|  |
| --- |
| PANSTAR DREAM |
| 선상 해커톤 빅데이터 경진대회 |
| 텍스트 마이닝을 통한 마케팅 솔루션 제안 |
|  |
| **정규면, 오치훈, 한승철, 박다예, 전민관** |
| **2019-02-19** |

|  |
| --- |
| 팬스타 드림에서 제공한 데이터와 인터넷 온라인에서 크롤링한 데이터를 활용하여 캔스타 드림의 마케팅 방향을 제공하고자 하였다. |

**목차**

**1. NAVER 팬스타 크루즈**

**-1.제거할 형태소 목록**

**-2. 빈출 어휘 출력 -3. 선택 형태소 목록 및 감성 분석**

**2. 선상 팬미팅**

**-감성분석**

**공통사항**

**- 형태소 분석 : KONLPY - Twitter**

**- '검색','문의','전화','후원'이라는 형태소는 광고이기 때문에 제거가 선행된다.**

**I. naver 팬스타크루즈**

**-** 네이버 블로그에 '팬스타크루즈'를 검색어로 데이터를 크롤링한 데이터

**1.제거할 형태소 목록**

'팬스타크루즈','팬','스타','크루즈','부산','오사카','여행','일본','바다','배,'타고','그램','오늘','한번','때','박','배타','번','오후','층','일','년','월','국제','터미널','항','호','수','박일','시','것','등','위']

(1)'팬스타크루즈','팬','스타','크루즈' -> **검색어**

(2)부산','오사카','여행','일본','바다','배', '국제','터미널 -> **팬스타 크루즈의 아주 기본적인 정보**

(3)'또','박','배타', '타고','번','오후','층','일','년','월', '항','호','수','박일','시','것','등','위' -> **의미 있는 명사가 아니거나 명사가 아님**

**2. 빈출 어휘 출력**

'출발': 297, '시간': 271, '여객': 184, '이용': 159, '도착': 150, '페리': 49, '투어': 149, '사진': 139, '선상': 133, '날': 131, '드림': 130, '관광': 130, '분': 128, '곳': 126, '왕복': 124, '상품': 124, '나잇': 123, '맞이': 120, '선박': 119, '저': 118, '공연': 110, '교토': 106, '한국': 105, '이벤트': 103

**3. 선택 형태소 목록(팬스타크루즈 평가 가능한 형태소 선정) 및 감성 분석**

**시간 (크루즈 이동시간)(부정)**

271개

Ex) 팬스타 크루즈 라는 부산-오사카 배편을 타고 갔는데 19시간이나 걸렸다 비행기로 가면 2시간 정도면 되지만 19시간이나 걸렸던 배라는 이유 때문에 배에서 인연도 만들었다. 아들과 같이 일본으로 들어가는 어머니 그리고 사업을 하고 있는 일본인도 만났다. 배는 비행기와 다르게 오랜시간동안 같이...

**공연 => 기분전환이 된다는 의미로 쓰임 (긍정)**

110개

Ex) 나 홀로 떠나는 여행 팬스타 크루즈 내부의 모습들 19시간이라는 장시간의 무료함을 달래주기 위해서 팬스타 크루즈 에서는 저녁시간에 작은 공연들을 한다. 유명하거나 인기가 큰 가수들은 아니나 배안에서 즐기는 공연도 즐겁기만 하다...."

**이벤트** **=> 거의 대부분 팬스타 크루즈의 이벤트에 대해서 알림(긍정)**

103 개

Ex)나가기가 추워서 그냥 선실 창문 통해서 찍었더니.. 사진이 - \_ -;; 뭐.. 어쨋든. \* \* \* 배에서는 저녁에 여러 이벤트도 벌이고 나름 볼거리들이 있기 때문에 크루즈를 이용해보지 않은 분들이라면 한번쯤 이용해보셔도 괜찮을거예요

공연과 이벤트가 고객이 긍정적으로 인식하고 있는지 확신할 수는 없었지만 본 분석으로 긍정적임을 알 수 있다. 공연과 이벤트가 같이 묶일 수 있는 컨텐츠를 고민해보니 팬미팅이 적절하다고 생각하여 선상 팬미팅에 대한 정보를 수집했다.

**II. 선상 팬미팅**

**-** 네이버 블로그에 '선상 팬미팅'를 검색어로 데이터를 크롤링한 데이터

- 네이버 블로그의 내부 내용은 크롤링이 되지 않기 때문에 사람 손으로 자료를 스크래핑

**선상팬미팅1-10.docx, 선상팬미팅11-20.docx**

- 엑셀의 경우 워드문서에서 엔터(\n)가 들어 간 경우에 다음 셀에 내용이 입력된다. 따라서 [출처]를 기준으로 글 하나로 취급하여 엔터로(\n) 구분한다. 그리고 팬미팅과 관련 없는 내용 삭제

- stopwords의 경우 1차적으로 팬스타크루즈의 단어를 이용한다.

**감성분석(대상 : '선상' 팬미팅)**

선상, 즉 배위나 배를 지칭하는 크루즈나 유람선을 포함하여 이러한 환경에서 하는 팬미팅에 대한 감성을 측정한다. 각 문장이 긍정이면 긍정 +1, 부정이면 -1점을 부여한다.

긍정 : 13 부정 : 3(**선상팬미팅- 긍부정.docx**참조)

따라서 '선상 팬미팅'은 이용객들에게 긍정적인 컨텐츠 라고 할 수 있다.