

Shopping Category classification AI model project

쇼핑중독

김우석

김정문

임민우

허환욱



Contents

01. Purpose

02. Data crawling

03. Preprocessing

04. Learning

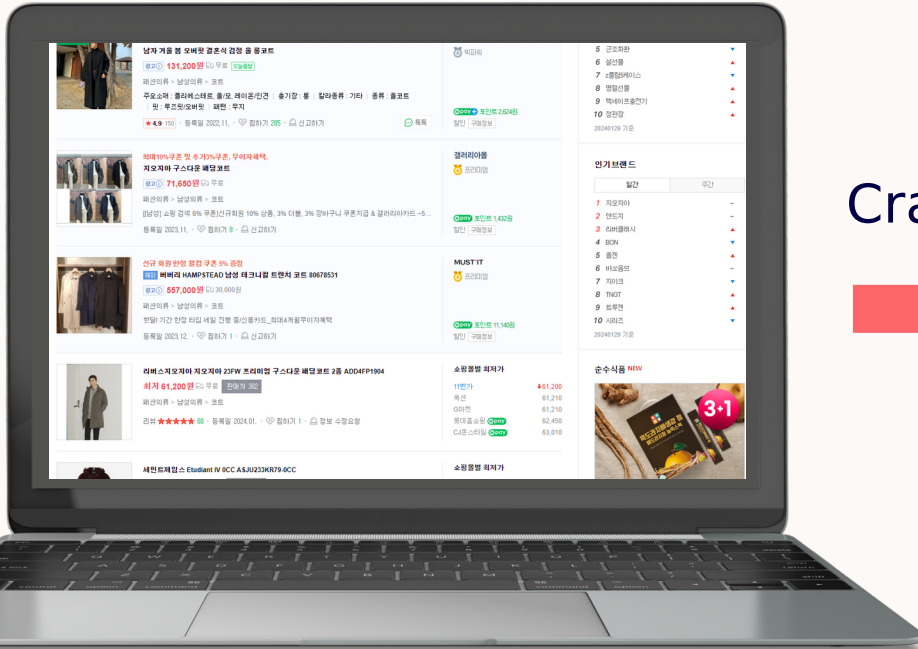
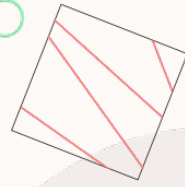
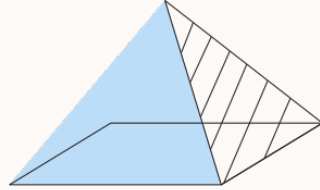
05. Result

06. Q&A

01. Purpose



01. Purpose



Crawling



Data preprocessing

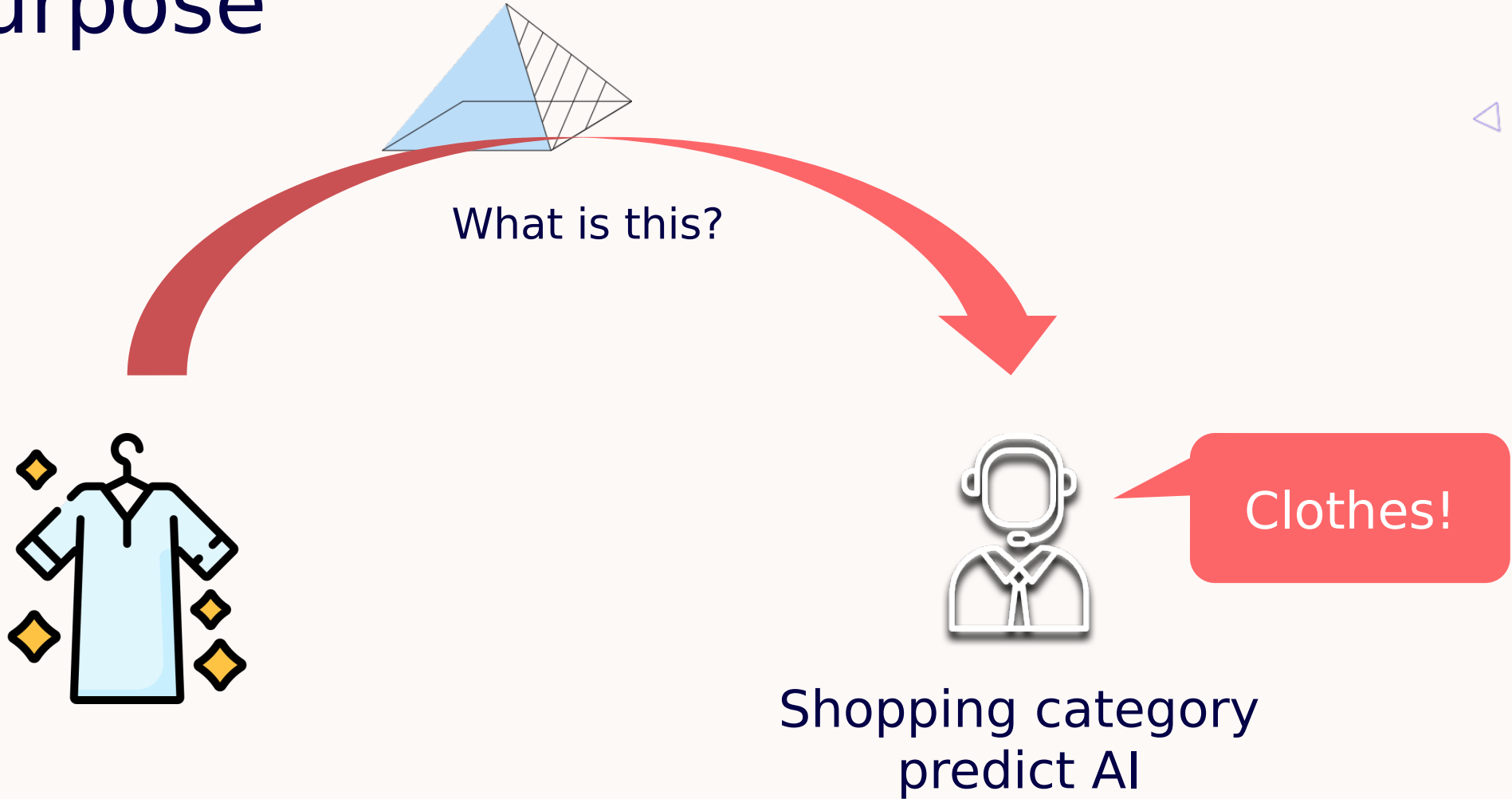


Learning



Shopping category
predict AI

01. Purpose



02. Data crawling



02. Data crawling

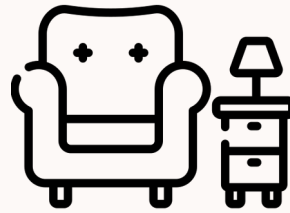
Category



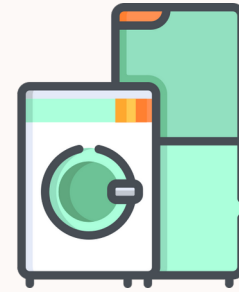
Fashion



Beauty



Furniture

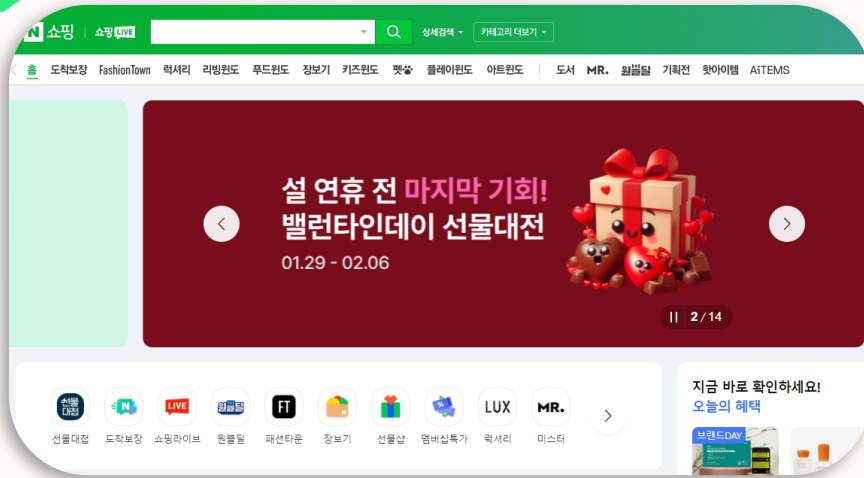


Digital

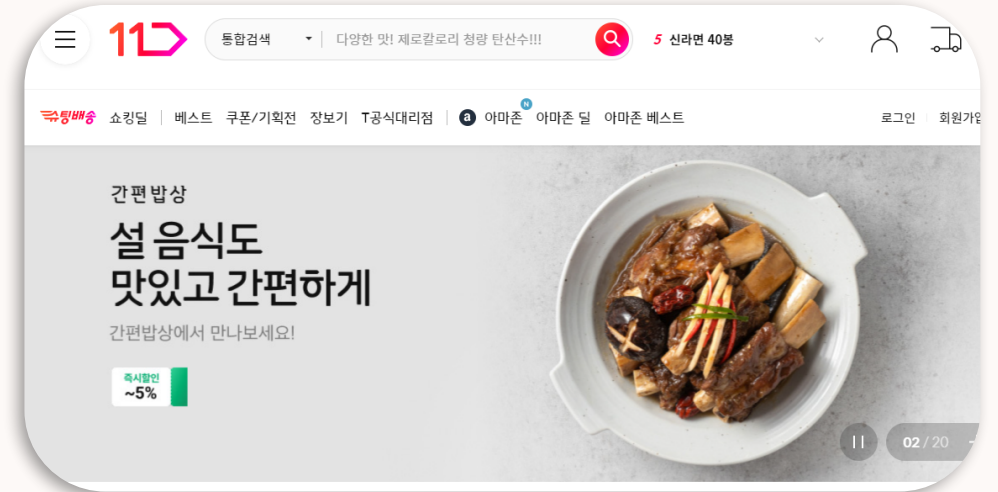


Food

02. Data crawling



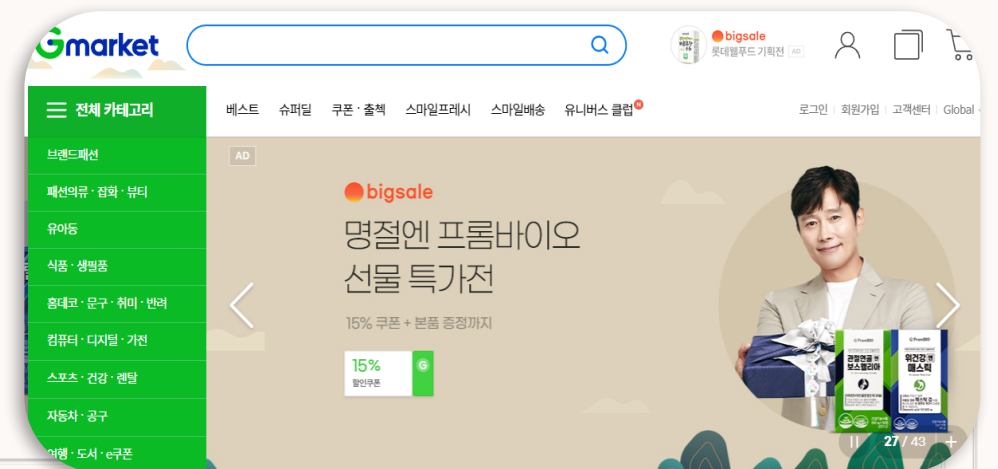
Naver



11st

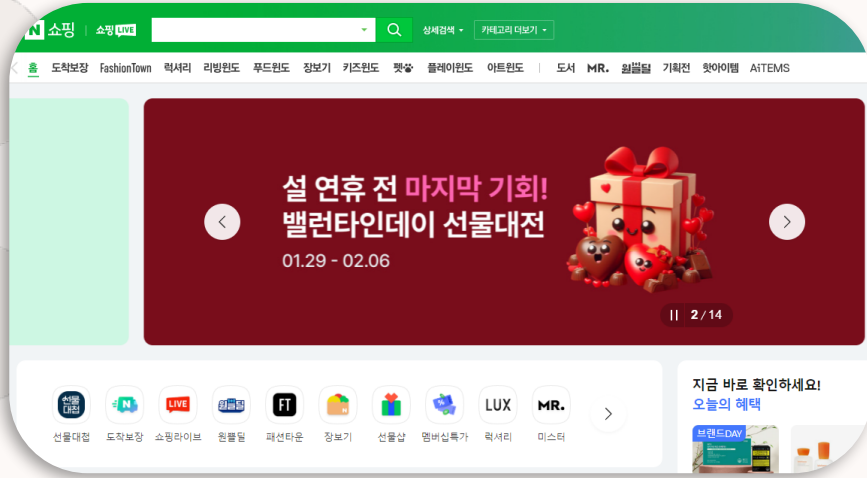


Coupang



Gmarket

02. Data crawling



Naver

Organized address and xpath



Coupang

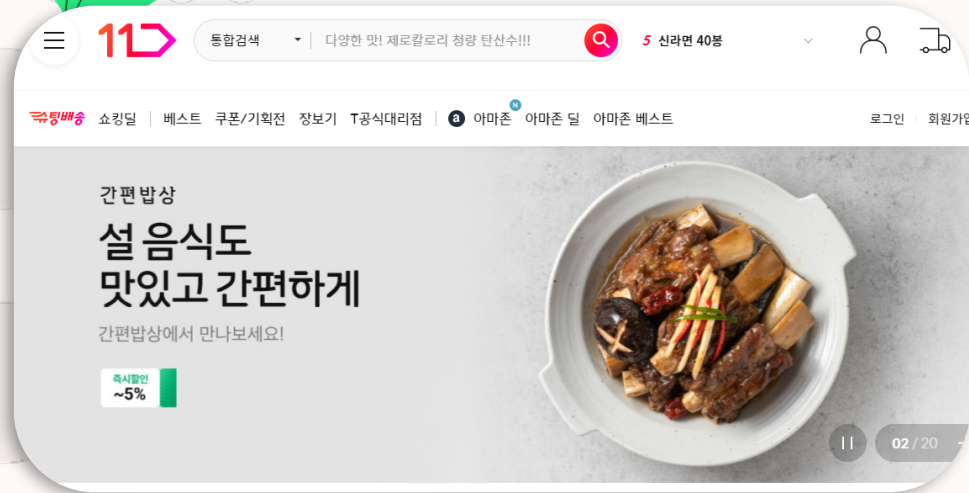
Block request from server



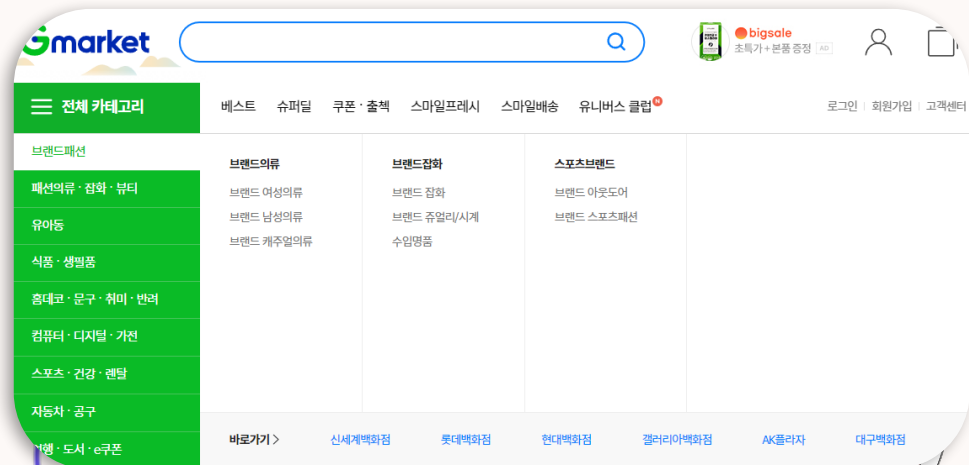
```
driver.delete_all_cookies()
```



02. Data crawling



11st



Gmarket

Irregular address and xpath



```
search_box = driver.find_element(by='xpath',  
                                value: '//*[@id="gnbCategory"]/  
actions = wb.ActionChains(driver).move_to_element(search_box)  
actions.perform()
```

```
title_tags = driver.find_elements(By.CLASS_NAME, value: 'pname')
```

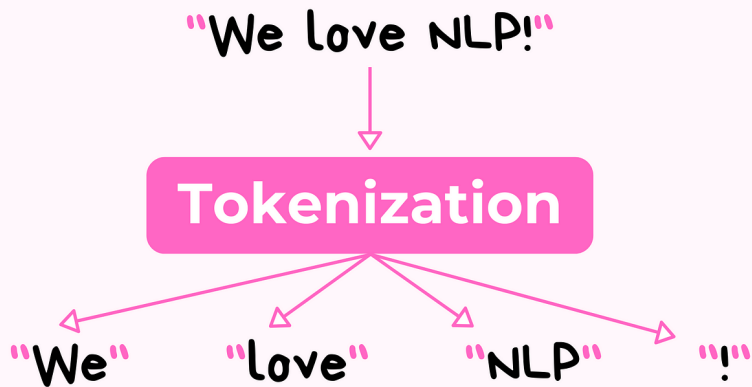


03. Preprocessing



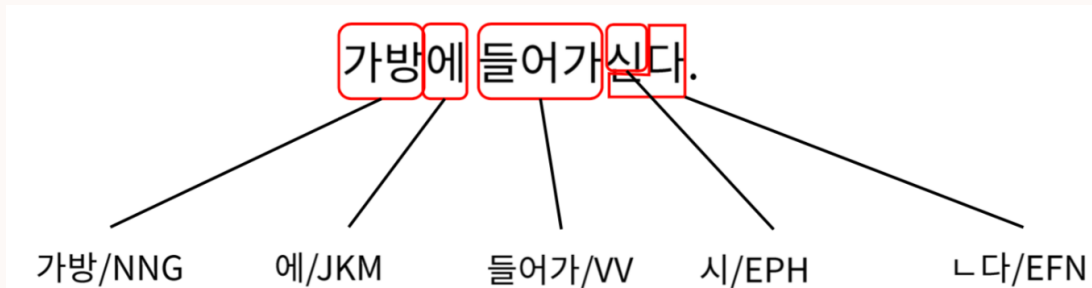
03. Preprocessing

Morpheme separation



Remove stop words

| | |
|---|-----------|
| 1 | ,stopword |
| 2 | 0,아 |
| 3 | 1,휴 |
| 4 | 2,아이구 |
| 5 | 3,아이쿠 |
| 6 | 4,아이고 |
| 7 | 5,어 |
| 8 | 6,나 |



03. Preprocessing

Remove stop words

| | | |
|-------|-----|-----------------|
| 17009 | 상품명 | 안전 무늬 유아 책상 |
| 17010 | 상품명 | 체스 공부상 대왕체스말 포함 |
| 17011 | 상품명 | 세 유아 공부 책상 의자 합 |
| 17012 | 상품명 | 체스 공부상 대왕체스말 포함 |
| 17013 | 상품명 | 롤페이퍼 키즈독서대 유아 의 |
| 17014 | 상품명 | 유베코 아기 유아 어린이 눈 |
| 17015 | 상품명 | 잡지진열장 잡지함 잡지꽂이 |
| 17016 | 상품명 | 옷장 아이옷장 아동옷장 원형 |
| 17017 | 상품명 | 오피스 책상 책꽂이 북스탠드 |

| | |
|-----|-----------|
| 789 | 788, 판매가 |
| 790 | 789, 정상가 |
| 791 | 790, 원 |
| 792 | 791, 무료배송 |
| 793 | 792, pay |
| 794 | 793, 최대 |
| 795 | 794, P적립 |
| 796 | 795, 상품명 |



Combining morphemes
And
Tokenization

04. Learning



04. Learning

```
11 model = Sequential()
12 model.add(Embedding(input_dim: 11911, output_dim: 300, input_length=24))
13 model.add(Conv1D(filters: 32, kernel_size=5, padding='same', activation='relu'))
14 model.add(MaxPooling1D(pool_size=1))
15 model.add(LSTM(units: 128, activation='tanh', return_sequences=True))
16 model.add(Dropout(0.3))
17 model.add(LSTM(units: 64, activation='tanh', return_sequences=True))
18 model.add(Dropout(0.3))
19 model.add(LSTM(units: 64, activation='tanh'))
20 model.add(Dropout(0.3))
21 model.add(Flatten())
22 model.add(Dense(units: 128, activation='relu'))
23 model.add(Dense(units: 5, activation='softmax'))
24 model.summary()
```

Learning using RNN

05. Result



05. Result

coupang



Coupang Data
AI model



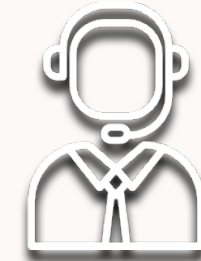
```
[3024 rows x 3 columns]
OX
0    2972
X     52
Name: count, dtype: int64
OX
0    0.982804
X    0.017196
Name: count, dtype: float64
```

First prediction

```
[3024 rows x 3 columns]
OX
0    2996
X     28
Name: count, dtype: int64
OX
0    0.990741
X    0.009259
Name: count, dtype: float64
```

+Second prediction

11



Coupang Data
AI model



```
Name: 4795, dtype: object
OX
0    3081
X    1715
Name: count, dtype: int64
OX
0    0.64241
X    0.35759
Name: count, dtype: float64
```

First prediction

```
Name: 4790, dtype: object
OX
0    3812
X     984
Name: count, dtype: int64
OX
0    0.794829
X    0.205171
Name: count, dtype: float64
```

+Second prediction

05. Result

coupang

N 쇼핑



11

Gmarket

Total Data AI model

```
Name: 52075, dtype: object
OX
0    51460
X     638
Name: count, dtype: int64
OX
0    0.987754
X    0.012246
Name: count, dtype: float64
```

First prediction

```
Name: 52070, dtype: object
OX
0    51742
X     356
Name: count, dtype: int64
OX
0    0.993167
X    0.006833
Name: count, dtype: float64
```

+ Second prediction

06. Q&A



The slide features a variety of decorative geometric elements. In the top left, there is a large light gray circle partially obscured by a red circle with a white border. A small green circle and a purple triangle are nearby. The top center has three light gray circles. The top right contains a green 3D pyramid with a white outline, surrounded by a small yellow dot, a red circle, and a blue dot. The bottom left shows a blue triangle with a white outline, a red dot, and a small blue circle. The bottom right has a light blue circle with a white outline, a red triangle, and a light gray rectangle. The text "Thanks !" is centered in a dark blue font.

Thanks !