



제 3회 투빅스 데이터 분석 컨퍼런스

피처링 네트워크 Featuring Network

가수 간 네트워크 시각화 및 음원 흥행도 예측

팀 '연결고리 (feat. MC상주니)'
이정민 이승은 최도현 김은서 문상준



INDEX

■ 주제 선정 배경

■ 프로젝트 소개

■ 데이터 수집 및 정제

■ 네트워크 분석

■ 시각화

■ 모델링

■ 결론 및 제언

1 주제 선정 배경

피처링? Featuring

다른 가수의 연주나 노래에 참여하여 도와주는 것

피처링이 대세!

- ✓ 서로 다른 장르 간의 결합으로 시너지 효과 창출
- ✓ 최근 대중음악에서 중요한 흥행 요인으로 평가



썸 - 정기고, 소유
feat. 릴보이



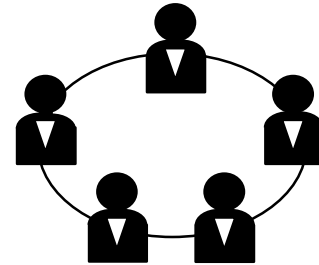
친구여 - 조피디
feat. 인순이

Synergy

~~1 + 1 > 2~~

그래서 우리는...

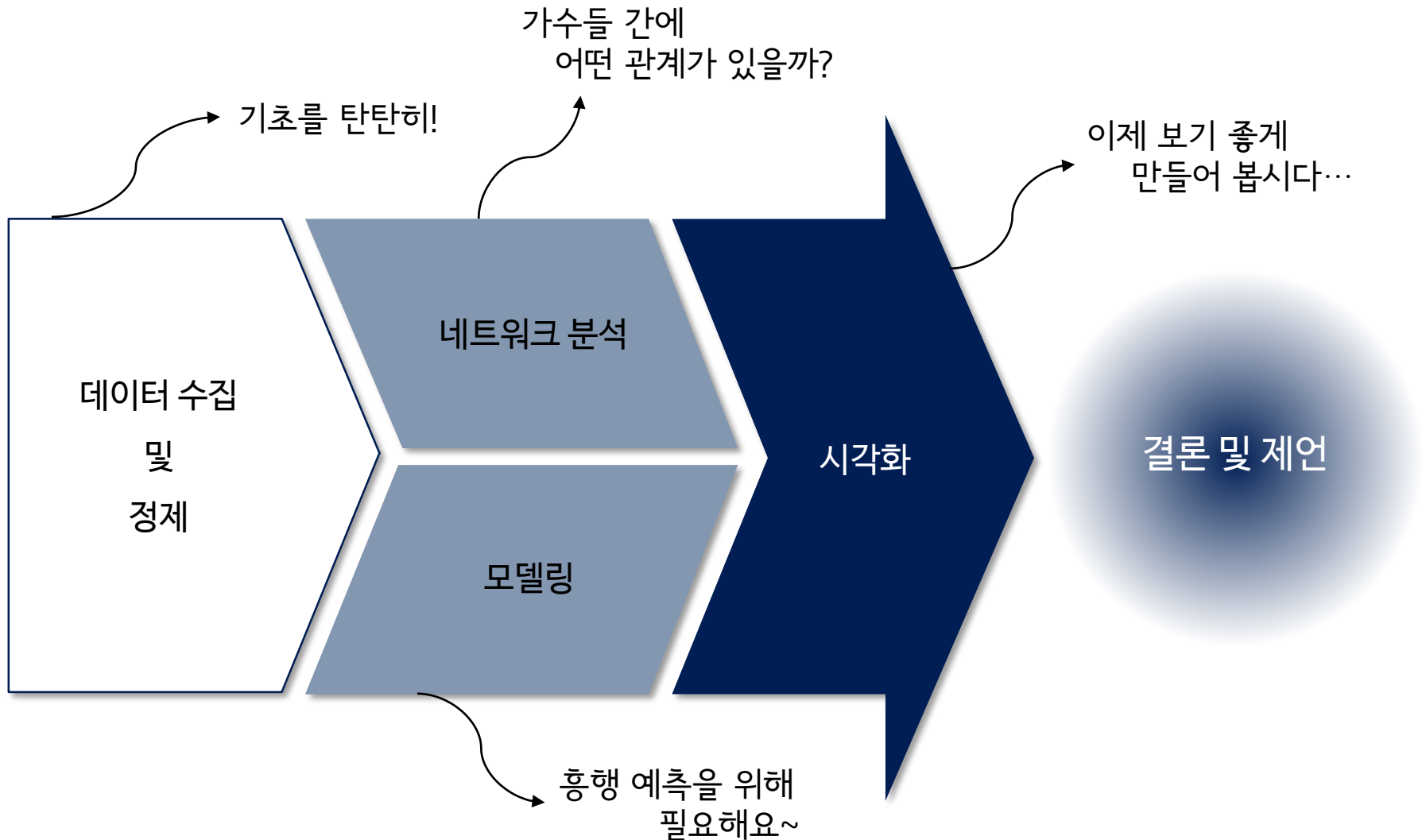
- ✓ 가수 간 피처링 네트워크 시각화



- ✓ 새로운 음원의 흥행 가능성 예측



2 프로젝트 소개



3 데이터 수집 및 정제

데이터 수집은 어디서?



네이버 뮤직 월간 TOP 100

2007년 5월 ~ 2015년 12월

네이버뮤직 총 101개월 간의 데이터 크롤링

데이터 정제, 왜 하는 걸까?



지코(ZICO)

지코 of 블랙비

Z.I.C.O

지코

데이터 예시를 살펴보자!

곡 제목	가수	피쳐링 가수	순위	년-월
소녀	오혁	NA	1	2015-12
DADDY	싸이	씨엘	2	2015-12
걱정말아요 그대	이적	NA	3	2015-12
나팔바지	싸이	NA	4	2015-12
혜화동 (혹은 쌍문동)	박보람	NA	5	2015-12
청춘	김필	김창완	6	2015-12
밤의 멜로디	브라운아이드소울	NA	7	2015-12
네게 줄 수 있는건 오직 사랑뿐	디셈버	NA	8	2015-12
I Remember You	싸이	자이언티	9	2015-12

4 네트워크 분석

네트워크의 기본 개념 알아보기

노드Node

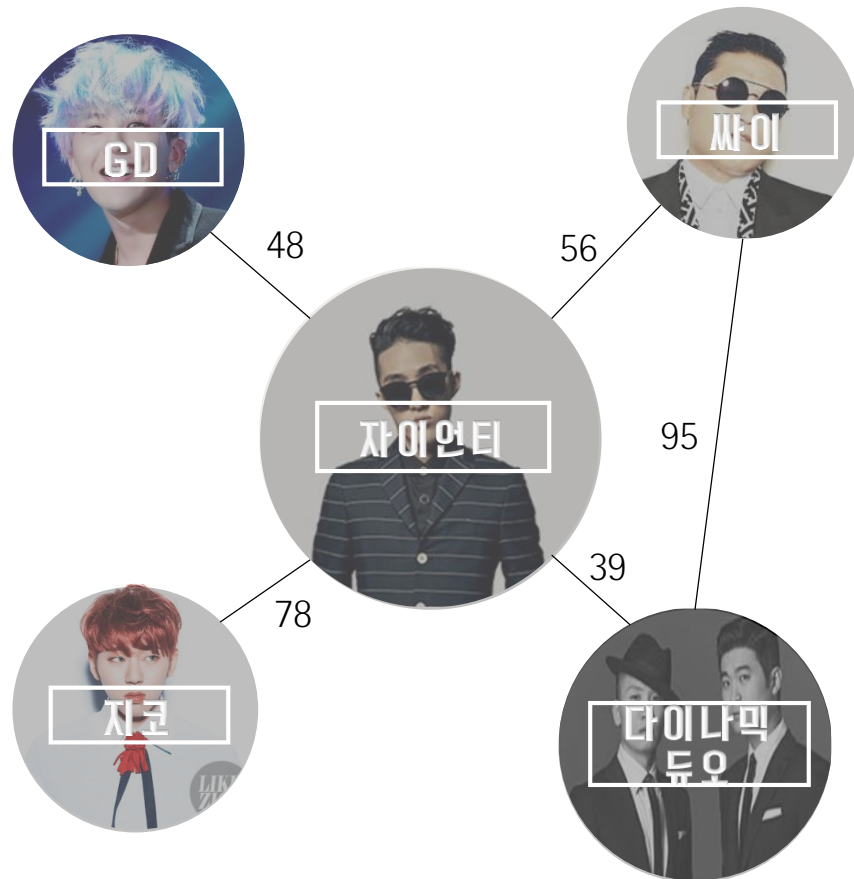
: 원으로 표현.

본 프로젝트에서는 각 가수들이
노드에 해당.

엣지Edge

: 노드들을 연결하는 선.

본 프로젝트에서 엣지는
두 가수가 공동작업 한 '곡'을 의미하며,
엣지 위에 각 곡의 **가중치**를 표시함.



$y = -x + 101$ 식을 이용해
차트 1위면 100점, 100위면 1점으로
가중치 적용!

뒤에 나올 시각화에서 사용될 개념이니 잘 알아두기 ~

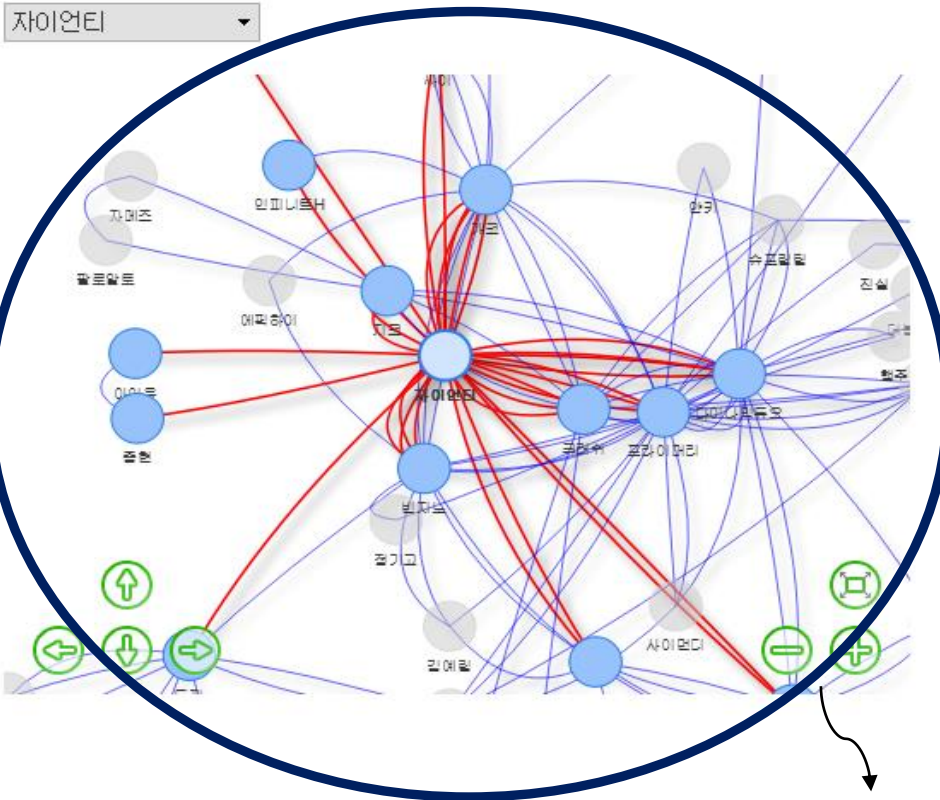
5 시각화

웹 상에서 R 결과를
시각화하는 도구

샤이니를 이용한 피처링 네트워크 시각화

네트워크 분석 결과를 샤이니를 이용해 시각화 하였습니다~
visnetwork 패키지를 사용하였습니다 ☺

Featuring networks



버튼 및 마우스 스크롤을 이용해
확대/축소할 수 있어요

가수

자이언티



자이언티

대표곡 : 양화대교

데뷔년도 : 2011년

Featuring 공합 Top3



1위 : 정인



2위 : 개코



3위 : 태양

6 모델링

모델링, 왜? 그리고 어떻게?

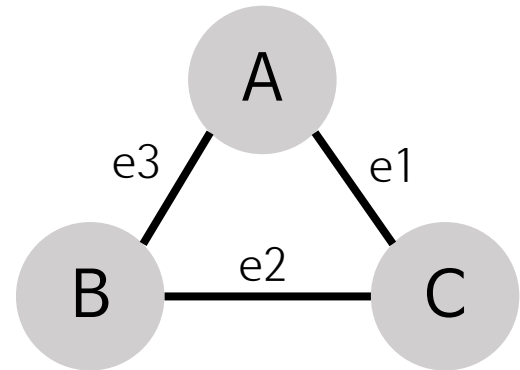
WHY? 두 가수가 공동작업을 했을 때, 차트에서 얼마나 좋은 성적을 거둘 수 있을지를 예측해보고자 본 모델링 수행

가수 A-B 간의 예상 성적을 중간다리 역할을 하는 가수 C와의 기존 성적(A-C, B-C)을 통해 예측한다는 아이디어로 '가수 삼각형'에 집중

HOW?

모델링 과정 이해하기 (1)

- STEP 1. 앞서 구현한 피처링 네트워크에서, 3번 만에 자기 자신에게 돌아오는 삼각형 1935개를 통해 컴퓨터를 학습시킴
- STEP 2. 학습 결과를 토대로 회귀분석, 인공 신경망, 회귀나무, 부스팅, 배깅, 랜덤 포레스트 등 다양한 기법을 이용하여 순위를 예측
- STEP 3. 실제 순위와 비교했을 때의 MAE (Mean Absolute Error, 평균 절대 오차) 값이 가장 낮은 **랜덤 포레스트 모형** 채택

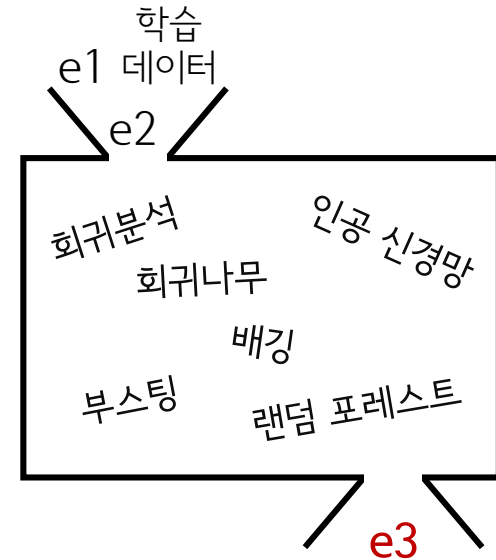
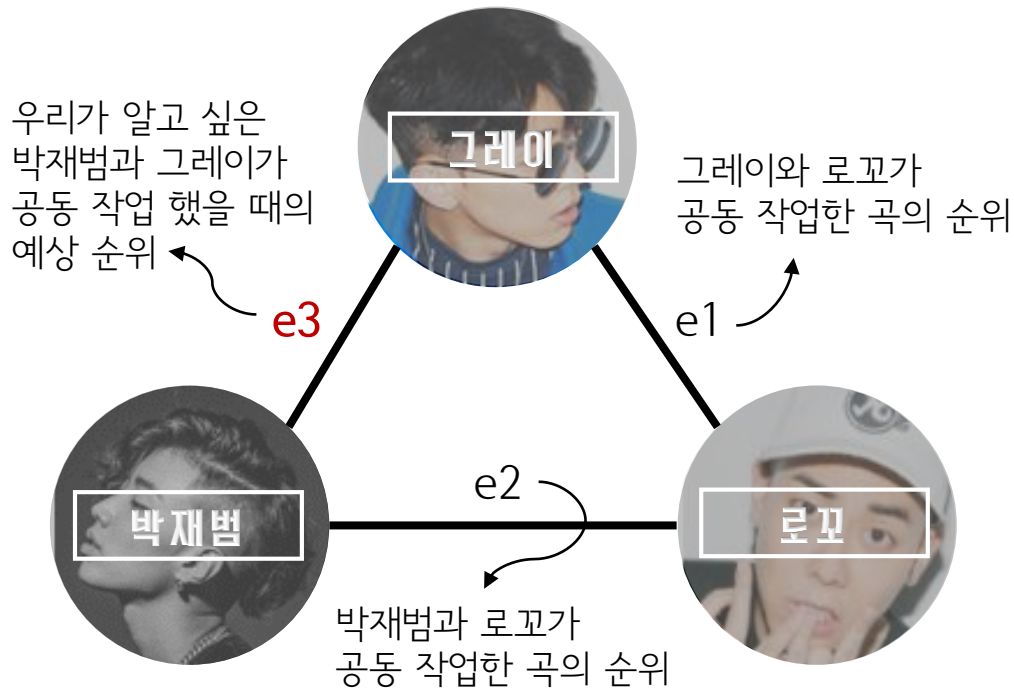


6 모델링

모델링 과정 이해하기 (2)

STEP 2.

학습 결과를 토대로 회귀분석, 인공 신경망, 회귀나무, 부스팅, 배깅, 랜덤 포레스트 등 다양한 기법을 이용하여 순위(e3)를 예측

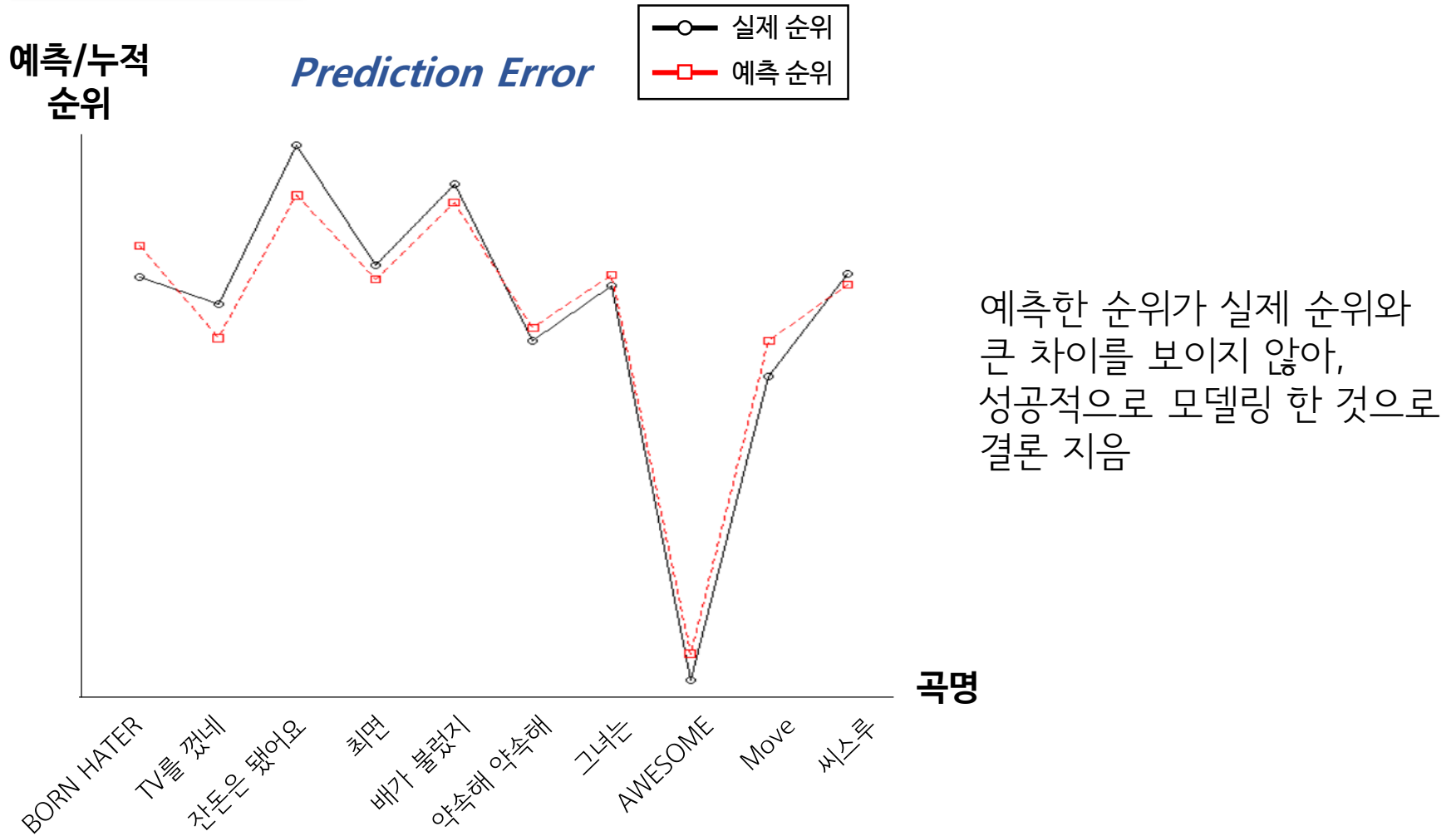


STEP 3.

실제 순위와 비교했을 때의 MAE(Mean Absolute Error, 평균 절대 오차) 값이 가장 낮은 **랜덤 포레스트 모형** 채택

6 모델링

모델링 결과



6 모델링

모델링 이상치 해석해보니...

다다익선? No!

공동작업을 많이 하는 가수들은 차트 순위의 편차가 크다.

곡 수가 많을 경우, 차트 안에서의 순위 편차가 크기 때문에 예측 오차도 크게 나타남

예를 들어 윤미래와 타이거JK의 사례

곡명	가수	피처링가수
Get It In	윤미래	타이거 JK, 정인
TV를 켜네	리쌍	윤미래, 권정열
이젠 너 없이도	유나킴	윤미래, 타이거JK

반짝스타는 예측이 힘들어요~

인지도가 낮은 가수가 이벤트성으로 곡을 발매해서 차트에 진입했을 때 오차가 크다.



예를 들어 기존에 차트에 오른 적이 없던 래퍼들이 쇼미더머니 등의 방송 출연으로 갑작스런 인기를 얻은 경우, 모델에 학습된 데이터가 없기 때문에 예측 오차가 크게 나타남

곡명	가수	피처링가수
이리와봐	마스타우	도끼, 바비
거북선	자메즈	팔로알토

웹 상에서 R 결과를 시각화하는 도구

네트워크 분석 결과를 샤이니를 이용해 시각화 하였습니다~
visnetwork 패키지를 사용하였습니다 😊

로깅 ▼



노드를 클릭하면,



해당 가수의 프로필과,
모델링 결과로 도출한
피처링 공합 Top 3 가수가 노출!



험해봅시다!



로고

대표곡 : 감아

데뷔년도 : 2012년

Featuring 공합 Top3



1위 : 스윙스



2위 : 윤미래



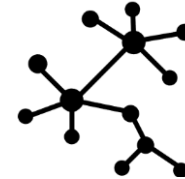
3위 : 크러쉬

7 결론 및 제언

이번 프로젝트의 의의!

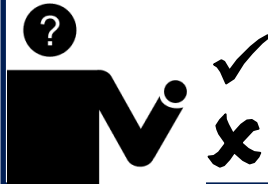
시각화 부문

피처링으로 맺어진 가수들 간의 관계를 쉽게 알아볼 수 있게
해주는 지도의 역할을 할 것으로 기대



모델링 부문

향후 가수나 소속사가 같이 작업할 상대를 결정할 때 참고할 수
있는 좋은 지표가 될 수 있을 것으로 기대



End of Document

Any Questions or Comments?