

2025년 1학기 회귀분석팀

김준영, 김재원, 이채빈, 조수희, 홍승범



# Index >>>



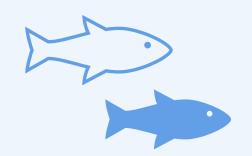
01 선정 배경

02 데이터 수집 및 전처리

03 감성점수 도출

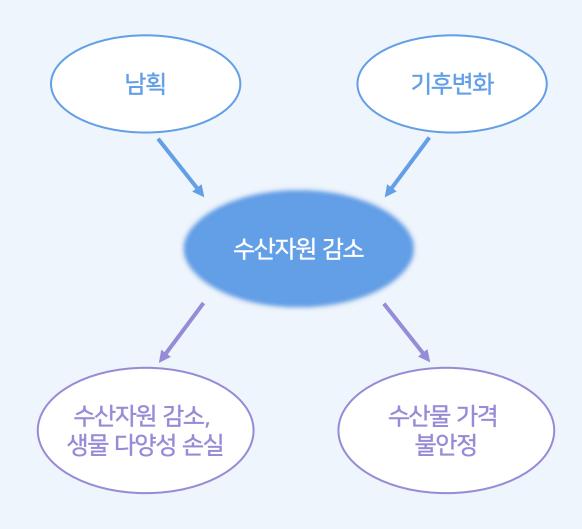
04 저평가 어종 도출

05 추천 시스템



# 01 선정배경

# 문제 상황





# 문제 상황





출처: 동아사이언스 21.06.05. 기사



#### 남획 | 남획의 문제점

남획의 문제점은 특정 어종을 **너무 많이 어획**하여 번식을 할 수 있는 성어가 충분히 남지 않아, 이러한 현상으로 **건강한 개체 수**를 유지할 수 없어 진다는 것

출처 : MSC 대한민국







#### 남획 | 초래하는 문제

어가 생계 위협

수산 자원이 감소해 국내 8만 7천여 명의 어가인구의 일자리 부족 위험

양질의 단백질 공급원 감소

수산물은 인간이 소비하는 1일 동물성 단백질 섭취량의 최소 20%를 차지

생물 다양성 파괴

남획은 기후, 깨끗한 대기, 물, 식량에 중요한 기여를 하는 생물 다양성과 서식지, 그리고 생태계에 영향





# 지속 가능한 어업 | 정부의 조치

해양수산부는 다양한 어업 관리 조치를 시행 중

산란기 내에 어획 금지(**금어기**), 치어를 보호하기 위한 크기제한 설정(금지체장)

위기종을 나라 어장에서 양식 후 대량 방류





생산량을 줄이는 해결방안

자원량을 늘리는 해결방안

# 선정 배경



# 지속 가능한 어업 | 정부의 조치



### 현실적인 문제점



어린 고등어인 갈고등어 어획량이 크게 증가

정부는 어획량 제한법 제정을 추진 중이지만 어민들은 비현실적이라 반발





### 해결방안 제시



데이터 분석을 통해, 생산이나 자원량이 아니라 수요 측면에서 해결방안을 제시



**저평가된 어종**을 제안함으로써 수요를 다각화



생산 다각화 유도



생산량이 몰리는 어종에 대한 남획을 줄이는 데 기여

# 선정 배경



### 기대효과

인기 어종의 **대체재**를 발굴, 인기 어종의 수요를 대체재로 분산



인기 어종의 수요 감소



가격, 거래량 감소







물가 안정 + 남획 감소



분석 흐름

최종 목표

데이터 분석을 통해 저평가된 어종을 도출

데이터 수집 및 전처리

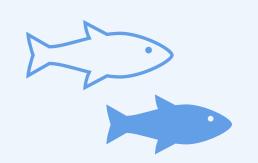
NLP 기반 감성 분석(점수 도출)

네트워크 분석으 로 사용자 신뢰도 산정

어종별 가중 평균 평가 점수 도출

위판 가격·거래량 을 통해 시장성 낮 은 어종 선별

AI 알고리즘을 통해 개인 맞춤 추천





#### 데이터 수집 | 필요한 데이터 설명

#### 네트워크 분석을 위해서는 **사용자와 어종**이 **매핑된 데이터**가 필요



✓ 단순히 '어종별 평가'가 아닌 '사용자별 어종별 평가'가 필요

#### 단순 어종별 평가

#### 고등**어**가 맛있어요

연어는 양파랑 같이 먹으면 진짜 맛있어요

나는 **방어**는 좋지만 **부시리**는 좀 별로던데

여러분 맛난 **광어** 드셔보세요

#### 사용자별 어종별 평가



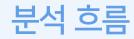
User 1: 고등어가 맛있어요

User 1: 연어는 양파랑 같이 먹으면 진짜 맛있어요

User 2: 나는 방어는 좋지만 부시리는 좀 별로던데

User 3: 여러분 맛난 **광어** 드셔보세요





#### 최종 목표

#### 데이터 분석을 통해 저평가된 어종을 도출

데이터 수집

데이터 전처리

감성점수 도출

수산물 유튜버 5인 댓글 크롤링

불용어 제거

해양수산부 산지위판 가격정보

해양수산부 '수산물(품종) 방언집'

어종 리스트 구축

어종 필터링

Kiwi 문장분리

유·무관 필터 링

감성점수 도출



분석 흐름

최종 목표

데이터 분석을 통해 저평가된 어종을 도출

데이터 수집

데이터 전처리

감성점수 도출

수산물 유튜버 5인 댓글 크롤링

불용어 제거

해양수산부 산지위판 가격정보

어종리

해양수산부 '수산물(품종) 방언집'

유튜브 댓글 약 100 만 행을 크롤링하여 불용어 제거













### 분석 흐름

#### 최종 목표

#### 데이터 분석을 통해 저평가된 어종을 도출

날짜	수협	어종
2025/05 /05	강진군수산업협동조합	꽃게(활)
2025/05 /05	강진군수산업협동조합	해삼(활)
2025/05 /05	강진군수산업협동조합	가리비류(활)

표 준 어	ł	방 언 명
가오리(류	!) フ	부리,가우리,참가오리, 물가오리,강재미, 갱게미,찰때기,나무쟁이
홍어(류)	2	재미,가부리,홍해,홍에,갱게미
가자미(류	!) 닡	세미,갈가자미,도다리,가재미,가삼,미주구리,까즈메기,납세
갈치	낕	치,풀치,은갈치
강달이(류	<u>!</u> ) ~	달이,황새기,깡치
고등어(류	<u>!</u> ) <u>I</u>	등어,고도리,소고도리,가라지
꽁치	슫	마(외래어),삼마이,공어,공미리,공멸
학공치	Y	요리(외래어),사이루,꽁치,청갈치,꽁치
넙치(광어	) 공	어,도다리
농어(류)	卢	에,까지맥이,깡다구,깔다구,깔대기,껄덕,연어병치,독도돔



표준명	방언
골뱅이	<del>큰구슬</del> 우렁이
아귀	
살오징어	오징어, 화살오징어
쥐치	쥐취
참돔	도미
청어	
고등어	
기름가자미	기름가재미
청어 고등어	·

해양수산부 산지위판 가격정보

해양수산부 '수산물(품종) 방언집'

어종 리스트 구축 장 위판장 데이터의 **어종명을** 기준으로 방언집을 참고하며 **어종 사전** 구축



분석 흐름

최종 목표

데이터 분석을 통해 저평가된 어종을 도출

데이터 수집

데이터 전처리

감성점수 도출

수산물 유튜버 5인 댓글 크롤링

해양수산부 산지위판 가격정보

해양수산부 '수산물(품종) 방언집'

불용어 제거

어종 리스트 구축 어종 필터링

정제된 어종 사전으로

○ 유튜브 댓글에서 **어종이 있는 문장**만 필터링

			TJ
작성자	댓글	영상	matched_fish
@jykim25		청어도 없네요ㅠ /watch?v=	ww.youtube.com :PSAT12Kl&pp= 돌가자미, 청어 AYcqlYzv
@jaewon99	꽁치도 학꽁? 소합니다	시도 정말맛있고 고 /watch?v=	ww.youtube.com ejD616ZU_MKI& 꽁치,학꽁치 YUJAYcqlYzv
@psat2007	대하 전어 꽃	걸 먹고 싶은데 잘 /watch?v=	ww.youtube.com g_kfTMJBc 꽃게, 대하, 전어





최종 목표

데이터 분석을 통해 저평가된 어종을 도출

데이터 수집

데이터 전처리

감성점수 도출

수산물 유튜버 5인 댓글 크롤링

해양수산부 산지위판 가격정보

해양수산부 '수산물(품종) 방언집'

불용어 제거

어종 리스트 구축 어종 필터링

Kiwi 문장분리

프 유 감성 Kiwi 형태소 분석기론 트 **문장 분리** 수행 도 불



분석 흐름

최종 목표

데이터 분석을 통해 저평가된 어종을 도출

데이터 프롬프트 러닝을 1차, 2차로 나누어서 진행처리

1차: 유/무관 필터링

수산물 유튜버

댓글 문장들 중 <mark>무관한 문장은 삭제</mark>하는 task

2차: 감성점수 도출

해양수산부 산

댓글 문장과 어종 사전을 제시하여

'사용자별 어종별 감성점수'를 **5점 척도로 도출** 

해양수산부 '수산물(금증) 방안입

Kiwi 문장뿐리

감성점수 도출

프롬프트 러닝

유·무관

필 터 링 감성점수 도출



일부 유튜브 채널 제거

최종 목표

데이터 분석을 통해 저평가된 어종을 도출

데이터 수집

데이터 전처리

감성점수 도출

수산물 유튜버 5인 댓글 크롤링

불용어 제거

 $( \lozenge )$ 

4주차에서는,

무관한 댓글이 많은 유튜브 채널 댓글을 모두 삭제

해양수산부 산지위판 가격정보

식당 관련 평가 多

제거

유튜버 관련 댓글 多

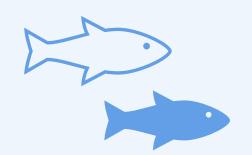
해양수산부 '수산물(품종) 방언집'

제거

수빙수

생선선생

입질의 추억TV 회사랑 수부해TV





### 감성점수 도출 방법 비교





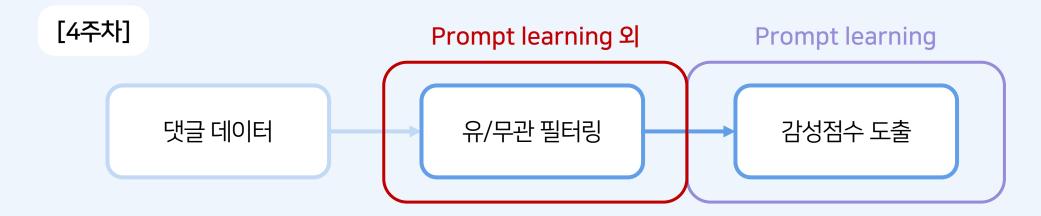
긍/부정 모델	고려사항	프롬프트 러닝	
0,1 척도	점수 척도	5점 척도 가능	
우수함(일부)	<b>성능</b> 우수함		
불가능	댓글 단위 맥락 고려	가능	
2+ 어종 2+ 감정 불가	문장 유형	2+ 어종 2+ 감정 가능	
학습 데이터셋 구축 필요 라벨링 비용		몇 가지 예시로 few-shot	



## 유/무관 필터링 | 필요성

유/무관 필터링 없이 프롬프팅을 통해 '유관한 문장을 구별하고 감성점수를 도출'할 경우, 두 가지 혼재된 작업을 동시에 잘 수행하지 못하는 것을 확인

> 2주차에서는 프롬프트 러닝을 통해 유/무관 필터링을 진행했으나, 4주차에는 프롬프트 러닝 외 방법으로 유/무관 필터링 시도





## 유/무관 필터링 | 파인 튜닝 모델

#### 모델 파인 튜닝 과정

긍/부정 모델 파인튜닝과 마찬가지로, **문장 단위로 유/무관 라벨링된 데이터**를 통해 **사전학습 모델**(KcELECTRA, KoBERT)를 각각 파인 튜닝함.





## 유/무관 필터링 | 파인 튜닝 모델



#### 라벨링 기준 정리

큰틀

- 1. '어종'에 대해
- 2. 실제로 먹어 보고 하는
- 3. '맛, 식감, 선도' 등에 대한 평가(이하 '맛 평가'라 칭함)

조건 1을 위배한 경우

유튜버에 대한 얘기 등

생선선생님 그러니 매니저님 에게 혼나조 보리새우를 까서 매니저님을 먼저 챙겼어야죠



## 유/무관 필터링 | 파인 튜닝 모델



#### 라벨링 기준 정리

먹어보고 싶다, 땡긴다, 부럽다, 먹으러 가야겠다, 궁금하다 등

꽃게 땡기네요

단, '또' 먹고싶다 혹은 '다시' 먹고싶다는 유관 처리

양미리구이도 또 한번 먹으러 가야겠어요 저도

조건 2를 위배한 경우

질문하거나, 추천을 요청하는 경우

요즘 꽃게 제철 맞나요?

자신이 내린 평가가 아니라 일반적으로 그러하다고 말하는 경우

방어가 그냥 입맛에 안맞는 사람도 많아요



#### 유/무관 필터링 | 파인 튜닝 모델



#### 라벨링 기준 정리

조건 3을 위배한 경우

가격에 대해서만 평가, 크기에 대한 평가, 단순 정보 제공 등 광어값 폭락했다더니 소비자가 구매하는 시세 보면 아닌가 봅니당. 단, 가격에 대한 얘기가 있더라도 1~3을 모두 충족했다면 유관



## 유/무관 필터링 | 파인 튜닝 모델



#### 라벨링 기준 정리

계절, 원산지, 조리법, 식당에 따른 조건부 평가와 비교 평가는 기본적으로 무관

저도 동해안 오징어가 서해안 오징어 보다 탱탱한 맛은 더 좋다고 생각합니다.

#### 추가 조건

단, 둘 이상의 어종 간 비교할 때 등장하는 조건부와 비교는 유관으로 허용 겨울 방어 보단 숭어가 최고죠.



#### 유/무관 필터링 | 파인 튜닝 모델



라벨링 예시

오징어 진짜 맛있죠

유관

조건 1 위배

생선선생님 오징어 닮으셨다더니

전혀 아니었어요!

조건부, 무관

오징어는 속초가 맛있는 듯…

조건 2 위배

오징어 먹고 싶네요

비교, 무관

오징어는 갠적으론 숙회보다도 회무침이 최고인 듯요

조건 3 위배

오징어 요새 가성비 좋죠

유관

회로 먹을 거면 새우보단 오징어죠



#### 유/무관 필터링 | 파인 튜닝 모델



#### 최종 데이터셋

데이터셋

전체 데이터 5,068개

무관 2,778개 / 유관 2,290개

클래스 불균형 지수

Imbalance Ratio: 1.2131004366812228

Shannon Entropy: 0.9933013995641776

Gini Index: 0.49536407463079857

Train/Test 분할

8:2 비율로 train / test 분할,

Train을 다시 8:2로 사용해 validation set으로 사용



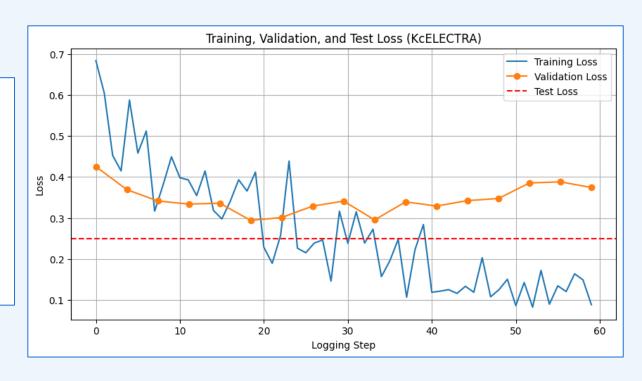
# 유/무관 필터링 | 파인 튜닝 모델



최종 fine-tuning 결과

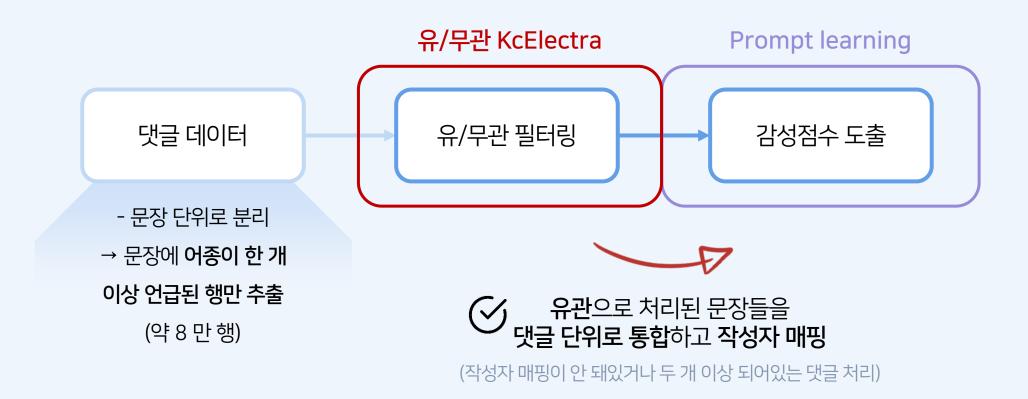
#### KcElectra

	precision	recall	f1-score	support
х	0.94	0.93	0.93	339
0	0.93	0.94	0.93	340
accuracy			0.93	679
macro avg	0.93	0.93	0.93	679
weighted avg	0.93	0.93	0.93	679



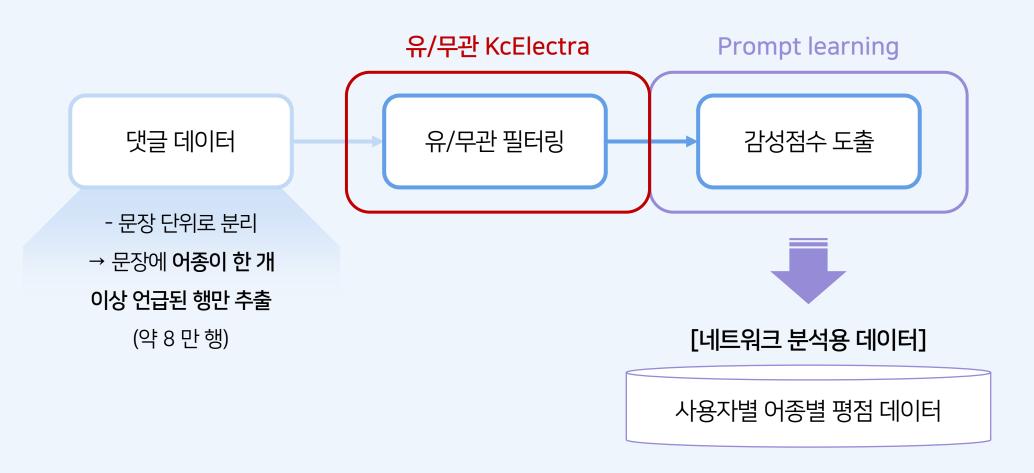


## 최종 흐름





## 최종 흐름





#### 프롬프트 엔지니어링 | G-Eval

#### G-Eval: NLG Evaluation using GPT-4 with Better Human Alignment

LLM을 '심사위원'으로 활용해 자연어 생성(NLG) 결과물을 정량·정성적으로 평가하는 프레임워크 사람 평가와 Spearman  $\rho$  = 0.514 를 달성해 기존 지표(BLEU, ROUGE)를 크게 능가



G-Eval을 통해 프롬프트 엔지니어링으로 구한 점수가 잘 생성되었는지 평가!





# 프롬프트 엔지니어링 | G-Eval

핵심 구성 요소	설명
Rubric Prompt	사람이 정의한 평가 기준(예: 사실성, 일관성 등)을 자연어로 기술
Chain-of-Thought (CoT)	GPT-4 등 심사 LLM이 각 항목별로 사고 과정을 전개하여 내부적으로 점수 산출
Form-Filling 출력	JSON 형식으로 점수와 근거를 구조화해 변환

#### 감성점수 도출



#### 프롬프트 엔지니어링 | G-Eval

LLM이 자신과 유사한 스타일에 가점을 주는 경향이 있다는 문제가 보고됨 그런 편향을 줄여주기 위해 직접 라벨링한 데이터를 gold로 추가

gold: 인간이 직접 붙인 정답 레이블



300개의 댓글에 대해 댓글별로 어종별 감성점수를 라벨링 Gold 추가를 통해 **모델-사람 일치도**와 **규칙 준수도**를 동시에 볼 수 있음

## 감성점수 도출



#### 프롬프트 엔지니어링

Kim, J. & Lee, S. (2023). 프롬프트 러닝을 활용한 NLG 평가 효율화

Zero-Shot 기반 LLM이 도메인 미학습 상태에서도 전통적 BERT 계열 **모델과 유사하거나 더 높은 성능**을 달성



따라서 이번 주제분석에서는 "BERT-Fine-Tuning" 대신 "Prompt Learning" 기반 점수 산출 방식을 채택

#### 감성점수 도출



#### 프롬프트 엔지니어링

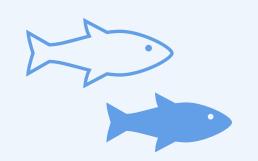
추가로 라벨링한 300행에 대해 일치 여부를 확인한 결과, 300행 중 261개가 직접 라벨링한 결과와 일치

일치하지 않은 것들은 주로 여러 어종 중 한 어종에 대한 점수가 잘못되거나 한 어종이 들어가 있는데 무관으로 처리한 경우 위주임을 확인

한 댓글에 많은 어종이 언급되는데 그 중 한 어종에 대한 점수만 틀린 경우



프롬프트 엔지니어링을 통해 얻은 어종별 점수를 통해 **네트워크 분석** 진행!





## 데이터

#### [네트워크 분석용 데이터]

#### 사용자별 어종별 평점 데이터

20046 행

9618 명의 사용자

사용자	어종	점수
@kjyblack	갑오징어류	4
@kjyblack	밴댕이	3
@kjyblack	자리돔	5
@kjyblack	한치	5
@-jaewoni	낙지류	4
@-jaewoni	문어	3
@-jaewoni	쭈꾸미	4
@-beomibutnotbert	미더덕	4
@-beomibutnotbert	대구	4
@-chaebini	고등어류	3
@-soooheee	광어	4
@-soooheee	굴	4
@-soooheee	도치	5



## 추천 어종 도출 흐름



추천 어종 도출 전체 흐름

사용자별 어종별 평점 데이터











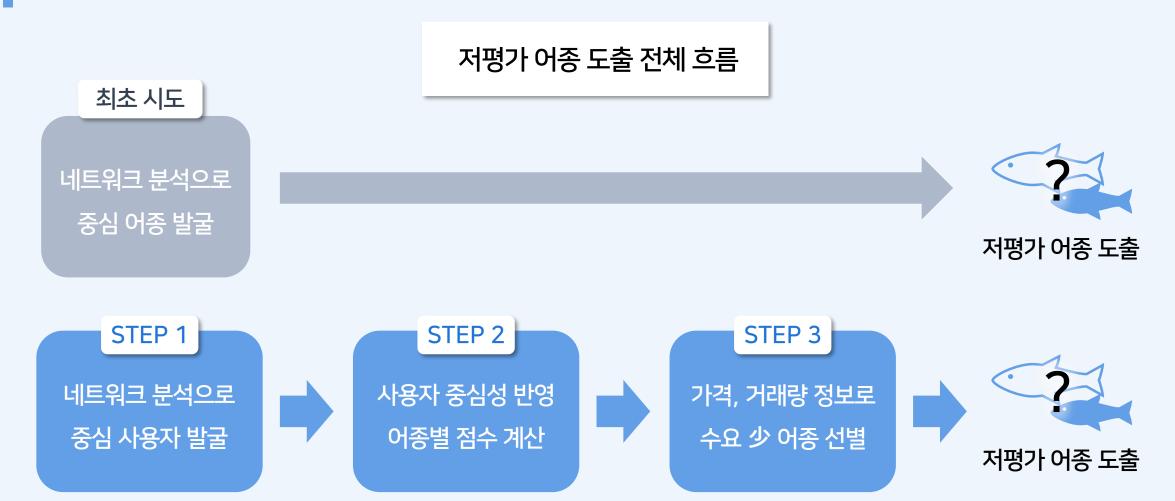


지속가능성 반영

먼저, **저평가 어종 도출 방법**부터 알아보자!



#### 저평가 어종 도출 흐름





## 최초 시도: 중심 어종 발굴 | 네트워크 분석

#### 네트워크 분석

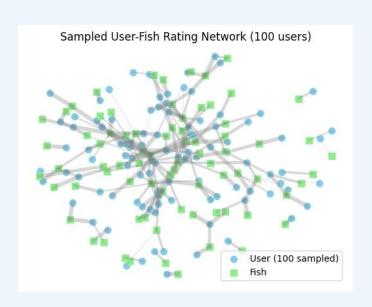
## **개체와 이들 사이의 관계로 구성된 네트워크**를 수학적으로 표현하고 시각화하여, **패턴, 구조, 동작 원리**를 이해하는 방법론

자세한 설명은 2주차 ppt 참고!

#### **사용자**와 **어종**을 각각 노드, **연결 여부**를 **엣지**로 설정해 네트워크를 구성

사용자	어종	점수
@kjyblack	갑오징어류	4
@kjyblack	밴댕이	3
@-jaewoni	낙지류	4
@-jaewoni	문어	3
@-beomibutnotbert	미더덕	4
@-beomibutnotbert	대구	4
@-chaebini	고등어류	3
@-soooheee	광어	4
:	:	:
@-soooheee	도치	5







#### 최초 시도: 중심 어종 발굴 | 네트워크 분석

#### 네트워크 분석

## **개체와 이들 사이의 관계로 구성된 네트워크**를 수학적으로 표현하고 시각화하여, **패턴, 구조, 동작 원리**를 이해하는 방법론

자세한 설명은 2주차 ppt 참고!

사용자와 어종을 각각 노드, 연결 여부를 엣지로 설정해 네트워크를 구성

사용자	어종	점수	Sampled User-Fish Rating Network (100 users)
@kjyblack	갑오징어류	4	· · · · · ·
@kjyblack	밴댕이	3	기계크 그러는 지표로 드세 조나 이즈 바그 나는!
@-jaewh###	문식에서 노	느간의 1	관계를 고려한 지표를 통해 <mark>중심 어종 발굴</mark> 시도!
@-jaewoni	문어	3	TS 알고리즘을 활용
@-beomibutnotbert	미더덕	<sub>4</sub> ΠI	
@-beomibutnotbert	대구	4	시각화한 것을 나중에 넣겠습니다
@-chaebini	고등어류	3	
@-soooheee	광어	4	
•	:	•	User (100 sampled)
@-soooheee	도치	5	



#### 최초 시도: 중심 어종 발굴 | HITS

#### HITS (Hypertext Induced Topic Selection)

웹페이지를 **허브(hub)와 권위(authority)**로 나누어 **각각의 점수를 반복적으로 계산**하는 알고리즘 허브는 권위 있는 페이지를 많이 연결하고, 권위는 허브 페이지에게 많이 연결되는 방식으로 서로를 강화

**좋은 출처(허브)** 와 **신뢰받는 문서(권위)**를 분리해 찾을 수 있도록 함.



Hub 점수 계산

$$h_{i+1}(p) = \sum_{q:(p\to q)} a_i(q)$$

Authority

$$a_{i+1}(p) = \sum_{q:(q \to p)} h_i(q)$$

$$\sum_{p} h_{i}(p)^{2} = \sum_{p} a_{i}(p)^{2} = 1$$



#### STEP 1: 중심 사용자 발굴 | 정의

STEP 1
네트워크 분석으로 중심 사용자 발굴

STEP 2

사용자 중심성 반영 어종별 점수 계신 중 STEP 3

가격이 낮고

거래량이 적은

어종 선별

2

저평가 어종 도출

본 분석에서 발굴하고자 하는 중심 사용자!

다음 사람들을 일명 '맛잘알' 로 여기고 점수를 높게 반영



여러 어종을 먹어본 사람

(= 여러 어종과 연결된)



대중적인 어종을 먹어본 사람

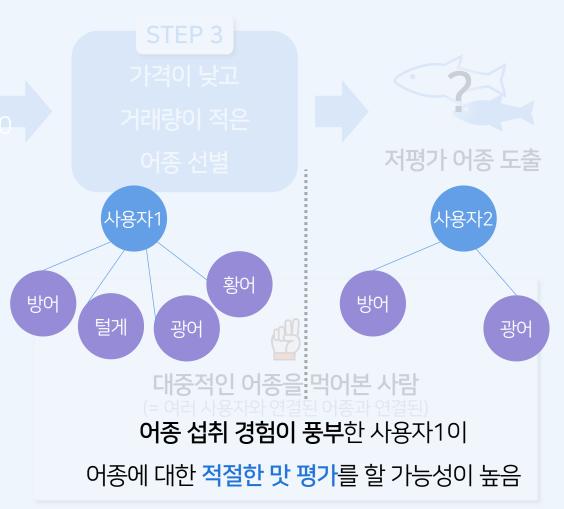
(= 여러 사용자와 연결된 어종과 연결된)



#### STEP 1: 중심 사용자 발굴 | 정의

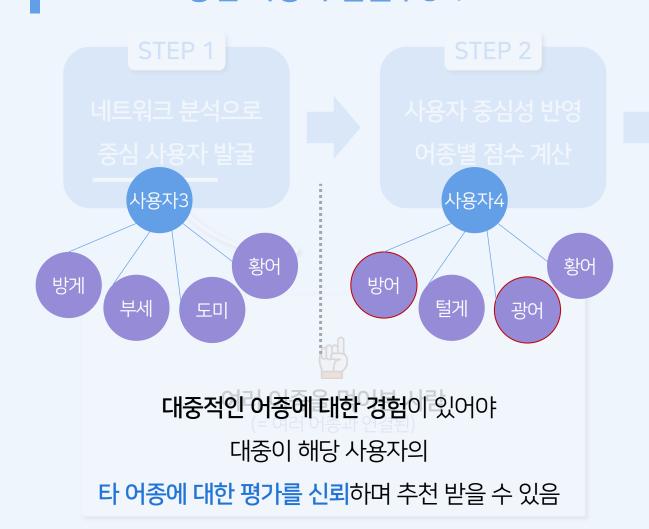








#### STEP 1: 중심 사용자 발굴 | 정의



#### ----- **여기서 잠깐!** 선정된 **저평가 어종**을 대중에게 추천해

선성된 서평가 어송을 내중에게 수선해 어종별 수요를 분산시키는 것이 최종 목표



서종 선별

저평가 어종 도출



#### 대중적인 어종을 먹어본 사람

(= 여러 사용자와 연결된 어종과 연결된)



#### STEP 1: 중심 사용자 발굴 | 적용



#### 여러 어종을 먹어본 사람

(= 여러 어종과 연결된)



#### 대중적인 어종을 먹어본 사람

(= 여러 사용자와 연결된 어종과 연결된)

두 가지 유형에 해당하는 사용자가 높은 중심성(허브)값을 갖도록 설계

#### 중심 사용자 발굴 알고리즘



Weight=1로 두고 평점을 반영하지 않음



HITS 알고리즘을 통해 사용자 허브값(hub) 계산

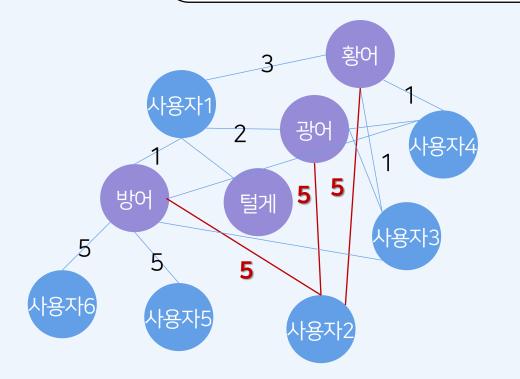


#### STEP 1: 중심 사용자 발굴 | 적용

#### 중심 사용자 발굴 알고리즘



Weight=1로 두고 평점을 반영하지 않음







평점이 **가중치로서 누적**되며 사용자 노드의 **중심(허브)값**이 크게 증가하는데 모든 어종에 <mark>높은 평점(5점)</mark>을 부여하는 사람이 **중심성이 높아지는** 문제

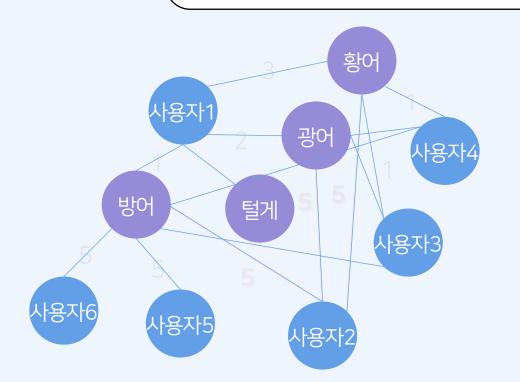


#### STEP 1: 중심 사용자 발굴 | 적용

#### 중심 사용자 발굴 알고리즘



Weight=1로 두고 평점을 반영하지 않음





명점이 **가중치로서 누적**되며



Weight=1로 설정함으로써 점수 미반영

모든 어송에 **마냥 높은 평점(5점)**을

부여하는 사람이 **중심성이 높아지는** 문제

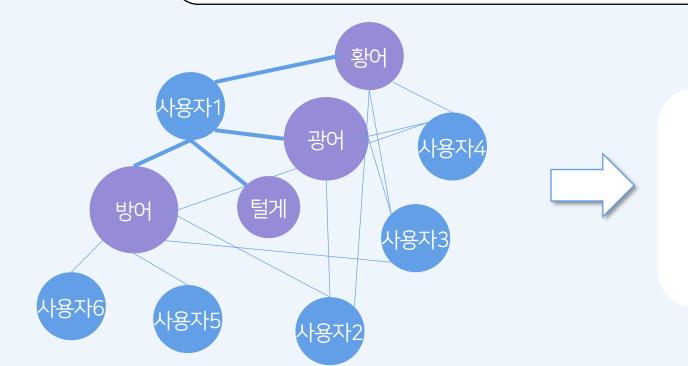


#### STEP 1: 중심 사용자 발굴 | 적용

#### 중심 사용자 발굴 알고리즘



HITS 알고리즘을 통해 사용자 허브값(hub) 계산





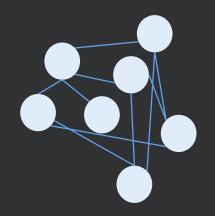
결과적으로 **방어, 광어와 같은**대중적인 어종을 섭취한 경험이 있고(점수X)

섭취한 어종의 수가 많을수록

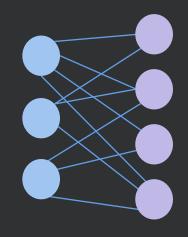
높은 허브값(hub)을 가짐



#### 왜 HITS로 결정했을까?



본 분석에서는 어종 - 사용자 사이에만 엣지가 있기 때문에 고유벡터 중심성과 HITS를 통한 결과 사이에 큰 차이가 존재하지 않음다만, 고유벡터 중심성에서는 모든 노드를 다 같은 것으로 보기 때문에, 분석 방법의 특성상 조금 더 직관적인 방법은 HITS라고 판단!



#### 고유벡터 중심성

사용자 
$$<\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$
 이번 분석에서는 사용되지 않음!

#### **HITS**

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$
 — 사용자  $\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$  — 어종



#### STEP 2: 사용자 중심성 반영 점수 계산



#### ✓ 어종의 내재적 가치 점수:

#### 단순 산술평균이 아닌 **사용자의 허브(hub)**가 반영된 **가중평균**으로 계산

사용자	사용자 허브 점수	어종에 대한 평점		
@kjyblack	0.3	1		
@-jaewoni	0.6	5		
@-beomibutnotbert	0.2	2		

단순 산술 평균

가중 평균

$$\frac{1+5+2}{3} \approx 2.7$$

$$\frac{0.3 \times 1 + 0.6 \times 5 + 0.2 \times 2}{1.1} \approx 3.4$$



#### STEP 2: 사용자 중심성 반영 점수 계산



#### ✓ 어종의 내재적 가치 점수:

단순 산술평균이 아닌 **사용자의 허브(hub)**가 반영된 **가중평균**으로 계산

사용자	사용자 허브 점수	어종에 대한 평점		
@kjyblack	0.3	1		
@-jaewoni	0.6	5		
@-beomibutnotbert	0.2	2		

'맛잘알' 사용자2의 평점이 더 크게 작용함!

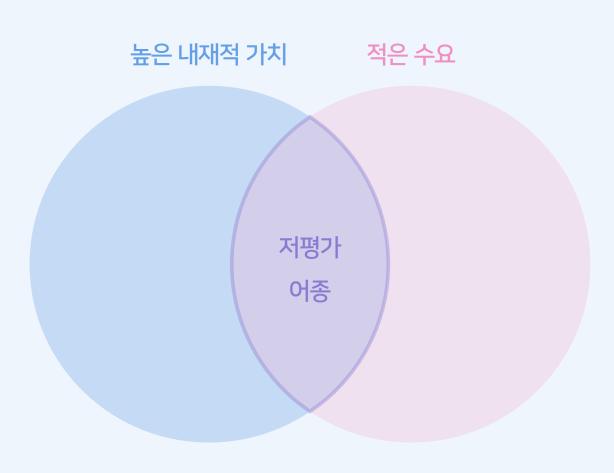


#### STEP 3: 가격 거래량을 통한 수요 파악

어종의 **내재적 가치**를 계산했으니, 이를 기반으로 **저평가 어종을 선별**해야 함

#### 저평가 어종

**내재적 가치**가 높으면서, **수요**가 적은 어종 즉, 현재 인기는 없지만 **잠재적으로 인기가 있을 만한** 어종





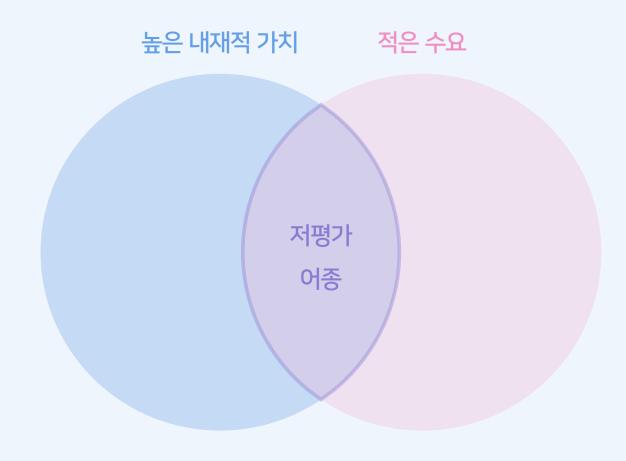
#### STEP 3: 가격 거래량을 통한 수요 파악

높은 내재적 가치

STEP 2에서 계산 완료!

적은 수요

수요·공급 법칙에 따른 가격과 거래량을 고려

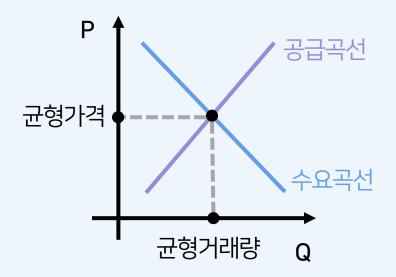




#### STEP 3: 가격 거래량을 통한 수요 파악 | 바탕 이론

#### 시장균형

가격의 자유로운 조정기능에 따라 어떤 가격에서 수요량과 공급량이 일치하는 상태로, 우리가 시장을 관찰할 때에는 대체로 **균형 상태에 있는 시장을 관찰** 





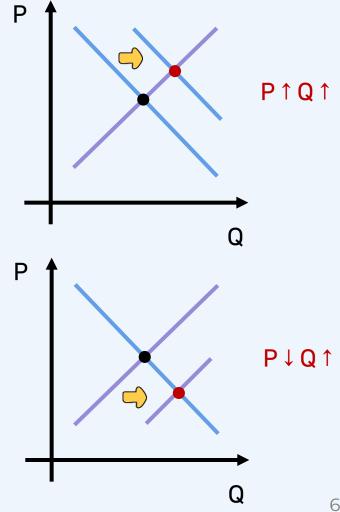
#### STEP 3: 가격 거래량을 통한 수요 파악 | 바탕 이론

#### 수요의 변동

가격 외 다른 요인에 의해 수요량이 변한 경우 (예. 백골뱅이 열풍) 균형은 공급곡선을 따라 움직이고 **가격과 수량은 같은 방향으로 변화** 

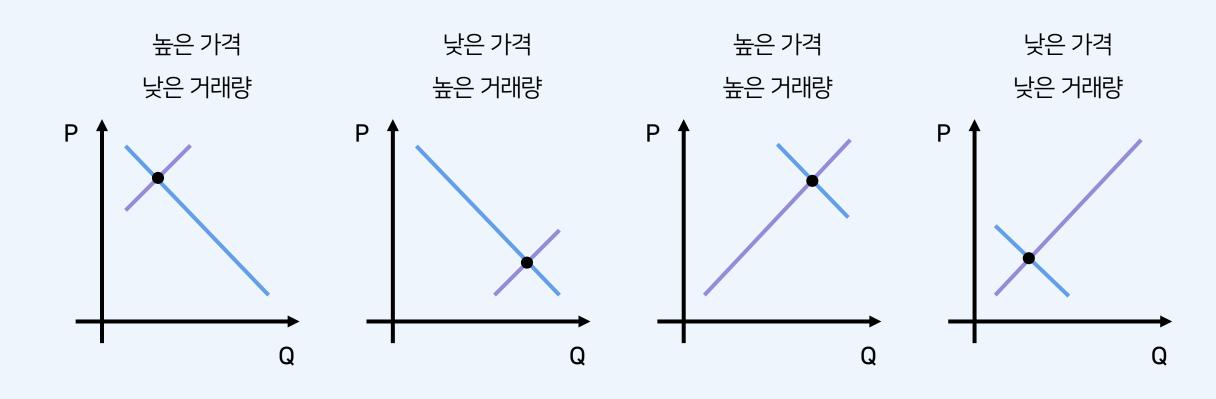
#### 공급의 변동

가격 외 다른 요인에 의해 공급량이 변한 경우 (예. 양식 기술 발전) 균형은 공급곡선을 따라 움직이고 **가격과 수량은 다른 방향으로 변화** 





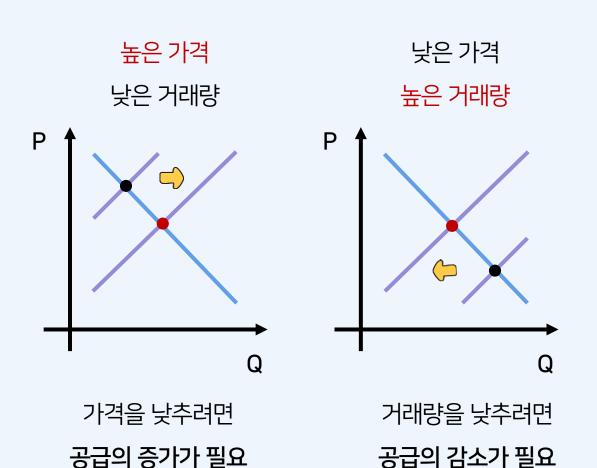
#### STEP 3: 가격 거래량을 통한 수요 파악 | 바탕 이론

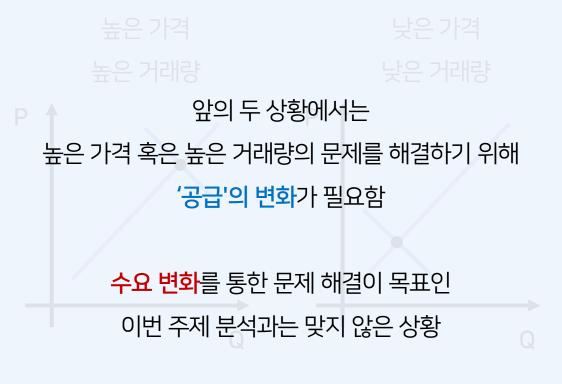


균형 상태를 가격, 거래량의 상대적 크기에 따라 네 가지 경우로 분류해볼 수 있음



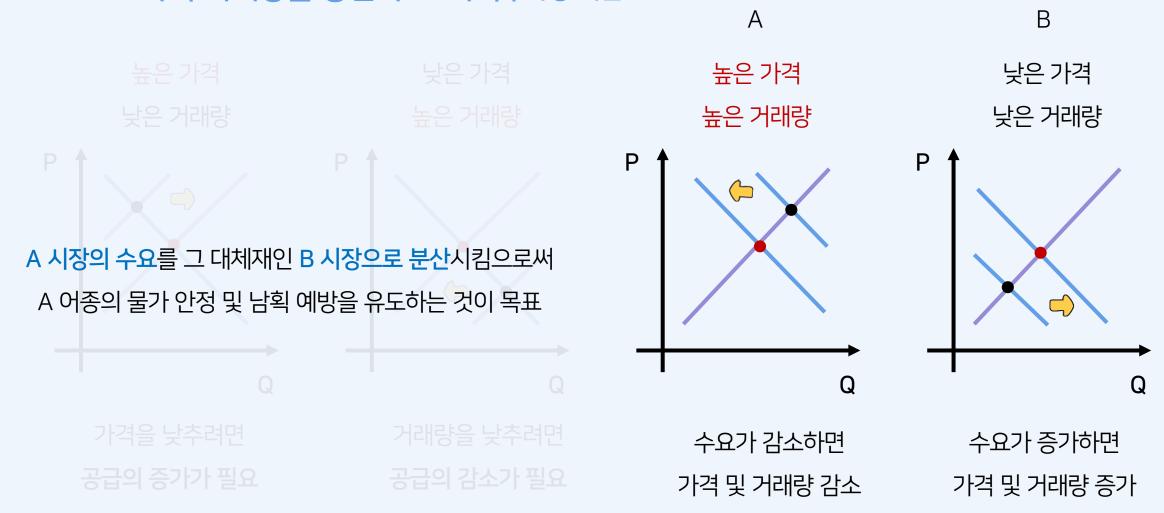
#### STEP 3: 가격 거래량을 통한 수요 파악 | 바탕 이론







#### STEP 3: 가격 거래량을 통한 수요 파악 | 바탕 이론





#### STEP 3: 가격 거래량을 통한 수요 파악 | 바탕 이론



이 어종의 수요를 늘림으로써 남획 위기 어종의 수요 분산 대상이 될 수 있는 상태!

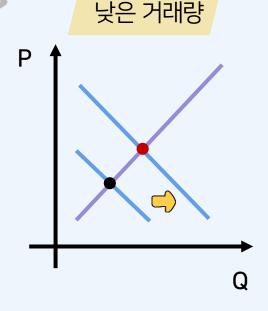
A 시장의 수요를 그 대체적, 본 분석에서 찾고자 하는 저평가 어종 후보군

A 어종의 물가 안정 및 남획 예방을 유도하는 7



어종들의 상대적인 가격, 거래량을 반영하여

현재 균형이 낮은 가격, 낮은 거래량에서 형성되어 있는 어종군을 찾자!



В

낮은 가격

수요가 증가하면 가격 및 거래량 증가



#### STEP 3: 가격 거래량을 통한 수요 파악 | 수요가 적은 어종

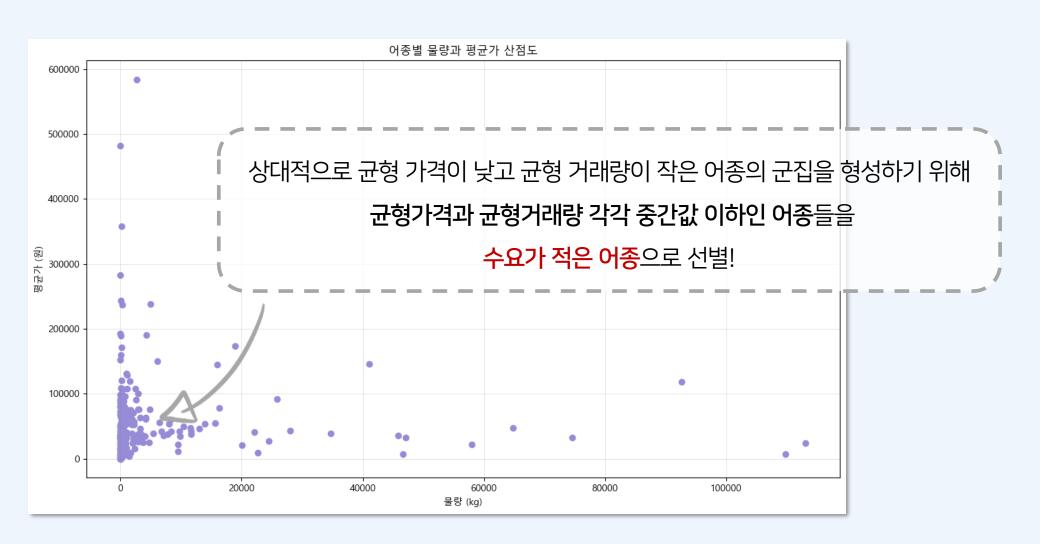
수산물은 계절에 따라 균형이 변하기 때문에, 다른 수산물과의 상대적인 가격과 거래량을 고려할 수 있는 최근 1년치 가격의 물량의 평균을 사용!



2주차에 미리 구해 두었던 **위판장 데이터를 거래량**으로 사용하고, 같은 출처에서 **가격 정보를 추가로 수집** 



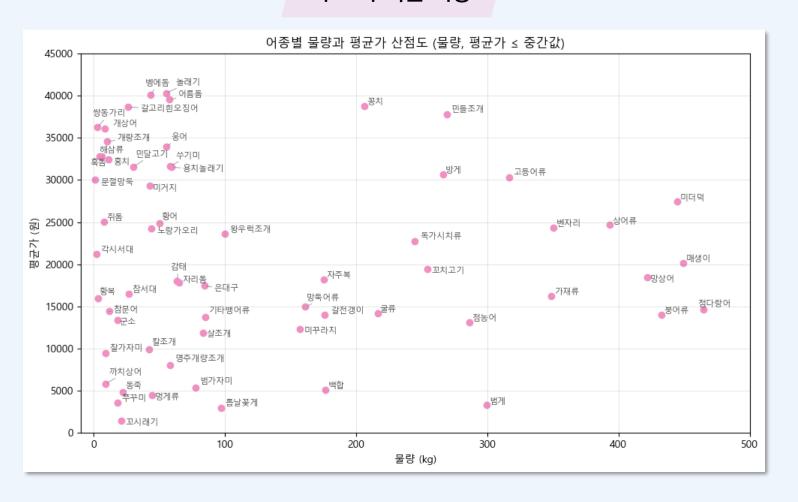
#### STEP 3: 가격 거래량을 통한 수요 파악 | 수요가 적은 어종





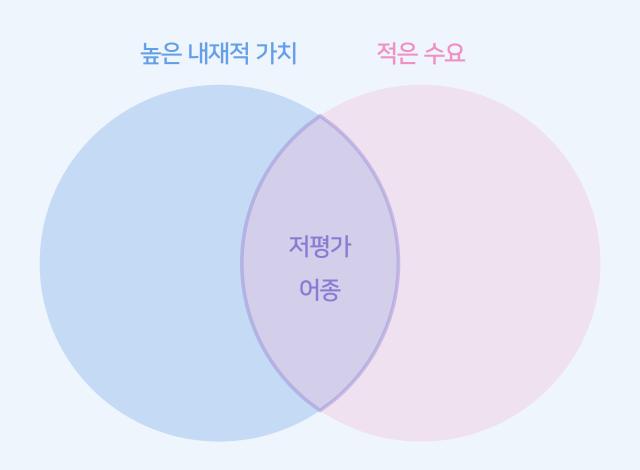
#### STEP 3: 가격 거래량을 통한 수요 파악 | 수요가 적은 어종

#### 수요가 적은 어종





최종: 저평가 어종



수요가 적은 어종군 중에서, 내재적 가치 점수가 높은 30개의 어종을 선정



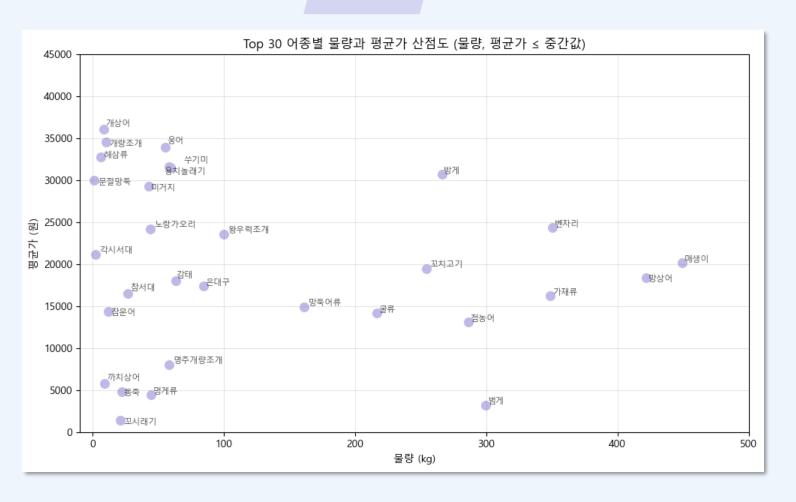
최종 저평가 어종 도출!





최종: 저평가 어종

#### 이 중 내재적 가치 점수가 높은 30개 어종을 선별하여 남은 최종 저평가 어종 TOP 30!





#### 추천 어종 선정



저평가 어종 탐색 목적이 '남획 감축을 지원하기 위한 수요 다각화'에 있었던 만큼, 알고리즘상 '저평가 어종'으로 나왔다고 해서 모두 최종 추천 어종으로 선정할 수 없음

고려해야 하는 이슈들

금어기/금지체장이 걸려있진 않은지

어획량(혹은 유통량)이 눈에 띄게 감소하진 않았는지

지나치게 가격이 비싼 희귀어종은 아닌지

보호종은 아닌지



## 추천 어종 선정 | 금어기, 금지체장



#### 금어기/금지체장이 걸려있는 어종은

#### 상위 저평가 어종이더라도 일괄적으로 제외

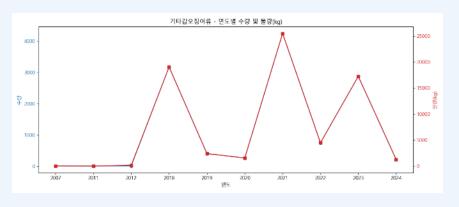
어종	언급_작성자수	가 <del>중</del> 평균점수	산술평균점수	가 <del>중</del> 평균순위	산술평균순위	순위차이	금어기/금지체장 어종
용가자미	8	4.845	4.375	13	23	-10	금어
참가자미	44	4.692	4.031	18	62	-44	금어
미거지	7	4.508	4.429	29	18	11	금어
볼락	176	4.365	4.387	40	22	18	금어
쥐노래미	47	4.142	3.915	67	104	-37	금어

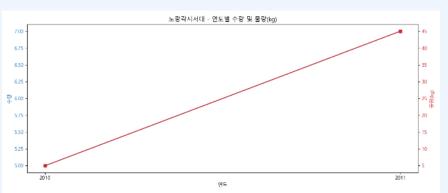


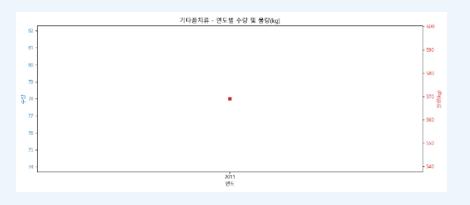


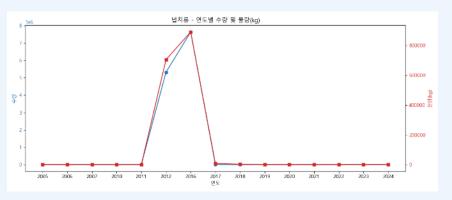
## 추천 어종 선정 | 어획량(유통량)

이때 2005년~2024년 각 **연도별 감소율의 평균**값, **2005년 대비 2024년 감소율** 등 다양한 기준을 고려하였으나 어종별로 각각 그래프를 확인해봤을 때 일괄적인 기준을 적용하기는 어렵다고 판단







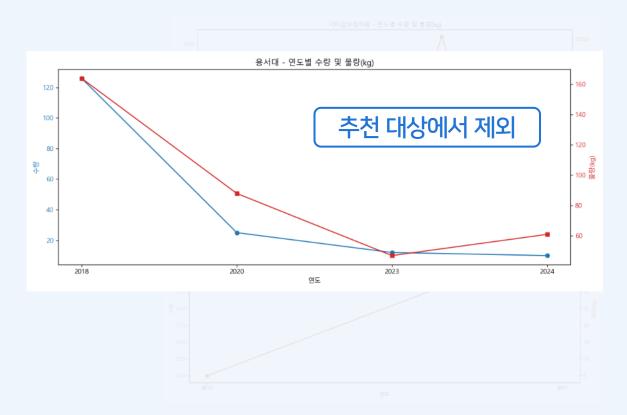




# 추천 어종 선정 | 어획량(유통량)

이때 2005년~2024년 감소율 등 추천 후보 어종들을 대상으로만 그래프를 확인하여,

명확한 감소 추세가 보일 경우엔 추천 대상에서 제외하는 방식 선택







# 추천 어종 선정 | 희귀어종

#### 어획량이 감소 추세가 아니더라도, **어획량 자체가 적거나** 가격이 매우 높은 어종은 추천 목적에 부합하지 않음

#### ■ 상류층도 먹기 힘든 명품 횟감, 노랑가자미

표준명: 노랑가자미(가자미목 가자미과)

방언: 데이터 없슴 (일본에선 마쓰, 타카노, 소나무 껍질 가자미 정도로 불림)

영명: Barfin flounder

일명: 마츠카와(マツカワ)

전장: 80cm

분포: 동해, 남해(부산), 일본 홋카이도, 사할린, 이바라키현 이북의 태평양, 타타르해협, 오호크츠해, 쿠릴열도

음식: 회, 초밥, 소금구이, 튀김

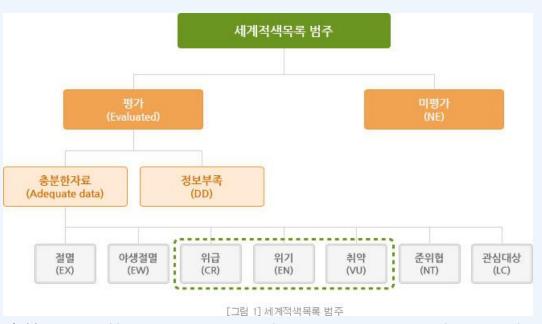
제철: 겨울(2~4월)

출처: '입질의 추억' 블로그 https://blog.naver.com/slds2/10188589189





## 추천 어종 선정 | 보호종



멸종우려(Threatened)에 해당하는 위급(CR), 위기(EN), 취약(VU) 범주에 속하는 어종은 추천에서 제외, 관심대상(LC)에 속하는 어종은 추천

출처: https://species.nibr.go.kr/endangeredspecies/rehome/redlist/redlist\_cate.jsp

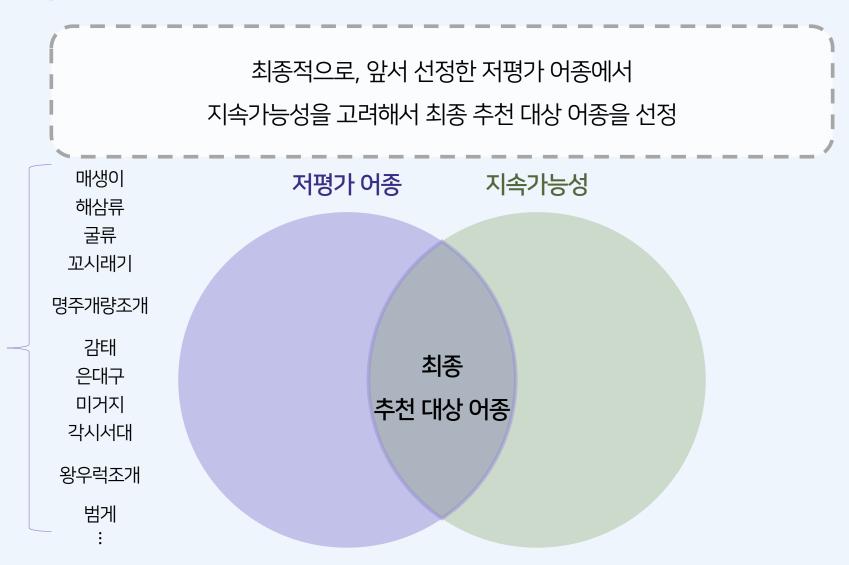
관심대상(LC)

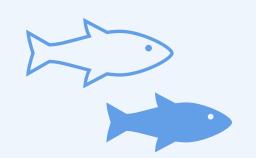
기준에 따라 평가했으나 위급, 위기, 취약, 준위협에 해당하지 않은 상태로 널리 퍼져 있고 개체수도 많은 분류군이 이 범주에 해당

번역 용어가 '관심대상'이라 보호 대상에 속하는 것 같지만 아님



# 추천 어종 선정







#### 추천 시스템 개요

그래프 분석을 통해 저평가 어종을 추려내고 필터링 기준도 선정 각 소비자의 선호에 맞는 저평가 어종을 추천해주는 추천시스템 서비스 아이디어 제안



∰ 메모리 기반 협업 필터링

가장 직관적이고 고전적인 추천 방식으로, 비슷한 특성을 가진 유저들이 선호하는 아이템을 바탕으로 추천을 생성

Ex) 코사인 유사도, 피어슨 상관계수 등을 이용한 유사도 지표 계산



🖞 모델 기반 협업 필터링

사용자-아이템 상호작용 데이터를 바탕으로 기계학습 모델을 학습하여 잠재요인 생성, 이를 기반으로 추천을 생성

Ex) 행렬분해, 딥러닝 기반 추천 모델



#### 추천 시스템 개요

그래프 분석을 통해 저평가 어종을 추려내고 필터링 기준도 선정 각 소비자의 선호에 맞는 저평가 어종을 추천해주는 추천시스템 서비스 아이디어 제안



∰ 메모리 기반 협업 필터링

가장 직관적이고 고전적인 추천 방식으로, 비슷한 특성을 가진 유저들이 선호하는 아이템을 바탕으로 추천을 생성





데이터가 희소하고 규모가 커질 경우 계산량, 품질 모두에서 한계를 보임

Ex) 코사인 유사도, 피어슨 상관계수 등을 이용한 유사도 지표 계산



### 추천 시스템 개요

그래프 분석을 통해 저평가 어종을 추려내고 필터링 기준도 선정 각 소비자의 선호에 맞는 저평가 어종을 추천해주는 추천시스템 서비스 아이디어 제안





🖞 모델 기반 협업 필터링

대규모 데이터에서도 효율적으로 작동, 희소한 데이터 상황에서도 일반화된 추천 가능



사용자-아이템 상호작용 데이터를 바탕으로 기계학습 모델을 학습하여 잠재요인 생성, 이를 기반으로 추천을 생성

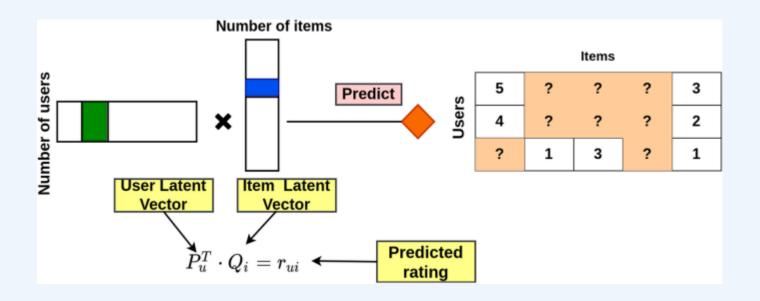
Ex) 행렬분해, 딥러닝 기반 추천 모델



### 방법 | SGD MF

#### SGD 행렬분해(Stochastic Gradient Descent Matrix Factorization)

(사용자 x 아이템) 평점 행렬을 사용자 잠재 요인( $p_u$ )과 아이템 잠재 요인( $q_i$ )으로 분해하여 관측되지 않은 평점을 예측하는 모델 기반 협업 필터링 기법





# 평가지표

추천시스템에서는 RMSE, MAE와 같은 기본적인 예측성능을 평가하는 지표 외에 다양한 평가 지표를 사용



추천시스템의 목적에 맞게 평가 지표를 설정하고 해당 평가 지표를 개선하는 방향으로 추천시스템을 개선해야!

### 평가지표

추천시스템에서는 RMSE, MAE와 같은 기본적인 예측성능을 평가하는 지표 외에 **다양한 평가 지표**를 사용

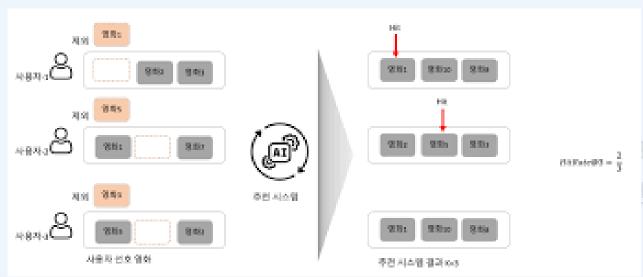




#### 평가지표

#### Hit Rate@k

- 1. 특정 유저의 선호하거나 클릭했던 모든 Item을 가져온다.
  - 2. 모든 Item가운데 하나만 의도적으로 제거한다.
- 3. 남은 Item들을 가지고 추천 모델을 학습한 뒤, Top K 추천 리스트를 추출
- 4. K개의 추천 리스트 가운데 아까 제거한 Item이 있다면 hit, 아니면 hit가 아니다.



평가 방식 변경 후 평가 방식에 맞게 하이퍼파라미터 튜닝 진행

표를 설정하고



시스템을 개선해야!

최종결과 0.3937로 준수한 성능을 보임



# 결과 예시

# 예시로, **@냥이와탄이** 유저명을 넣었을 때, '추천 어종 리스트' 중 **높은 예측 점수**를 가진 어<del>종</del>을 출력

추천 어종	예측 점수	가중평균순위
참서대	0.433836	59.0
벤자리	0.334147	99.0
명주개량조개	0.297248	19.0
꼬시래기	0.269619	17.0
점농어	0.166806	90.0
매생이	0.107156	8.0
노랑가오리	0.102562	66.0