

캡스톤설계프로젝트

Kingo Manager

Team AIU

목차

1. Objective
2. Milestone
3. Role
4. Frontend
5. Backend
6. Dataset
7. AI
8. Challenge & Evolution
9. Limitation & Future Work

Objective



Kingo Manager

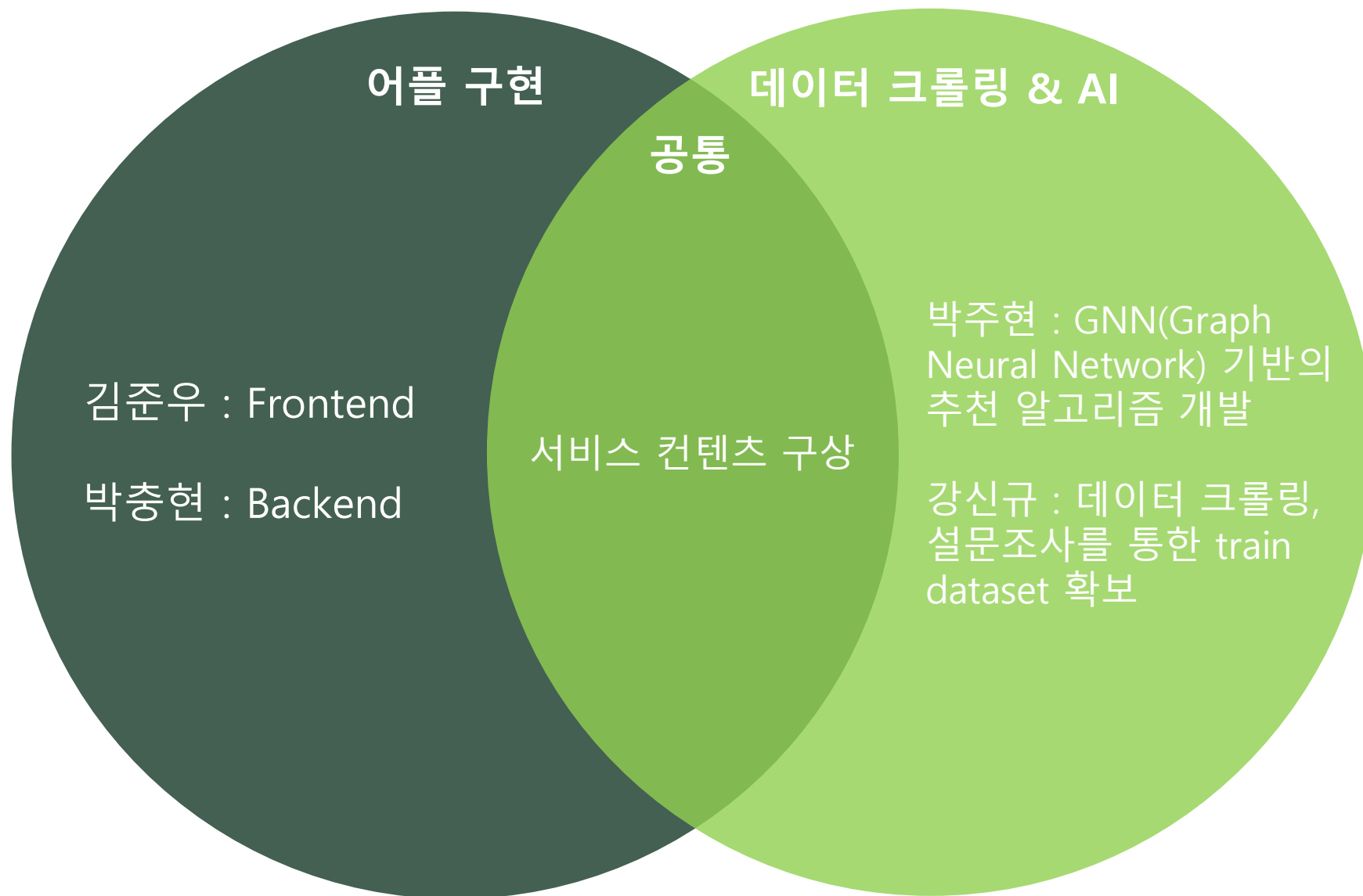
성균관대학교 진로활동 추천 서비스

1. 학사생활에 대해서 잘 모르는 새내기들 시기에 맞는 진로활동을 제공
2. 3품 충족 여부, 진로 관심사를 고려하여 맞춤 콘텐츠를 추천
3. 진로활동 정보를 찾기위해 다양한 웹사이트를 방문해야 하는 수고를 덜어줌

Milestone

No.	분야	수행내용/기간	3~4주	5~6주	7~8주	9~10주	11~12주	13~14주	15주
1	Frontend	회원가입/로그인							
2	Frontend	메인페이지 구현							
3	Frontend	Backend 통신							
4	Backend	Lamda 서버 deploy							
5	Backend	RDS 테이블 구축							
6	Dataset 전처리	설문조사							
7	Dataset 전처리	Data Augmentation							
8	Dataset 전처리	Crawling							
9	AI	모델 선정							
10	AI	모델 개발							
11	AI	Backend 연동							
12	디버깅	디버깅							

Role of Each Member



데이터 범위

1. 학점관리
2. 자기이해와 진로탐색
3. 봉사
4. 동아리, 학회
5. 외국어 점수 획득(국제품)
6. 자격증
7. 공모전/외부활동
8. 졸업요건
9. 기업인턴십/산학협력
10. 취업캠프/서류,면접 클리닉
11. 대학원/학석연계

크롤링

1. 성균관대학교 교내봉사 ur1004
2. 성균관대학교 교내봉사프로그램 ur1004 공지사항
3. 1365 수원 봉사활동 리스트
4. vms 수원 봉사활동리스트
5. 잡코리아 매출 1000대기업
6. 잡코리아 인턴 채용공고
7. 인크루트 신입 100대기업
8. 인크루트 인턴 채용공고
9. 사람인 정규 + 인턴 채용공고
10. 성균관대학교 소프트웨어대학 취업, 인턴십 공고
11. 챌린지스퀘어 비교과프로그램
12. 톨게이트 인턴/산학협력 모집공고
13. 캠퍼스몬 공모전
14. 위비티 공모전

크롤링 계속

- 15. 대학원 연구소별 소식, 정보
- 16. 동아리 정보
- 17. 토익, 오픽 안내
- 18. 학생성공 게이트웨이 해외유학생/교환학생 정보
및 상담 연계
- 19. 글로벌 하계학기 등 다양한 외국어 프로그램
- 20. 학생인재 개발팀 자기이해와 진로탐색을 위한
가이드북
- 21. 챌린지 스퀘어 자기 이해와 진로탐색
- 22. 카운슬링센터
- 23. 교육개발센터 학습컨설팅
- 24. 레드커넥트 헌혈 어플리케이션

카테고리	항목 수(개)
학점관리	6
자기이해와 진로탐색	8
봉사	47
동아리, 학회	98
외국어 점수 획득(국제품)	11
자격증	5
공모전/외부활동	37
기업인턴십/산학협력	18
졸업요건	5
취업캠프/서류,면접 클리닉	51
대학원/학석연계	65
합계	351

Contents

Title	Category	contents
학습 컨설팅	학점 관리	효과적인 학습 방법을 찾고자하는 고민을 덜기 위한 학습 진단 및 코칭 서비스
성균 튜터링	학점 관리	학습자 주도의 상호작용적 협동학습의 기회를 제공하고 자기 주도적 학습에 기반한 학습
성균 공부방	학점 관리	방학 기간 중 재능을 나눠주는 활동
FSD그룹코칭	학점 관리	학생들의 학습역량을 신장시키기 위해 다양한 특강이나 워크숍 개최
교육개발팀 공지	학점 관리	학습 관련 활동 공지사항 확인
카운슬링센터	학점 관리	정신적으로 힘든 것과 관련해서(더 나은 대학 생활을 위해) 상담을 하는 공간

수작업으로 입력된 정보

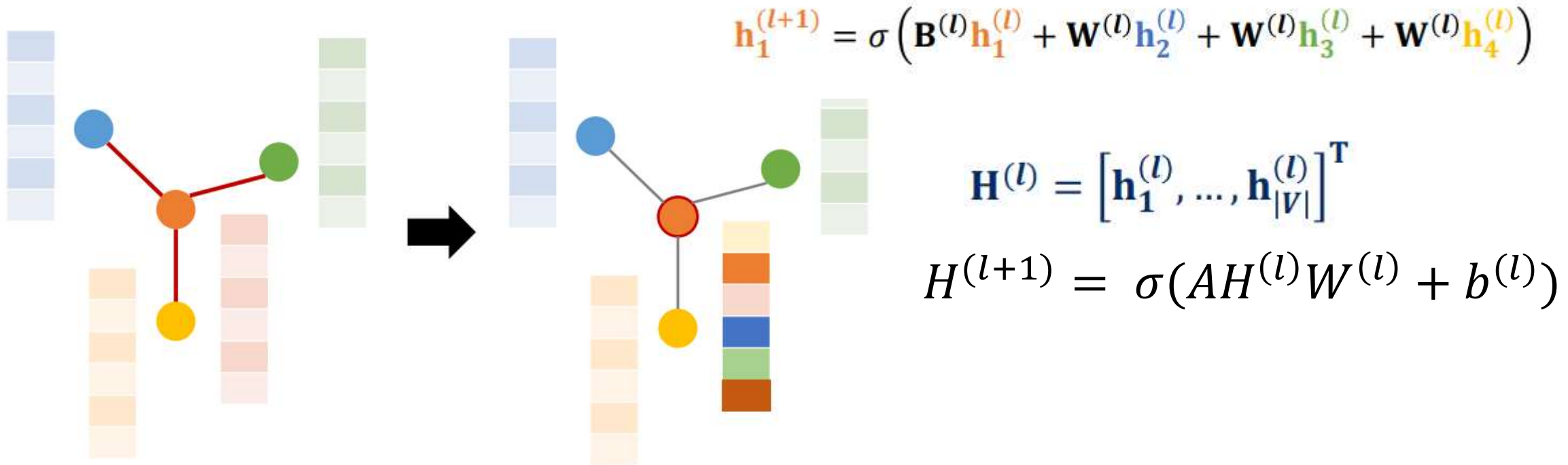
	A	B	C	D
1	title	category	contents	link
2	title	category	contents	link
3	헌혈도 앱으로 하는 시대 레드 커넥트	봉사	가까운 헌혈의집 확인부터 예약까지!	https://play
4	성균관대학교 UR1004 봉사활동	봉사	379 행사보조 NEW 이윤미2022-11-21 흥사단교육운동본부 2022.11.26~2022.11.26 3 0	https://ur10
5	성균관대학교 UR1004 봉사활동	봉사	378 도산의밤 상패, 메달 전달 요원 봉사(여자 2명) 이지영 차장2022-11-02 도산아카데미 2022.11.09~2022.11.09 2 0	https://ur10
6	성균관대학교 UR1004 봉사활동	봉사	377 창경궁 나들이 엄주랑2022-10-12 창경어린이집 2022.10.28~2022.10.28 6 1	https://ur10
7	성균관대학교 UR1004 봉사활동	봉사	376 제20회 이곳만은 꼭지키자 시민공모전 시상식 김화정2022-10-04 한국내셔널트러스트 2022.10.22~2022.10.22 2 0	https://ur10
8	성균관대학교 UR1004 봉사활동	봉사	375 후원물품 정리(창고정리) 자원봉사자 모집 문소영2022-09-28 사단법인 사랑의친구들 2022.10.05~2022.10.07 4 0	https://ur10
9	성균관대학교 UR1004 봉사활동	봉사	374 사무보조 이윤미2022-09-27 흥사단교육운동본부 2022.09.30~2022.09.30 1 1	https://ur10
10	성균관대학교 UR1004 봉사활동	봉사	373 '2022 사랑의 친구들 바자' 행사 물품 운반 자원봉사자 모집 ① 문소영2022-09-06 사단법인 사랑의친구들 2022.09.06~2022.09.06 1 0	https://ur10
11	성균관대학교 UR1004 봉사활동	봉사	372 '2022 사랑의친구들 바자' 행사 진행보조 자원봉사자 모집 ② 문소영2022-09-06 사단법인 사랑의친구들 2022.09.06~2022.09.06 1 0	https://ur10
12	성균관대학교 UR1004 봉사활동	봉사	371 '2022 사랑의친구들 바자' 행사 진행보조 자원봉사자 모집 문소영2022-09-06 사단법인 사랑의친구들 2022.09.06~2022.09.06 1 0	https://ur10

데이터 크롤링을 통해서 입력된 정보

11개의 진로활동 카테고리로 구분하고, 351개의 콘텐츠 정보를 수집

AI – Basic of GCN

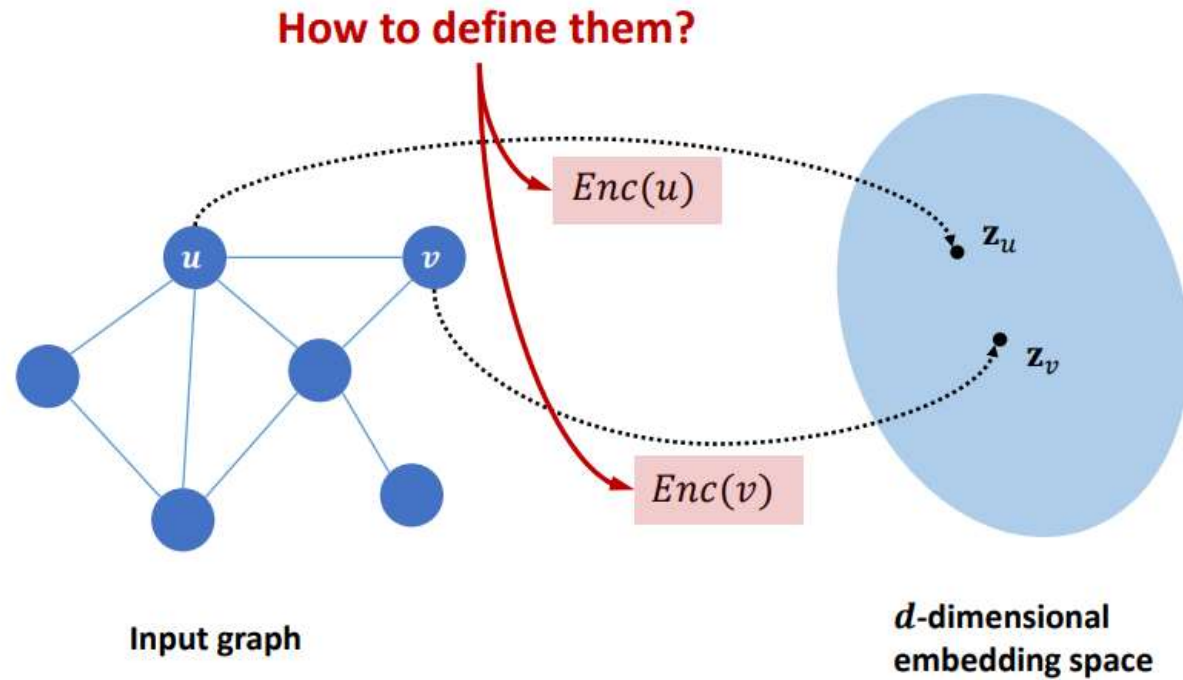
- GCN : Compute the messages for adjacent nodes and a target node.
- Nodes are updated differently by neighbor nodes.



AI – Basic of GCN(1)

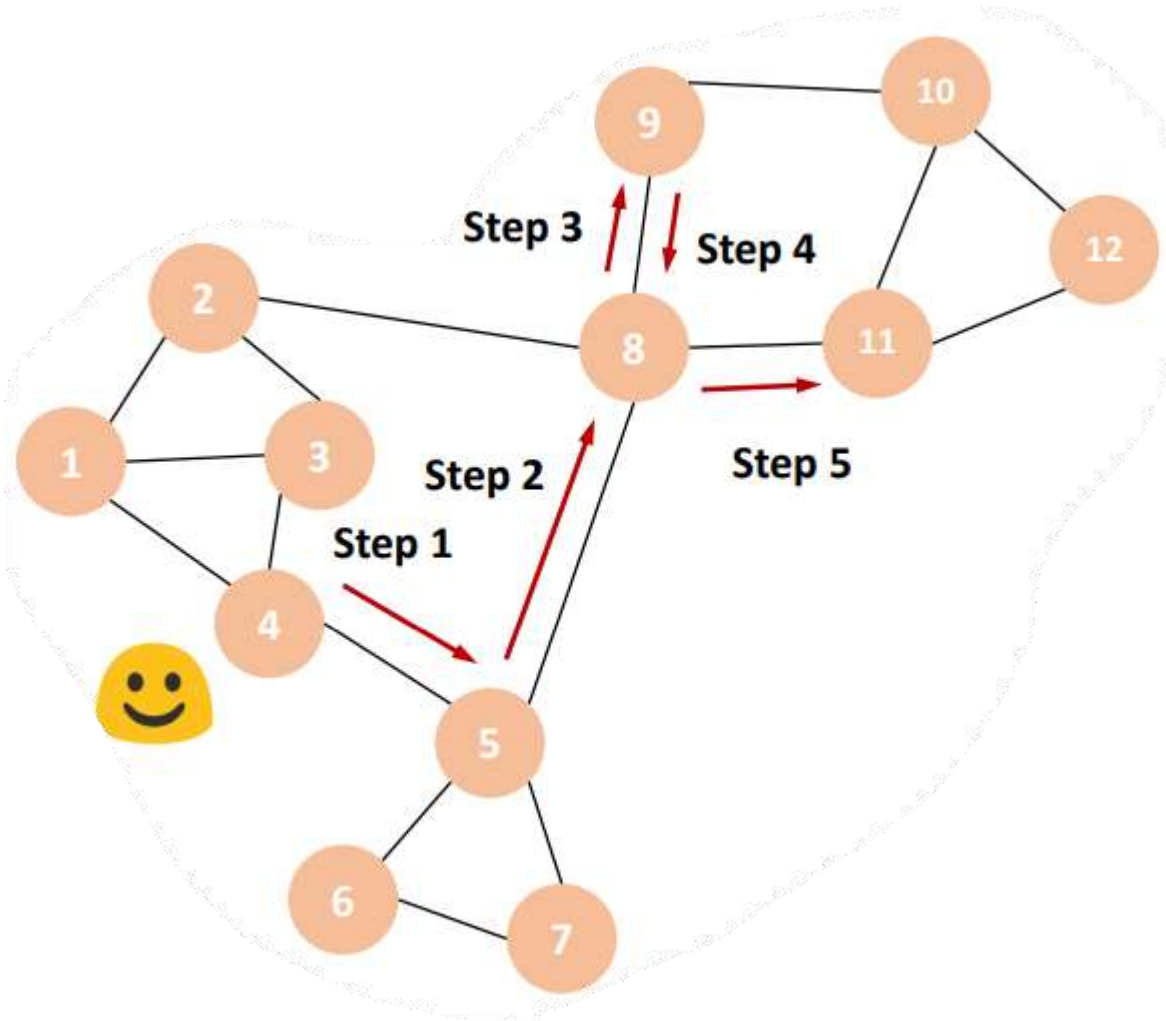
- Node Embeddings

$$\text{similarity}(u, v) \approx \mathbf{z}_v^T \mathbf{z}_u$$



AI – Basic of GCN(2)

- Random Walk

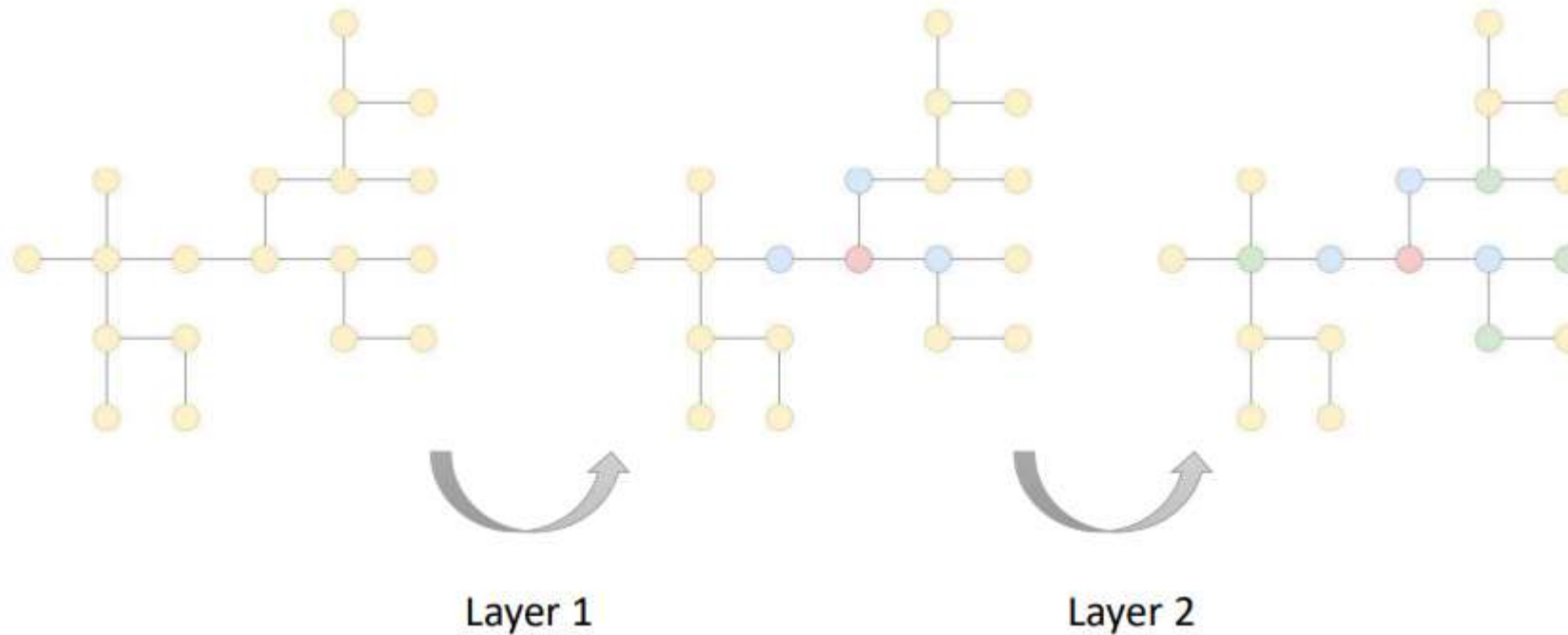


$$\mathbf{z}_v^T \mathbf{z}_u \approx$$

Probability that u and v
co-occur on a random
walk over the graph.

AI – Basic of GCN(3)

- Average neighborhood information and stack neural network



Dataset Example

uid(node)	feature	target(y)
0	(0,0)	0
1	(0,1)	2
2	(0,1)	3
3	(0,4)	4
4	(0,6)	1

```
dataset > features.txt
1 {"0": [0, 0], "1": [0, 0], "2": [0, 0], "3": [0, 0], "4": [0, 0], "5": [0, 1], "13": [0, 1], "14": [0, 2], "15": [0, 2], "16": [0, 3], "17": [0, 3], "24": [0, 3], "25": [0, 3], "26": [0, 4], "27": [0, 5], "28": [0, 6], "36": [1, 0], "37": [1, 0], "38": [1, 0], "39": [1, 0], "40": [1, 1], "48": [1, 1], "49": [1, 1], "50": [1, 1], "51": [1, 1], "52": [1, 3], "59": [1, 3], "60": [1, 3], "61": [1, 3], "62": [1, 3], "63": [1, 3], "64": [1, 3], "65": [1, 3], "66": [1, 3], "67": [1, 3], "68": [1, 3], "69": [1, 3], "70": [1, 3], "71": [1, 6], "72": [1, 6], "73": [1, 7], "74": [1, 7], "75": [1, 7], "76": [1, 7], "77": [1, 7], "78": [1, 7], "79": [1, 7], "80": [1, 7], "81": [1, 7], "82": [1, 7], "83": [2, 2], "84": [2, 3], "85": [2, 3], "86": [2, 3], "87": [2, 3], "88": [2, 3], "89": [2, 3], "90": [2, 3], "91": [2, 3], "92": [2, 3], "93": [2, 3], "94": [2, 8], "95": [2, 8], "96": [2, 8], "97": [2, 8], "98": [2, 9], "99": [2, 9]}
```

Feature = {id : [학년, 관심사]}

```
dataset > target.csv
1 id,target
2 0,0
3 1,0
4 2,0
5 3,0
6 4,0
7 5,0
8 6,1
9 7,1
10 8,1
11 9,1
12 10,1
```

Target = {id, y값}

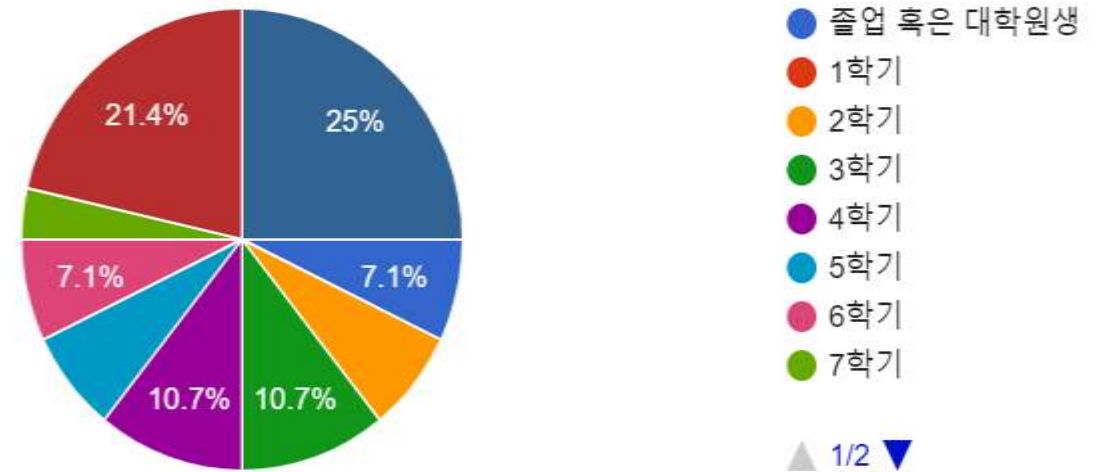
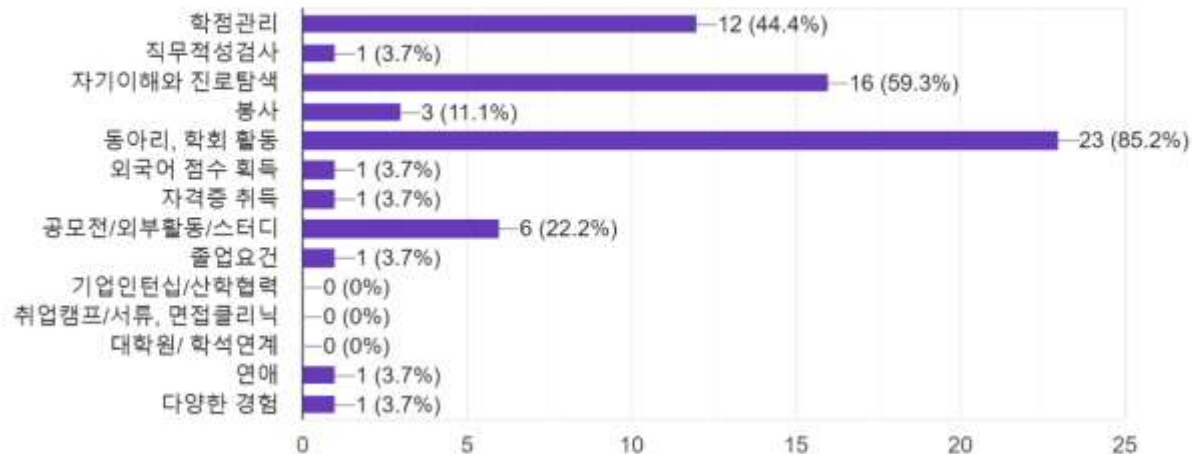
```
dataset > edge.csv
1 id_1,id_2
2 0,1
3 0,2
4 0,3
5 0,4
6 0,5
7 0,6
8 0,7
9 0,8
10 0,9
11 0,10
12 0,11
13 0,12
14 0,13
15 0,14
16 0,15
17 0,16
```

Edge = 같은 feature끼리 연결되도록

Train Dataset

1. 1학년에게 가장 추천하는 활동은 무엇인가요? (최대 3개)

응답 27개



성균관대학교 학부생 및 졸업생을 대상으로 학년별 관심 진로활동과 자주 방문하는 진로활동 관련 페이지를 설문조사

Train Dataset & Augmentation

16	후배관심사	학년	1	2	3	4
17	카테고리	학점관리	1	4	2	1
18		자기이해와 진로탐색	2	2	3	5
19		봉사	0	0	0	3
20		동아리, 학회	1	5	3	1
21		외국어 점수 획득(국제품)	0	0	0	0
22		자격증	0	1	0	2
23		공모전/외부활동	0	3	2	7
24		졸업요건	0	0	2	6
25		기업인턴십/산학협력	0	2	2	3
26		취업캠프/서류,면접 클리닉	0	1	1	6
27		대학원/학석연계	1	0	0	1
28		sum	5	18	15	35

	A	B	C	D	E	F
1	선배추천	학년	1	2	3	4
2	카테고리	학점관리	12	16	7	4
3		자기이해와 진로탐색	17	9	4	3
4		봉사	3	2	2	0
5		동아리, 학회	23	15	6	0
6		외국어 점수 획득(국제품)	1	2	2	1
7		자격증	1	0	0	2
8		공모전/외부활동	6	10	11	8
9		졸업요건	1	3	1	7
10		기업인턴십/산학협력	0	2	10	8
11		취업캠프/서류,면접 클리닉	0	0	1	6
12		대학원/학석연계	0	1	1	1
13		sum	64	60	45	40

설문조사에서의 관심도 비율을 고려하여 1286개의 feature node를 제작하였다.

학년별 feature node 계산 : $250 * (\text{후배 관심사 비율}) * (\text{선배추천 비율})$

Ex) 1학년, 관심사는 '자기이해와 진로탐색' 이라고 대답한 경우 feature [0,1] $250 * (2+1)/(5+11) = 46.875$, 즉 47개

Target 의 비율은 선배추천에 따라 47개중 target 이 봉사인 경우는 $47 * (3+1)/(64+11)=2.50666... = 3$ (올림)

즉 feature = [0,1, y = [2] 인 쌍이 3개가 생긴다.

총 1286개의 node 를 생성했다.

AI – Input, output, evaluation

id	email	password	name	grade	insung	changeui	gukje	interest
1	capstone@gmail.com	capstone	캡스톤	4	True	True	False	8

```
key = 'capstone_teamH_000'}
{'email': 'capstone@gmail.com', 'title': '교육개발팀 공지', 'category': '학점 관리', 'contents': '학습 관련 활동 공지사항 확인', 'link': 'https://ctl.skku.edu/ctl/'}
{'email': 'capstone@gmail.com', 'title': '성균 공부방', 'category': '학점 관리', 'contents': '방학 기간 중 재능을 나눠주는 활동', 'link': 'https://ctl.skku.edu/ctl/'}
{'email': 'capstone@gmail.com', 'title': '성균 튜터링', 'category': '학점 관리', 'contents': '학습자 주도의 상호작용적 협동학습의 기회를 제공하고 자기 주도적 학습을'}
ng.do', 'admin_key': 'capstone_teamH_000'}
{'email': 'capstone@gmail.com', 'title': '카운슬링센터', 'category': '학점 관리', 'contents': '정신적으로 힘든 것과 관련해서(더 나은 대학 생활을 위해) 상담을 하는'}
teamH_000'}
```

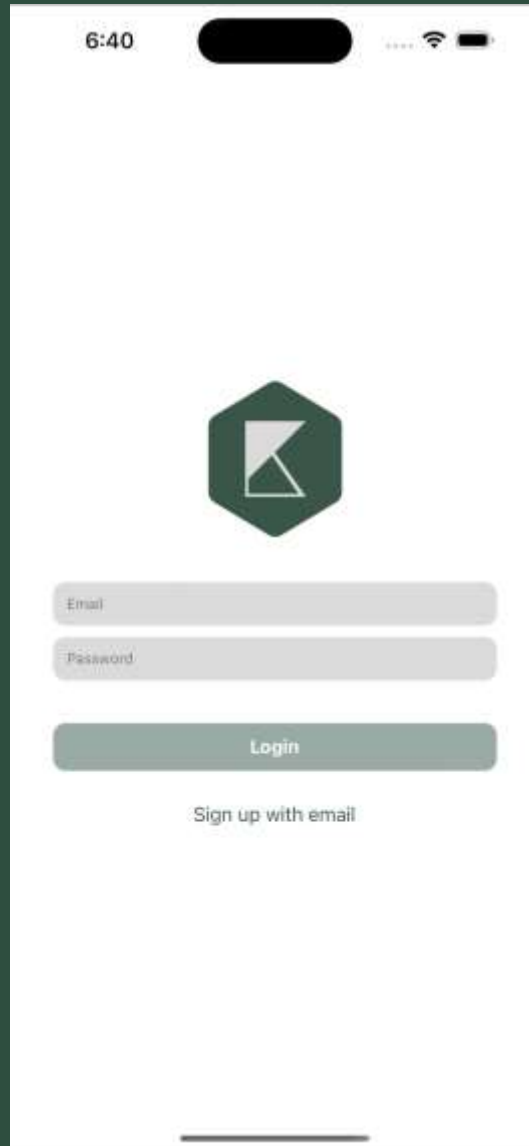
F1 score: 0.8679098679098679

Test accuracy: 0.8746

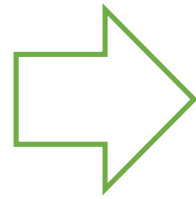
Learning Rate: 0.01, epoch = 200,

Training_data = 750 , Validation_data = 250

Application Demo

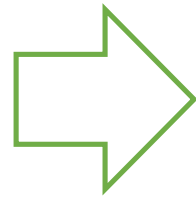


Application Functions



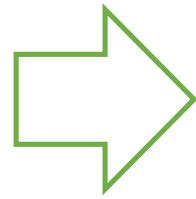
- 선택된 게시물을 다시 렌더링하여 본문과 하이퍼링크를 보여준다.

Application Functions



- 링크를 클릭하면 바로 해당 페이지로 넘어가게 된다.

Application Functions



- 제목, 내용, 링크, 카테고리
고리에 검색어가 들어가
있는 게시물만 골라서
보여준다.

Application Functions

7:15

설정

학년

☐ 1학년

☐ 2학년

☐ 3학년

☒ 4학년 이상

인성품

☐ 인성품

☐ 참의품

☒ 국제품

학점관리

☐ 자기이해와 진로탐색

☐ 봉사

☐ 동아리, 학회

☐ 외국어 점수 획득(국제품)

☐ 자격증

☐ 공모전/외부활동

☐ 졸업요건

☐ 기업인턴십/산학협력

☒ 취업캠프/서류, 면접 클리닉

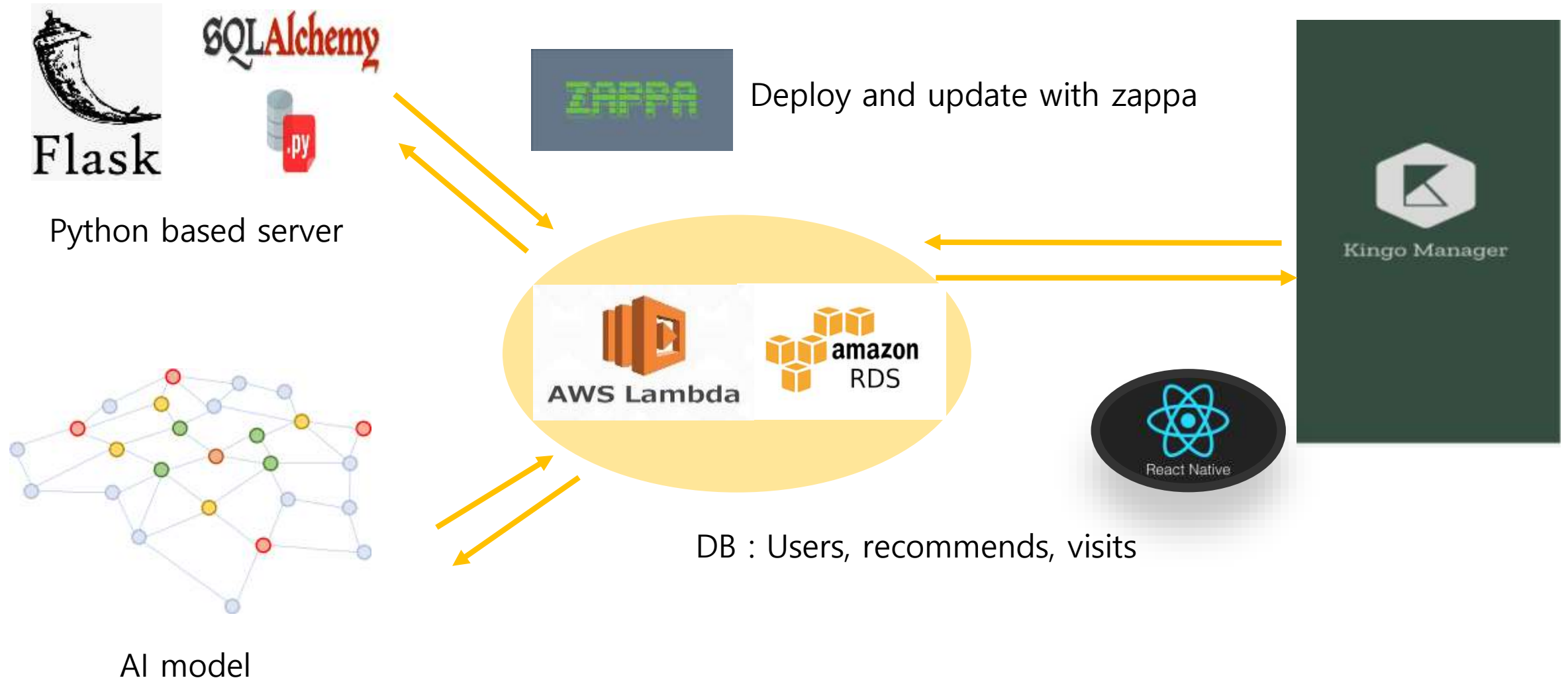
☐ 대학원/학석연계

Save

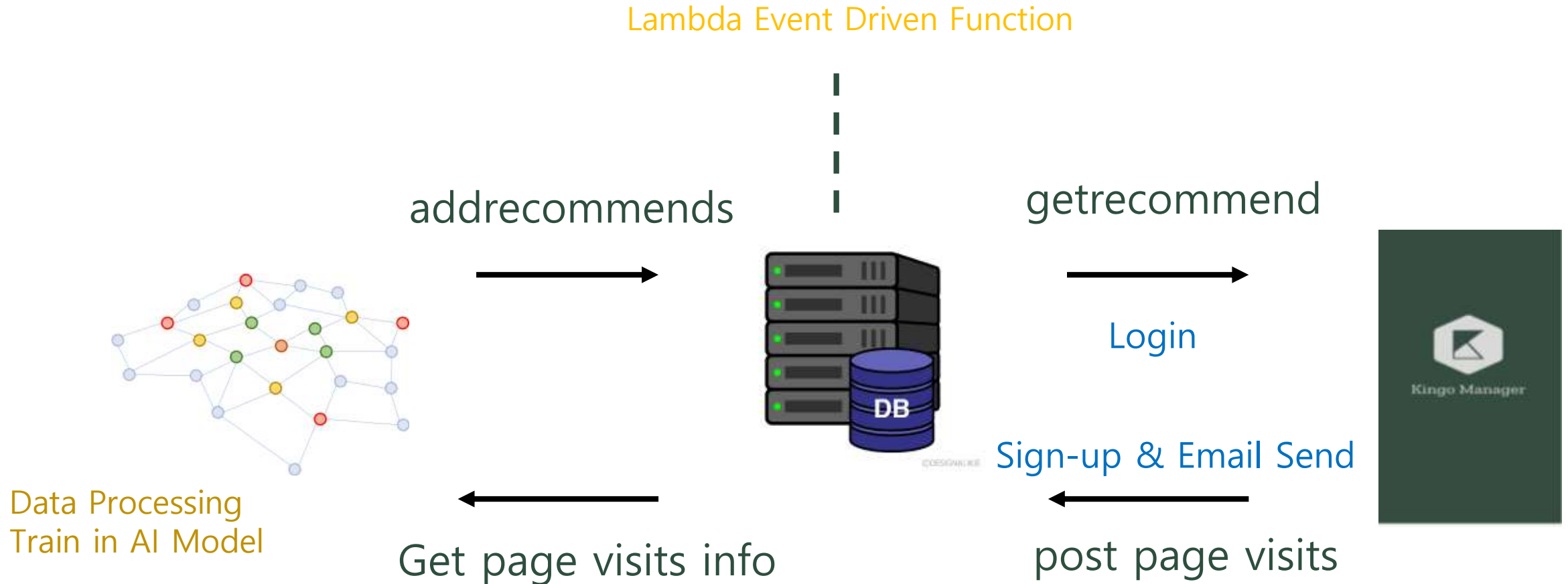
Sign out

- 계정 정보인 학년, 상품 취득 현황, 관심사를 변경할 수 있다.

Backend : Architecture



AI Model & UI Communication



Challenge & Evolution



Dataset

- 데이터 크롤링 과정에서 외부 사이트가 서로 다른 포맷을 가지고 있어서, 전처리와 크롤링 자동화가 다소 어려웠다.
- Selenium 과 beautifulsoup4를 적절히 사용해서 해결할 수 있었다.



AI

- 데이터 전처리 자동화를 원활하게 하고 원하는 형태의 결과를 출력하는 모델을 선정하는 것에 시간을 오래 소요했다.
- 결국 찾아서 필요한 dataset을 구축하고 모델을 활용할 수 있었다.

Challenge & Evolution

A diagram of a frontend component represented as a white rectangle with a black border and a black arrow pointing to the right.

Frontend

- 로그인 할 때 서버에서 정보를 가져와서 로컬 저장소에 넣어서 사용하는 방식으로 구현하였는데 로컬 저장소에 넣는 과정보다 다음 페이지로 넘어가 이전 정보를 불러오는 속도가 더 빨라서 그 부분을 수정하는것이 어려웠다.
- 로그인 부분의 알고리즘을 변경하여 해결하였다.

A diagram of a backend component represented as a white rectangle with a black border and a black arrow pointing to the right.

Backend

- 성균관대학교 학적정보와 GNN 모델의 학습 파라미터, 추천정보를 UI와 상호 연계 및 관리할 수 있는 서버를 구축하는 과정이 어려웠다.
- RDS db에 4개의 table을 정의하여서 핵심 데이터를 관리하고, Event-Driven-Function을 제공하는 aws Lambda 를 사용하여 필요한 기능들을 구현하였다.

Limitation & Evolution



Dataset

- Feature 의 개수가 너무 적다. 따라서 학습이 빨리 끝나지만, 조금 더 구체적인 추천이 불가능하다.
- 학과나 전공, 세부관심사에 대한 정보가 늘어남에 따라서 Feature 를 추가할 수 있을 것이다.



AI

- Category가 아니라 Content를 연관정도에 따라 추천해주는 것은 성공하지 못했다.
- Feature의 정보가 다양하지 않아 추천해주는 정확도에 한계가 있다.

Limitation & Evolution



Frontend

- 어떤 페이지 구성으로 하는 것이 사용자가 처음 앱을 사용할 때 더 쉽게 사용할 수 있는지 잘 몰랐는데 이전에 앱 개발 경험이 있었다면 더 나은 앱을 만들었을것 같다.
- 카카오, 네이버, 구글 등의 로그인 API를 가져와서 로그인 하는 기능을 추가하지 못하였다.



Backend

- Login 과정에서 데이터 보안에 더 신경쓰지 못했다.
- GCN모델을 서버상에 업로드하지는 않았기 때문에 로컬 컴퓨터가 항상 켜져있어야 컨텐츠 정기 업로드가 가능하다.
- Aws EC2 모델을 사용해서 모델 정보를 서버에 업로드한다면 서버를 더욱 자동화된 구조로 설계할 수 있을 것 같다.