



팀명 : 아나바다  
제작 : 201311198 김용태  
DB 제작 : 201411099 김웅기  
201311159 이현우  
ppt 제작 : 201611115 최병권



# 파인더

(finder)

요즈음 국내의 수많은 쇼핑몰들이 소비자들에게 **상품을 추천**하는 방법은 **정확도**가 매우 낮은 경우가 많습니다.

**HOT 인기 멋남 아이템**

**Best** 주문복수

【London Slacks】  
팬츠 스타일 슬랙스ver.2  
53,000원  
32,000원  
리뷰 : 4045

주문복수

코지 오버핏 니트폴라  
34,000원  
리뷰 : 110

**Best** 주문복수

【London Slacks】  
베이직 히팅 치노팬츠  
31,000원  
리뷰 : 563

▲ 베스트 아이템 추천

**RECOMMEND ITEM : 이런 상품들도 구매했어요**

슬림블랙 허리밴딩팬츠  
36,000원

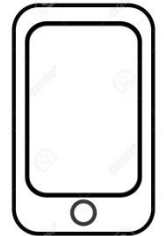
두루두루 울베니  
14,000원

주구잠참 블랙데님팬츠  
45,000원

▲ 이 상품을 구매한 다른 회원  
들이 구매한 상품 추천



RECOMMEND



pixtstock.com - 34767082

저희는 쇼핑몰 이용자들 개개인의 관심과 선호도에 따른  
상품 추천 알고리즘을 구현해 보고자 합니다.

# 프로젝트 목표

- 사용자(user)의 관심사와 취향 데이터를 수집, 분석해 데이터베이스 구축
- 그에 따른 추천 상품들을 도출하는 알고리즘 개발
- 사용자의 추천 상품을 추천 목록에 출력
- 쇼핑몰과 추천 상품을 연결
  - 이 부분은 네이버 open api를 통해 연결하였습니다.
- 이용 횟수에 따른 추천목록의 신뢰성이 증가

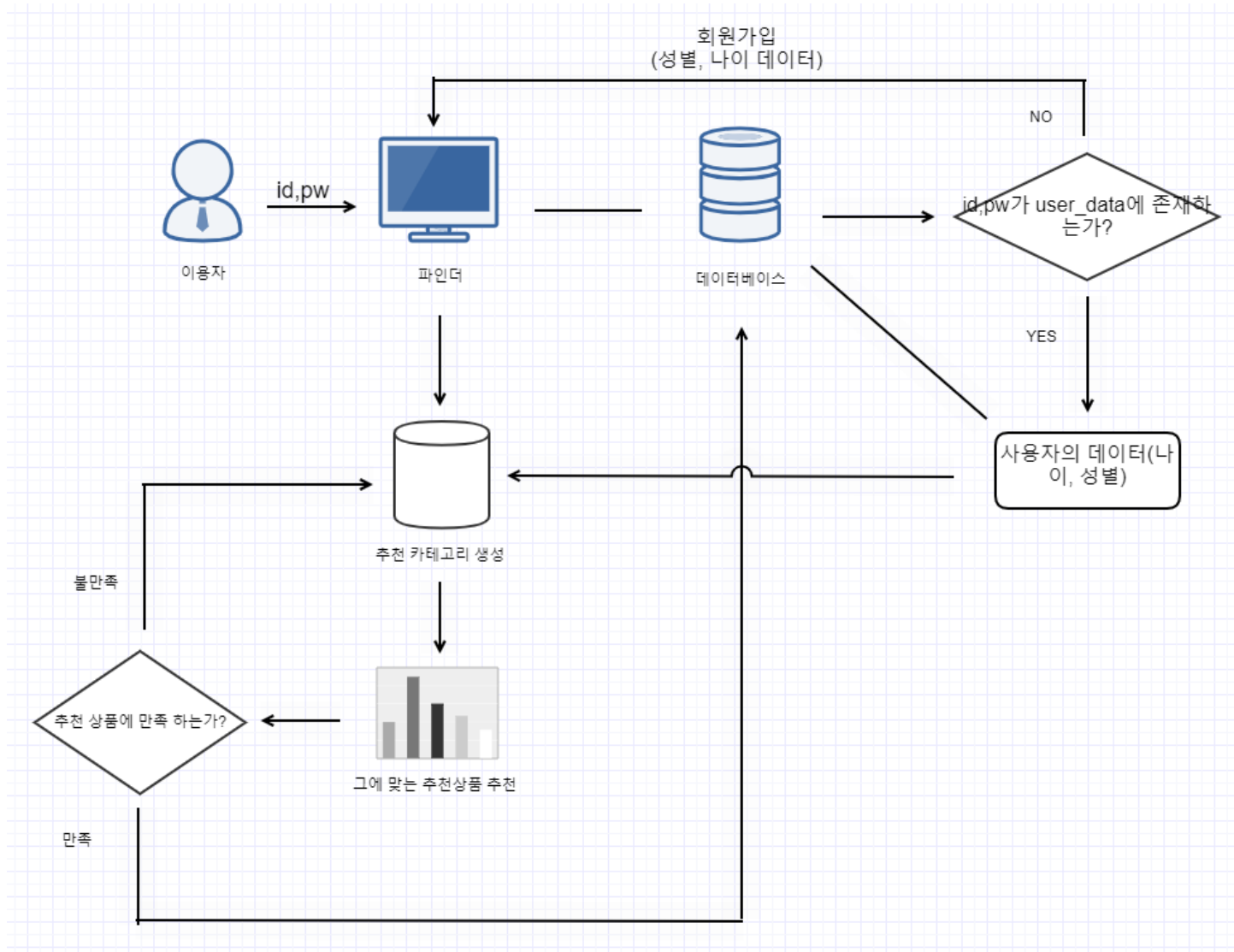


# 구현 방향

- 이용자의 선택에 따라 **가중치(선택횟수)**를 갱신, 신뢰성 높은 **데이터베이스** 구축
- **연관 규칙 마이닝과 협업 필터링** 기반, DB 특정 항목의 가중치와 이용자의 선택에 따라 **맞춤 추천 목록** 제공
- **개인화를 통한 신뢰성의 증가**
  - 사용자가 선택한 물품에 대한 **상관분석**, 1순위로 선택한 물품의 가중치가 높은 물건을 추천하여, **신뢰성**을 높이도록 하였습니다.



# 시스템 구성도



# 데이터베이스 구축

- 개인 프로필의 **성별,연령**에 따라 데이터를 구분
- 이용자의 **성별,연령대**에 따라 데이터들이 다른 **가중치(선택횟수)**를 가짐
- 알고리즘과 이용자들의 **선택**으로 인해 DB의 **가중치들이 갱신** 됨.
- **콜드 스타트(cold start)**를 방지하기 위해 첫 DB는 **통계적 자료**를 근거로 **연관수** 설정.
  - 통계청 2020년 2월 온라인쇼핑 동향조사
  - 2018 컨슈머인사이트 상품구입 및 행태 및 변화 추적조사
  - 네이버 데이터 랩



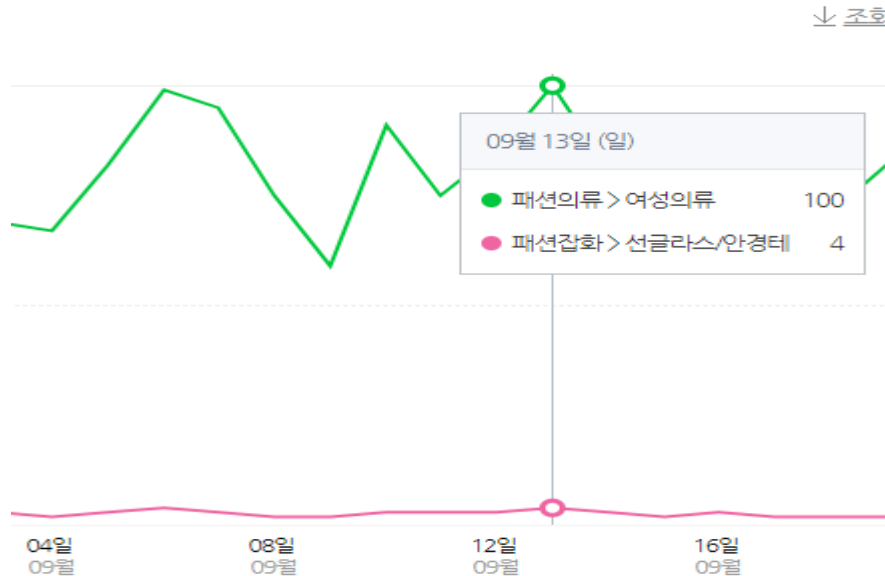
# 데이터베이스 구축

## 가중치 설정 근거:네이버 데이터 랩

쇼핑인사이드 검색결과가 의류 쪽이 제일 많았습니다.

각 성별 의류를 가중치 10000으로 삼고, 나머지는 비율을 보면서 결정하였습니다.

Ex) 패션의류가 10000이므로, 선글라스/안경테는 400입니다.





| id        | 가<br>종<br>치 | 여성<br>의류 | 여성언더웨어/<br>잠옷 | 남성<br>의류 | 남성언더웨어/<br>잠옷 | 여성<br>신발 | 남성<br>신발 | 신발<br>용품 | 여성<br>가방 | 남성<br>가방 | 여행용가방/<br>소품 | 지갑  | 벨트  | 모자  |
|-----------|-------------|----------|---------------|----------|---------------|----------|----------|----------|----------|----------|--------------|-----|-----|-----|
| 거실가구      | 190         | 3        | 3             | 3        | 3             | 3        | 3        | 3        | 3        | 3        | 3            | 3   | 3   | 3   |
| 광학기기/용품   | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 남성가방      | 480         | 9        | 9             | 9        | 9             | 11       | 14       | 11       | 11       | 0        | 14           | 14  | 11  | 11  |
| 남성신발      | 3400        | 65       | 65            | 65       | 65            | 81       | 0        | 89       | 86       | 86       | 86           | 81  | 81  | 81  |
| 남성언더웨어/잠옷 | 180         | 14       | 14            | 14       | 0             | 3        | 3        | 3        | 3        | 3        | 3            | 3   | 3   | 3   |
| 남성의류      | 10000       | 420      | 420           | 0        | 420           | 190      | 190      | 190      | 190      | 190      | 190          | 190 | 190 | 190 |
| 남성화장품     | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 네일케어      | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 노트북       | 1100        | 20       | 20            | 20       | 20            | 20       | 20       | 20       | 20       | 20       | 20           | 20  | 20  | 20  |
| 마스크/러     | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 모자        | 340         | 6        | 6             | 6        | 6             | 6        | 10       | 10       | 6        | 10       | 10           | 10  | 10  | 0   |
| 바디케어      | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 베이스메이크업   | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 벨트        | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 4   | 0   | 4   |
| 뷰티소품      | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 색조메이크업    | 70          | 1        | 1             | 1        | 1             | 1        | 1        | 1        | 1        | 1        | 1            | 1   | 1   | 1   |
| 생활/건강     | 8000        | 160      | 160           | 160      | 160           | 160      | 160      | 160      | 160      | 160      | 160          | 160 | 160 | 160 |
| 선글라스/안경테  | 140         | 2        | 2             | 2        | 2             | 2        | 2        | 2        | 2        | 2        | 2            | 9   | 9   | 9   |

## 구축한 데이터베이스(위) 근거한 자료(아래,우측)

| 구분    | 조사시점    | 구축한 데이터베이스(위) 근거한 자료(아래,우측) |           |          |          |          |                   |                  |                          |          |              |           |        |  |
|-------|---------|-----------------------------|-----------|----------|----------|----------|-------------------|------------------|--------------------------|----------|--------------|-----------|--------|--|
|       |         | 식품/<br>음료                   | 의류/<br>잡화 | 생활<br>용품 | 뷰티<br>용품 | 건강<br>식품 | 문파/<br>디지털<br>컨텐츠 | 가전/<br>디지털<br>기기 | 스포츠/<br>레저/<br>자동차<br>용품 | 여행<br>상품 | 가구/홈<br>인테리어 | 유아동<br>용품 | 기타     |  |
|       | 전체      | 77.9                        | 72.4      | 59.9     | 50.2     | 31.4     | 30.3              | 30.2             | 22.5                     | 19.1     | 12.7         | 8.1       | 10.8   |  |
| 성별    | 남성      | 73.2 ▽                      | 66.9 ▽    | 55.3 ▽   | 33.1 ▽   | 26.7 ▽   | 28.2 ▽            | 36.9 ▲           | 30.8 ▲                   | 17.2 ▽   | 11.7 ▽       | 8.7 ▲     | 8.3 ▽  |  |
|       | 여성      | 82.6 ▲                      | 78.1 ▲    | 64.7 ▲   | 67.9 ▲   | 36.3 ▲   | 32.5 ▲            | 23.2 ▽           | 14.1 ▽                   | 21.2 ▲   | 13.7 ▲       | 7.5 ▽     | 13.4 ▲ |  |
| 연령대   | 20대     | 70.1 ▽                      | 68.8 ▽    | 42.7 ▽   | 54.6 ▲   | 22.0 ▽   | 43.2 ▲            | 26.6 ▽           | 13.5 ▽                   | 18.8     | 10.9 ▽       | 3.3 ▽     | 12.0 ▲ |  |
|       | 30대     | 77.5                        | 73.2      | 61.7 ▲   | 54.7 ▲   | 29.3 ▽   | 33.4 ▲            | 33.8 ▲           | 24.6 ▲                   | 21.9 ▲   | 16.0 ▲       | 21.3 ▲    | 11.7   |  |
|       | 40대     | 82.0 ▲                      | 75.7 ▲    | 67.6 ▲   | 51.3     | 32.9 ▲   | 32.2 ▲            | 32.1 ▲           | 27.2 ▲                   | 18.2     | 14.5 ▲       | 9.2 ▲     | 10.0   |  |
|       | 50대 이상  | 79.2 ▲                      | 71.7      | 62.4 ▲   | 44.7 ▽   | 36.3 ▲   | 20.9 ▽            | 28.6 ▽           | 22.8                     | 18.4     | 10.5 ▽       | 2.2 ▽     | 10.3   |  |
| 성BY연령 | 남20대    | 65.2 ▽                      | 56.6 ▽    | 37.6 ▽   | 32.0 ▽   | 18.4 ▽   | 39.3 ▲            | 32.4 ▲           | 17.8 ▽                   | 15.0 ▽   | 8.7 ▽        | 2.9 ▽     | 8.6 ▽  |  |
|       | 남30대    | 73.3 ▽                      | 67.8 ▽    | 57.2 ▽   | 38.1 ▽   | 24.8 ▽   | 31.0              | 41.8 ▲           | 34.1 ▲                   | 19.6     | 15.2 ▲       | 23.2 ▲    | 8.4 ▽  |  |
|       | 남40대    | 77.5                        | 70.4 ▽    | 61.7     | 35.0 ▽   | 27.6 ▽   | 30.1              | 39.9 ▲           | 37.0 ▲                   | 17.3 ▽   | 14.1 ▲       | 10.8 ▲    | 7.3 ▽  |  |
|       | 남50대 이상 | 74.5 ▽                      | 69.6 ▽    | 59.2     | 29.3 ▽   | 31.5     | 19.3 ▽            | 34.5 ▲           | 31.5 ▲                   | 16.8 ▽   | 9.5 ▽        | 1.9 ▽     | 8.8 ▽  |  |
|       | 여20대    | 75.7 ▽                      | 82.4 ▲    | 48.3 ▽   | 79.8 ▲   | 26.0 ▽   | 47.6 ▲            | 20.1 ▽           | 8.6 ▽                    | 23.1 ▲   | 13.4         | 3.6 ▽     | 15.7 ▲ |  |
|       | 여30대    | 81.9 ▲                      | 78.8 ▲    | 66.3 ▲   | 72.1 ▲   | 33.9 ▲   | 35.9 ▲            | 25.5 ▽           | 14.6 ▽                   | 24.3 ▲   | 16.7 ▲       | 19.4 ▲    | 15.2 ▲ |  |
|       | 여40대    | 86.5 ▲                      | 81.1 ▲    | 73.7 ▲   | 68.1 ▲   | 38.4 ▲   | 34.4 ▲            | 24.1 ▽           | 17.2 ▽                   | 19.1     | 14.9 ▲       | 7.6       | 12.7 ▲ |  |
|       | 여50대 이상 | 83.8 ▲                      | 73.7      | 65.6 ▲   | 59.8 ▲   | 40.9 ▲   | 22.4 ▽            | 22.9 ▽           | 14.3 ▽                   | 19.9     | 11.4 ▽       | 2.6 ▽     | 11.8   |  |

|          | 가<br>방 | 패션용품<br>의류/신발 | 패션<br>스포츠<br>액세서리 | 화장품     | 아동<br>유아용품 | 식품      |         |        | 생활      |         |
|----------|--------|---------------|-------------------|---------|------------|---------|---------|--------|---------|---------|
|          |        |               |                   |         |            | 음·식재료   | 농축수산물   | 생활     | 생활용품    |         |
| 2017     | 19,245 | 20,284        | 32,016            | 91,172  | 33,872     | 104,216 | 79,970  | 24,246 | 74,636  | 87,228  |
| 2018     | 22,272 | 23,368        | 37,305            | 98,521  | 36,153     | 134,349 | 104,944 | 29,405 | 137,050 | 100,461 |
| 2019     | 26,072 | 25,396        | 41,833            | 123,822 | 40,053     | 169,629 | 134,287 | 35,342 | 157,777 | 100,461 |
| 2017.1/4 | 4,755  | 4,560         | 6,659             | 20,987  | 8,222      | 25,113  | 18,906  | 6,207  | 28,859  | 18,189  |
| 2/4      | 4,612  | 4,967         | 8,652             | 18,587  | 8,272      | 24,606  | 19,279  | 5,327  | 28,117  | 17,731  |
| 3/4      | 4,804  | 5,186         | 8,245             | 19,891  | 8,004      | 28,244  | 21,587  | 6,657  | 30,245  | 18,990  |
| 4/4      | 5,074  | 5,571         | 8,461             | 21,706  | 9,374      | 26,253  | 20,199  | 6,054  | 31,332  | 19,926  |
| 2018.1/4 | 5,589  | 5,838         | 7,153             | 23,489  | 8,627      | 32,588  | 24,884  | 7,705  | 32,847  | 20,858  |
| 2/4      | 5,526  | 5,838         | 9,544             | 23,935  | 9,032      | 31,217  | 24,703  | 6,514  | 33,351  | 21,137  |
| 3/4      | 5,496  | 5,737         | 9,648             | 24,306  | 8,369      | 35,414  | 27,721  | 7,693  | 33,493  | 21,469  |
| 4/4      | 5,661  | 5,956         | 10,961            | 26,791  | 10,125     | 35,130  | 27,637  | 7,493  | 37,359  | 23,764  |
| 2019.1/4 | 6,293  | 5,812         | 8,438             | 27,830  | 9,252      | 39,750  | 30,919  | 8,831  | 37,427  | 24,004  |
| 2/4      | 6,394  | 6,545         | 11,138            | 28,948  | 9,940      | 40,076  | 32,094  | 7,982  | 37,971  | 24,346  |
| 3/4      | 6,443  | 6,474         | 10,545            | 31,238  | 9,432      | 44,642  | 35,345  | 9,297  | 39,448  | 25,140  |
| 4/4      | 6,942  | 6,565         | 11,712            | 35,806  | 11,430     | 45,161  | 35,929  | 9,232  | 42,931  | 26,971  |
| 2018. 2  | 1,782  | 1,784         | 2,007             | 7,464   | 2,686      | 11,044  | 8,147   | 2,888  | 10,040  | 6,360   |
| 3        | 1,942  | 2,101         | 2,838             | 8,317   | 2,989      | 10,557  | 8,269   | 2,288  | 11,689  | 7,393   |
| 4        | 1,855  | 1,963         | 2,986             | 8,024   | 3,204      | 10,254  | 8,100   | 2,154  | 11,236  | 7,099   |
| 5        | 1,906  | 1,981         | 3,308             | 8,185   | 3,108      | 10,708  | 8,490   | 2,217  | 11,239  | 7,160   |
| 6        | 1,765  | 1,894         | 3,250             | 7,725   | 2,719      | 10,255  | 8,112   | 2,143  | 10,876  | 6,879   |
| 7        | 1,820  | 2,099         | 3,604             | 7,790   | 2,874      | 10,111  | 8,669   | 2,242  | 11,549  | 7,402   |
| 8        | 1,807  | 1,944         | 3,002             | 8,020   | 2,758      | 11,402  | 9,051   | 2,351  | 11,157  | 7,149   |
| 9        | 1,769  | 1,694         | 3,042             | 8,496   | 2,737      | 13,101  | 10,001  | 3,100  | 10,787  | 6,918   |
| 10       | 1,847  | 1,826         | 3,893             | 8,943   | 3,089      | 11,485  | 9,104   | 2,381  | 12,433  | 7,887   |
| 11       | 1,800  | 1,922         | 3,759             | 9,116   | 3,101      | 11,843  | 9,277   | 2,567  | 12,620  | 7,998   |
| 12       | 2,013  | 2,208         | 3,329             | 8,732   | 3,936      | 11,802  | 9,256   | 2,546  | 12,306  | 7,878   |
| 2019. 1  | 1,975  | 1,893         | 2,702             | 8,993   | 3,090      | 15,930  | 11,971  | 3,959  | 12,582  | 8,121   |
| 2        | 2,068  | 1,830         | 2,475             | 8,611   | 2,826      | 10,891  | 8,678   | 2,213  | 11,532  | 7,407   |
| 3        | 2,250  | 2,089         | 3,261             | 10,236  | 3,336      | 12,929  | 10,270  | 2,659  | 13,304  | 8,476   |
| 4        | 2,137  | 2,109         | 3,545             | 9,617   | 3,503      | 13,313  | 10,598  | 2,715  | 12,548  | 8,006   |
| 5        | 2,201  | 2,310         | 4,019             | 9,892   | 3,429      | 13,796  | 11,086  | 2,700  | 13,068  | 8,411   |
| 6        | 2,057  | 2,126         | 3,575             | 9,439   | 3,008      | 12,978  | 10,410  | 2,568  | 12,355  | 7,930   |
| 7        | 2,049  | 2,288         | 3,845             | 9,849   | 3,156      | 13,572  | 10,897  | 2,675  | 12,587  | 8,245   |
| 8        | 2,173  | 2,094         | 3,311             | 10,398  | 3,060      | 15,179  | 12,130  | 3,049  | 13,119  | 8,514   |
| 9        | 2,221  | 2,092         | 3,389             | 11,010  | 3,215      | 15,891  | 12,318  | 3,573  | 13,742  | 8,381   |
| 10       | 2,296  | 2,033         | 4,021             | 11,555  | 3,461      | 14,813  | 11,854  | 2,959  | 13,962  | 8,840   |

# 알고리즘 구현 내용(1)

- 첫 사용자의 **관심사와 취향**을 알아보기 위해 총 50개의 카테고리를 10개씩 5번 사용자에게 제시합니다.
  - 이때 제시되는 카테고리는 네이버 쇼핑 데이터랩을 활용하였습니다.
- 사용자는 제시된 50개의 카테고리에서 **총 5개**를 선택하게 됩니다.
- 50개의 카테고리를 모두 제시하는 이유는 **필터버블**을 방지하기 위해서 입니다.

# 필터버블이란?

- **필터 버블(filter bubble)**이란 사용자가 개인화 알고리즘에 의해 생기는 **정보편식 현상**으로 사용자가 제한된 주제의 제한된 정보만을 접하게 되는 것을 말한다.
- 저희는 필터버블을 방지하고자, 1~5번째의 선택에서는 **모든 카테고리**에서 선택하도록 하였습니다.

## 알고리즘 구현 내용(2)

- (1)에서 선택된 5가지의 카테고리, DB의 가중치를 기반으로 **지지도가 가장 높은 5가지** 카테고리, 총 **10개**를 제시합니다.
  - 지지도 : 특정항목의 가중치 / 가중치의 총 합
- 10개의 카테고리 중 선택한 카테고리는 1순위 카테고리가 됩니다.

## 알고리즘 구현 내용(3)

- **상관분석**을 이용하여, 1순위의 카테고리 and 선택된 카테고리들 간의 관계를 구해, 각 카테고리의 추측(신뢰도 $\times$ 지지도)를 구합니다.
  - 신뢰도 : 특정 항목과 연관된 품목의 가중치 / 특정항목의 가중치
- 상위 5개와 추측(신뢰도 $\times$ 지지도) 5개를 선택지로 제시합니다.
- 이 과정에서 선택된 카테고리는 **2순위 카테고리**입니다.

# 상관분석 이란?

- 상관분석은 변수들 간의 선형적 관계에 대한 분석입니다.
- 상관분석에는 피어슨 상관계수를 사용하였습니다.
- 선택되어진 5개의 상품의 상관계수를 구하고, 그에 따른 추측(지지도x신뢰도)의 총합 / 유사도의 총합의 값을 구하여, 이게 높은 5상품을 추천하였습니다.

- 피어슨 상관계수 구하는 식

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{(\sum X^2 - \frac{(\sum X)^2}{N})(\sum Y^2 - \frac{(\sum Y)^2}{N})}}$$

## 알고리즘 구현 내용(4)

- 알고리즘을 마치고, 사용자에게 **다시하기**와 **만족버튼**을 제시합니다.
- **다시하기**를 선택 시, 프로그램이 처음부터 동작합니다.
- **만족**을 선택시, 1순위와 2순위 카테고리가 **DB에 저장**됩니다.
  - 1순위의 가중치 값이 1 증가하고, 1순위의 2순위 연관숫자의 값이 1 증가합니다.

## 알고리즘 구현 내용(5)

- 만족하기를 선택시 UPDATE\_LIST 에 id, 테이블명, 1순위, 2순위 선호도 데이터를 삽입합니다.
- UPDATE\_LIST의 값은 프로시저에 사용됩니다.
- 프로시저가 끝나면 UPDATE\_LIST의 값을 모두 지워줍니다.
- 이벤트 스케줄러를 통해 프로시저를 하루에 한번 동작하도록 설정 해 프로그램 실행시 매번 데이터를 받아오는 단점을 보완했습니다.



# 그 후 로그인

- 이용자가 **다시 로그인시** 이용자에게 추천되는 상품은 다음과 같습니다
  - 현재의 1순위 카테고리
  - 1순위 카테고리 상품 중 신뢰도 1순위 카테고리
  - 1순위와 2순위 카테고리를 상관분석 후 나온 1순위 카테고리
- 선택지가 마음에 들지 **않다면, 다시하기**를 통해 처음과 같은 과정으로 돌아가게 됩니다.



# 마치며

- 아쉬웠던 점

- 실 거래 데이터를 구하지 못해, 연관수치가 다소 주관적이고 이 알고리즘의 정확도 판단을 못해본 점이 아쉽습니다.
- 상관분석의 경우, 원인-결과가 아닌 유사도만을 검사하는 분석이기 때문에, 추천의 정확도가 떨어지는 점이 아쉽습니다.

# 참고문헌

- 인터넷 쇼핑몰에서 구매품목을 이용한 고객의 예측모델 설계 지혜영1 · 조완현2 12전남대학교 통계학과 접수 2008년 12월 8일, 수정 2008년 12월 23일, 게재확정 2008년 12월 26일
- 네이버 데이터랩
- 통계청 2020년 2월 온라인쇼핑 동향조사
- 2018 컨슈머인사이트 상품구입 및 행태 및 변화 추적조사