터가 수집될 때 기존 클러스터 구조를 유지하면서 변화하는 사용자 행동을 신속하게 반영할 수 있는 유연성을 제공한다. 이는 빠르게 변화하는 스마트폰 시장의 소비자 트렌드와 신기술에 적응해야 하는 상황에서 중요하다. 클러스터링을 통해, 시스템은 사용자의 현재

선호도뿐만 아니라 시간에 따라 변화하는 선호도도 포착할 수 있으며, 이를 통해 동적이고 개인화된 사용자 경험을 제공한다. 본 연구에서 클러스터링 기법의 이러한 유연성을 활용하여 사용자의 동적인 선호도를 반영할 수 있는 추천 시스템을 개발하고자 한다. 이 시스템은 변화하는 사용자의 요구와 시장 트렌드를 반영하여 개인에게 맞춤형 스마트폰 추천을 제공함으로써, 소비자의 의사 결정 과정을 간소화하고 그들의 만족도를 극대화할 것을 기대한다.

### 2.1. K-MEANS

K-MEANS 클러스터링은 주어진 데이터를 'K'개의 클러스터로 나누는 알고리즘이다. 이 방법은 각 클러스터의 중심 (centroid)을 계산하고, 각 데이터 포인트를 가장 가까운 중심에 할당함으로써 작동한다. 클러스터의 중심은 할당된 포인트들의 평균 위치로 정의되며, 알고리즘은 중심과 데이터 포인트 간의 거리의 제곱합이 최소화되는 방향으로 클러스터의 중심을 반복적으로 업데이트한다. K-MEANS는 클러스터의 수가 사용자에게 의해 사전에 결정되어야 하며, 각 클러스터는 비슷한 크기의 분산을 가정한다.

### 2.2. DBSCAN

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)은 밀도 기반 클러스터링을 수행하여 밀집도가 높은 영역을 클러스터로 식별하고, 밀도가 낮은 영역을 노이즈로 분류하는 알고리즘이다. 이 방법은 각 데이터 포인트를 중심으로 정의된 반경 내에 위치한 이웃의 수에 따라 그 포인트가 클러스터의 핵심 포인트, 경계 포인트, 또는 노이즈인지를 결정한다. 핵심 포인트들이 서로 연결되어 클러스터를 형성하며, 이 알고리즘은 클러스터의 형태나 크기에 제한을 두지 않는다.

### 2.3. Hierarchical Clustering

Hierarchical Clustering은 데이터를 중첩되지 않는 계층 구조로 조직하는 알고리즘이다. 이 방법은 데이터 포인트 간의 유사성에 기반하여 각각의 포인트를 개별 클러스터로 간주하고, 가장 유사한 클러스터끼리 병합해 나가는 방식으로 진행된다. 이 과정은 덴드로그램이라 불리는 트리 형태의 다이어그램으로 시각화될 수 있으며, 사용자는 이 트리를 특정 높이에서 자르는 것으로 원하는 수의

클러스터를 얻을 수 있다. 계층적 클러스터링은 거리 또는 유사도 측정 방법에 따라 다양한 결과를 생성할 수 있으며, 미리 클러스터의 수를 결정할 필요가 없는 장점이 있다.

## 2.4. Clustering Performance Evaluation

클러스터링의 성공을 평가하기 위해서는 클러스터 내의 응집도와 클러스터 간의 분리도를 측정하는 다양한 지표들이 사용된다. 응집도는 클러스터 내의 데이터 포인트들이 얼마나 서로 가까운지를 나타내는 지표이며, 분리도는 서로 다른 클러스터들이 얼마나 잘 구분되는지를 나타낸다. 이러한 평가 지표들은 클러스터링 알고리즘의 설정을 조정하고, 최적의 클러스터링 결과를 도출하는 데 도움을 준다.

## 2.5. Silhouette Score

Silhouette Score 는 클러스터링의 효과를 평가하는 지표로, 클러스터 내의 데이터 포인트들이 얼마나 밀접하게 모여 있는지 (응집도), 그리고 다른 클러스터와는 얼마나 잘 구분되는지 (분리도)를 측정한다. 스코어는 각 데이터 포인트에 대해 계산되며, 해당 포인트가 자신이 속한 클러스터 내의 다른 포인트들과의 평균 거리 (응집도)와 가장 가까운 클러스터까지의 평균 거리 (분리도)를 비교한다. Silhouette Score 에서는 -1 에서 +1 사이의 값을 가지며, +1 에 가까우면 클러스터 내의 포인트가 서로 가깝고 다른 클러스터와는 멀리 떨어져 있다는 것을 의미한다. 값이 0 에 가까우면 클러스터 간의 구분이 명확하지 않음을, -1 에 가까우면 클러스터 할당이 부적절함을 나타낸다. 전체 데이터 셋에 대한 Silhouette Score 의 평균값은 클러스터링의 전반적인 적합도를 반영한다.

## 3. Optimization of Recommender System

추천 시스템의 성공은 데이터의 질과 크게 의존한다. 이는 사용자 경험을 향상하게 시키고, 개인화된 제안을 제공하여 고객 만족도를 높이는 데 핵심적인 역할을 한다. 데이터 수집 단계에서는, 신뢰할 수 있고 포괄적인 데이터를 확보함으로써, 시스템이 더 정확한 추천을 할 수 있는 토대를 마련한다. 데이터 전처리 단계에서는 수집된 데이터 전처리 단계에서는, 수집된 데이터를 클러스터링 알고리즘에 적합한 형태로 정제하고 변환하는 작업이 이루어진다. 이

과정은 불필요한 변수를 제거하고, 결측치를 처리하며, 중복을 배제하여 데이터의 정확성을 보장한다. 또한 연속형 데이터를 범주형 데이터로 변환함으로써, 더 복잡한 사용자의 선호와 요구를 보다 잘 이해하고, 그에 따른 추천을 할 수 있도록 한다.

## 3.1. Data Collection

본 연구에서 사용된 'S'사의 스마트폰의 제품 스펙은 2014년부터 제조된 제품 스펙부터 2023년 9월까지 출시된 제품 스펙까지 'S'사의 공식 홈페이지에서 웹 크롤링하여 데이터를 수집하였다 (Yoo, 2023 & Jo, 2023; Yoo, 2023). 가격 부분에서는 'S'사의 뉴스룸을 통해 직접 수집하였다. 크롤링은 Python 라이브러리인 BeautifulSoup 을 활용하였고 Pandas 와 Numpy 를 사용하여 하나의 데이터 프레임으로 병합하여 실험 데이터 셋을 구축하였다 (Yoo, 2023; Jo, 2023; Yoo, 2023).

## 3.2. Data Preprocessing

수집한 데이터 중에서 색상만 다르고 같은 기종일 경우는 색상을 제외한 동일한 스펙 부분과 결측값을 제거하여 최종으로 사용할 데이터 셋을 만들었고 클러스터링에 학습에 사용된 최종 변수는 Table 1 과 같다 (Jeong, 2021).

**Table 1: Variants used for Clustering**

Feature	Meaning
Price	Released Price(Won)
Battery	Battery Capacity(mAh)
Weight	Product Weight(g)
Camera Resolution	Camera Resolution(MP)
ROM	Product Storage(GB)

'Price'는 해당 제품의 출고가를 나타내는 데이터다. 'Battery'는 해당 제품의 총 배터리 용량을 나타낸 데이터이다. 'Weight'는 스마트폰의 무게를 나타낸다. 'Camera Resolution' 스마트폰의 전면 카메라 화소를 나타내는 데이터이다. 마지막 'ROM'은 스마트폰의 저장 용량을 나타내는 데이터이다. 모든 변수들은 연속형 데이터의 속성을 가지고 있어서 이 속성을 범주화를 통해 연속형 데이터에서 범주형 데이터로 변환하는 과정은 Table 2 와 같다.

**Table 2:** Process of converging categorical data from continuous data

Feature	Continuous	Categorical
Price	100,000 <= Price < 300,000	1
	300,000 <= Price < 500,000	2
	500,000(Won) <= Price < 700,000	3
	700,000 <= Price < 900,000	4
	900,000 <= Price	5
Battery	Battery < 3,000	1
	3,000 <= Battery < 3,500	2
	3,500 <= Battery < 4,000	3
	4,000 <= Battery < 4,500	4
	4,500 <= Battery	5
Weight	Weight < 160	1
	160 <= Weight < 200	2
	200 <= Weight	3
Camera Resolution	Camera Resolution < 10	1
	10 <= Camera Resolution < 20	2
	20 <= Camera Resolution	3
ROM	ROM <128	1
	128 <= ROM <= 256	2
	256 < ROM	3

'Price', 'Battery', 'Weight', 'Camera Resolution', 'ROM'의 변수들을 특정 구간으로 나눔으로써, 각범주 내의 데이터 포인트들은 유사한 특성을 공유하게 되어, 클러스터 분석에서 의미 있는 그룹을 형성할 수 있다. 이는 클러스터링 알고리즘의 식별 능력을 향상하게 시키고 정확한 패턴과 경향을 더 잘 포착할 수 있게 해주며, 결과적으로 사용자에게 개인화된 추천을 제공하는 시스템의 성능을 향상하게 시킨다.

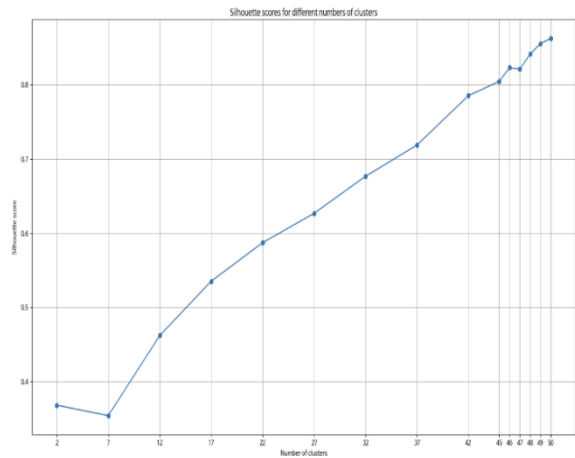
**3.3. For Each Clustering used in the Experiment, Hyperpareters**

Table 3 은 각 클러스터링마다 최적의 하이퍼파라미터에 관한 내용이다. 하이퍼파라미터를 획득 방법은 Silhouette Score 를 통해 획득했다. K-MEANS 는 2 에서 50 까지 다양한 클러스터 수에 대해 클러스터링을 수행하고 각각에 대한 Silhouette Score 를 계산한 결과 클러스터 수 50 개가 0.8625 로 가장 높게 나왔다. DBSCAN 은 'eps'(이웃을 정의하는 거리 임계값)와 'min\_samples'(핵심 포인트를 정의하는 이웃의 최소 수)에 대한 최적값을 찾는 과정은 eps 값을 0.1 부터 2.9 까지 min\_samples 값을 2 부터 14 까지

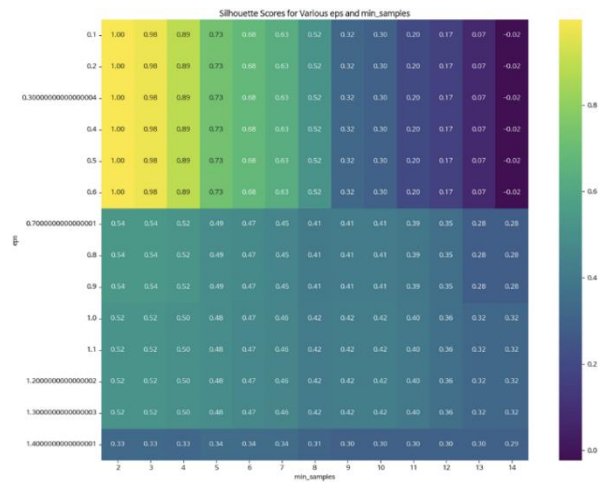
변화시키면서 실험을 수행한 결과 eps 는 0.1 ~ 0.6, min\_samples 는 2 에 대한 Silhouette Score 점수 0.9999 로 가장 높게 나왔고 노이즈 포인트가 1 개 검출되었다. Hierarchical Clustering 에서는 클러스터 수를 2 부터 100 까지 변화시키면서 각 클러스터링 결과에 대해 Silhouette Score 를 평가한 결과 69 개의 클러스터의 수가 Silhouette Score 0.9981 로 가장 높게 나왔다.

**Table 3:** Hyperparameters for each Clustering

Model	Hyperparameter
K-MEANS	n_clusters = 50
DBSCAN	eps = 0.1 ~ 0.6, min_samples = 2
Hierarchical Clustering	n_clusters = 69



**Figure 1:** Silhouette Score for Hyperparameters of K-MEANS



**Figure 2:** Silhouette Score for Hyperparameters of DBSCAN

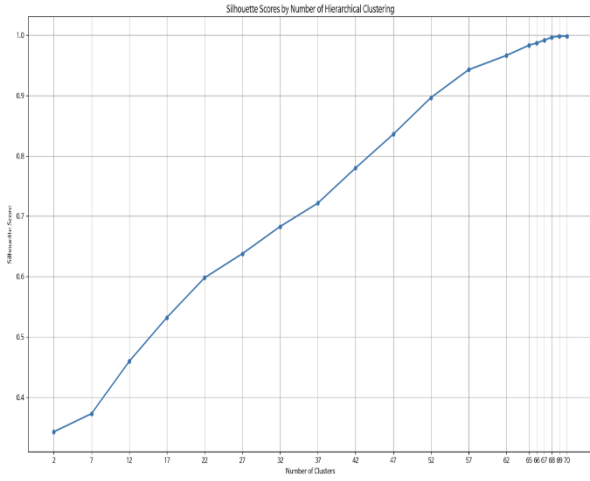


Figure 3: Silhouette Score for Hyperparameters of Hierarchical Cluster

## 4. Experimental Results and Performance Evaluation

### 4.1. Experiment Environment

본 실험에서 결과와 사용된 클러스터링 모델은 Google Colab 환경에서 Pandas, Numpy, scikit-learn 등의 라이브러리를 이용하여 구현하였다.

### 4.2. Actual Test Results and Performance Evaluation

각 클러스터링 모델을 사용하여 'S'사의 스마트폰 제품 스펙, 가격을 활용하여 사용자가 출고가, 배터리 용량, 무게, 카메라 화질, ROM 을 직접 선택하여 추천 결과가 나오도록 설정했다. 추천 결과는 다음과 같은 3 가지 조건이 있다. 첫 번째 사용자가 선택한 5 가지 특성에 모두 맞는 추천 스마트폰 상품을 출력한다. 두 번째 추천 스마트폰 상품이 5 개 미만일 경우 4 가지 특성에 맞는 추천 스마트폰 상품이 출력된다. 세 번째 1 순위, 2 순위 추천 스마트폰 상품 개수를 합친 개수가 5 개 미만인 경우 1 순위, 2 순위, 3 순위 상품과 함께 3 가지 특성에 맞는 3 순위 상품을 함께 출력한다. 3.3 에서 언급했던 DBSCAN 에서 노이즈 포인트가 한 개 검출된 제품 스펙과 가격의 경우만 제외하고 사용자가 어떤 사양을 선택을 하던 3 개의 클러스터링에 대한 추천 스마트폰 제품 출력은 동일하게 나왔다. 노이즈 포인트가 검출된 제품 스펙과 가격 부분은 'Price' 부분은 300,000 <=

Price < 500,000, 'Battery' 부분은 Battery < 3,000, 'Weight' 부분은 160 <= Weight < 200, 'Camera Resolution' 부분은 Camera Resolution < 10, 'ROM' 부분은 ROM <128 의 범위가 나왔다.

Priority	1순위	Galaxy Grand Max	최미트	310000	2500	161	CMOS 5.0 MP	16
Priority	기종명 (모델명)	색상	출고가	배터리 용량(mah, typical)	무게(g)	전면 카메라	· 해상	스토리지(저장 용량) (GB, TB)
2순위	Galaxy A5	분백	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	파이트	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	골드	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	핑크	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	분루	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	실버	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	분백	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	파이트	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	골드	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	핑크	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy J7	분백	369600	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy J7	파이트	369600	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy J7	골드	369600	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy Grand 2	분백	517000	2600	166	CMOS, 1.9 MP	16	
2순위	Galaxy Grand 2	파이트	517000	2600	166	CMOS, 1.9 MP	16	
2순위	Galaxy Grand 2	핑크	517000	2600	166	CMOS, 1.9 MP	16	
2순위	Galaxy Wide	파이트	319000	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy Wide	골드	319000	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy Wide	분백	319000	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy On7 (2016)	분백	399300	3300	167	CMOS 8.0 MP	16	
2순위	Galaxy On7 (2016)	골드	399300	3300	167	CMOS 8.0 MP	16	
2순위	Galaxy Folder2(4G LTE 모델)	분백	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(4G LTE 모델)	레드	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(3G WCDMA 모델)	분백	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(3G WCDMA 모델)	레드	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(3G WCDMA 모델)	골드	297000	1950	164	CMOS 5.0 MP	16	

Figure 4: Results of K-MEANS and Hierarchical Clustering applying DBSCAN noise points

Priority	기종명 (모델명)	색상	출고가	배터리 용량(mah, typical)	무게(g)	전면 카메라	· 해상	스토리지(저장 용량) (GB, TB)
2순위	Galaxy A5	분백	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	파이트	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	골드	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	핑크	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	분루	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A5	실버	484000	2300	123	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	분백	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	파이트	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	골드	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy A3 (2016)	핑크	352000	2300	132	CMOS 5.0 MP	16	
2순위	Galaxy J7	분백	369600	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy J7	파이트	369600	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy J7	골드	369600	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy Grand 2	분백	517000	2600	166	CMOS, 1.9 MP	16	
2순위	Galaxy Grand 2	파이트	517000	2600	166	CMOS, 1.9 MP	16	
2순위	Galaxy Grand 2	핑크	517000	2600	166	CMOS, 1.9 MP	16	
2순위	Galaxy Wide	파이트	319000	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy Wide	골드	319000	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy Wide	분백	319000	3000	171	CMOS 5.0 MP	16	
2순위	Galaxy On7 (2016)	분백	399300	3300	167	CMOS 8.0 MP	16	
2순위	Galaxy On7 (2016)	골드	399300	3300	167	CMOS 8.0 MP	16	
2순위	Galaxy Folder2(4G LTE 모델)	분백	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(4G LTE 모델)	레드	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(4G LTE 모델)	골드	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(3G WCDMA 모델)	분백	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(3G WCDMA 모델)	레드	297000	1950	164	CMOS 5.0 MP	16	
2순위	Galaxy Folder2(3G WCDMA 모델)	골드	297000	1950	164	CMOS 5.0 MP	16	

Figure 5: Results of DBSCAN applying noise points of DBSCAN

K-MEANS 와 Hierarchical Clustering 의 결과는 모든 제품을 포함한 스마트폰에 대해 각각 0.8625, 0.9981 의 높은

Silhouette Score 를 기록하며, 상대적으로 균일한 클러스터를 형성했다. 반면, DBSCAN 은 'Galaxy Grand Max'를 유일한 노이즈 포인트로 분류하였으며, Silhouette Score 는 0.9999 로 K-MEANS 와 Hierarchical Clustering 보다 우위를 보였다.

## 5. Conclusion and Future

본 연구는 'S'사의 스마트폰 제품 스펙, 가격을 활용하여 클러스터링을 활용한 스마트폰 사용자 추천 시스템을 목표로 하였다. 총 3 개의 클러스터링을 사용하여 Silhouette Score 를 통한 평가를 하였다. DBSCAN 은 특정 기준에 부합하지 않는 제품들을 섬세하게 구별해내는 능력을 보여주었다. 이러한 세밀한 분류 능력은 사용자에게 맞춤형, 더욱 정밀한 추천을 가능하게 함으로써, 사용자 만족도를 극대화할 잠재력을 내포하고 있다. 반면에, K-MEANS 와 Hierarchical Clustering 은 포괄적인 데이터 집합에 대해 넓은 범위의 옵션을 제공하는 데에 특화되어 있다. 이는 사용자들에게 다양한 대안을 제시함으로써 선택의 폭을 넓히고, 개인의 취향이나 요구를 넘어서는 다양한 선택지를 고려할 수 있게 한다. 본 연구의 한계점으로는 스마트폰 데이터에 대한 부족으로 인해서 다양한 변수를 넣을 수 없다는 점이다. 추후 연구에 서는 더 많은 스마트폰 데이터를 확보하여서 클러스터링 뿐만 다른 비지도 학습, 지도 학습을 이용해 예측, 강화 학습, 등등 다양한 연구로 확장되기를 기대한다.

## References

- Yang, J., & Seo, H. (2020). The influence of consumer attribute preference and purchase motivation on product choice. *Korean Journal of Management Education*, 30(6), 189-204.
- Ha, H. (2007). The influence of perceived difficulty of choice and anticipated regret on satisfaction and loyalty among internet shopping mall users. *Journal of Service Management*, 8(3), 85-107.
- Lee, H., & Lim, G. (2023). A study on hybrid music recommendation system utilizing music and playlist metadata. *Journal of Korean Intelligent Information Systems Society*, 29(3), 145-165.
- Yoo, J., Jo, S., & Yoo, S. (2023). Performance evaluation of machine learning algorithms for predicting bestseller books. *Journal of the Korea Information Technology Society*, 21(7), 1-6.
- Jeong, J. (2021, May 25). *Smartphone, the biggest complaint is battery life, 30%*. *ZDNet Korea*. Retrieved from <https://zdnet.co.kr/view/?no=20210522145609>