# **Shopping Category classifi-cation** AI model project

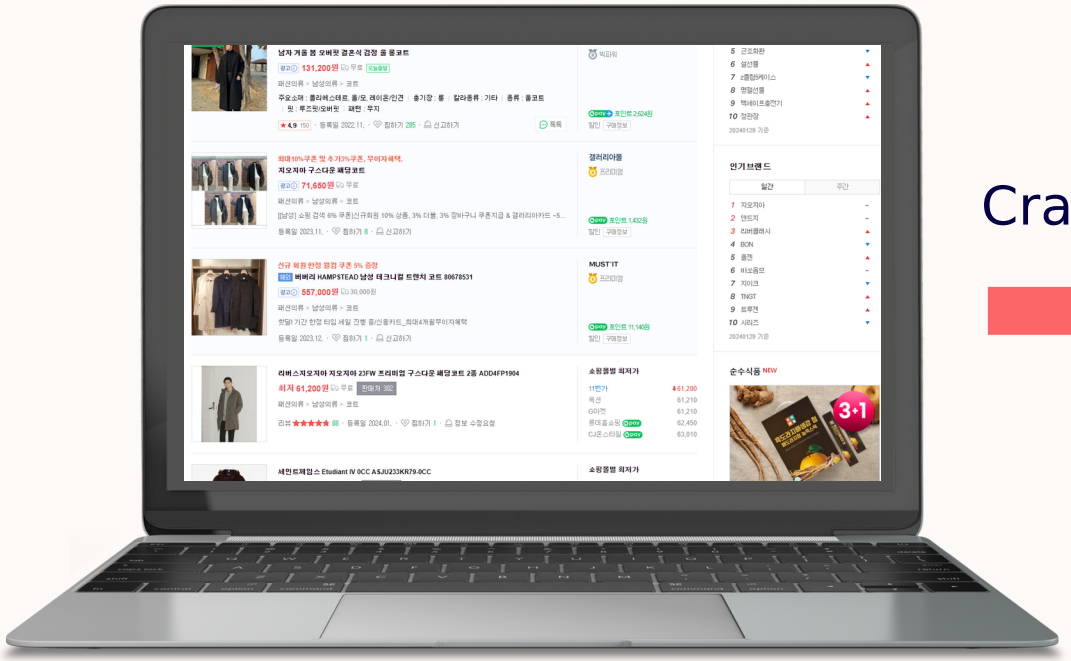
쇼핑중독

- 김우석
- 김정문
- 임민우
- 허환욱

# Contents

# 01. Purpose

# 01. Purpose



Crawling → Data prepro-cessing → Learning

Shopping category predict AI

# 01. Purpose

What is this?

Clothes!

Shopping category
predict AI

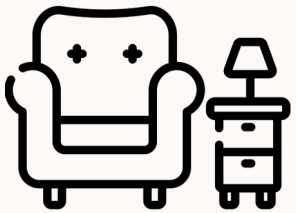
# 02. Data crawl-ing

# 02. Data crawling

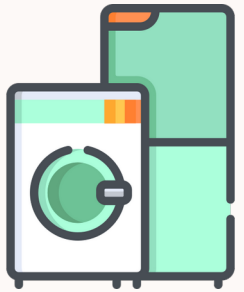
## Category

Fashion

Beauty



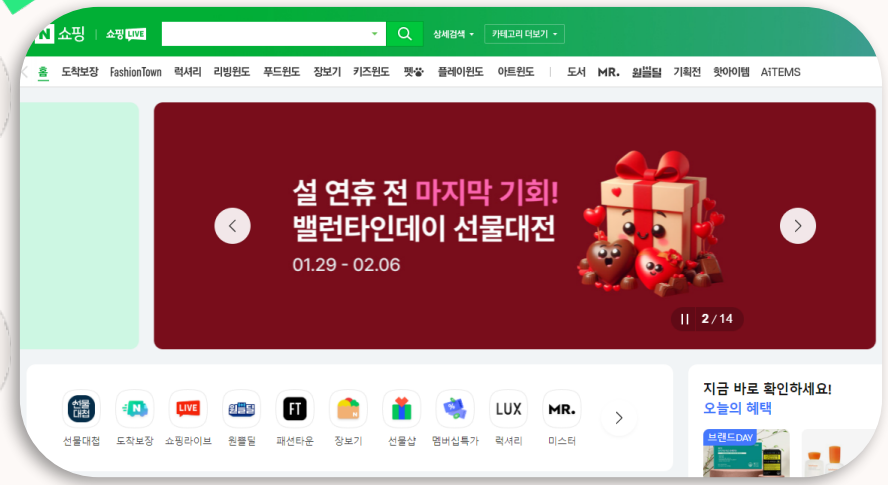
Furniture



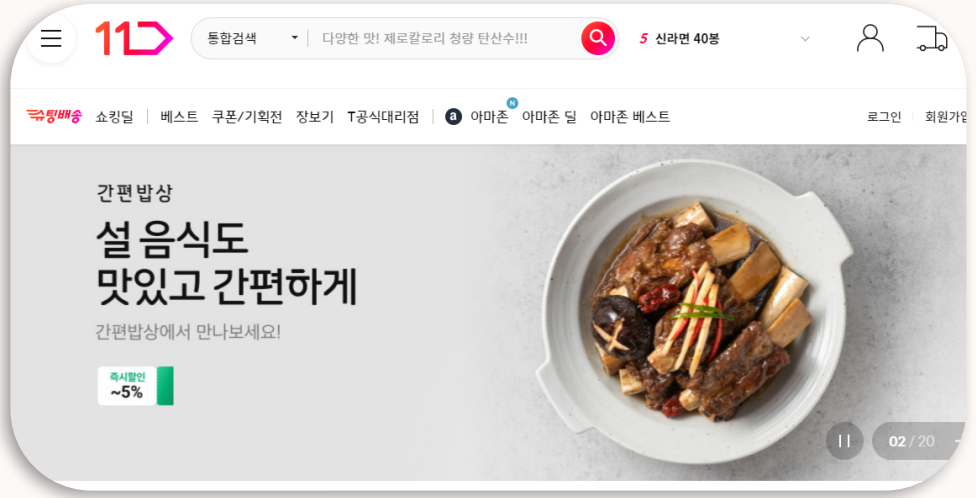
Digital

Food

# 02. Data crawling

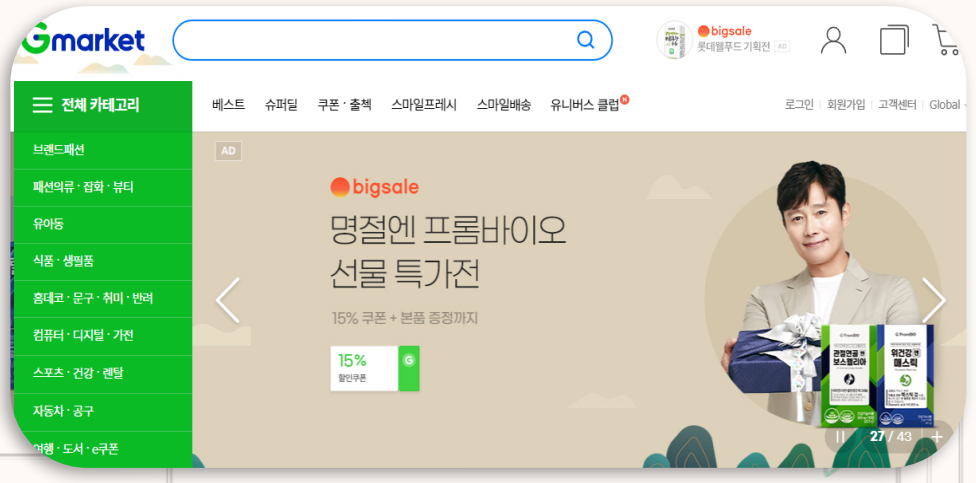



Naver



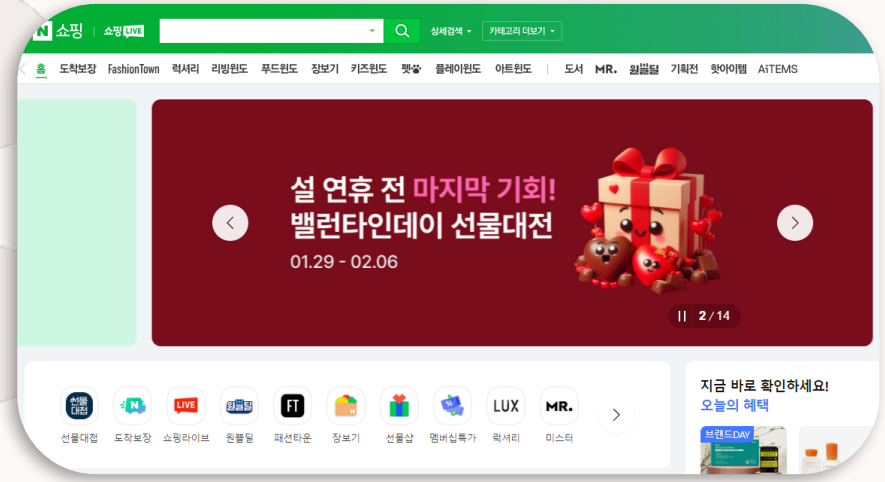

11st


Coupang




Gmarket

# 02. Data crawling




Naver

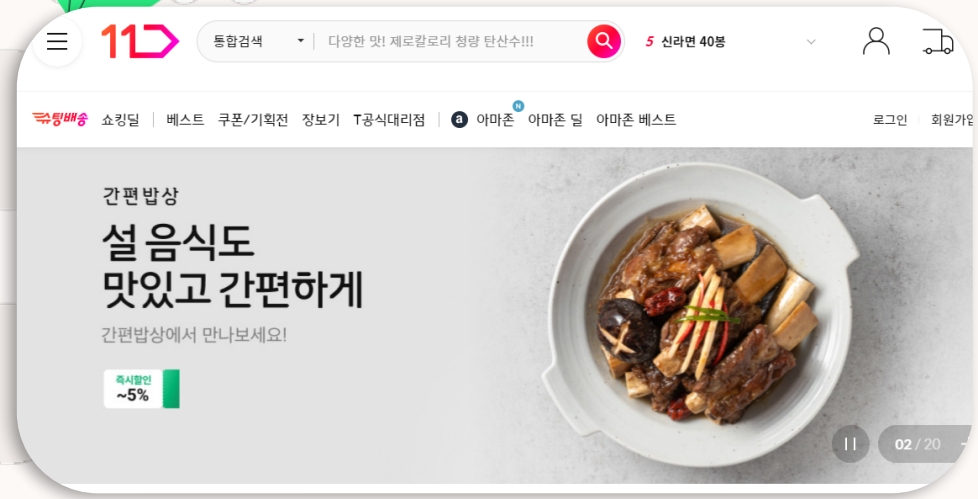

Coupang

Organized address and xpath

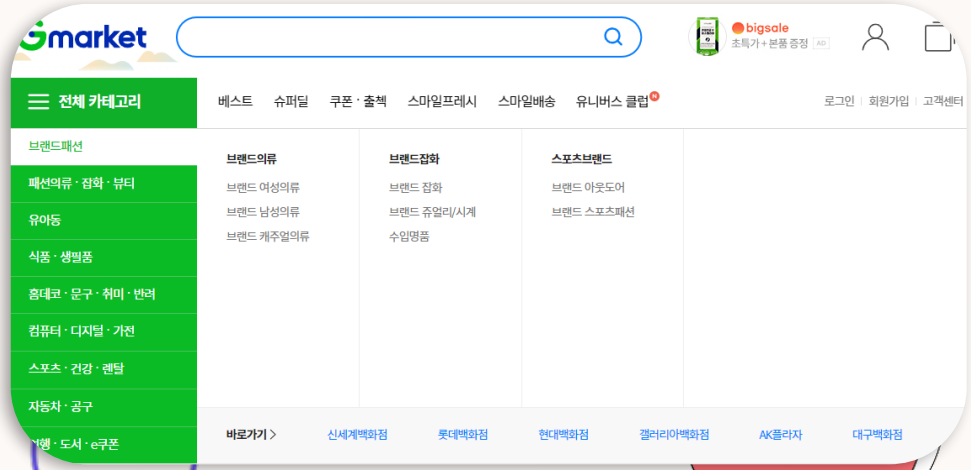
Block request from server

```
driver.delete_all_cookies()
```

# 02. Data crawling




11st




Gmarket

Irregular address and xpath

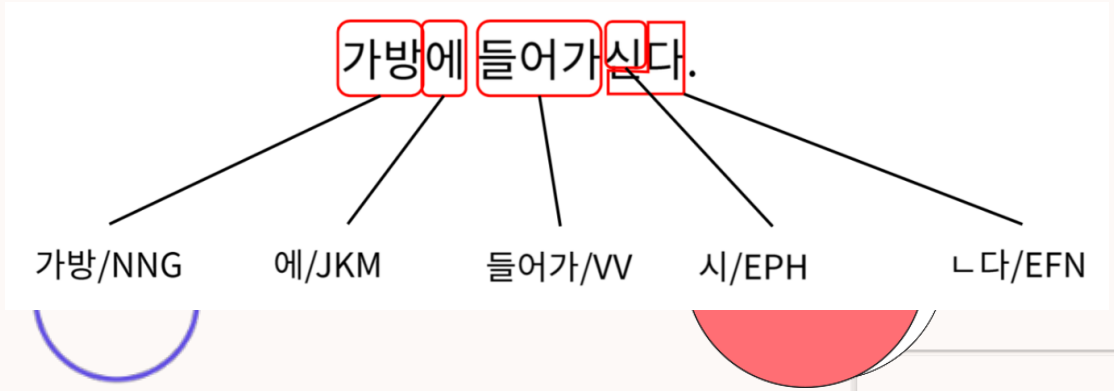
```
search_box = driver.find_element( by: 'xpath',
                                  value: '//*[@id="gnbCategory"]/
actions = wb.ActionChains(driver).move_to_element(search_box)
actions.perform()
```

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

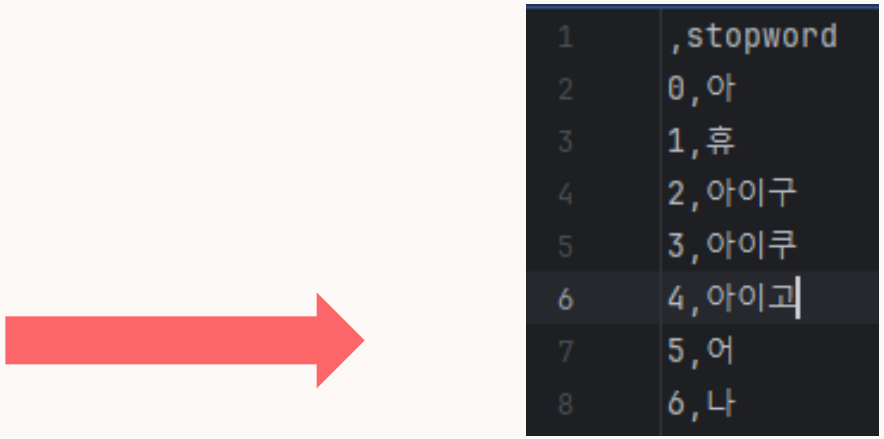
# 03. Preprocessing

# 03. Preprocessing

## Morpheme separation

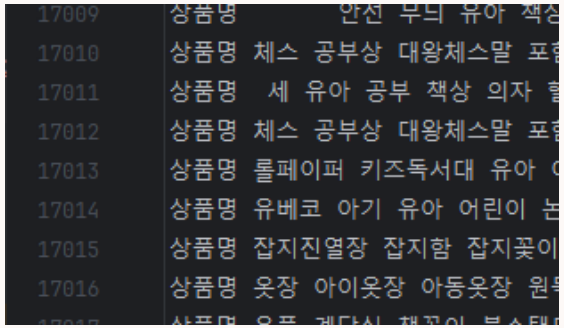

## Remove stop words



| | ,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





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 Data
AI model

Coupang Data
AI model

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

First prediction

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

+Second prediction

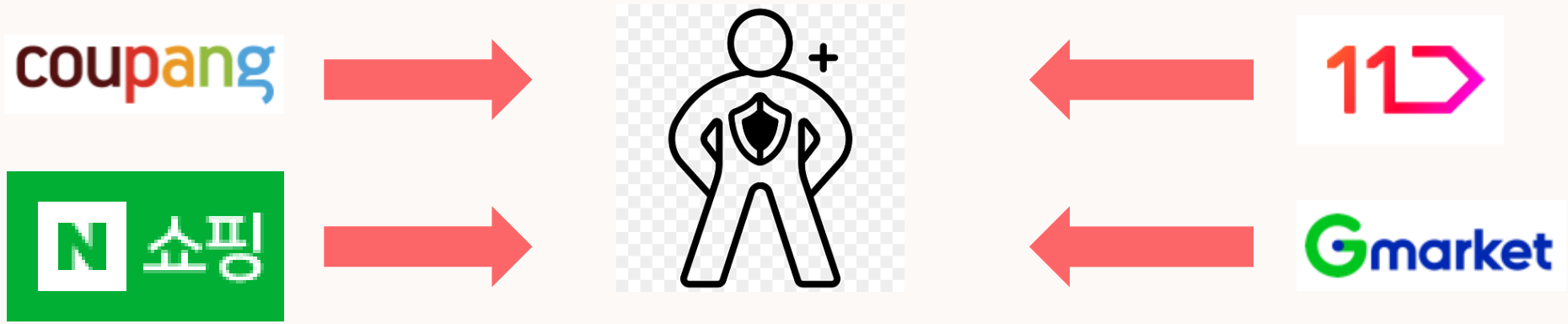
```
Name: 4795, dtype: object
OX
O    3081
X    1715
Name: count, dtype: int64
OX
O    0.64241
X    0.35759
Name: count, dtype: float64
```

First prediction

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

+Second prediction

# 05. Result




Total Data AI model

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

First prediction

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

+Second prediction

# 06. Q&A

# Thanks !