



# 라디오 사연 기반 노래 추천 시스템

텍봇

전민재 20132706

민향숙 20152645

박주란 20152648

# 목 차

1. 주제선정 배경
2. 선행 연구
3. 라디오 사연기반 음악추천 시스템
4. 데이터 수집 및 전처리
5. 모델1 – 유사한 사연 기반 노래 추천 시스템
6. 모델2 – 유사한 사연 기반 노래 추천 시스템
7. 모델3 – 유사한 사연 기반 노래 추천 시스템
8. 모델평가 - 설문조사
9. 활용방안



## 주제선정 배경

수많은 콘텐츠들 중에 사용자에게 적합한 것을  
효율적으로 제공하기 위한 **개인 맞춤형 서비스 시대**가 도래하였다.  
소비자들이 직접 나서서 콘텐츠를 찾아보는 데에는 한계가 있는데,  
방대한 전체 콘텐츠의 양에 비해 그들이 접하는 부분은 다소 한정적이기 때문이다.  
**소비자가 아직 경험 또는 인식하지 못했지만,**  
**흥미를 느낄 만한 새로운 콘텐츠(Long-tail에 속하는)**를  
추천해 주기 위해 다양한 추천 시스템들이 제안, 연구되고 있다.



## 선행연구

### ① 라디오 사연 분석을 통한 음악 추천시스템(2012)

: 라디오 사연을 분석하여 사연 간 유사도에 기반한 노래 추천 알고리즘

### ② MusicSense: Contextual Music Recommendation using Emotional Allocation Modeling (2007)

: 웹 페이지의 텍스트를 읽을 때 텍스트와 관련이 있는 노래를 추천해주는 모델

### ③ Social Tagging and Music Information Retrieval (2008)

: MIR(Music Information Retrieval) 문제 해결을 위한 소셜태깅(Social Tagging)에 대한 연구

### ④ InCarMusic: Context-Aware Music Recommendations in a Car(2011)

: 상황의 유사성에 기반하여 확률론적 모델을 기반으로 노래를 추천해주는 모델



## 라디오 사연 기반 음악 추천 시스템

최종적으로 2개의 모델을 제안하고자 한다.

1

텍스트의 감정과 상황을 분석하여 노래를 추천해주는 시스템을 만들고자 한다. 그러나 한국어는 영어와 다르게 제대로 된 감정사전이 구축되어 있지 않다. 따라서 우리는 텍스트 자체의 감정과 상황을 분석하는 것이 아니라 비슷한 감정과 상황을 가진 텍스트를 찾아서 노래를 추천해주는 시스템을 구축해보고자 한다. 그렇기 때문에 우리는 텍스트와 노래가 하나의 쌍으로 묶여 있는 라디오 사연 데이터를 기반으로 모델링을 진행하고자 한다.

2

기존의 방식은 제대로 된 한국어 감정사전이 존재하지 않기 때문에 고안한 방법이었다. 여기에 우리는 우리만의 감정사전을 구축하여 새로운 모델링을 시도해보았다. 노래 정보를 수집하면서 노래 태그와 가사가 노래의 감정을 반영해준다는 사실에서 고안한 것이다. 그러나 감정 단어라는 것은 의미론적으로 유사함을 뜻하지 단어 적으로 일치성을 의미하는 것이 아니다. 여기서 사연과 노래 태그/가사 간에 단어 적인 일치성을 기대할 수 없다는 문제가 있다. 이것을 해결하기 위해서 단어의 의미를 벡터로 표현해주는 Word2vec 모델을 사용하고자 한다. 사연과 태그/가사를 바탕으로 Word2vec 모델을 생성함으로써 비슷한 감정을 가지는 단어들은 비슷한 벡터 값을 가지도록 하였다. 따라서 우리는 해당 모델을 한국어 감정사전으로 대체하여 모델링을 진행하고자 한다.


# 데이터 수집 및 전처리

## 1) 노래 데이터 : 지니, 멜론, 엠넷에서 DJ 앨범 크롤링

### ① DJ앨범에서 DJ앨범 태그와 DJ앨범 수록곡 크롤링

휴식 X


태그 추가



나른한 주말, 나지막이 흐르는 재즈 선율

#오늘의선곡 #DJ 157 #재즈 #휴식 #오후 #집

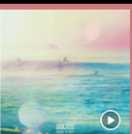
♡ 165 15곡



스타플레이리스트 : 미야오의 '버스타고 집에는 걸, 감성 가득...

#오늘의선곡 #스타DJ #휴식

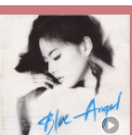
♡ 24 15곡



6월의 눈부신 아침부터 별헤는 밤까지

#오늘의선곡 #DJ 樂PD #발라드 #휴식 #휴가/여행 #해변

♡ 62 15곡



문득 떠나보는 2000년대로의 추억 여행

#오늘의선곡 #DJ 지용 #발라드 #휴식

♡ 85 15곡

6월의 눈부신 아침부터 별헤는 밤까지

구성

태그

#오늘의선곡 #DJ 樂PD

#발라드 #휴식 #휴가/여행 #해변 #기분전환 #부드러운 #달콤한 #국내음악

♡ 62 46,586 15곡

이 음악들과 함께 추억을 만들어 보아요~

수록곡



듣기

+

추가

담기

다운

번호	곡정보	듣기	추가	담기	다운	유비	더보기
1	 오렌만예 김현철   김현철 Vol.1						
2	 시간 0158 ▾ 5집 Big 5						
3	 그대와 단둘이서 김현식 ▾ 비처럼 음악처럼						
4	 설렘 조용필   Hello						
5	 아름다워 윤수일   골든 히트곡 모음집(사랑...						
6	 아마추어 윤종신 & 장재인 ▾ 2017 월간 윤종신 9월호						

## 데이터 수집 및 전처리

### ② 각 사이트별로 노래X태그 빈도수 행렬 생성

songid	songebid	songebnm	title	singer	가을	감사	감성	겨울	설레는	추억	차분	외로운	맑은	사랑	즐거움	분위기	용
10015192	2760765	Gran Bufe Dizzi Laica Shmaltz!			0	0	0	0	0	0	0	0	0	0	0	0	0
1001880	310104	Greatest H Pop	N Sync		0	0	0	0	0	0	0	0	0	0	3	0	0
1001881	310104	Greatest H Gone	N Sync		0	0	1	0	0	0	0	0	0	0	0	0	0
10024197	2761728	Some Oth Lucky To	Bill Evans		0	0	1	0	0	0	2	0	0	0	0	0	0
10024584	2761766	Moonriver Moon Rivi	Audrey He		0	0	2	0	0	0	0	0	0	0	2	0	0
1002536	310155	Next Wavi Blaze It Uj	Mondo Gi		0	0	0	0	0	0	0	0	0	0	0	1	0
10027254	2762033	YYY ACA Summer F Yoyoyo	Av		0	0	0	0	0	0	0	0	0	0	1	0	0
1002851	310185	김현철 Be 그대니까요	김현철		0	0	3	0	0	0	0	0	0	0	0	0	0
1002853	310185	김현철 Be 왜그래	김현철		0	0	0	0	0	2	0	0	0	0	0	0	0

### ③ 3개 음원 사이트에서 수집한 노래 데이터를 가수/제목을 기준으로 JOIN

- 동일한 태그는 값을 더해 줌.

### ④ 멜론에서 노래 가사, 장르, 인기도 값 크롤링

- Pop, Animation, New Age, Classic 등의 장르는 제거

### ⑤ 노래X태그 빈도수 행렬을 하나의 연결된 문장으로 바꾸어 줌.

songtitle	singer	가을	감사	감성	겨울	눈	맑은	봄	...	여유	외로 운	응 원	이 별	즐거 움	차 분	추억	그디 스마 스	행복	호 린
그대니까요	김현철	0.0	0.0	5.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0	0.0	4.0	0.0	0.0	0.0

songtitle	singer	tag																	
그대니까요	김현철	감성	감성	감성	감성	감성	여유	추억	추억	추억	추억								



## 데이터 수집 및 전처리

- ⑥ 가사가 있는 노래는 노래 태그 뒤에 한 문장으로 이어서 붙여 줌.
  - 개인의 주관으로 생성한 태그만으로는 노래 정보를 모두 파악할 수 없다고 판단하여 가사를 추가적으로 사용함.
- ⑦ 노래별로 태그와 가사로 이루어진 문장을 형태소 분석을 통해 명사, 형용사, 숫자만 추출.
  - 형태소 분석을 통해 불필요한 단어들을 제거해줌.
  - 감성힙합, 힙합재즈, 감성재즈를 감성, 힙합, 재즈로 각각 분리해주기 위함.

## ⑧ 결과물 예시

```
['잔잔 새벽 잔잔',  
'잔잔 새벽 잔잔',  
'스트레스 기분 기분 전환 사랑 괜찮 아름 예전 사랑 전하 사랑 자리 사랑 전하 사랑 자리 자리',  
'명곡 명곡 추억 추억 댄스 댄스 댄스 댄스 댄스 댄스 화창 화창 화창 화창 년대 년대 편안함 편안함 드라이브 드라이브 가  
요 가요 가요 가요 가요 가요 사랑 사랑 년대 년대 행복 행복 행복 행복 운동 운동 안타 습관 주위 후회 종일 기억 혼자 사  
실 기억 의미 후회 기회 준비 기대 종일 기억 혼자 사실 기억 의미 후회 혼자 사실 기억 의미 후회',  
'발라드 댄스 캐롤 설렘 드라이브 가요 사랑 풍경 마이 얼굴 매일 아침 창문 상상 눈빛 수필 목소리 파랑 선을 가지 표정  
상념 행복 어젯밤 꿈속 꿈결 눈빛 수필 목소리 길목 첫발 지향 깊이 신비 가까이 행복 소망 인연 미련 현실 천상 꿈속 눈빛  
수필 목소리 컷가 어지 목소리',
```



## 데이터 수집 및 전처리

### 2) 사연 데이터 : KBS, SBS, MBC 약 30개 라디오에서 사연 수집

#### ① 라디오 사연 게시판에서 사연 데이터 수집

##### ▶ 참여코너 > 사연과 신청곡

번호	제목
<b>[공지]</b> 나누고 싶은 <이야기>와 듣고 싶은 <신청곡>을 남겨주세요 (1)	
249	노래 들려주세요~
246	이별과 후회
245	딥 톱스를 적으려다...
244	외로운 한때 길냥이들과의 추억
243	이럴 때
242	미국에서의 정
241	시가지의 마미오 코트

15227 찬란샘이 언니 착하다 했는데 역시.. T.T

15226 군대가기전 현장에서 누나목소리들으며!!

15225 이별을 준비하네요

15224 어린이날 네임핀, 머리띠를 만들어주고 기뻐하는 ...

15223 이 나라의 첫으로 조 마트에서봐요

##### ▶ 매일코너 > 사연과 신청곡

푸른밤에서 함께 나누고 싶은 이야기,  
푸른밤 가족들과 함께 듣고 싶은 신청곡, 이곳에 남겨주세요 :)

게시판 운영 원칙

번호	제목	작성자	작성일	조회
1378	내년에도 만나요!	주여진	2018.06.11	12
1377	오늘 하루는 정말 정신 없는 하루 였습니다.	박기석	2018.06.10	1
1376	마눌!힘내라..	김동수	2018.06.09	6
1374	오늘도 재기만을 위해	홍순찬	2018.06.07	9
1373	비밀독서단에서 만났던 동진님.	최성지	2018.06.06	9
1372	안녕하세요! 곧 새출발을 앞두고 있습니다.	김태용	2018.06.06	5
1371	널고 푸른밤	백승엽	2018.06.05	12

2016-05-02 K 김태인

2016-05-02 K 이훈

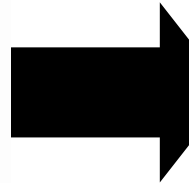
2016-05-02 K 김수경

2016-05-02 K 송하나

## 데이터 수집 및 전처리

### ② 엑셀을 이용해 사연과 신청곡 분리 (350자 이상의 사연만 사용)

안녕하세요~오늘이 저희 엄마의 마지막 출근날입니다.늦각이 대학생으로 젊은애들 사이에서도 밤새 공부하셔서 장학금에 여러 자격증까지 섭렵하며 우수한 성적으로 졸업하여 예전에 컬투쇼에도 방청신청이 되어 오빠들에게 무한 칭찬을들으셨는데요~^^그 만학도 학생이 오늘 5년 넘게 다니던 정든 직장의 마지막 출근길입니다.엄마는 한의원에서 간호조무사로 근무하시는데요.늦은 나이에 취업하셔서 첫직장이라 정이 많이 들었는지 오늘 아침에도 눈가가 축축해지시며 환자분들께 다 인사 못하고 가서 어찌나하는 걱정을 하시더라고요. 외할머니가 아프셔서 병간호겸 엄마도 나이가 많으신데 일을 오래 하셔서 힘이 드시는지 잠시 쉬고싶어하셨는데 오랫동안 정든 환자들 생각에 한달만 한달만 하다가 오늘까지 버텨오신거거든요.마지막 출근날까지 다 인사 못하고 못보고 가는 정든 환자분들 걱정하시는 엄마를 보면 그동안 직장,그리고 환자분들에게 얼마나 애정을 가지고 일하셨는지 마음에 새삼 와 닿으면서 뭉클했습니다.혹시 인사 못하고간 옥수동 ㅇㅇ한의원 강채원선생님 환자분들~많이들 서운해 마시고요. 저희 엄마한테도 정말 그동안 너무 너무 수고했다고 말해주고 싶습니다. 옥상달빛의 수고했어 오늘도 신청합니다^^



	D	E	F
	singer	songtitle	text
	다비치	8282	정민누아 오랜만에 사연을
	2NE1	Come Ba	너무나 사랑스런 그녀(축하
	EXO	12월의 기	부모님이 맞벌이 하시는데
	바이브	1년 365일	내생에 첫번째 생일 축하는
	러블리즈	Ah-Choo	사장님 생신축하드려요 정
	서인국	All For You	퀴즈임파서블 신청합니다~
	Crush	Beautiful	명절에 일하는 서비스직분
	Crush	Beautiful	나이는 숫자에 불과.. 랩을
	조용필	Bounce	덕분에 산격샷 파티했어요

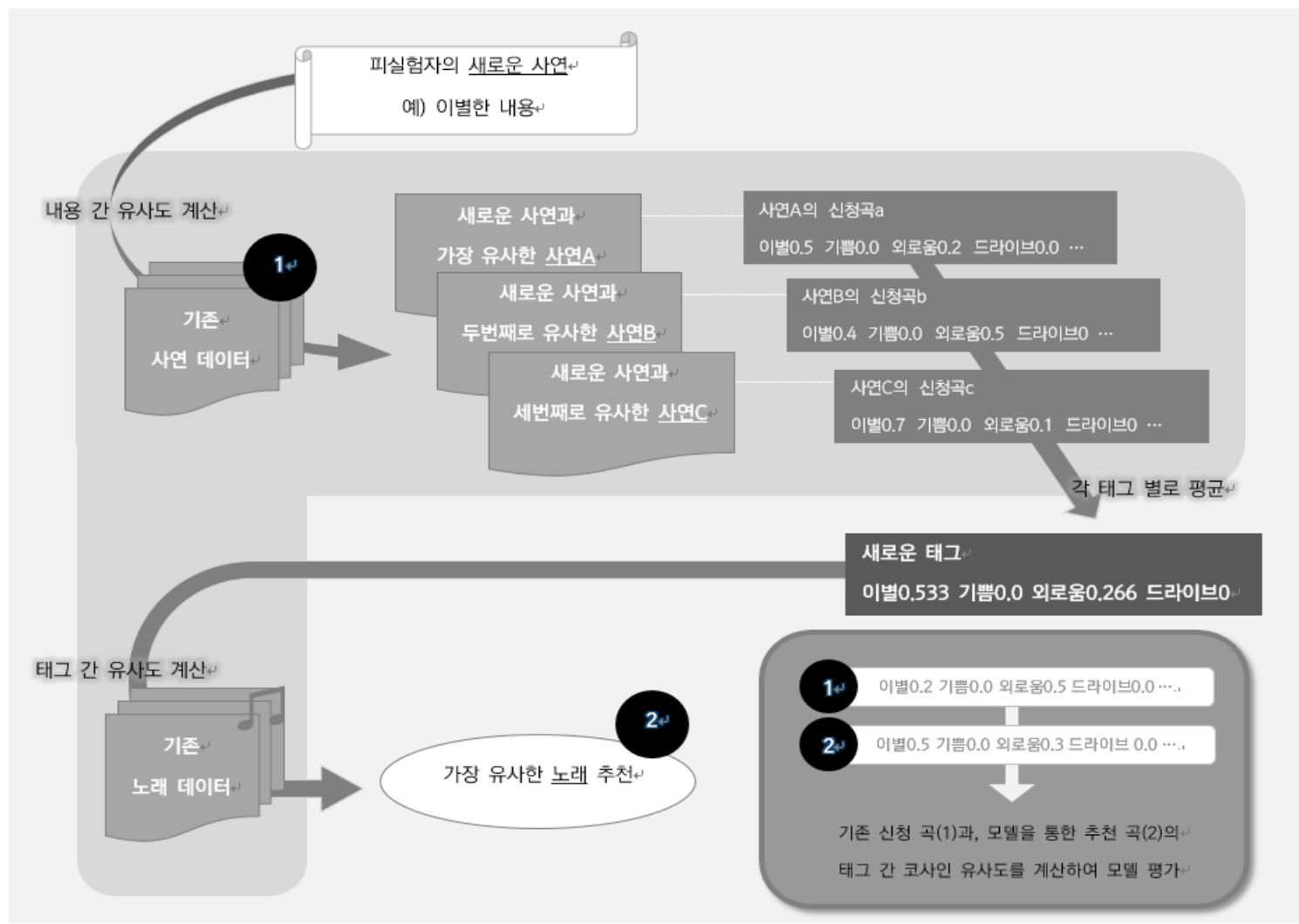
### ③ 신청곡 분리할 때 노래 데이터와 텍스트 동일하게 맞춤.

### ④ 신청곡이 없는 데이터는 따로 모아 둬.

- 신청곡이 있는 데이터 : 1905개
- 신청곡이 없는 데이터 : 4552개



# 모델1 – 유사한 사연 기반 노래 추천 시스템 (모델구성)





## 모델1 – 유사한 사연 기반 노래 추천 시스템 (모델구성)

### ① 라디오 사연 형태소 분석

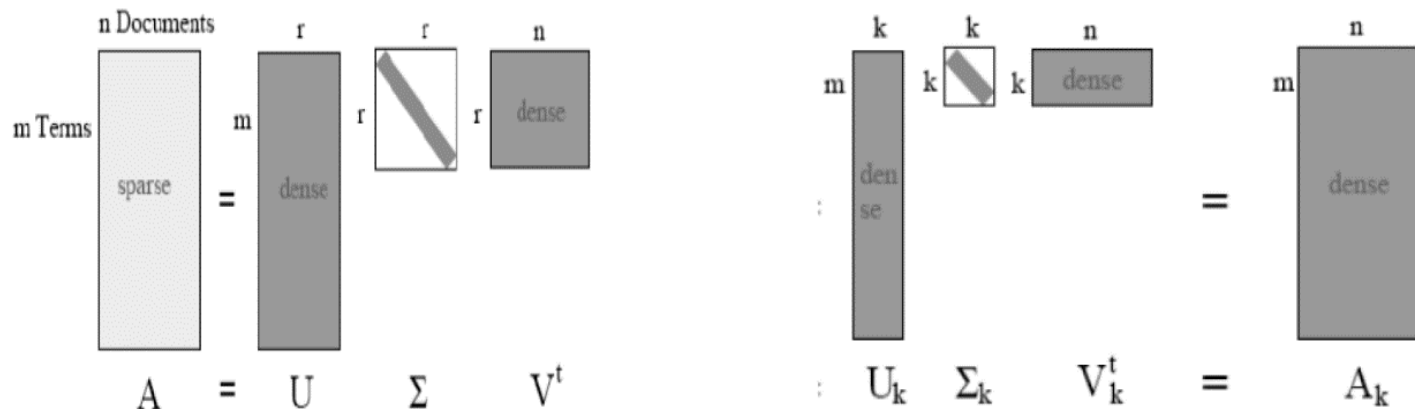
- Python의 Twitter 분석기를 이용해 명사, 형용사만을 추출함.
- 총 단어 개수는 14324개.

### ② TF-IDF를 활용한 문서 분석

- Python의 sklearn TfidfVectorizer 함수를 이용해 TF-IDF 행렬 생성.
- 출현 빈도수가 5이하인 단어는 제거
- 총 단어 개수는 3084개
- 사연의 특징을 나타내는 단어에 더 가중치를 줌.

### ③ SVD를 통한 차원 축소

- 고유값의 개수를 K로 축소하여 TF-IDF 행렬 차원 축소.
- 최적의 K값은 모델 평가 방법에서 설명하고자 함.





## 모델1 – 유사한 사연 기반 노래 추천 시스템 (모델구성)

### ④ 유사한 사연 선정

- 사연간 코사인 유사도 값을 구함.
- 대각행렬은 0으로 만들어 줌.
- 코사인 유사도 값이 높은 N개의 사연을 선정.
- 최적의 N값은 모델 평가 방법에서 설명하고자 함.

### ⑤ 노래 태그 행렬 생성

- 태그와 가사가 문장으로 이루어져 있는 노래 데이터를 행렬로 바꾸어 줌.
- Sklearn의 CountVectorizer 함수를 이용해 2번 이하로 등장한 단어는 제외.
- 40439(노래) X 1727(태그) 행렬 생성.

### ⑥ 유사한 사연들 기반 새로운 노래 벡터 생성

- 4번에서 선정한 사연들의 신청 곡들을 추출.
- 각 신청곡들의 노래 태그 벡터를 찾아서 태그별로 평균값 도출.

### ⑦ 노래 태그 벡터간 유사도 기반 음악 추천

- 새로운 노래 태그 벡터와 기존 노래 태그 행렬 간 코사인 유사도 값을 구함.
- 코사인 유사도 값이 가장 높은 노래를 추천해 줌.



## 모델1 – 유사한 사연 기반 노래 추천 시스템 (객관적 모델평가)

➔ 기존 신청곡과 모델의 추천곡 간의 코사인 유사도 값

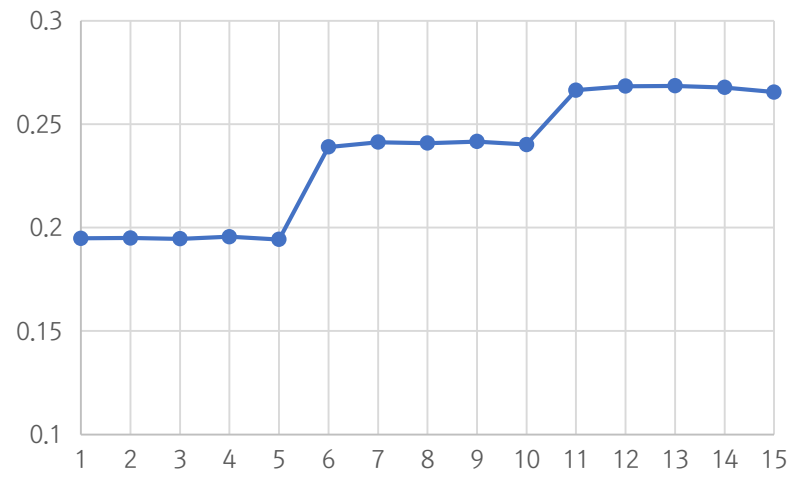
모델의 성능을 평가하는 객관적 지표로는 사연의 기존 신청곡과 모델의 추천곡 간의 코사인 유사도 값으로 한다. 기존의 사연과 유사한 노래를 신청해 줄 수록 해당 모델의 성능이 높다고 판단할 수 있기 때문이다. 즉, 전체 1905개 사연의 신청곡과 추천곡 간 각각의 코사인 유사도 값을 구한 후 평균값이 높을 수록 좋은 모델이다.

따라서 해당 지표를 바탕으로 SVD 차원수 K값과 사연 개수 N값을 정하고자 한다.

K : 500,700,900,1100,1300    N: 3,5,7    ➔ 총 15개의 경우의 수 발생

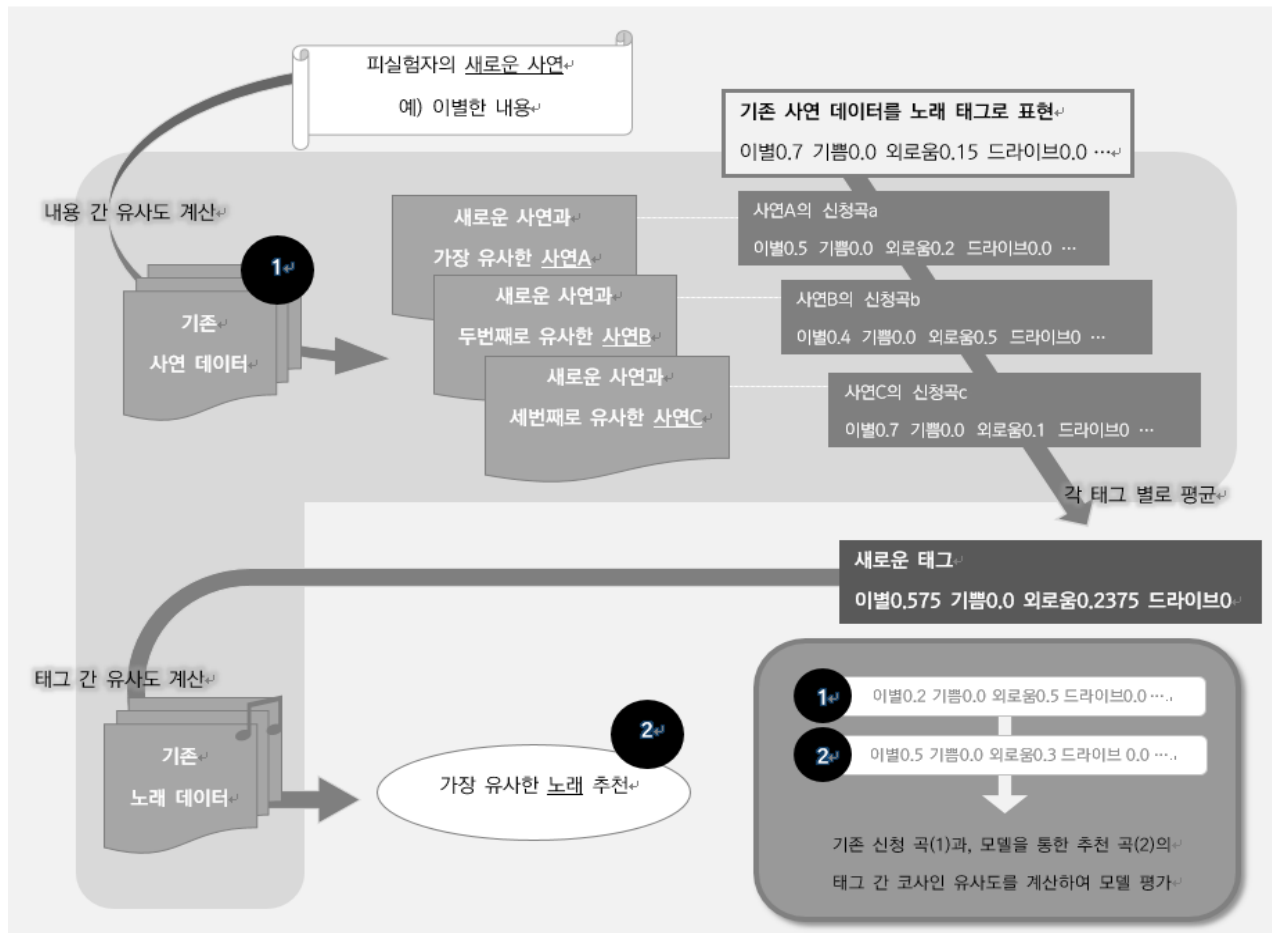
사연 개수	차원 축소	사연 개수	차원 축소	사연 개수	차원 축소
3	500	5	500	7	500
3	700	5	700	7	700
3	900	5	900	7	900
3	1100	5	1100	7	1100
3	1300	5	1300	7	1300

K값-900  
N값-7



## 모델2 – 유사한 사연 기반 노래 추천 시스템 (모델구성)

- 기존 모델에서 사연의 텍스트 내용을 반영한 모델.
  - 사연의 텍스트 내용을 1727개 태그 단어로 표현하는 것.
  - 텍스트 내용을 추가 했지만 동일한 평균값을 보여줌.
- ➔ 노래 태그1727개가 사연 텍스트 내용을 잘 반영하지 못하기 때문에 새로운 태그 벡터 생성시 큰 영향을 미치지 않음.



따라서,  
1번 모델을  
사용하고자 함.



## 모델1 – 유사한 사연 기반 노래 추천 시스템 (모델 결과 사례)

1) 추천곡과 신청곡 간 유사도 값이 0인 것.

태그 간  
유사도

추천곡

추천곡과  
신청곡간 유사도

신청곡

per	songid	text_i	title	artist	rst	songid	songtitle	singer
0.549904	3385862	786	Click Me	Zion.T	0.000000	2104465	마주치지 말자	SUPER JUNIOR

양디. 저는 스무살 여학생이에요. 저희 친할머니께서는 조그마한 구멍가게를 운영하시는데 오늘은 제가 가게를 보았어요. 근데 6시쯤에 아저씨 3분이 오시더니 가게 안에 있는 작은 테이블에 자리하고는 아무것도 사지 않고 수다를 떨더군요. 20분정도가 지나자 저한테 아가씨, 소주 2병만. 하면서 당연한듯이 반말로 술을 가져오라는거예요. 잠시후에는 "아가씨, 라이터 하나." 하면서 또 저를 부르더라구요. 그러더니 저희 가게안에서 담배를 피우는거예요. 작은 구멍가게에서 3명이서 담배를 피우니까 미쳐버리겠더라구요. 그냥 제가 가게를 나와버렸어요. 그 아저씨들은 8시가 넘어서 갔는데, 갈 때도 아가씨, 이거 다 계산해서 내이름으로 외상걸어놔. 하고 자기이름만 말하고 가는거있죠. 너무 화가 치밀어 올랐어요. 그 아저씨들이 자연스럽게 외상을 한 걸 보면 자주 오는 사람들인 것 같은데, 할머니는 이런것들을 매번 익숙하게 겪었을 생각을 하니까 눈물이 나더라구요. 남들한테는 낱아빠진 구멍가게로 보일지 몰라도 저랑 할머니한테는 추억이 많은 소중한 공간인데. 양디, 저 정말 너무 속상해요. 그 아저씨들뿐만 아니라 민폐손님들 모두 저희 가게에 발을 안 들었으면 좋겠어요.





## 모델1 – 유사한 사연 기반 노래 추천 시스템 (모델 결과 사례)

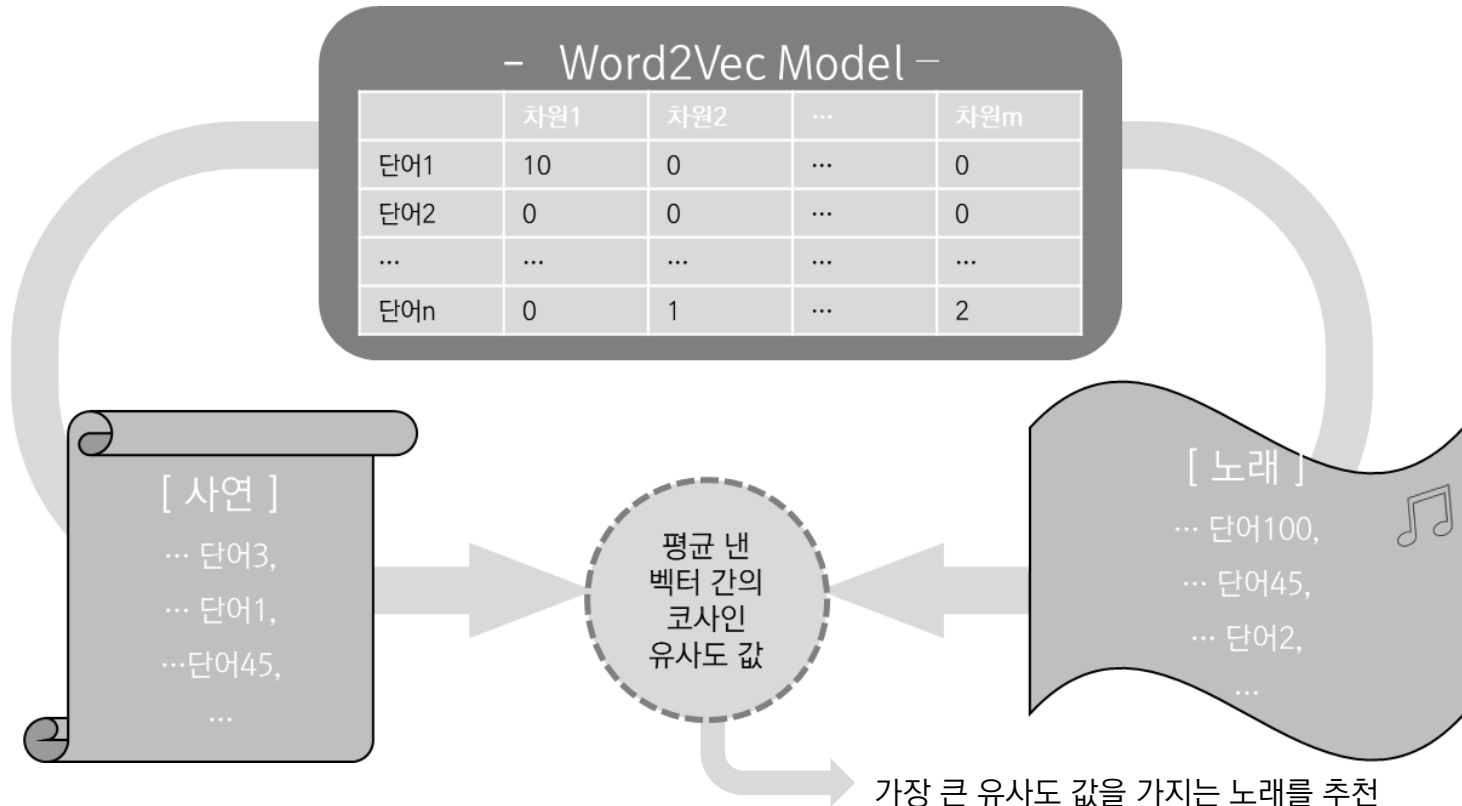
2) 추천곡과 신청곡 간 유사도 값이 1인 것(추천곡과 신청곡이 동일)

태그 간 유사도	추천곡		추천곡과 신청곡간 유사도		신청곡	
0.868454 3849494 1536 이등병의 편지 김광석			1.000000 3849494 이등병의 편지 김광석			

엄마의 바람대로 멋진 군인이 되어라.  
중학교 1학년때부터 제 곁을 지켜주던 친구가 지난달 초 세상을 떠났습니다. 35년동안 제 친구가 되어준 고마운 친구는 많은 사람들의 말대로 하늘나라에서 착한 사람이 부족해서 데려간 거라 믿고 있습니다. 천사처럼 퍼주기 좋아하고 명랑하고 바지런한데다 음식 솜씨까지 좋았던 그래서 아깝고 아까운 친구에겐 입대를 앞둔 아들 녀석이 있습니다. 온몸에 암이 전이되어 투병하면서도 아들이 입대해서 멋진 모습의 군인이 되길 바랐지요. 이제 그 아들이 군대에 갑니다. 눈물 흘려줄 엄마가 함께 할 수 없어 마음이 아픕니다. 엄마 닮아 선하고 너그러운 녀석은 제게 "이모, 저 이제 군대가요" 명랑하게 보고를 합니다. 부디 건강하고 무사히 군생활 마치고 돌아오길 기도합니다. 사랑하는 내 친구 입대하는 모습 지켜보며 흐뭇해할 거예요. 어릴 때 몸이 약해 초등학교때까지 업고 다녔던 녀석이 현역으로 입대를 하게 되었으니 얼마나 자랑스럽겠어요. 엄마의 사랑을 기억하며 아들은 진짜 남자가 되어 돌아오리라 믿습니다.

### 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

- 기존 모델의 형식(사연과 신청 곡 쌍)과 다른 접근 방법이다.
- 사연과 노래 정보(가사, 태그 등)를 Word2Vec으로 워드 임베딩 한다.
- 생성된 모델을 바탕으로 사연과 노래의 유사도(내용 및 감정)를 기준으로 노래를 추천한다.
- 사연 내에서 주제가 명확하게 제시되었을 때, 사연의 '내용' 과 '가사' 간 관계를 중심으로 추천하는 경향이 강함
- 모델 성능의 객관적 평가 척도가 부재하기 때문에, 사용자를 대상으로 설문조사가 필수적임





## 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

### 1. 기본적인 텍스트 처리 방법

- Konlpy의 한글 형태소 분석기 중, 속도가 가장 빠르고 품사 분류 능력이 준수한 편인 Twitter를 사용하였으며 명사, 동사, 형용사를 사용하였다.
- 사연(사연 내용), 노래 정보(가사, 태그, 제목, 장르)를 통합하여 사용하였다.
- 동시 전처리하지 않고, 각각의 특징을 고려하여 기본적인 전처리를 한 후,
- 동사·명사·형용사를 추출한 데이터를 통합하였다.

전체 태그 내  
출현 빈도  
상위 20개 단어

word	count
드라이브	17421
밤	17354
기분	16546
기분전환	16075
가요	15624
발라드	13852
힙	11092
사랑	10838
힙합	10642
팝	10476
새벽	10211
휴식	10056
20대	9946
댄스	9343
비	8044
국내	7466
국내음악	7070
일	7006
2010	6686
2010년대	6629

### 2. 텍스트 전처리 - ① 노래 태그

- 전체 노래 데이터에 대한 설명을 가능하게 하는 속성이다.
- 41148곡 포함 / 고유 개수: 1536개
- 감정 뿐만 아니라 상황까지 나타낼 수 있다.
- 듣는 사람의 입장이 반영되었다는 것이 가장 큰 특징이며,
- 여러 사람들의 '태깅'을 통해 누적된 값을 가진다.

예) 싸이 - 연예인 #신나는#신나는#신나는#댄스...



## 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

### 2. ② 노래 가사

- 전체 노래 데이터의 19666곡만이 가사 데이터를 가진다. (팝송 제외 후)
- 가사가 등록되지 않은 곡들이 많고 전체 노래 데이터 중 제거하지 못한 팝송이 있기 때문
- 이 때 팝송을 어느 정도 제거해보고자 장르를 기준으로 필터링을 거쳤는데, 제거한 장르로는 [Pop, Foreign, Animation, New Age, Adult Contemporary, 동요, World, CCM, Blues, Instrumental, 워십, 찬송가, 기타]가 있다.
- 명사, 동사, 형용사를 기준으로 형태소 분석을 진행하였으며, Komoran을 사용하였다.
- 출현 빈도 수를 기준으로 정렬한 상위 100개 단어 중, 감성이나 상황을 나타낸다고 보기 어렵거나, 대부분의 가사에 등장하는 등의 지나치게 일상적인 단어는 불용어 처리했다.

```
word_except = ['펼치', '모르', '사람', '마음', '시간', '생각', '오늘', '이월',  
              '이제', '그러', '보이', '세상', '가슴', '하루', '모습', '위하',  
              '노래', '순간', '부르', '지나', '그때', '나오', '지금', '애기',  
              '미치', '남자', '여자', '바람', '친구', '변하', '조금', '만들',  
              '거리', '머리', '기분', '시작', '내일', '이러', '어리', '어제',  
              '나가', '동안', '영화', '다음', '이야기', '계절']
```



## 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

### 2. ③ 노래 장르

- 12792개 곡, 39개 장르로 구성되어 있다.
- 앞의 ‘가사’를 처리할 때, 필터링 후의 장르(텍스트 자체)를 추가하였다.

### 2. ④ 노래 제목

- 34612개 곡, 9669개의 단어로 구성되어 있다.
- 노래의 내용을 함축적으로 표현하는 단어/어구/짧은 문장이다.
- 길이 자체가 매우 짧기 때문에 불용어 처리를 따로 하지 않았다.
- 대부분의 노래를 포괄하는 속성이라는 점에서 추가하는 데에 의의를 두었다.
- 영어가 포함된 제목은 구글 스프레드 시트의 ‘translator’ 함수를 사용하여 한국어로 자동 번역 후 사용하였다.
- Twitter와 Komoran의 형태소 분석을 모두 해보고 분석 후 남는 텍스트의 길이가 더 긴 것을 사용



## 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

### 2. ⑤ 사연

- 기존 사연 데이터(사연-신청곡 쌍) 2275개에 신청곡이 없어 기존 모델에 사용하지 못했던 사연 데이터를 추가한 4552개의 사연으로만 이루어진 데이터셋이다.
- 불용어에는 높은 등장 빈도에 비해 의미가 거의 없는 어구 등을 중심으로, 라디오 진행자를 지칭하는 말인 ' x디(D)) ' 을 추가하였다.
- 'stem' 의 동의어 및 유사어 통합 기능에서 착안하여, 형태소 분석 함수 정의를 할 때 여러 조건문들을 사용해 더욱 섬세한 형태소 분석을 가능하게 했다. 예를 들면 '앓다', '앓는', '앓았다' 등은 '앓' 으로 통일시켜 3개 단어로 분산될 것을 1개 단어로 축소시킨다. 또한 '감성힙합' 이라는 장르 관련 단어는 명사+명사 형태의 단어로서, 기본적인 형태소 분석 시 '감성' 이라는 단어만 남지만 감성과 힙합 두 단어로 나뉘지게 하였다. 이를 통해 차원 수를 축소하고, 모델 효율성을 재고하였다.



# 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

## 3. 데이터 통합

: 노래 태그 + 노래 가사 + 노래 장르 + 노래 제목 + 노래 사연

tag	lyric	songtitle	genre	story
2237 [밤 잔잔한 새벽 잔잔]		제목		
2238 [밤 잔잔한 새벽 잔잔]		번개 무서워		
2397 [스트레스 기본 기본전환]	[사랑 만나 괜찮 아름다 예전 사랑 만나 전하 사랑 만나 자리 기다리 사랑 만나 전...		Ballad	
2466 [팝 팝 40대 40대 40대 40대 명곡 명곡 10대 10대 힘내 힘내 20대 2...	[언타깝 습관 주위 후회 종일 해매 박히 기억 혼자 사실 잠들 기억 의미 후회 지내...			
2940 [발라드 댄스 캐롤 선텐 셀 드라이브 가요 사랑]	[종경 아이 얼굴 줄기 매일 아침 만나 날아가 깨어나 창문 상상 빠져 눈빛 수필 들...	천연	Dance	
3428 [까페 가을 비오는날 30대 행복 비]		동화		
3680 [마이크 마이크 팝 팝 팝 팝 팝 발라드 발라드 발라드 발라드 발라드 발...	[인연 영기 연인 남남 살아가 괜찮 잘못 이별 유일 후회 살아가 붙잡 거지 불안 눈...	위해	Ballad	
3737 [발라드 20대 슬픔 비오는날 가요 비]	[기다리 떠나 눈물 지지 살아가 가요 아프 인가 가시 머물 떠나 기나길 의미 사랑 ...	가시		
3739 [발라드 발라드 발라드 발라드 발라드 발라드 2000 20대 슬픔 추억 달콤 달콤 ...	[사랑 손짓 힘들 모질 아프 눈물 얼룩지 포기 우리 허락 건디 힘들 모질 아프 눈물...	약속		
3743 [팝 팝 40대 발라드 발라드 20대 슬픔 재즈 추억 추억 늦은 밤 영상 휴식 댄스...	[인사 가요 잘못 떠나 이별 감당 어렵 약속 지우 얼마 걸어가 뒤돌 보아주 그리 에...	그런	R&B / Soul	
3822 [출퇴근 출퇴근 발라드 공부 공부 2000 2000 추억 1990년대 1990년대 ...	[얼굴 두렵 사진 한장 조작 망설이 지우 가요 잠시 가요 마주치 스치 뒷모습 한참 ...		R&B / Soul	
3824 [발라드 발라드 발라드 10대 20대 20대 슬픔 슬픔 노래방 댄스 댄스 댄스 댄스...	[돌아서 떠나가 이유 어렵 슬프 눈물 돌아서 멈추 사랑 두렵 돌아서 떠나가 이유 부...		R&B / Soul	
3827 [40대 10대 2000 20대 추억 영상 휴식 휴식 댄스 댄스 댄스 댄스 힐링 화창한날...	[이거 이상 상관없 끝나 게임 감당 수없 오해 변명 돌이키 사랑 속삭이 아쉽 생각나...	타임 아웃		
3859 [추억 선텐 셀 사랑 회상]	[머칠 설치 느낌 눈빛 지우 어찌 좋아하 다르 만나 기다리 매력 발취 사로잡 간직 ...	모든 권리 그것		
4067 [슬픔 별 선텐 셀 사랑 이별]	[별을 무너지 보내 어둠 보내 영혼 대신 바지 영혼 대신 바지 말기 가지 어둠 하늘 ...	아가페	Rock	
4077 [20대 추억 댄스 기본 기본전환 행복 회상]	[만나 사이 보통 사이 포기 사랑 여백하 막히 떠나 어떨 사랑 약속 잊어버려 오가 ...	가니	Dance	
4078 [마트에서 20대 추억 회 화창한날 편안함 R&B/힙합 가요 기본 힙합 사랑 30대...	[머릿결 사랑 빠져 사랑 빠져 예쁘 달려오 순정 가슴속 내리 사랑 마지막 우리 나누...	키스		
4104 [휴식 주말 힐링 새벽 꽃 가을 따뜻한 드라이브 오후 겨울]	[사랑 멈추 회색빛 하늘 사이 여름 희망 사랑 이름 목소리 날카를 입맞춤 기억 거짓...	서천 화무		
4112 [명곡 20대 추억 신나 댄스 댄스 드라이브 신나는 사랑 운동]	[떠르오 자신 끝나 흔들리 이해 미워하 전부 두려워하 놓치 떠나 떠나 끝나 흔들리 ...	내 사랑		
4660 [밤 추억 새벽 회상]				
5112 [밤 발라드 2000 지난날 추억 늦은 밤 댄스 댄스 1990년대 90's 트렌드 ...	[사랑 지기 이거 사랑 영일 지기 보내 사랑 기억 사랑 행복 기다리 사랑 사랑 영일...		Dance	
5156 [첫사랑 라디 발라드 발라드 발라드 발라드 발라드 유재하 추억 눈 눈 이영훈 종교음...	[아프 계시 흐리 하늘 내리 지나가 거스르 눈물 계시 함박눈 하얗 덮이 들리 기억 ...	기억 사랑		
5367 [밤 영상 눈 휴식 휴식 겨울밤 선텐 셀 편안함 카페 비오는날 RnB 다양화창트 감...		거기 도착함		
5390 [팝 팝 R&B/힙합 휴식 힙 댄스 힐링 일렉 새벽 일 가요 힙합 일렉트로닉]	[아이 뿔히 눈을 의미 어쩔 슬픔 아름다 꿈속 인형 미소 하얗 아워 서글프 힘들 소...	힘겨워 아이 위하여		

t4522	열 - 태그, 가사, 장르, 노래제목, 사연 행		[이야기, 직장, 선택, 이야기, 대신, 울리, 직장, 선택, 자식, 아이, 직장,...
t4523			[생기, 누님, 누님, 얼마, 금요일, 비행기, 다녀오, 생기, 에피소드, 얘기, ...
t4524			[아빠, 이야기, 직장, 선택, 이야기, 대신, 울리, 직장, 선택, 자식, 아이, ...
t4525			[친구, 이야기, 고등학교, 졸업, 지나, 다리, 삼겹살, 목살, 개탄, 평소, 고...
t4526			[사랑, 이별, 이유, 얼마, 인터넷, 모음, 파업, 경험, 재미, 도움, 이별, ...
t4527	- 노래 데이터는 노래 고유 아이디(songid, 불연속형의 숫자) - 사연 데이터는 문자 't' 와 사연 인덱스를 결합한 textid 사용		[일반, 패션, 업계, 밥벌이, 청년, 사실, 짜증, 이야기, 지인, 사랑, 계시, ...
t4528			[고통, 신음, 소리, 전문가, 해외, 파견, 생활, 증인, 직장인, 국내, 해외, ...
t4529			[요즘, 분위기, 오래되, 얘기, 지인, 울리, 얘기, 인사, 성의, 그릴, 시간, ...
t4530			[남자, 휴가, 성수기, 시즌, 포함, 친구, 주전, 계획, 생각, 들뜨, 토요일, ...
t4531			[직장, 다니, 여성, 오늘, 이야기, 내용, 관하, 얘기, 얼마, 성수기, 여행, ...
t4532			[목소리, 회사, 다니, 풍일, 회사, 회사, 우어곡절, 백화점, 회사, 사할, 크...



## 모델3 – 사연 및 노래 정보의 텍스트를 이용한 Word2Vec 모델

### 4. Word2Vec

- 여러 가지 파라미터 조합으로 모델을 생성해 보았는데, 시도해본 가장 낮은 차원의 모델(50차원)과 가장 높은 차원의 모델(500차원)의 추천 곡 간 유사도는 크게 차이 났지만 추천 곡이 동일한 경우가 많았다.
- 따라서 모델을 통한 노래 추천 속도의 향상을 위해 낮은 차원의 모델(50차원)을 사용했으며 min\_count는 10으로 설정하였다.
- Word2Vec 모델의 한계점은 학습을 하는 데에 필요한 목표값(예를 들면 실제 신청곡)이 없었기 때문에 모델 파라미터 조정이나 모델 성능 향상을 위해 필요한 학습 및 테스트 등의 과정을 거치지 못했다.
- 사연을 새로 입력 받고, 입력된 사연 벡터와 노래 벡터 간 코사인 유사도를 기반으로 가장 높은 유사도를 갖는 노래를 추천 곡으로 출력하는 알고리즘을 갖는다.
- Word2Vec 기반 모델의 평가 방법은 설문조사에 기반하여 이루어 졌지만, 해당 설문조사에 대한 피드백의 과정으로 모델 수정 등의 과정을 갖지 못한 것 등을 아쉬운 점으로 들 수 있다.



## 두 모델의 장점/단점

### 1번 모델

- 상황과 감정이 한 가지이고, 상황과 감정에 대한 단어가 확실히 드러날 때, 노래 추천이 좀 더 잘됨.
- 유사한 감정과 상황을 가진 사연이 많을 수록 노래 추천이 잘 됨.
- 유사한 감정과 상황을 가진 사연이 적다면 노래 추천이 잘 되지 않음.
- 기존 모델은 사연 당 하나의 신청곡이 있기 때문에 사연의 내용을 신청곡이 다양한 측면에서 반영하지 못해 비슷한 여러 사연의 신청곡을 함께 고려해야 한다는 단점이 있음.

### 3번 모델

- 기존 모델은 상황과 상황에 맞는 노래에 대한 사례가 있지만 추가 임베딩 모델은 사례가 없고 가사 기반으로 스토리나 감정선이 비슷한 노래를 추천해 줌.
  - 추가 임베딩 모델은 성능 평가에 대한 객관적인 지표가 없을 경우 모델 자체의 성능 향상을 시키기 매우 어렵다는 한계가 있음.
- 
- 이러한 두 모델은 서로 보완 가능한데, 임베딩 모델을 통해 사연 당 신청곡을 늘리고, 늘어난 신청곡을 바탕으로 사연에 대해 더 정밀한 노래 추천이 가능할 수도 있음.



## 모델평가- 설문조사

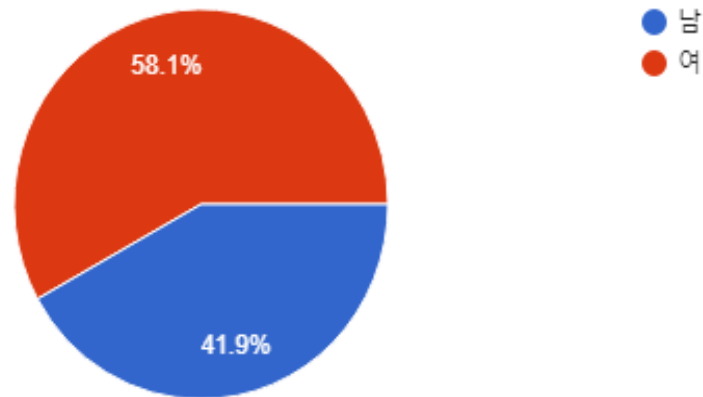
### 설문 진행 방식

- ① 사연 데이터에서 랜덤하게 뽑은 사연 1개를 피설문자에게 제시
- ② 사연을 읽은 후, 1번 추천 곡을 듣고 사연과 어울리는지 1점~5점으로 평가
- ③ 2번 추천 곡을 듣고 사연과 어울리는지 점수로 평가

설문 참가자(62명 \_ 남자 26명, 여자 36명)

당신의 성별은 무엇입니까?

응답 62개





## 모델평가- 설문조사

### ※ 사연 ※

끝까지 연락 한통 없네 한 번을 붙잡지를 앓는구나 매번 내가 먼저 잡았는데 너는 그냥 잡혀준 거 뿐이었어 이젠 너가 언제부터 날 사랑하지 않게 됐는지도 모르겠어 이번부터일까 사실은 예전부터 였는데 그냥 지내왔던걸까 함께하면서도 매번 나는 너에게 서운해하며 혼자 울었어 너와 만날 때 나 혼자 많이 울었으니까 헤어지면 너가 더 힘들거라고 너가 나 처럼 울게 될거라고 그렇게 생각했는데 여전히 나만 우는구나 시간이 더 지나고 너가 아파했으면 좋겠어 나처럼 내생각에 울면서 후회했으면 좋겠어 이제 나는 너의 이름은 입에도 울리지 않을거야 너가 돌아올거라는 생각은 여기서 끝낼게

### 1번 모델

#### 임창정 – Love Affair

1. 사연과 첫 번째 추천 곡이 얼마나 어울리는지 점수로 평가해주세요.(1점~5점) \*

	1	2	3	4	5	
1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	5

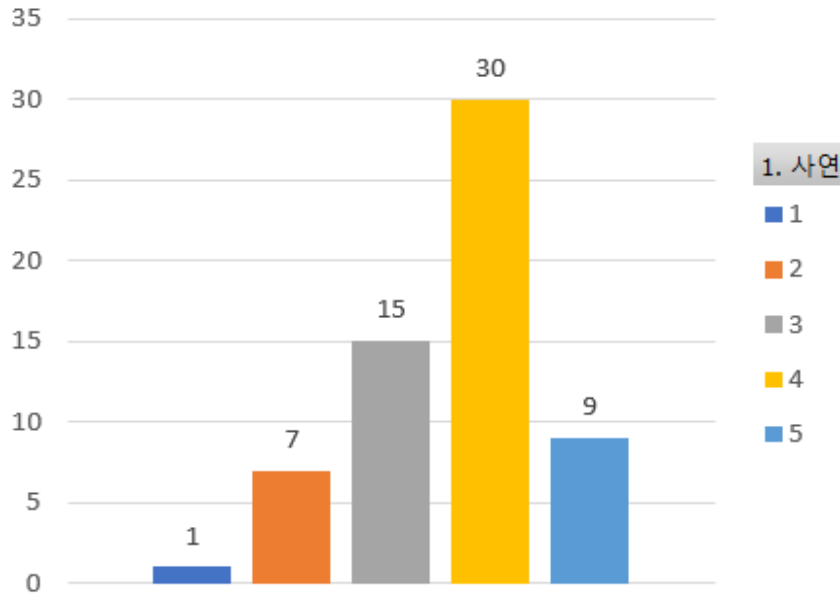
### 3번 모델

#### 피터팬 콤플렉스 – 내 맘을 알기나 해

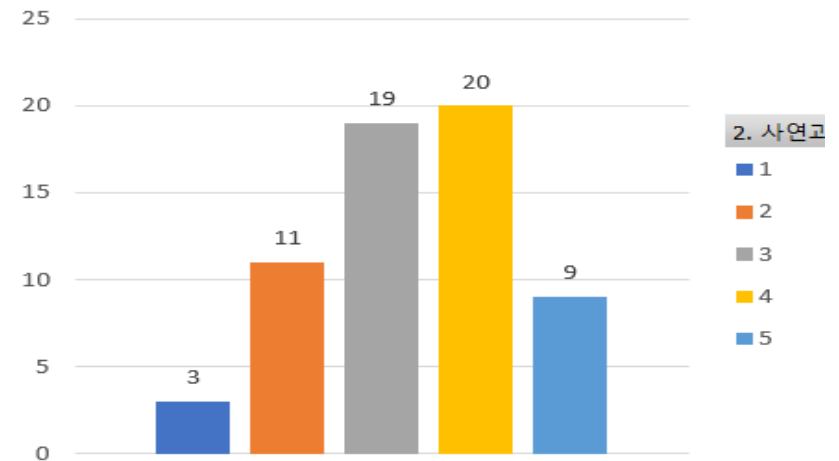
2. 사연과 두 번째 추천 곡이 얼마나 어울리는지 점수로 평가해주세요.(1점~5점) \*

	1	2	3	4	5	
1	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	5

## 모델평가- 설문조사



추천 곡 1에 대한 평가 점수



추천 곡 2에 대한 평가 점수

모델평가	기존모델 추천 곡1	임베딩 추천 곡2
사연 (사랑, 이별)	3.63	3.34

두개의 모델 모두 평균 3점 이상의 점수를 보여줌.  
 ➔ 사연과 보통 이상으로 어울리는 노래를 추천해 줌.

# 활용방안

## (1) 조사대상 및 방법

구분	내용
① 조사목적	○ 음성인식 스피커 이용실태 현황 및 소비자 만족도 조사
② 조사대상	○ 국내외 음성인식 스피커*를 사용하고 있는 이용자 300명 * 기가지니(KT), 누구(SKT), 예코(아마존), 홈(구글) (제품명 기준 가나다순)
③ 조사방법	○ 온라인 설문조사기관 이용, 신뢰도 95%, 표본오차 ±5.65% (6% 이하의 대표성 있는 신뢰수준)
④ 조사기간	○ 17. 6. 21. ~ 17. 6. 30.
⑤ 조사내용	
1) 사용자 이용실태 조사	○ 구매동기, 사용기간, 자주 사용하는 기능 및 전반적인 제품 만족도 수준

## (2) 응답자 현황

- 음성인식 스피커를 사용하는 이용자 300명을 대상으로 온라인 설문조사를 실시함.
- 인공지능(AI) 관련 기술이 적용되어 기존에 없었던 신제품이라 조사시점 당시에 널리 사용되고 있지 않는 실정을 고려하여 설문대상자를 추출

[표 3-1] 조사 대상자 인구통계학적 특성

구분		빈도(명)	구성비(%)	구분		빈도(명)	구성비(%)
성별	남	114	38%	지역	서울	119	39.7%
	여	186	62%		경기	55	18.3%
연령	20대 이하	74	24.7%		인천	20	6.7%
	30대	143	47.7%		강원	3	1%
	40대	59	19.6%		충청	17	5.7%
	50대 이상	24	8%		전라	19	6.3%
					경상	65	21.7%
					제주	2	0.6%

사람들은 인공지능과의 일상대화 기능에 관심도가 높음.

□ (구매 이전 기대한 제품의 특성) 응답자의 46.3%(139명)이 '쉽고 간편한 음성인식 시스템의 이용환경' 이라고 응답함

○ ▲ 기기와의 일상대화(23%, 69명), ▲ 일정 및 스케줄 관리(13%, 39명), ▲ 가정용 사물인터넷 기능(12.4%, 37명), ▲ 생활편의(날씨·교통정보 등) 기능 (5.3%, 16명) 순

[표 3-4] 구매 이전 기대한 음성인식 스피커의 특징

[단위: %(명)]

구분	인공지능에 대한 호기심	지인의 소개	음성인식 기능 필요	선물용	기타
응답자(300명)	67.7%(203명)	20.0%(60명)	6.0%(18명)	5.0%(15명)	1.3%(4명)

[표 3-5] 자주 이용하는 음성인식 스피커의 기능(중복응답)

[단위: %(명)]

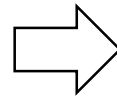
항목	응답자(300명)
① 음악재생	71.3%(214명)
② 날씨·교통정보	41.0%(123명)
③ 인터넷 정보검색	40.3%(121명)
④ 타이머·스케줄 관리	35.7%(107명)
⑤ 라디오·뉴스·팟 캐스트	31.0%(93명)
⑥ IPTV 연동 서비스	29.0%(87명)
⑦ 쇼핑 및 음식주문	17.7%(53명)
⑧ 가정용 사물인터넷	14.3%(43명)
⑨ 일상대화	14.0%(42명)
⑩ 기타 편의기능	5.7%(17명)

일상적인 대화를 바탕으로 노래를 추천해주는 서비스를 제공한다면, 이용자들에게 높은 호응을 얻을 수 있을 것으로 기대.

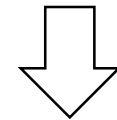
## 활용방안

### 기존 음성인식 스피커 음악 추천 시스템

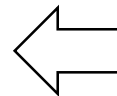
- 1) 클로바 \_ NAVER
  - 협업 필터링(CF) / 머신러닝
- 2) 카카오 미니 \_ KAKAO
  - 협업 필터링(CF) / 콘텐츠 기반 필터링(CB)
- 3) 에코 \_ AMAZON
  - 협업 필터링(CF)



이전 스트리밍 기록 혹은  
비슷한 사용자의 영향을 통해  
음악 추천



상황 혹은 감정이 포함된 사연 데이터  
분석을 통해 그 상황 혹은 감정에  
듣고 싶은 음악을 추천



현재 사용자가 처한 상황을  
반영하지 못 함  
과거 기록이 없어도 추천할 수 있어  
콜드 스타트 문제 해결 가능

## 활용방안

### 유사 서비스 현황

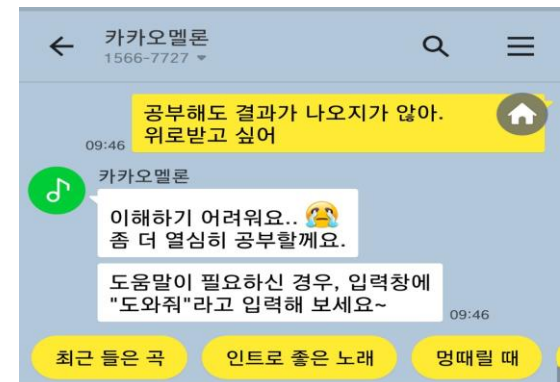
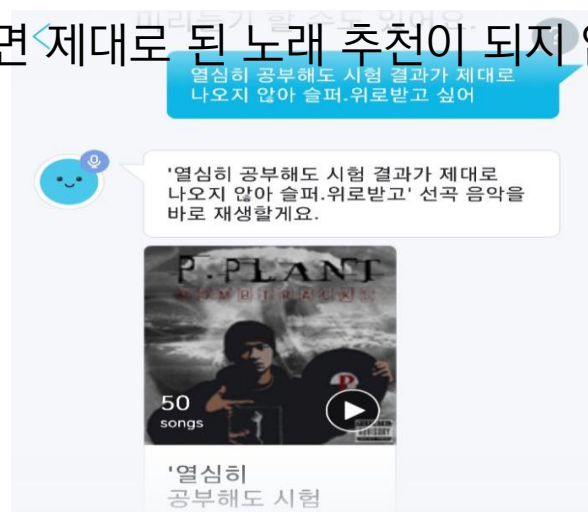
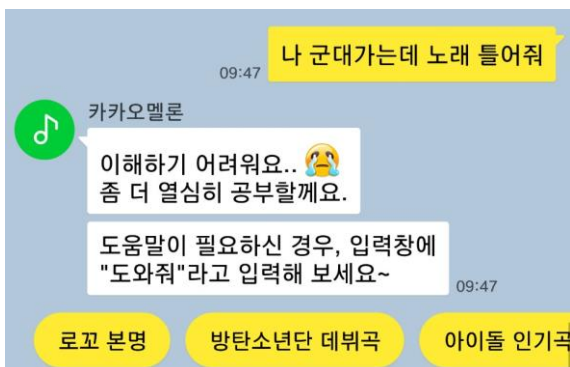
- 현재 음악 플랫폼 멜론과 지니에서 비슷한 서비스를 제공 중

#### 1) Genius \_ Genie

- 음성 명령을 내용을 분석하여 DJ 앨범을 추천해줌.
- 노래 태그와 같이 '비오는 날', '신나는 노래'와 같은 명령어만 처리 가능.
- 원하는 명령어가 길어지면 제대로 된 노래 추천이 되지 않음.

#### 2) 로니 \_ Melon

- 카카오톡 대화방에서 채팅으로 음악들을 추천해줌.
- 노래 태그와 같이 '비오는 날', '신나는 노래'와 같은 명령어만 처리 가능.
- 원하는 명령어가 길어지면 제대로 된 노래 추천이 되지 않음.







안녕하세요, 9704님!

저희의 설문조사에 응해 주셔서 감사합니다.

작성해주신 사연에 추천해 드릴 곡은 YouAndLovE & MAMA의 ' 당신의 손 ' 이라는 노래입니다.

아래는 저희의 추천 모델에 사연을 입력해, 추천 곡을 출력한 결과입니다.

사연: 스텔라장의청춘이나 광진언김필오늘하루같은 노래들으면서 많은위로를 받는편이에요! 죽을것같이 힘들었던 재수기간을 나름대로 최선을다해 보내고 대학에왔는데 2년이지나는 동안 너무 뛰어나고 범접하는사람들이 많다는 생각에 전공이 적합한지와같은 고민과 조금하고 불안함의 감정들이 지친시험기간에는 더많이 들어요.맘편히 듣고 공부부담을 조금이라도 내려놓을수있게 도울 노래있을까용?

추천곡: YouAndLovE - 당신의 손



## 실제 활용

YouAndLovE &  
MAMA-01-당신의 손.mp3

유효기간: ~06/24

용량: 7.32MB



열기 · 폴더 열기

오후 9:57



317은진

헉.....음악선물 진짜쌍좋다.....

프뮤설정할래

ㅠㅏㅠㅏㅠㅏㅠㅏㅠㅏㅠㅏㅠㅏ

오후 10:03

오후 10:05

헐...

오후 10:06

다행이야...ㅠㅏ



317은진

영영나교회다니자나 약간찬양뽐링  
임

프로그램이분석해서

추천해주능거?

오후 10:21

웅 근데 너가 교회다니는 것 까지 변  
수로 넣진 않았어

오후 10:21

방~긋^^...



317은진

ㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋㅋ  
ㅋㅋㅋㅋ

오후 10:22



317은진



저장

지차신기하네

들으면서잔다

오후 11:54

## 실제 활용



안녕하세요, 9169님!

저희의 설문조사에 응해 주셔서 감사합니다.

작성해주신 사연에 추천해 드릴 곡은 Supreme Team의 'Step Up' 라는 노래입니다.

사연: 복학 했는데 하려는 공부는 잘 안되고 맨날 술만 쳐먹으러가는 것 같네요... 근데이게 잘못된거는 아니라고 생각하는데 뭔가가 잘못된거 같기도 하고 이럴때 아니면 언제 또 돌아보겠냐 라고 생각하는 데도 한편으로는 미래에 후회하지 않을까하기도 합니다. 사람이 다 성향이 다르고 생각도 다르기 때문에 남들하는 말에 귀담아야 할까 아니면 내가 가는 길이 맞다고 확신하고 가야할까 고민중입니다. 제가 힘 한쪽을 좋아해서 그 쪽으로 추천해줬으면 좋겠습니다.

추천곡: Supreme Team - Step Up



손을 번쩍 든 무지

ㅋㅋㅋㅋㅋㅋ이노래알고있었어 ㅋㅋㅋㅋ

오후 11:32



손을 번쩍 든 무지

노래좋음 !ㅋㅋㅋ

오후 11:33

## 실제 활용



안녕하세요, 7147님!

저희의 설문조사에 응해 주셔서 감사합니다.

“공부가 제대로 되지가 않아요. 열심히 한다고는 했는데 결과가 항상 좋지 않네요.

아무리해도 결과가 안나오니 열심히 할 힘이 안나네요. 점점 지쳐만가구요ㅠㅠ 어떻게 하면 좋을까요.”

작성해주신 사연에 추천해 드릴 곡은 걱정말아요 그대라는 노래입니다.

아래는 저희의 추천 모델에 사연을 입력해, 추천 곡을 출력한 결과입니다.

	similarity	singer	songid	songtitle
0	0.847665	이적	7921208	걱정말아 요 그대
1	0.691496	하현우	8050532	걱정말아 요 그대
2	0.665698	들국화	4350118	걱정말아 요 그대