

머신러닝을 활용한 텍스트 마이닝

# 간편 결제 어플리케이션 리뷰 분석

# 목차

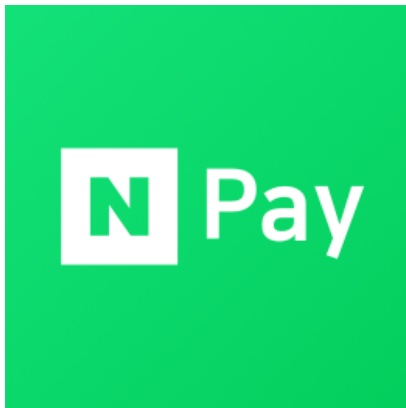
---

1. 주제 소개 및 Data설명
2. 빈도분석
3. 감성분석
4. Clustering
5. 아쉬운 점



# 1.주제 및 DATA

주제 : 텍스트마이닝을 활용한 간편 결제 어플리케이션의 리뷰 비교분석



## 사용한 DATA

- 구글플레이스토어에서 네이버페이, 삼성페이, 카카오페이의 앱 리뷰, 기간, 별점 크롤링
- 앱스토어에서 네이버페이, 카카오페이의 앱 리뷰, 기간, 별점 크롤링

# 1.DATA 수집방법 : Google Play Store

Selenium을 사용한 구글플레이스토어 리뷰 크롤링 실패  
데이터양이 적은 어플의 리뷰는 크롤링 가능하지만 데이터 양이 많은 어플의 리뷰는 에러발생

```
# 크롤링하고자 하는 리뷰 페이지의 url
# url = 'https://play.google.com/store/apps/details?id=com.samsung.android.spay&hl=ko&gl=US&showAllReviews=true'

driverPath = "C:/chromedriver_win32/chromedriver"
driver = webdriver.Chrome(driverPath)
driver.get(url)
SCROLL_PAUSE_TIME = 2
SCROLL_TIMES = 4 # 4번 스크롤 두 더보기 버튼 생성
CLICK_PAUSE_TIME = 2

last_height = driver.execute_script("return document.body.scrollHeight")

# 스크롤 가장 아래까지 내려가 ([더보기] 누르면서)
while True:
    for i in range(SCROLL_TIMES):
        driver.execute_script("window.scrollTo(0, document.body.scrollHeight);")
        time.sleep(SCROLL_PAUSE_TIME)
    # 중간중간 [현재 리뷰] 버튼 누르기
    spread_review = driver.find_elements_by_xpath("//button[@jsaction='click:TigIPc']")
    for i in range(len(spread_review)):
        isTrue = spread_review[i].is_displayed()
        time.sleep(5)
        if isTrue:
            driver.execute_script("arguments[0].click()", spread_review[i])
            time.sleep(CLICK_PAUSE_TIME)

    more_button = driver.find_elements_by_xpath("//span[@class='RveJvd snByac']")

    if more_button:
        more_button[0].click()
        time.sleep(CLICK_PAUSE_TIME)

    # 더이상 내려가는 곳이 없으면 break
    new_height = driver.execute_script("return document.body.scrollHeight")
    if new_height == last_height:
        break
    last_height = new_height
```



## Google Play Scraper 이용!

```
from google_play_scraper import Sort, reviews_all

result = reviews_all(
    'com.kakaopay.app',
    sleep_milliseconds=1000, # defaults to 0
    lang='ko', # defaults to 'en'
    country='us', # defaults to 'us'
    sort=Sort.MOST_RELEVANT, # defaults to Sort.MOST_RELEVANT
    # filter_score_with=5 # defaults to None (means all score)
)
```

**Google Play Store App Review Crawler**  
→ 카카오페이, 네이버페이, 삼성페이 모든 리뷰 크롤링

# 1.DATA 수집방법 : App Store



문제 : 앱스토어 리뷰 페이지에서는 10개의 리뷰만 확인 할 수 있음

XLM을 Parsing하여 리뷰데이터 추출

App Store 미리보기

네이버페이

평가 및 리뷰

4.8

1만개의 평가



★★★★★

jjoao0, 2022. 04. 18.

매출워져 결제!!  
토스, 카카오뱅크 쓸수있게해서 너무 좋습니<sup>ㅏ</sup>!!  
스마트워치 사용하는사람이 많아진만큼 매출워져로  
결제가능하면 정말 좋을것같습니<sup>다</sup>!!

★★★★☆

로소로소, 2021. 11. 06.

어쨌든계 상당히맘음  
네이버페이 사용 의무화 할 필요가 있나<sup>ㅓ</sup>? 카카오뱅크 쓰다가  
리워드프로그램이 네이버페이가 우월해서 넘어왔는데 뭐  
하하하하! 일단 네이버페이 가져다쓰게 너무 심세<sup>다</sup>. 가<sup>ㅓ</sup>이<sup>ㅓ</sup>  
배달아들 시집때도 생체인증 안 받아주고요.

★★★★★

lgh0116, 2021. 11. 05.

정말 편하고 좋은것같습니<sup>다</sup>  
저는 리워드같은거 안올리는 사람인데로  
아이들은 장알종고 편한것 같습니<sup>다</sup>  
아이들은 NFC기능이 없어서 결제할 때마다 현찰을  
지불하거나 카드결제를 해야되는 번거로움 있는데<sup>ㅓ</sup>!

★★★★★

아오유아유, 2021. 08. 31.

잘쓰고있어요~  
너무 잘쓰고 있어서 너무편하고 다른 앱을 대체할 수 있을 것  
같습니<sup>다</sup>.  
다만 처음 뽀빠시 카드 추가할때 하나 추가하면 다시 목록으로  
나가지고 오류있다고 실패하면 다시 카드 목록으로 나<sup>ㅓ</sup>!

★★★★☆

whatever\_#, 2021. 10. 05.

환대카드=0  
네이버페이가 좋아요. 환대카드도 좋아요. 근데 네이버<sup>ㅓ</sup>!  
개발자 답변,  
whatever\_# 님, 네이버페이앱을 이용해주셔서

★★★★★

종혜고나눔계로, 2022. 03. 28.

네이버 페이 너무 좋ونة 송금기능등에 하나만 고쳐주세요.  
네이버페이가 너무 좋ونة 송금기능을 전환부터 받고 일관부터  
송금 가능해졌으면 좋겠네<sup>ㅓ</sup>!!

★★★★★

GwanakMan, 2021. 08. 10.

편리한 UI  
카카오뱅크, 페이코 보고 오니까 정말 편하<sup>ㅓ</sup>!  
토스나 뽀빠시에도 동등 대해 가능해<sup>ㅓ</sup>!  
자산 수입·지출에서 카테고리별로 지정하는 기능부 기능도  
있으면 좋겠네<sup>ㅓ</sup>!

★★★★★

해뽀, 2022. 03. 11.

네이버페이 어플 꽤 이해와 생겼는지  
모르겠지만 생겨서 좋네<sup>ㅓ</sup>. 분명 작년 하반기에도 안보였는데  
언제 만들었나<sup>ㅓ</sup>?  
검색해보니 작년 8월달에 출시했네<sup>ㅓ</sup>.  
네이버 어플 하나도 안알았는데 이 어플은 바로 알지<sup>ㅓ</sup>!

★★★★★

santacome, 2021. 08. 11.

제발 부탁요  
이 앱하나로 전세계 카드 가맹점에서 결제할 수 있었으면  
좋겠네<sup>ㅓ</sup>. 일단 국<sup>ㅓ</sup>외<sup>ㅓ</sup>도.  
아이폰 사용자 지갑 들고 다니기 심운데 방법이 없네<sup>ㅓ</sup>.  
ㅠㅠ 요즘 코로나로 오프라인 결제도 거의 안하고 기<sup>ㅓ</sup>!

★★★★☆

우희비름, 2021. 09. 06.

결제할 때  
타사 온라인 쇼핑몰이나 사이트에서 네이버페이로 결<sup>ㅓ</sup>!  
개발자 답변,  
우희비름 님, 네이버페이앱을 이용해 주셔서 감사합니<sup>다</sup>!



```
#1464496236 카카오페이
#1554807824 네이버페이
res_dict = []
for j in range(1,11):
    url = "https://itunes.apple.com/kr/rss/customerreviews/page=%i/id=1464496236/sortby=mostrecent/xml?urlDesc=customerreviews/id=1464496236/sortBy=mostRecent/xml"% j
    response = urlopen(url).read()

    xml = xmltodict.parse(response)
    XmlToJson = json.loads(json.dumps(xml))

    for i in range(len(XmlToJson['feed']['entry'])):
        res_dict.append({
            'DATE' : XmlToJson['feed']['entry'][i]['updated'],
            'STAR' : XmlToJson['feed']['entry'][i]['im:rating'],
            'LIKE' : XmlToJson['feed']['entry'][i]['im:voteSum'],
            'DISLIKE' : int(XmlToJson['feed']['entry'][i]['im:voteCount']) - int(XmlToJson['feed']['entry'][i]['im:voteSum']),
            'TITLE' : XmlToJson['feed']['entry'][i]['title'],
            'REVIEW' : XmlToJson['feed']['entry'][i]['content'][0]['#text']
        })
res_df = pd.DataFrame(res_dict)
res_df['DATE'] = pd.to_datetime(res_df['DATE'], format='%Y-%m-%dT%H:%M:%S-07:00')
# res_df.to_csv('./data/IOS_ReviewData.v1.0.csv')
```

네이버페이 리뷰의 경우 7페이지까지 있었고 카카오페이의 경우 10페이지까지 밖에 추출이 되지 않음

# 1.DATA

	DATE	STAR	REVIEW
0	2022-05-20 00:13:02	4	결제 방식을 바코드 결제가 아니라 패이코처럼 삼성 페이로(mfc 방식으로) 결제할 ...
1	2022-05-17 16:49:48	5	굿
2	2022-05-16 03:26:43	1	2단계 인증 로그인 설정시 로그인때 몇번씩 반복해도 알림이 오지않고 OTP번호로도 ...
3	2022-05-14 17:46:40	1	웹보다 접근성 ui 다떨어짐 쓸이유가 없는 앱
4	2022-05-14 08:56:01	3	먼처음에 앱 다운로드하고 본인인증을 하라 해서 했더니 인증번호가 안 옵니다... 몇...
...	...	...	...
657	2021-08-03 12:38:12	5	굿굿굿-bbb
658	2021-08-03 11:10:32	5	아호 네이버 페이 앱이네요
659	2021-08-03 00:32:59	5	편하고 좋아요!
660	2021-08-01 21:53:31	5	별리만들어주세요
661	2021-08-01 08:52:51	5	이거왜 리뷰가없죠?네이버맞조?!!

네이버페이 앱 리뷰 크롤링 (662개)

	DATE	STAR	REVIEW
0	2022-05-20 03:10:17	1	주민등록증이 없는데 어케 민증을 발급기간을 입력하라는건지...
1	2022-05-19 23:02:15	5	페이중앙은 장인듯
2	2022-05-19 22:53:53	1	지은 변경한 적 없는데 지은 변경되었다고 재등록하라고 나와요. 그래서 재등록 계속 하...
3	2022-05-19 22:05:08	3	비밀 최형욱 성선허
4	2022-05-19 10:54:18	2	Mymin7@18
...	...	...	...
4425	2019-05-28 18:36:01	3	아직 카톡에서 벗어나지 못해서 거기서 처리하는 메뉴들도 많네요. 차차 해줄거라고 믿...
4426	2019-05-28 18:33:48	5	간결하니 좋네요
4427	2019-05-28 17:32:35	5	카카오페이 독립축하합니다
4428	2019-05-28 16:28:27	5	너무 간편하고 ui가 손에 걸리네요 이런 금융앱은 처음이에요
4429	2019-05-28 16:07:02	5	와~ 드디어 페이앱이 나왔군요. 기대가 큼니다! 자주 애용할게요~

카카오페이 앱 리뷰 크롤링 (4430개)

2020년 이후 데이터를 사용

	DATE	STAR	REVIEW
0	2022-05-20 12:38:26	5	많은분들이 공감하시겠지만 "편하다" 이죠. 그럴 어떻게 편하나? 간단한 본인인증 후 ...
1	2022-05-20 12:23:35	5	항상 잘 쓰는데
2	2022-05-20 11:56:03	5	잘 사용하고 있어요
3	2022-05-20 11:40:17	5	편해요
4	2022-05-20 11:22:38	5	사용하기 참 편해요^^
...	...	...	...
176151	2015-10-23 23:50:38	5	속만 올려대네요 ㅋㅋ
176152	2015-10-23 22:25:02	1	ㅈㄱㄴ
176153	2015-10-22 11:05:35	1	사용안하고 싶다
176154	2015-10-22 03:17:13	5	대박이네요.
176155	2015-10-22 00:22:43	5	Works very well

삼성페이 앱 리뷰 크롤링 (176156개)

- 훨씬 오래전 출시 된 삼성페이에 비해 네이버페이의 리뷰 수가 상대적으로 부족
- 오래된 리뷰는 도움이 되지 않을 것이라고 예상
- 기간을 줄여 삼성페이와 카카오페이의 리뷰 수를 줄임

기간 조정 후 네이버페이 리뷰수: 662  
기간 조정 후 삼성페이 리뷰수: 47419  
기간 조정 후 카카오페이 리뷰수: 4106

## 2.빈도분석

### WHY?

- 앱 리뷰에는 1점부터 5점까지 5개 척도의 별점이 존재
- 앱을 이용하는 유저가 어떤 이유에서 해당 별점을 주는지 궁금해짐
- 간편결제 어플마다 어떤 차이점이 있을지 혹은 별점을 주는 양상이 비슷한지 알아보고 싶음

### How!

- 5개의 별점마다 빈도분석을 진행하여 각각의 간편결제 어플에 그러한 별점을 준 이유를 파악해보자!

## 2. 빈도분석

### 전처리 과정 ; Data cleaning

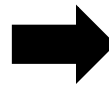
```
for data in [kakao, naver, samsung]:  
  
    data['cleaning_REVIEW'] = data.REVIEW.str.replace("[^가-힣 ]", "") #한글, 공백만 남기기  
    data['cleaning_REVIEW'] = data['cleaning_REVIEW'].str.replace('^ +', "") #emptyvalue로 변경  
    data['cleaning_REVIEW'].replace('', np.nan, inplace=True) #nan으로 변경  
    data.dropna(axis=0, inplace=True)
```

#### ① 한글, 공백만 남기고 모두 제거

```
#py-hanspell 사용 => 맞춤법 검사 및 띄어쓰기  
for data in [kakao, naver, samsung]:  
    lists=[]  
    for i in range(len(data)):  
        check=spell_checker.check(data.cleaning_REVIEW.iloc[i]).checked  
        lists.append(check)  
    data.cleaning_REVIEW=lists  
  
dfkakao=kakao  
dfnaver=naver  
dfsamsung=samsung
```

#### ② Py-Hanspell을 사용하여 맞춤법 검사를 진행하고 띄어쓰기를 알맞게 수정

카카오페이처럼 위젯 가능하게 업데이트 해주시면w 아주 좋을거같아요. 위젯 추가해주세요~  
장바구니도 없음 생각 좀 하고 **앱만 드세요**  
조아용



카카오페이처럼 위젯 가능하게 업데이트해주시면 아주 좋을 거 같아요 위젯 추가해주세요  
장바구니도 없음 생각 좀 하고 **앱만 드세요**  
좋아요

일부 문장의 경우 제대로 띄어쓰기가 진행되지 않음



## 2. 빈도분석

### 전처리 과정 ; Tokenization

```
from konlpy.tag import Okt
okt=Okt()

def pos_filtering(text):
    pos_word_list = okt.pos(text, stem = True) #원형으로
    pos_list = ['Noun', 'Verb', 'Adjective'] #명사 - 동사 - 형용사 추출
    pos_filtered_word_list = []

    for word, pos in pos_word_list: # 품사 필터링
        if pos in pos_list:
            pos_filtered_word_list.append(word)

    return pos_filtered_word_list

dfkakao['token_REVIEW']=dfkakao.cleaning_REVIEW.map(lambda x : pos_filtering(x))
dfsamsung['token_REVIEW']=dfsamsung.cleaning_REVIEW.map(lambda x : pos_filtering(x))
dfnaver['token_REVIEW']=dfnaver.cleaning_REVIEW.map(lambda x : pos_filtering(x))
```

#### ① Okt 라이브러리 이용하여 명사, 형용사, 동사를 추출

Mecab을 사용하지 않은 이유

Okt에서 제공하는 Stem=True와 같은 기능을 찾지 못함  
→ 별점별 빈도분석 비교에서 제대로 된 분석을 위하여  
리뷰데이터의 동사를 원형으로 바꿔줘야 한다고 판단

```
#한글자 지우기
dfkakao['token_REVIEW'] = dfkakao['token_REVIEW'].map(lambda x : [word for word in x if len(word) != 1])
dfsamsung['token_REVIEW'] = dfsamsung['token_REVIEW'].map(lambda x : [word for word in x if len(word) != 1])
dfnaver['token_REVIEW'] = dfnaver['token_REVIEW'].map(lambda x : [word for word in x if len(word) != 1])

C:\User\user\anaconda3\lib\site-packages\ipykernel_launcher.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
This is separate from the ipykernel package so we can avoid doing imports until

#불용어 제거
stopwords = ['하다', '있다', '되어다', '그렇다', '되다', '없다', '안되다', '해주다', '아니다', '가능하다', '됐다', '같다', '주사', '해보다', '수가', '보이다', '차다', '피로', '정말', '진짜',
'씨다', '싫다', '어떻다', '자다', '보다', '이다', '이렇다', '나다', '오다', '가다', '하란', '하나요', '술다', '양다', '이제', '갑자기']
dfkakao['token_REVIEW'] = dfkakao['token_REVIEW'].map(lambda x : [word for word in x if word not in stopwords])
dfsamsung['token_REVIEW'] = dfsamsung['token_REVIEW'].map(lambda x : [word for word in x if word not in stopwords])
dfnaver['token_REVIEW'] = dfnaver['token_REVIEW'].map(lambda x : [word for word in x if word not in stopwords])
```

#### ② 한 글자 단어들은 지워주고 불용어 제거

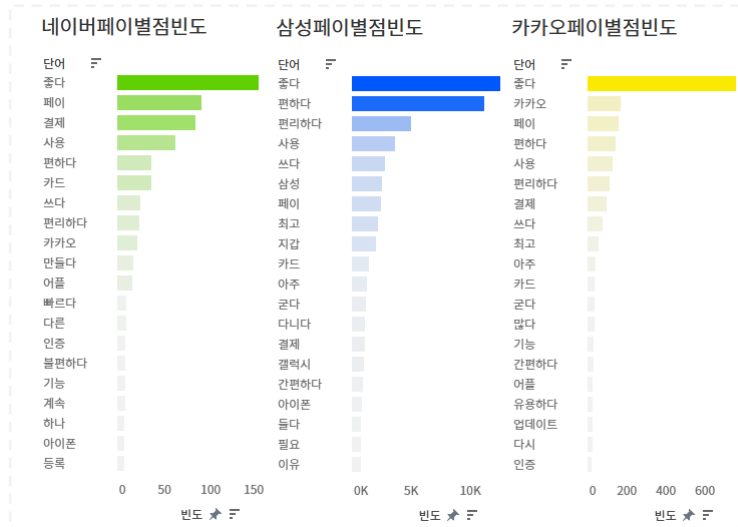
```
# stopwords = ['네이버', '페이', '삼성', '카카오']
wordlist=[]

for data in [naver,samsung,kakao]:
    lists=[]
    for i in range(1,6):
        df=data[data.STAR==i]
        word_list=sum(df['token_REVIEW'].to_list(),[])
        # if word_list not in stopwords:
            c = Counter(word_list)
            lists.append(c.most_common(50))
    wordlist.append(lists)
```

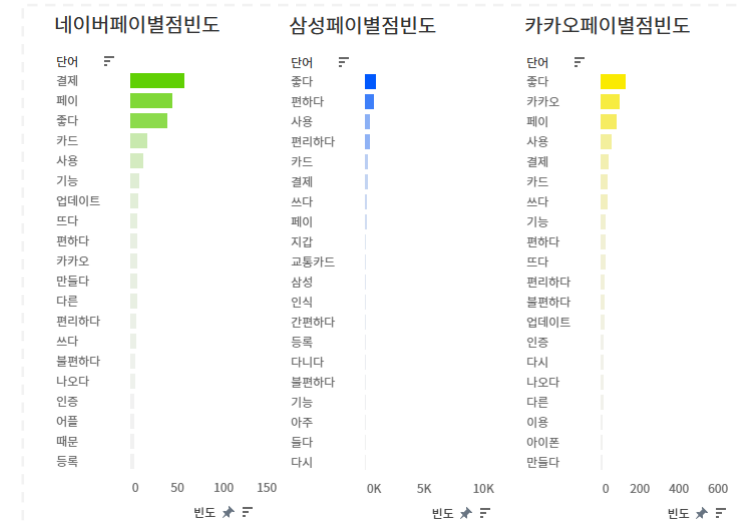
#### ③ Counter을 이용하여 빈도분석

## 2.빈도분석

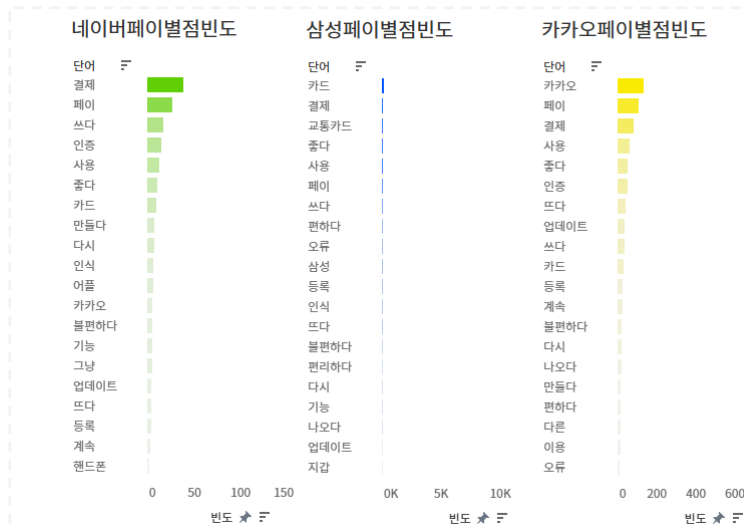
★★★★★



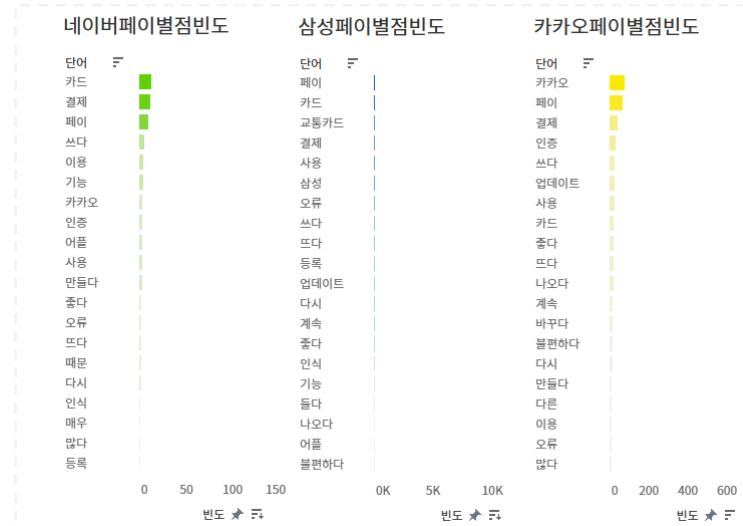
★★★★☆



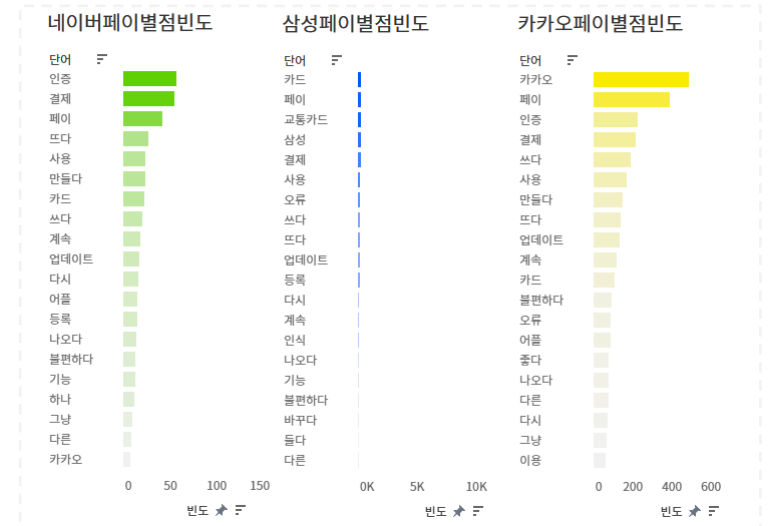
★★★★☆



★★★☆☆



★★☆☆☆



## 3. 감성분석

### WHY?

- 앞선 분석에서는 단순히 별점마다 빈도가 높은 단어를 통해 별점별 어플 리뷰를 분석해 보았음
- 이 과정에서 별점별 비교는 잘 되었으나 어플별 비교는 잘 안됨
- 더 구체적으로 어떤 단어가 들어있는 문장일수록 긍정적인 리뷰이고 부정적인 리뷰인지 파악하고 싶음

### How!

- 각 어플마다 긍정, 부정 리뷰를 분류할 수 있는 감성분석 모델을 만들어 어떤 키워드가 긍부정 리뷰에 영향을 미치는지 비교해보자

# 3. 감성분석

## DATA 전처리

```
def pos_filtering(text):  
    pos_word_list = okt.pos(text)  
    pos_list = ['Noun'] #명사 추출  
    pos_filtered_word_list = []  
  
    for word, pos in pos_word_list: # 품사필터링  
        if pos in pos_list:  
            pos_filtered_word_list.append(word)  
  
    return pos_filtered_word_list
```

#명사토큰화

```
kakao['token_REVIEW']=kakao.cleaning_REVIEW.map(lambda x : pos_filtering(x))  
samsung['token_REVIEW']=samsung.cleaning_REVIEW.map(lambda x : pos_filtering(x))  
naver['token_REVIEW']=naver.cleaning_REVIEW.map(lambda x : pos_filtering(x))
```

#한글자 지우기

```
kakao['token_REVIEW'] = kakao['token_REVIEW'].map(lambda x : [word for word in x if len(word) != 1])  
samsung['token_REVIEW'] = samsung['token_REVIEW'].map(lambda x : [word for word in x if len(word) != 1])  
naver['token_REVIEW'] = naver['token_REVIEW'].map(lambda x : [word for word in x if len(word) != 1])
```

이전에 data cleaning 작업을 거친 데이터에서 명사만 추출  
나머지 전처리과정은 이전과 동일

#감성 라벨링 (1~3 => 0; 4~5 => 1)

```
naver.STAR=naver.STAR.apply(lambda x : 1 if x>3 else 0)  
kakao.STAR=kakao.STAR.apply(lambda x : 1 if x>3 else 0)  
samsung.STAR=samsung.STAR.apply(lambda x : 1 if x>3 else 0)
```

별점 1점~3점은 부정 리뷰, 4점~5점은 긍정 리뷰로 라벨링

### 3. 감성분석

#### DATA 분포 확인

##### 네이버 페이

```
naver.STAR.value_counts()
```

```
1    309
0    267
Name: STAR, dtype: int64
```

Data 불균형 x

Train Data, Test Data → 8:2

##### 카카오 페이

```
kakao.STAR.value_counts()
```

```
0    1733
1    1276
Name: STAR, dtype: int64
```

Data 불균형 x

Train Data, Test Data → 7:3

##### 삼성 페이

```
samsung.STAR.value_counts()
```

```
1    21316
0     1853
Name: STAR, dtype: int64
```

심한 Data 불균형 존재

Train Data, Test Data → 7:3

SMOTE 이용  
데이터 불균형 해결

1	14954	→	1	14954
0	1264		0	14954

Train Data

#TF-IDF

```
vectorizer = TfidfVectorizer(min_df=1, norm='l2', ngram_range=(1,1))
train_features = vectorizer.fit_transform(train_texts)
```

## 3. 감성분석

### #최적모델찾기

```
clf = LogisticRegression(random_state=0)
params = {"C": [0.1, 0.5, 1, 5, 10]}
model = GridSearchCV(clf, param_grid=params, cv = 10, scoring='accuracy')
model.fit(train_features, train_labels)
```

```
model = LogisticRegression(C=5)
```

```
model.fit(train_features, train_labels)
test_features = vectorizer.transform(test_texts)
pred_labels = model.predict(test_features)
print('잘못 분류된 강의평: {} out of {}'.format((pred_labels != test_labels).sum(), len(test_labels)))
print('Accuracy: %.2f' % accuracy_score(test_labels, pred_labels))
```

잘못 분류된 강의평: 28 out of 116  
Accuracy: 0.76

```
model = LogisticRegression(C=1)
```

```
model.fit(train_features, train_labels)
test_features = vectorizer.transform(test_texts)
pred_labels = model.predict(test_features)
print('잘못 분류된 강의평: {} out of {}'.format((pred_labels != test_labels).sum(), len(test_labels)))
print('Accuracy: %.2f' % accuracy_score(test_labels, pred_labels))
```

잘못 분류된 강의평: 190 out of 903  
Accuracy: 0.79

```
model = LogisticRegression(C=10)
```

```
model.fit(train_features, train_labels)
test_features = vectorizer.transform(test_texts)
pred_labels = model.predict(test_features)
print('잘못 분류된 강의평: {} out of {}'.format((pred_labels != test_labels).sum(), len(test_labels)))
print('Accuracy: %.2f' % accuracy_score(test_labels, pred_labels))
```

잘못 분류된 강의평: 655 out of 6951  
Accuracy: 0.91

- Logistic Regression 사용
- Grid Search 사용하여 최적의 하이퍼파라미터 구함
- 후보 C : 0.1, 0.5, 1, 5, 10
- 10 cross validation

### 네이버페이

- C = 5
- Accuracy : 0.76

### 카카오페이

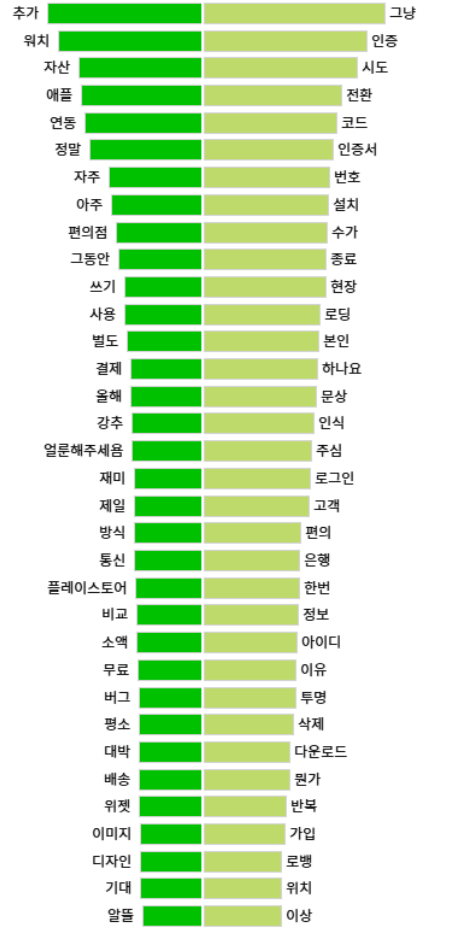
- C = 1
- Accuracy : 0.79

### 삼성페이

- C = 10
- Accuracy : 0.91

### 3. 감성분석

네이버긍정



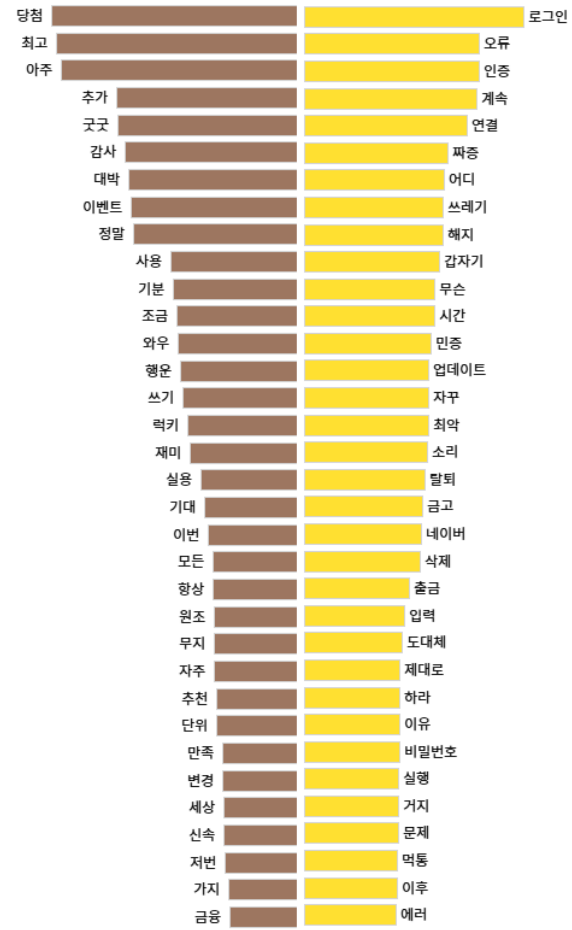
계수 ㄹ 4 3 2 1 0 0 -1 -2 -3 -4

삼성긍정



계수 ㄹ 15 10 5 0 0 -2 -4 -6 -8

카카오긍정



계수 ㄹ 3 2 1 0 0 -1 -2 -3

## 4. Clustering

비슷한 리뷰끼리 묶어 유저들의 요구사항을 파악하고자  
Clustering 진행

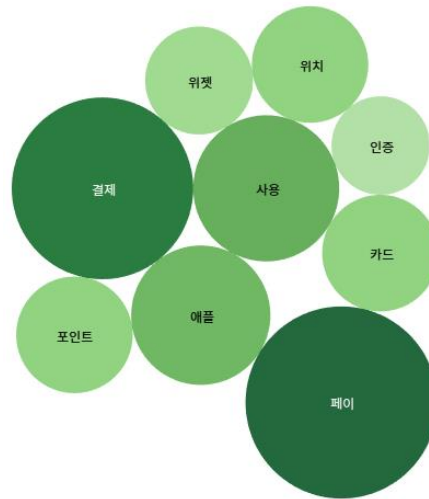
### DBSCAN 알고리즘 사용

- 밀도 기반의 클러스터링 기법
- 클러스터 개수를 정할 필요 x
- 노이즈 벡터를 따로 분류하여 노이즈 처리에 강함

```
tfidf_vectorizer = TfidfVectorizer(min_df = 5, ngram_range=(1,2))
tfidf_vectorizer.fit(samsung.token_REVIEW)
vector = tfidf_vectorizer.transform(samsung.token_REVIEW).toarray()
vector = np.array(vector)
model = DBSCAN(eps=0.3,min_samples=5, metric = "cosine")
result = model.fit_predict(vector)
samsung['result']=result
```

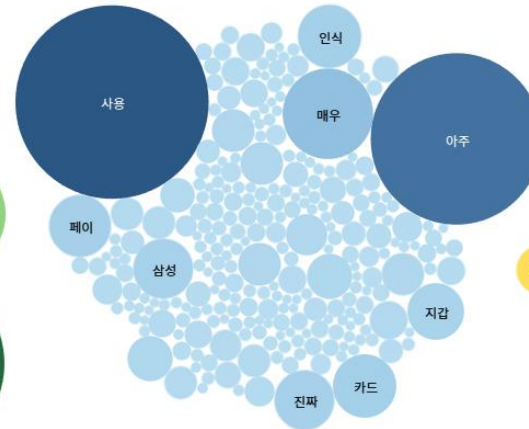
각 클러스터별 대표 키워드 도출 → 빈도분석활용

네이버페이



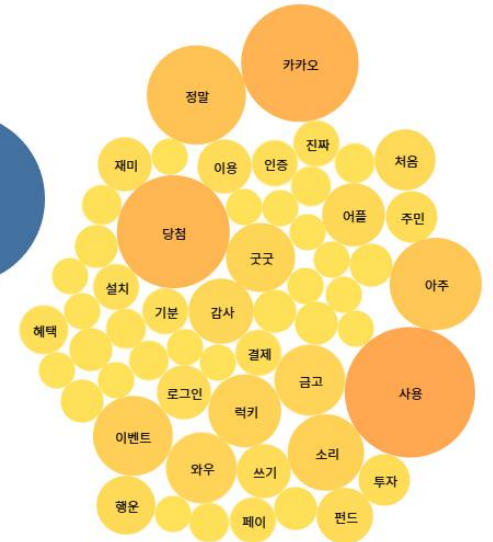
군집 수 : 11

삼성페이



군집 수 : 406

카카오페이



군집 수 : 61



# 4. Clustering

## Keyword “결제” in NAVER PAY

DATE	REVIEW	cluster_num	키워드
2022-01-02 20:30:02	카드결제 먼저 가능하게 워젯 있었으면 더 좋겠어요. 그거 말고는 빠르고 편리합니다.	6	결제
2022-01-02 01:00:55	카드결제 드디어되네요	6	결제
2021-10-18 07:09:01	아. 카드결제 기능이 없어서 아직 아니네. 카드결제 빨리 만들어주세요. 현기증 난단 말이에요.	6	결제
2021-08-28 21:43:01	설치했는데 내일되서 eu에서 결제하려 하는데 아직도 대기중 이라고 떠서 카드로 결제했어요 ㅠㅠ	6	결제
2021-08-24 01:24:20	카드결제도 가능하게 해주세요	6	결제
2021-08-09 19:42:40	나왔네요!! 기대하며 써보겠습니다. 카드 등 록한 카드로는 결제가 안되지만 기다립니다. 많이 업그레이드 부탁드립니다.	6	결제
2021-08-08 22:26:41	네이버 앱 들어갈 필요없고 앱이 확실히 빠르네요. 오프라인에서 카드결제만 되면 완벽하겠군요.	6	결제
2022-01-14 13:21:59	그동안 신용카드만 결제해서 복잡하고 할 게 많아서 너무 화났는데 네이버페이 결제하고 아무리 길어도 2분도 안되서 끝났네요 너무 좋아요. 앞으로도 자주 쓸게요.	7	결제
2021-12-24 14:58:28	네이버 페이앱은 결제나 실행 잠금이 되는데 네이버 앱은 그런게 없어보이네요. 앱 뜯다 쓰면 보안 무용지를 아닌가요?	7	결제
2021-12-24 12:54:20	타 앱에서 네이버페이 결제할 때 네이버앱 말고 이걸로 결제할 수 있게 해주심 참 좋겠습니다. ㅠ	7	결제
2021-08-30 05:45:09	오프라인 매장에서 네이버페이 결제할 때 편리해서 좋아요. 그런데 네이버페이 결제내역 확인이 좀 메인에 있으면 좋을텐데 숨겨놔서 한창 찾음요. — 네이버페이를 온라인에서 많이 쓰는데 너무 오프라인 전용같이 만들어놓게 아쉽네요.	7	결제
2021-08-10 16:04:49	웹페이지에서는 결제내역이 있는데 앱에서는 아무리 찾아봐도 없네요. 네이버에서 결제한 내역이요.	7	결제
2021-12-05 01:23:30	결제할 수 있는곳이 많이생겼음 좋겠어요.	9	결제
2021-09-30 00:23:41	결제잘되네요	9	결제
2021-09-26 11:49:06	다이소 결제 안되요	9	결제
2021-08-07 14:50:36	결제하기 넘 편리하네요.	9	결제
2021-08-06 08:05:03	결제만 온전히 쓸 수 있으니 편하네요	9	결제

### Insight

- 카드결제기능을 원하는 리뷰들이 많음
- 카드결제 기능이 추가됨
- 카드결제 시스템에 대해 추가적인 요구

## 4. Clustering

Keyword “결제” in KAKAO PAY

	DATE	REVIEW	cluster_num	키워드
106	2021-11-22 00:37:06	결제가 안되요	33	결제
114	2021-11-12 18:27:42	간단히 신속히 결제가 편리 해요.	33	결제
119	2021-11-01 17:01:17	왜 결제가 안되죠...?...	33	결제
181	2021-10-01 07:07:44	편하구 쉽게 결제가 가능해서 좋아요	33	결제
284	2021-04-13 17:30:12	결제 편해요	33	결제
368	2021-03-09 16:53:30	편안한 결제	33	결제
523	2021-01-12 15:23:57	결제할때 짱편해용	33	결제
667	2020-08-05 15:28:34	잘 사용하고 있습니다. 결제 간편하게 할 수 있어서 좋아요.	33	결제

### Insight

- 결제와 관련된 리뷰는 적음
- 소수의 유저가 결제 오류가 나서 리뷰를 작성

# 4.Clustering

## Keyword “결제” in SAMSUNG PAY

DATE	REVIEW	cluster_num	키워드
2022-05-12 16:43:14	간편결제 좋음	32	결제
2022-04-29 23:25:16	간편결제	32	결제
2022-04-28 00:20:31	쉽고 간편 결제 🍌	32	결제
2022-04-24 19:50:23	사용하기 쉽고 결제 잘되고 편함이 최고의 무기인듯 잘쓰고있습니다~	32	결제
2022-04-21 19:57:05	간편결제 가 좋다.	32	결제
2022-04-07 23:54:52	간편결제가 좋아요	32	결제
2022-03-02 22:22:33	간편결제 좋아요	32	결제
2022-03-02 11:34:48	즐거운 세상 편리한결제 최고	32	결제
2022-03-01 11:09:04	너무편하고 좋아요 간편결제는 삼성페이 굿	32	결제
2022-02-17 07:11:00	간편결제의 수장	32	결제
2022-02-12 06:13:10	간편결제	32	결제
2022-01-26 20:41:22	간편결제 너무편해요~	32	결제
2022-01-19 19:44:11	대한민국의 간편결제 시장이 크지 못하는 강력한 이유.	32	결제
2022-01-03 09:17:57	간편결제의 역사는 삼성페이의 등장 전후로 나뉜다	32	결제
2021-10-30 00:45:09	간편결제~굿^^	32	결제
2021-10-22 08:56:54	빠른간편결제	32	결제
2021-09-29 11:15:49	간편하고 편리해요 안드로이드의 최애템입니다 카드사별로 간편결제앱 안 깔아도 되서 좋아요	32	결제
2021-09-13 19:49:43	플립3 지문 간편결제 최고	32	결제
2021-09-11 20:44:53	최고 간편결제 짱	32	결제
2021-09-06 03:44:40	간편결제 사용에 편리합니다.	32	결제
2021-09-01 05:34:18	간편결제 최고	32	결제

2828	2021-08-17 12:38:45	간편결제하기 좋아요	32	결제
3088	2021-07-09 21:40:49	간편결제에서 가장 편리하고 오류가 없음	32	결제
3250	2021-06-16 11:24:09	간편결제 편리하고 좋아요	32	결제
3722	2021-04-24 10:21:50	모바일 간편결제 너무 편해서 좋아요	32	결제
4192	2021-03-01 14:44:49	간편결제	32	결제
4755	2021-01-18 10:59:09	간편결제 편리하네요	32	결제
4952	2020-12-24 13:02:26	간편결제 감사	32	결제
5244	2020-11-20 11:03:42	간편결제 좋아요	32	결제
5620	2020-10-16 16:50:13	언제나 간편 결제 참 편해요.	32	결제
5749	2020-10-06 22:18:43	결제가 빠르고 좋아요 최고!	32	결제
5949	2020-09-21 16:40:58	모바일결제시 삼성페이 되는건 정말 최고예요. 간편결제보다 더편한 삼성페이 주변 모르는사람 많은데 널리알리고싶을정도 ㅎㅎㅎ 잘쓰고있어요!	32	결제
6268	2020-08-31 22:38:37	간편결제 넘 편해요	32	결제
6676	2020-08-05 10:20:19	간편결제 🍌	32	결제
7013	2020-06-26 20:09:58	간편결제 좋아요	32	결제
7543	2020-05-06 14:16:12	전자결제 최고의앱	32	결제
7673	2020-04-27 00:05:59	모바일결제하기 편하고 간편결제까지 한번에 해결할수있어 편해여 결제한번으로 할인도 받고 놀 쓰게되네여	32	결제
7888	2020-04-06 20:31:17	간편결제는 당연 삼성페이가 최고네요!	32	결제
8286	2020-03-08 23:43:14	말할 필요없는 간편결제~!!	32	결제
8463	2020-02-29 15:41:58	간편결제 편하고 좋아요	32	결제
8790	2020-01-30 23:54:46	편리한 간편결제 앱입니다	32	결제
331	2022-04-15 10:34:44	최고의 편리한 스마트 결제 시스템	113	결제
359	2022-04-12 08:07:47	이보다 편한 결제 시스템은 대한민국에 없다	113	결제
361	2022-04-11 21:47:06	간편한결제시스템 좋아요	113	결제
491	2022-03-29 18:52:07	항상 쓰는 결제 시스템	113	결제
789	2022-03-02 22:43:19	세상최고로 편한 결제시스템 대 박 만 족	113	결제

### Insight

- 유저들이 결제와 관련된 리뷰를 타 어플보다 많이 남김
- 대부분의 유저가 결제가 편하다는 긍정적인 댓글을 남김

## 5. 아쉬운 점

날짜	내용
0	2022.05.25. 오후 5:53 [Wn테라 지불결제 서비스 제공한 관련 업체들 줄줄이 점검할 듯테라-루나 사태 금융...
1	2022.05.26. 오전 9:48 [WnWnWnWnWn[서울=뉴시스]최원석 BC카드 사장(왼쪽)과 패트리코 배런 알토...
2	2022.05.25. 오전 9:54 [WnWnWnWnWn롯데카드가 카카오뱅크와 '카카오뱅크 롯데카드' 회원을 위해 고객...
3	2022.05.26. 오전 8:17 [WnWnWnWnWn@CJ프레시웨이[데일리안 = 임유정 기자] CJ프레시웨이가 간편...
4	2022.05.26. 오전 4:07 [WnWtWtWt◆ 한국의 우수콜센터 ◆ WnWnWnWn 현대홈쇼핑이 한국능률협회컨...
...	...
622	2021.11.16. 오후 5:01 [Wn포항사랑카드 개인한도 30만 원이내 충전 후 가맹점에서 모바일 결제 경북 ...
623	2021.10.22. 오전 1:02 [Wn"아이폰 유저분들~ 이제 세븐일레븐에서 터치만으로 옛지있게 간편결제하세요" ₩...
624	2022.05.19. 오후 3:33 [WnWnWnWnWn서울 여의도 금융감독원 @금융감독원[데일리안 = 김효숙 기자] ...
625	2021.12.15. 오전 10:36 [WnNHN페이코, '연말결산 리포트' 공개영화관, 캠퍼스, 유희미족 등 오프라인 ...
626	2021.10.20. 오후 1:33 [Wn"동일 행위, 동일 규제" vs "비교 불가" 팽팡금융위, 빅테크 간편결제 수...

기존에 진행하려던 뉴스데이터를 활용한 간편결제 시스템 동향파악을 진행하지 못함

- 네이버 뉴스 기사 크롤링에 어려움을 겪어 데이터 수집이 상대적으로 오래 걸림
- 전처리까지 모두 진행하고 토픽모델링까지 시도했으나 파라미터 조정을 하지 못해 토픽 분류가 제대로 되지 않음

군집 별 키워드를 도출할 때 빈도 기준으로 키워드를 뽑았더니 만족스러운 결과는 아님

크롤링에 시간을 너무 많이 쏟아 데이터 전처리에 투자할 시간이 적었음

시각화 Tool 태블로를 활용하여 결과물을 정리해서 보여주고 싶었으나 시간 부족



감사합니다