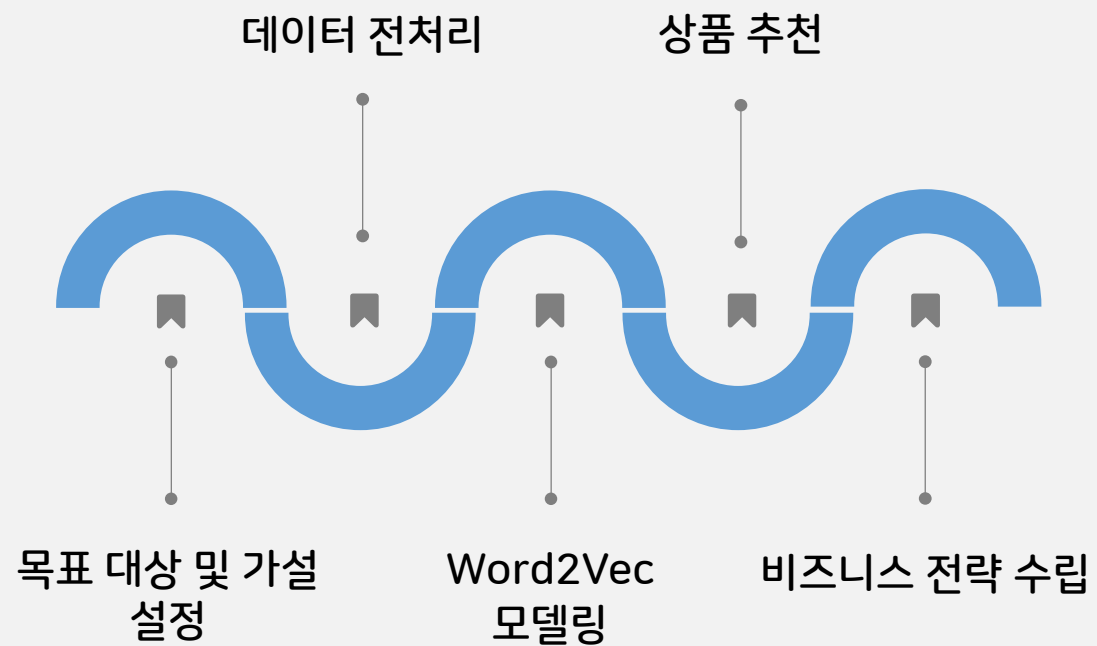


제6회 L.POINT
Big Data Competition

Audience Targeting

팀명: DnA

PROJECT STEPS



1. 목표 대상 및 가설 설정

잠재고객 설정

주제 1

Audience Targeting

(부제 : 잠재고객 맞춤 콘텐츠 추천)

- 세부과제

- 1) 비식별 온라인 잠재고객의 행동/소비 패턴 생성
- 2) AI 기반 고객 맞춤 상품 추천 알고리즘 도출
- 3) 1), 2) 를 활용한 비즈니스 전략 수립

가장 적절한 잠재고객은 누구일 것인가?



잠재고객이란?

현재는 고객이 아니지만 잠재적으로
고객이 될 가능성이 있는 구매자



구매를 한번도 하지 않은 고객을
잠재고객으로 설정

가설 설정

Action_Type

검색

제품 목록

제품 세부정보 보기

장바구니 제품 추가

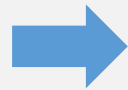
장바구니 제품 삭제

결제 시도

구매 완료

구매 환불

결제 옵션



책이나 컴퓨터에서,
목적에 따라 필요한 자료들을 찾아내는 일



온라인 쇼핑몰에서의 목적!! 제품 구매
목적에 대한 필요한 자료!! 제품



가설 설정

즉, **검색**은 곧 **구매**로 이어질 것이다!!

2. 데이터 전처리

데이터 전처리



데이터 분할

거래 정보의 clnt_id를 활용해
온라인 행동 정보 데이터에서
구매를 한 고객과 하지 않은
고객으로 나눔으로써
Word2Vec 모델에 대한
학습 데이터와 적용 데이터로
데이터 구분



크롤링

기존의 상품 분류 정보가
영문이었던 점과 **검색 키워드
(한글)와 구매 카테고리(영문)의**
의미 비교를 쉽게 확인하기
위해서 **네이버 파파고 크롤링**
진행으로 **구매 카테고리 번역**



단어 추출

데이터 셋 마다 clnt_id 별로
검색 및 구매에 대한 단어를
추출하고 각 clnt_id의 **검색과
구매를 하나의 문자열로** 제작
함으로써 Word2Vec 학습을
진행할 때 **동시발생 횟수가
많은 키워드에 대해 유사한
의미를 부여**

* 상품 분류 정보에 nan 값이 존재하고 온라인 행동
정보에서 검색 없이 구매하는 고객이 있어 데이터
클렌징을 진행하였습니다.

3. Word2Vec 모델링

Word2Vec 모델링



```
from gensim.models import Word2Vec
from gensim.utils import simple_preprocess

wv = Word2Vec(size=20,
              min_count=1,
              workers=4,
              sentences=[simple_preprocess(p) for p in y_text]
)
```

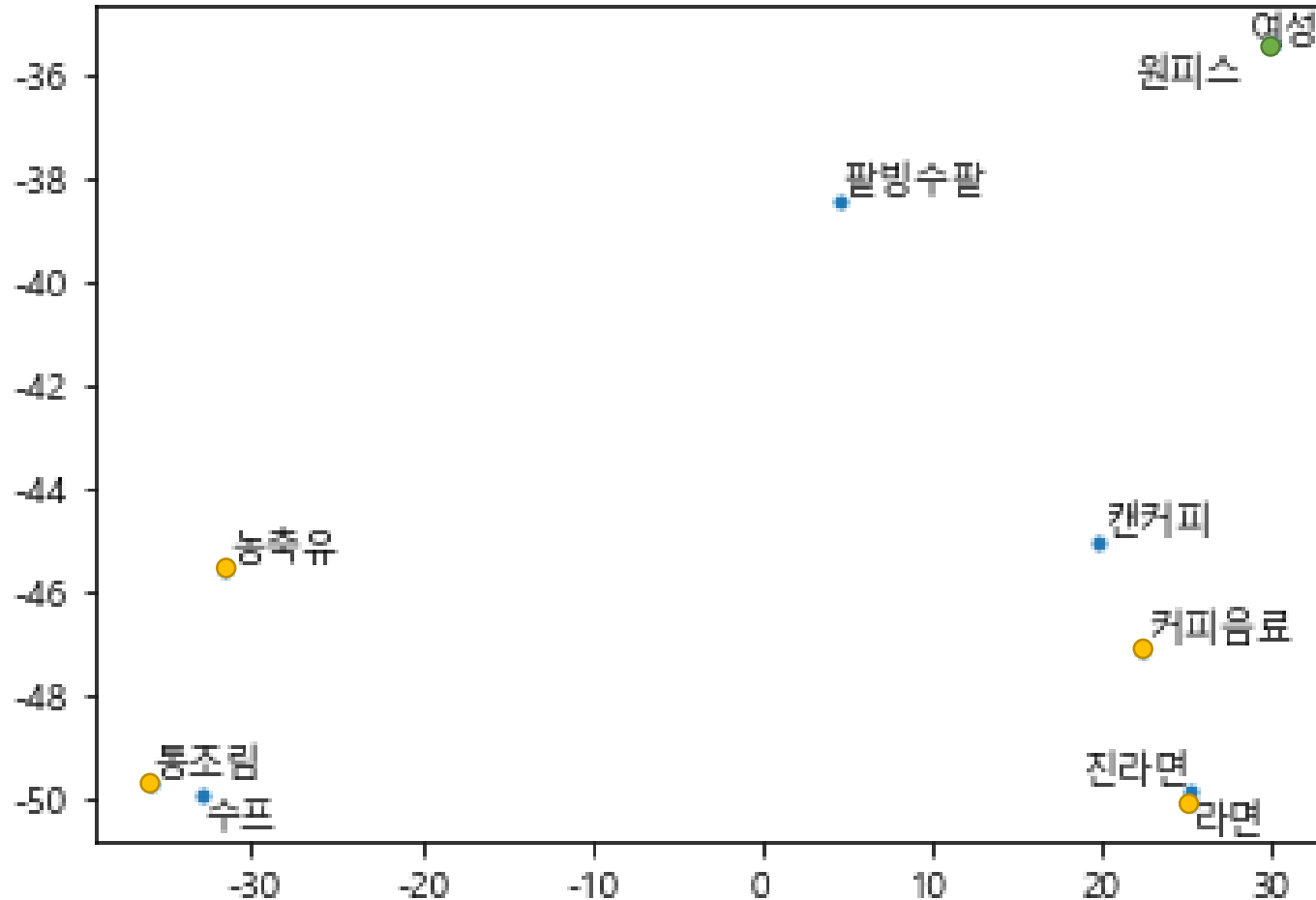
```
wv.wv['커피']
```

```
array([-2.287933 ,  0.6657695, -0.5850172,  1.4788895,  3.2032442,
        1.4812545, -1.3358833,  1.5167226,  0.9947786, -4.9584084,
       -2.6560576,  2.001303 ,  1.9721632, -1.560138 ,  4.424811 ,
        0.9674497, -0.56011146,  3.7695029, -0.4452537, -4.424564 ],
      dtype=float32)
```

```
wv.wv.most_similar('커피')
```

```
[('부드러운', 0.9325265884399414),
 ('콜라', 0.9117951393127441),
 ('푸딩과', 0.9076296091079712),
 ('젤리', 0.9031435251235962),
 ('믹스', 0.8967173099517822),
 ('사과주', 0.8941375017166138),
 ('스파클링', 0.8932083249092102),
 ('음료수', 0.8901629447937012),
 ('주머니', 0.8861982226371765),
 ('믹스커피', 0.8849079608917236)]
```

TSNE를 통한 Word2Vec 시각화(가설 증명)



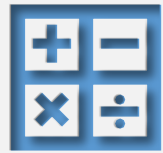
Word2vec을 통해 구해진 좌표 값

- ?번 고객이 검색한 키워드
- ?번 고객이 구매한 카테고리
- ??번 고객이 검색/구매한 내역



주로 검색한 키워드는 구매로 이어지고 연결되는 단어들은 의미상으로 가까이 배치되어 있으며, 다른 고객과는 멀리 배치되는 것을 확인할 수 있다.

고객에 대한 고유 벡터 생성



고객 간의 유사도

앞서 보인 TSNE 시각화에서 설명한 것처럼 다른 키워드를 가지는 고객과의 구분이 가능하기 때문에 고객에 대한 키워드들의 평균 좌표를 통해 의미를 잃지 않는 선에서 고유한 벡터 생성 가능



데이터 제거

구매를 하지 않은 고객에 대한 고유 벡터 생성에서 학습 데이터를 통해 Word2Vec 모델을 구성하였기 때문에 기존에 없던 검색어로 검색한 경우 벡터 생성이 불가능하고 Word2Vec에 함께 학습하는 경우에는 오버피팅의 위험이 있기 때문에 해당 고객 제거 후 진행

Cosine Similarity

```
index_t = []
index_c = []
for i in tqdm.tqdm_notebook(range(0, len(n_wv))):
    a = i
    cos = []
    for j in range(0, len(y_wv)):
        cos.append(cosine_similarity(n_wv[a].reshape(1,-1), y_wv[j].reshape(1,-1)))

    cos_df = pd.DataFrame({
        'cos_sim': cos
    }).sort_values(by='cos_sim', ascending = False).reset_index()

    sim_c = cos_df.iloc[1,0]
    index_c.append(sim_c)

    b = clean_y_t[sim_c]
    sim_t = new_product.query('kor_clac_nm3 in @b').iloc[:,1].values
    index_t.append(sim_t)
```

```
final = pd.DataFrame({
    'clnt_id': n_df.clnt_id.unique(),
    'trans': index_t
    'sim_id': index_c
})
```

구매를 하지 않은 고객과 구매한 고객 간의
코사인 유사도 계산



가장 유사한 고객의 구매내역 추출

4. Item Based Recommendation

Co-occurrence matrix

Co-occurrence(동시 발생 횟수) matrix: pd_c 간의 동시 발생 횟수를 각 고객의 구매한 pd_c 수 행렬을 내적해서 생성

* Co-occurrence matrix 계산은 MS에서 만든 SAR(Simple Algorithm for Recommendation)에서 참고함



고객별 구매 제품 수 행렬(고객 수 * pd_c 수)

```
[ [1, 0, 0, ..., 0, 0, 0],  
  [1, 0, 0, ..., 0, 0, 0],  
  [2, 0, 0, ..., 0, 0, 0],  
  ...,  
  [0, 0, 0, ..., 0, 0, 2],  
  [0, 0, 0, ..., 0, 0, 1],  
  [0, 0, 0, ..., 0, 0, 2]]
```

Co-occurrence matrix

```
[ [ 52, 2, 2, ..., 0, 3, 3],  
  [ 2, 5, 1, ..., 0, 0, 0],  
  [ 2, 1, 7, ..., 0, 0, 0],  
  ...,  
  [ 0, 0, 0, ..., 6, 2, 0],  
  [ 3, 0, 0, ..., 2, 3103, 148],  
  [ 3, 0, 0, ..., 0, 148, 564]]
```

Co-occurrence matrix

Co-occurrence matrix의 n번째 행이나 열은
n번째 pd_c와 다른 pd_c가 얼마나 많은 고객들이
같이 구매했는지 알려줌



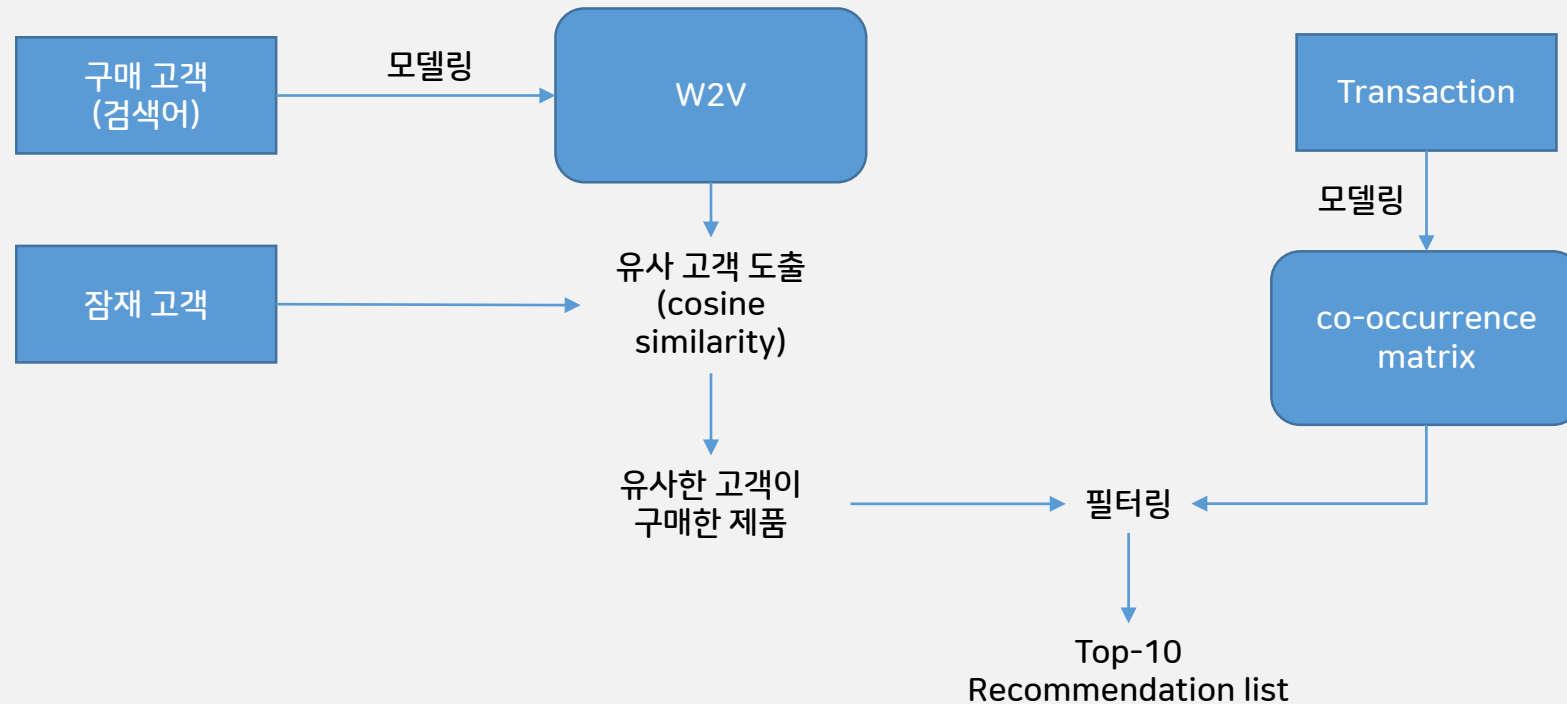
W2V 모델링 결과 나온 가장 유사한 고객의 구매내역을
Co-occurrence matrix와 비교해서
값이 높은 pd_c를 추려서 Top-K 리스트를 생성함

(유사한 고객의 구매내역이 여러 개인 경우
랜덤 추출을 통해서 K개를 뽑음)

Top-10 list

	clnt_id	pd_c	clac_nm1
0	65991	893	Liquors / Alcoholic Beverages
1	65991	723	Health Foods
2	65991	333	Cosmetics / Beauty Care
3	65991	841	Kids' Clothing
4	65991	1212	Snack Foods
5	65991	1018	Others (Non-Products)
6	65991	565	Fruits
7	65991	1616	Vegetables
8	65991	221	Cleaning / Laundry / Bathroom Accessories
9	65991	114	Beverages

전체 과정 도식화



	clnt_id	pd_c	ciac_nm1	ciac_nm2	ciac_nm3
0	65991	893	Liquors / Alcoholic Beverages	Beer	Domestic Beer
1	65991	723	Health Foods	Nutritional Supplements	Omega3 Oils / Other Extracted Oils
2	65991	333	Cosmetics / Beauty Care	Skin Care	Beauty Soaps
3	65991	841	Kids' Clothing	Preschoolers' Lower Bodywear / Bottoms	Infant / Toddlers' Leggings
4	65991	1212	Snack Foods	Snacks	Fruit Snacks
5	65991	1018	Others (Non-Products)	Income from leasing	Income from leasing
6	65991	565	Fruits	Imported Fruits	Bananas
7	65991	1616	Vegetables	Tofu / Bean Sprouts	Soybean Sprouts
8	65991	221	Cleaning / Laundry / Bathroom Accessories	Cleaning Accessories	Trash Bags
9	65991	114	Beverages	Water	Water

5. 비즈니스 전략 수립

현재 사항

다른 고객들이 함께 본 상품

<p>매장 배송</p> <p>한우 사태 (1등급)(100G)</p> <p>판매가 4,980원</p> <p>- 1 + </p>	<p>매장 배송</p> <p>한우 양지 국거리용 (1등급)(150G)</p> <p>판매가 10,200원</p> <p>- 1 + </p>	<p>매장 배송</p> <p>한우 정육 국거리용 (1등급)(150G)</p> <p>판매가 7,470원</p> <p>- 1 + </p>	<p>매장 배송</p> <p>한우 정육 접채용 (1등급)(150G)</p> <p>판매가 7,470원</p> <p>- 1 + </p>
---	---	--	---

● 0000

같은 카테고리의 상품 추천
구매에 대한 부분 고려하지 못함

BEST

베스트 더보기 >

롯데마트물 인기상품만 골라올라서!

<p>FOOD LIFE</p> <p>전체 쌀/잡곡 채소</p> <p>과일 정육/계란 생선/건해산물</p> <p>견과/산식 냉장/생동/간편식 가공식품/대용식</p> <p>친환경식품 건강식품 조리식품</p>	<p>매장 배송</p> <p>바나나(송이)</p> <p>판매가 1,980원</p> <p>- 1 + </p>	<p>매장 배송</p> <p>제주 서귀포 감귤(3.8KG/박스)</p> <p>판매가 8,980원</p> <p>- 1 + </p>	<p>매장 배송</p> <p>Only Price 1등급 우유(930ML*2입)</p> <p>판매가 3,000원</p> <p>- 1 + </p>
--	---	---	---

고객별 추천이 아닌
가장 많이 팔리는 상품 일괄적 추천

개선 사항

같은 카테고리의 상품 추천
구매에 대한 부분 고려하지 못함



유사한 고객이 함께 구매한 상품을
보여줌으로써 고객 간의 유기적인
상품 추천이 가능하도록 개선

고객별 추천이 아닌
가장 많이 팔리는 상품 일괄적 추천



알고리즘을 통해 고객별 추천 카테고리 생성 가능
추가적으로 구매 상품을 정확하게 알 수 있다면
카테고리별 추천 상품을 제시하는 것도 가능

Thank you for Reading!