

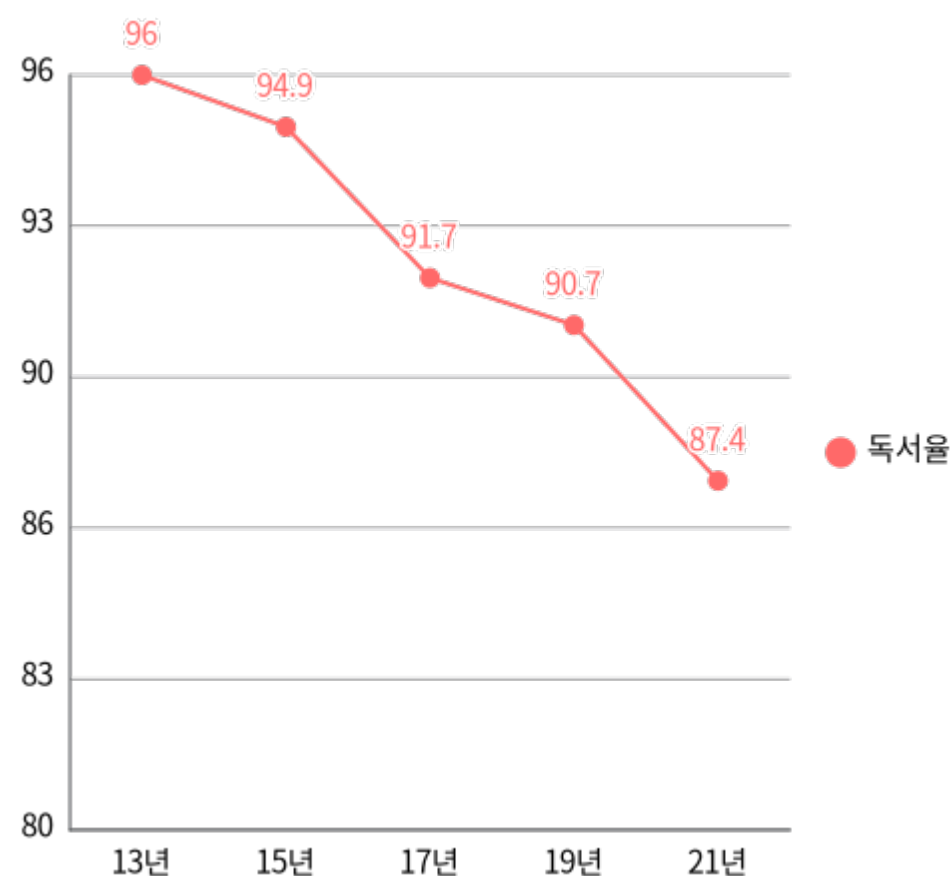
고등학생의 독서 흥미 유발을 위한 소셜 플랫폼: 다독다독

Team 프리텐다드



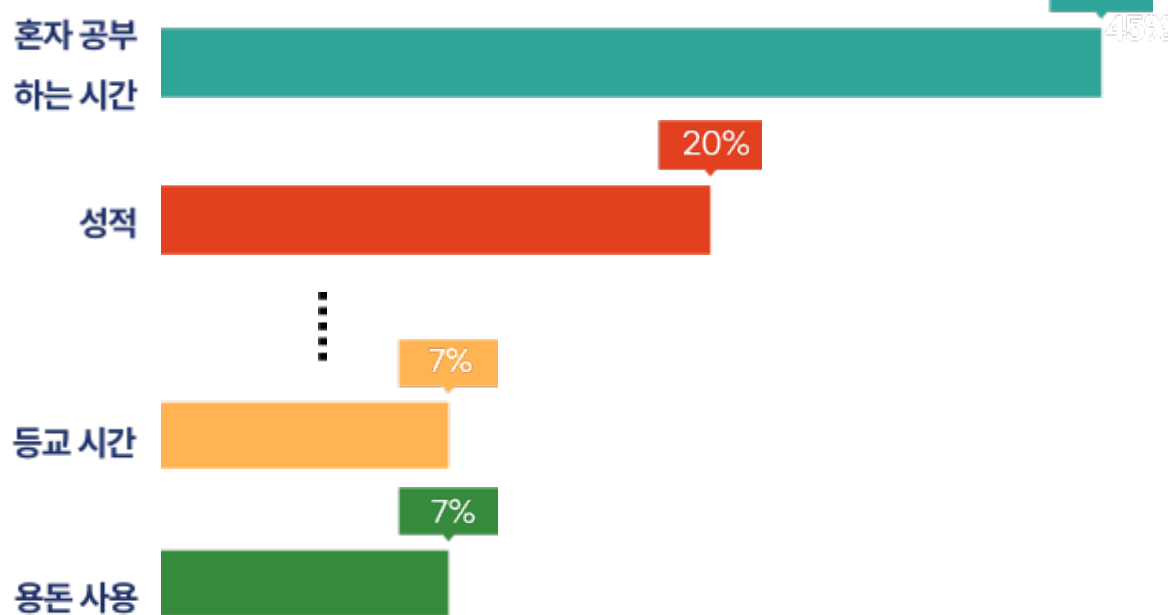
문제 상황 1. 전체적인 독서율과 독서량이 꾸준히 감소하며, 이는 학생들의 공부 시간, 성적, 자기효능감, 진로 성숙도, 다문화수용성에 유의한 영향을 준다.

독서율, 독서량: 꾸준히 감소 추세



- 독서율: 13년 이후로 지속적으로 감소하였으며,
- 10명 중 책을 한권이라도 안 읽은 학생이 10명 중 1명 이상 (21년)

공부 시간, 성적, 자기효능감, 진로 성숙도, 다문화수용성에 영향



*RandomForest Feature Importance 0.1 상대적도 기준

	진로성숙도	자기효능감	다문화 수용성
Coef	1.1697	0.5855	0.663
Std	0.059	0.054	0.049
P> t	0.000	0.000	0.000

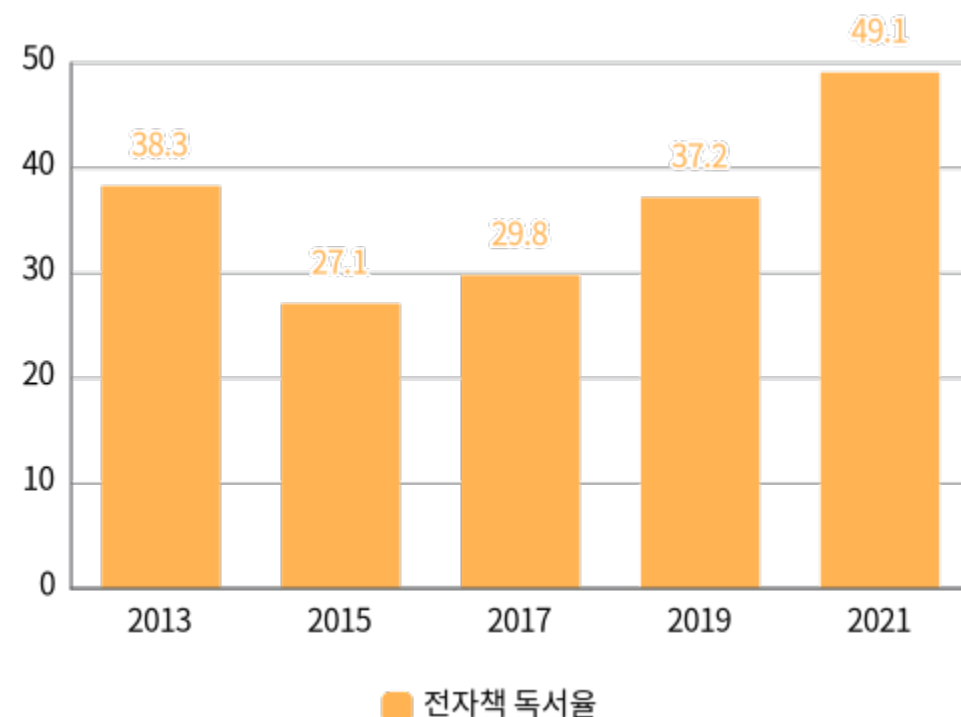
*OLS Regression fitting 기준 Coefficient, Standard Deviation, P-value

2021년 독서실태조사(학생) 데이터를 기준으로 RandomForest의 feature importance를 내림차순으로 정렬했을 때, '혼자 공부하는 시간' 및 '성적'이 가장 높은 중요도를 가졌다. 이는 독서와 성적 및 자습시간 사이에 유의한 연관이 있음을 의미함

또한, 한국고용패널조사 21차년도(2016) 자료를 활용하여 진로성숙도(관련 변수 14개 평균), 자기 효능감(관련 변수 6개 평균), 다문화 수용성(관련 변수 5개)에 대해 OLS Regression을 취하였을 때, Coefficient가 유의한 설명력을 보여줌. 이는 독서가 진로성숙도, 자기 효능감, 다문화 수용성에 긍정적인 영향을 미친다는 것을 시사함.

문제 상황 2. 전자책 독서율이 증가하고 디지털 매체가 나날히 발달해가지만,
독서를 다루는 고등학생 대상의 디지털 플랫폼이 부재하다. (e-book 제외)

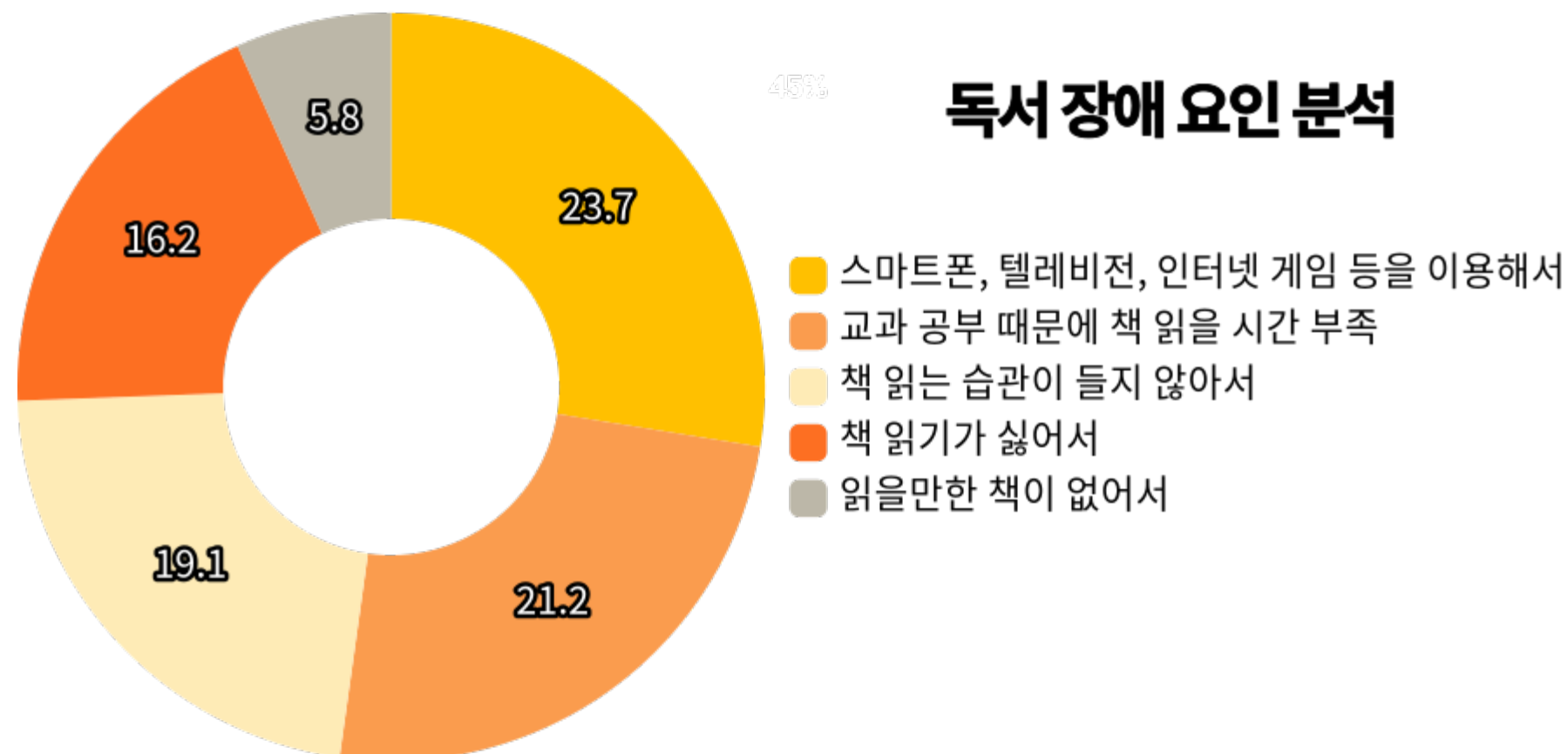
전자책 독서율: 꾸준히 증가 추세



전자책 독서율이 꾸준히 상승하여 21년도 독서실태조사 데이터 기준
50% 수준까지 도달

➡ 전자책 활용한 정책 및 어플리케이션 필요

독서 장애 요인 분석



학생들이 독서에 가장 어려움을 겪는 이유로 스마트폰, 텔레비전, 인터넷 게임 등을 이용해서가 23.7%로 가장 높게 나타남. (21년 독서실태조사)
OECD 국제 학업성취도평가(PISA)에서 한국의 청소년층 읽기 영역 수준이 현저히 낮아졌는데, 특히 문장이나 짧은 단락의 의미를 이해하고 필요
한 정보를 찾을 수 있는지를 나타내는 '축자적 의미 표상'의 정답률도 2009년 61.54%에서 2018년 46.54%로 15%p 낮아짐.

이는 단순히 독서 장애 요인을 넘어서, 과도한 스마트폰 콘텐츠 소비가 중/고교 수학 능력에 필요한 문해력을 저하시킨다는 것을 시사함.

➡ 전자책을 소비하는 학생층에게 편의를 제공하고, 문해력을 효과적으로 증진시키고 관리해주는 '독서' 기반 플랫폼이 절실

문제 상황 3. 고등학생들의 독서 활성화를 위해 설계된 '독서' 교과 학습 활동이 학생들의 흥미를 충분히 자극하지 못함.

독서교과 2단원의 학습 목표



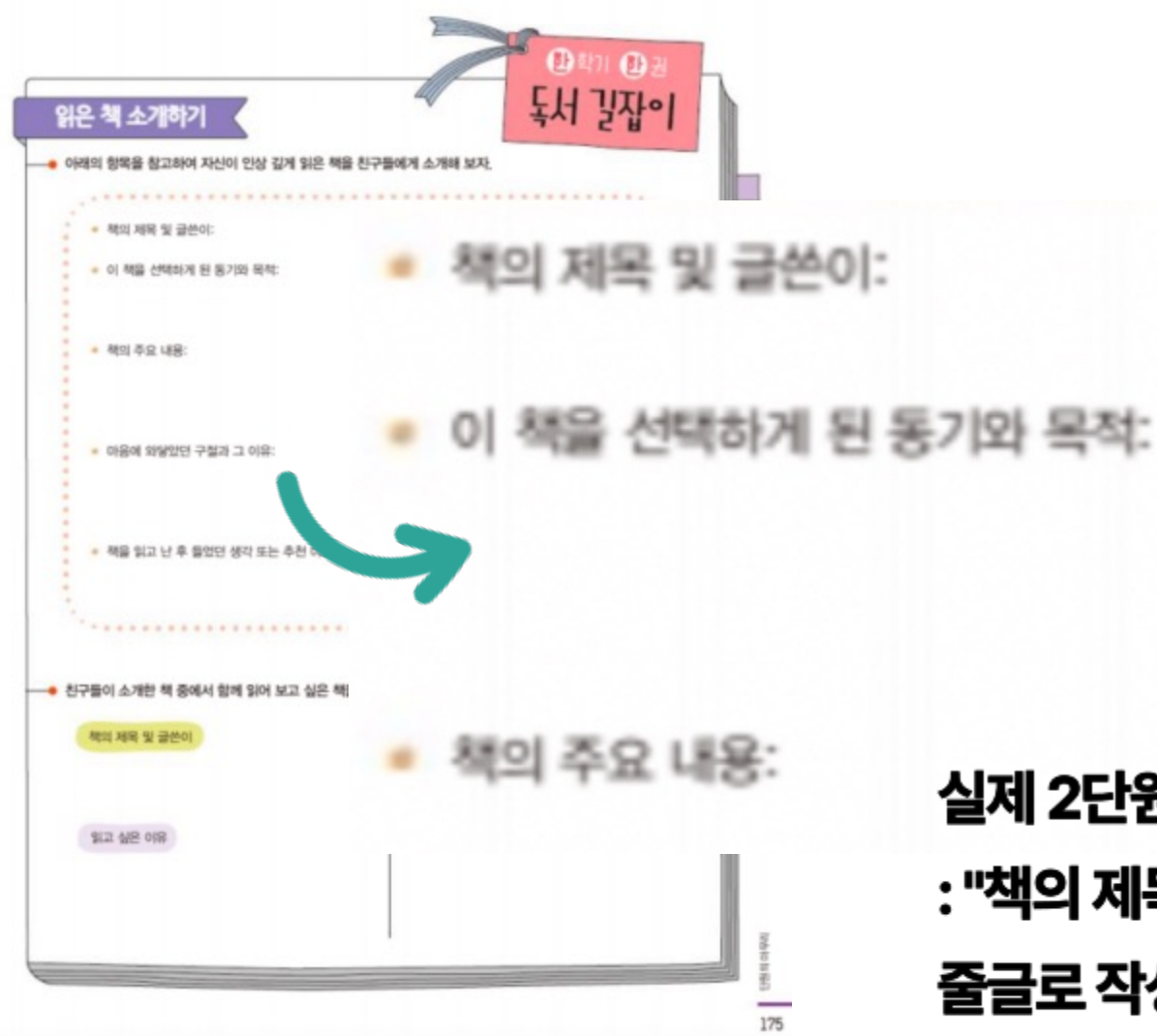
2단원의 실제 학습 활동 = 학습 목표 & 활동의 미스매치

이 단원에서는 글에 드러난 정보를 바탕으로 글의 중심 내용, 주제, 글의 구조와 전개 방식 등 사실적인 내용을 파악하며 읽는 방법을 익힐 것이다. 나아가 글쓴이의 의도나 글의 목적, 숨겨진 주제까지 추론하며 읽는 방법 역시 알아볼 것이다.

우리는 이러한 독서 활동을 하면서 필요한 자료나 정보를 수집·분석·평가하여 문제를 해결하는 방법과, 매체에 담긴 문자 언어의 의미를 이해하며 자신과 타인, 세계의 관계를 점검하는 방법을 배울 수 있다.

2단원 "소통하는 독서"의 학습 목표

- 글의 목적을 추론하며 읽기
- 자신과 타인, 세계의 관계를 점검하기



실제 2단원의 학습 활동

:"책의 제목, 글쓴이, 선택 동기를
줄글로 작성하여 친구들에게 소개하기"



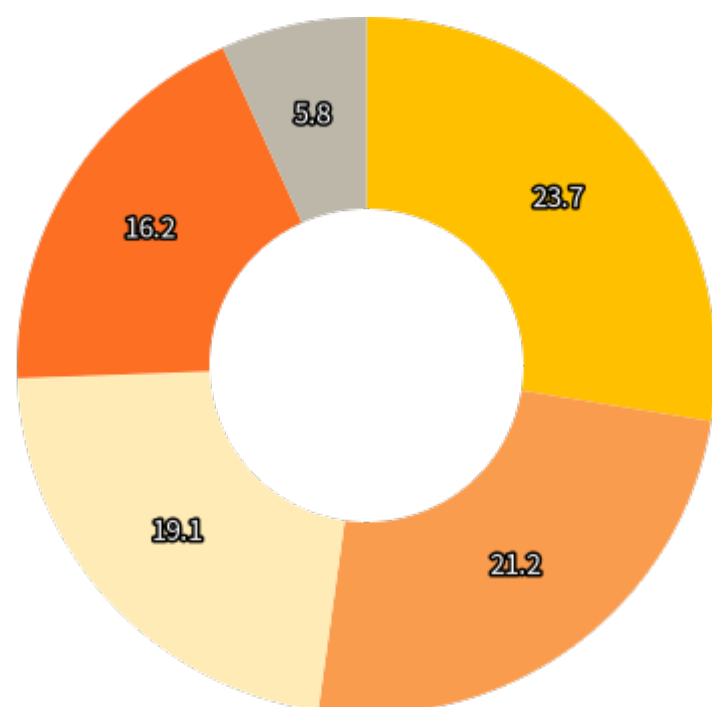
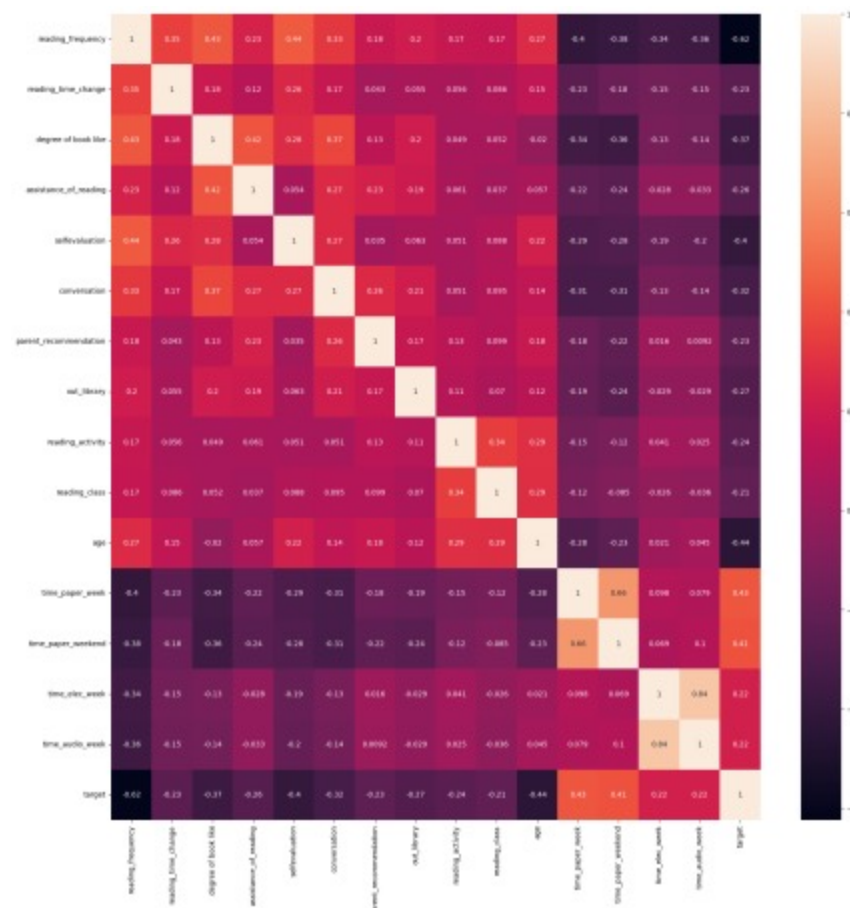
“ 독서에 대한 인식을 바꿀
다독 다독 (다같이, 多독서) ”

Application 설계 방향

1. 독서량과의 상관관계 분석

Spearman Correlation 기준 상관계수

- **독서 노출 빈도**(=0.62): 독서 노출 빈도 강화
- **자기 평가**(=0.4) : 자기 평가 기능 강화
- **책 장르 선호도**(=0.32) : 도서 추천 기능 강화
- **대화**(=0.32) : 소통 기능 탑재
- **나이**(=-0.44) : 고등학생을 위한 플랫폼 개발



2. 흥미 위주의 기능

- (책 읽는 것이 싫고) 습관이 들지 않아서 (19.1%)
- 책 읽기가 싫어서 (16.2%)

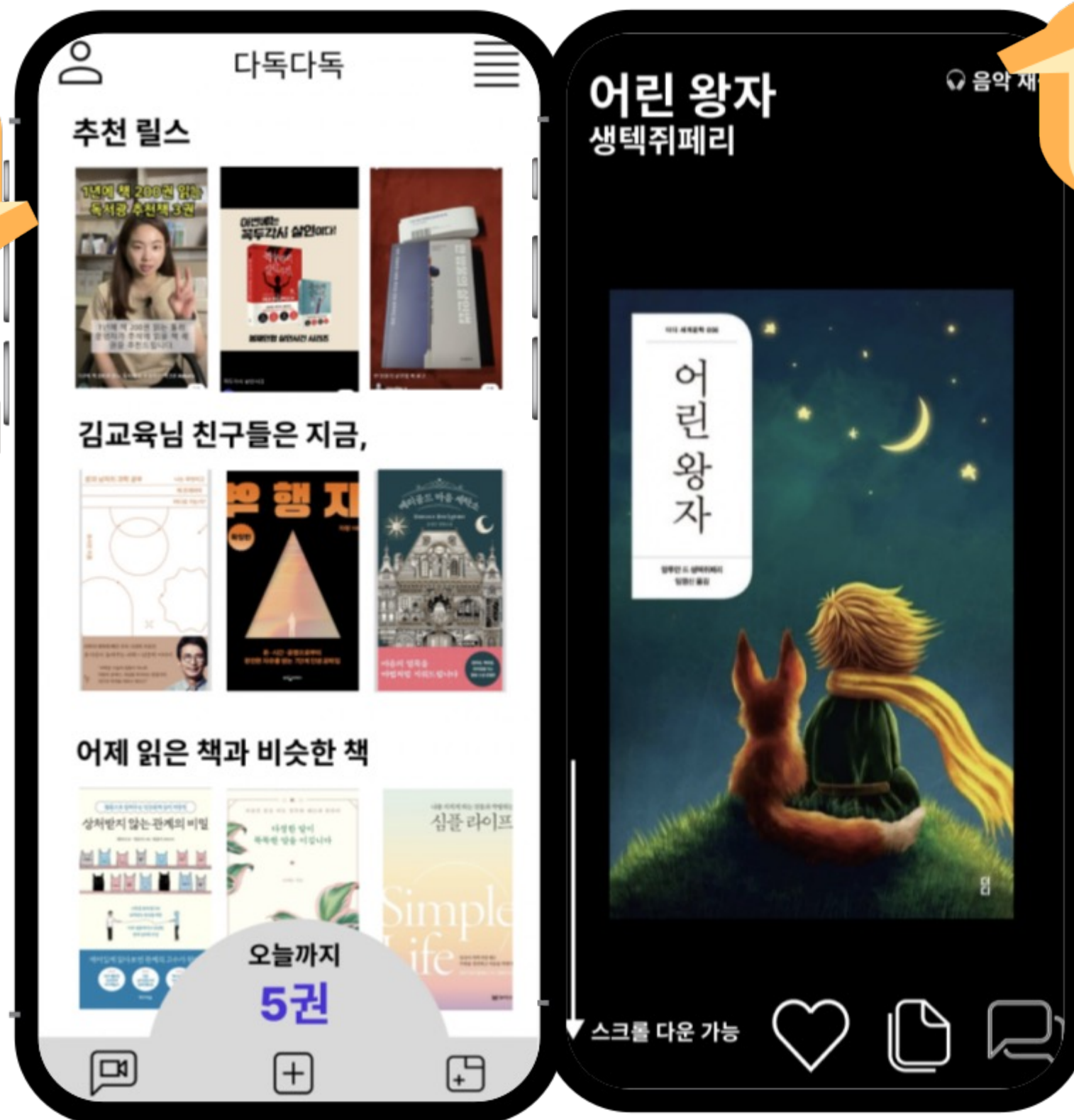
→ 책 읽기가 싫다 = 책에 대한 막연한 거부감이 있다는
응답이 총 35.3%에 달함

→ 독서에 대한 인식 변화를 위해서는
흥미, 재미 요소가 전제되어야 함

1. 도서 관련 릴스, 숏폼 등 콘텐츠 업로드

나한테 어울리는 **숏폼과 도서**는 무엇이 있을까?

- 청소년 인스타그램 사용률 90% 이상
(2018년 한국아동청소년패널조사 기준)
- 어플 구동 즉시 메인 화면에
도서 관련 숏폼 및 추천 도서 표시
- 도서 관련 숏폼 좋아요, 댓글 기능 설정
- 숏폼 콘텐츠: 책 추천 영상, 책 광고



오늘의 숏폼에서 추천해주는
도서는 뭘까?

- 본인을 표현하는 채널로서의 숏폼

1) 숏폼 콘텐츠

- 도서 내용/추천 위주의 콘텐츠로 게시
- 숏폼 게시 전 운영진 측에서 검토
- 광고 등 유해한 게시물 업로드 사전 제재,
담임선생님께 내용 인계

2) 좋아요/댓글

- 싫어요 버튼은 X
- 댓글 상 유해한 내용 이상치 탐지 및 제재
- 컬렉션 버튼을 통해 관련된 도서 담기 가능

2. 오늘의 글귀

오늘 올라온 글귀는 뭘까?
책에서 찾아 읽어봐야겠다!

- **독서 노출 빈도 & 독서 상관관계: 0.62**
- **어플 노출 빈도를 늘리는 홈 위젯 기능**

1) "오늘의 글귀" 기능

- 시집, 소설 등 다양한 매체에서 글귀 수집
> 어플 시작 시 팝업으로 제공

2) 시기맞춤형 글귀 제공

- 수능 등 고등학생에게 위로가 필요한 시기
적절한 글귀 및 도서를 추천

3) 어플 실행 유도를 위한 출석 체크 기능

오늘은 6월 모의고사,
힘이 나는 글귀로
마음을 달래볼까요?

그래도 먼지가 된 나를
하루종일
찬란하게 비춰주셔서 감사합니다.

- 정호승, 햇살에게
<내가 사랑하는 사람>, 정호승 시선집

내가 사랑하는 사람
정호승 시선집

오늘이 없는 사람
사랑하지 않는다
오늘이 사랑하는 사람
사랑하지 않는다
한 그루 나무의
늘 사랑한다
바닷물이
바닷물이



처음 오셨군요, 000님의
독서 MBTI를 알아볼까요?

I 혼자 읽기 **성향** 같이 읽기 **E**

S 몰아 읽기 **습관** 나눠 읽기 **D**

L 문학 **장르** 비문학 **N**

I 책 읽고 끝 **기록** 기록 남기기 **E**

3. 독서 MBTI, 자기평가

이번 달에 몇 권을 읽었더라?

- **자기평가 & 독서량 상관관계: 0.4**
- 스스로 독서 습관 점검하는 기능 필요

1) 독서 MBTI

- 독서 성향을 파악하고, 비슷한 유저끼리 도서를 추천해주는 DB로 활용
- 매월 1일 본인의 독서 취향 변화 확인

2) 셀프 체크리스트: 독서 습관 변화 확인

- 목표한 독서량을 이루셨습니까?
- 자투리 시간을 활용하여 독서를 하셨습니까?
- 독서를 기록으로 남기셨습니까?
- 본인의 독서 점수를 평가해볼까요?

4. 또래 소통 기능

다른 친구들은
어떤 책을 읽고 있을까?

- 소통 및 독서량 상관관계: 0.32

1) 또래 피드를 통한 소통 기능

- 친구들이 읽은 도서 목록 확인
- 친구들이 남긴 평가 확인 가능

2) 슷폼, 오늘의 글귀 공유 기능

- 학생이 직접 만든 슷폼 & 오늘의 글귀에 대해 좋아요, 댓글 기능을 통해 또래 소통 활성화
- 또래가 읽고 있는 책 컬렉션 기능을 통해 보고싶은 도서 목록에 담기 가능



5. 기타 기능

1) 목표권수 설정 가능

2) 독서종합시스템 연계하여 독후감 작성

- 복사-붙여넣기 방지 + 글자 수 제한 (500자) 등으로 간소화
- 생활기록부와 연계하여 등재
- 숏폼이나 오늘의 글귀 등 다른 형식으로 독후감 작성 가능 (담임 확인 필)
- 수행평가 대체, 공개여부 설정 기능

3) 도서 담기 기능: 읽고 싶은 도서 관리

4) 전자책 플랫폼과 연계 가능 (추후 제공)

5) 특수교육 대상 학생들을 위한 채널

오늘의 독후감은 어떤 형식
으로 작성해서 제출할까?

개인별 독서 설문 (독서 MBTI) 데이터

지금까지 읽은 도서 목록(독서종합시스템 연계)

앞으로 읽을 도서 데이터 반영

도서 섹션 이용 내역

서열 척도화 (User Feature)

NLP [Stemming, Vectorizer]

LightGBM 등 머신러닝 모델

코사인 유사도

- 1) 학생이 읽은 도서와 유사한 도서 추천
 - 2) 학생이 읽은 도서와 분야가 반대인 도서 추천
 - 3) 학생이 본 섹션과 비슷한 콘텐츠 추천
- ➔ 도서 취향 기반 도서 추천 플랫폼 구축



Start



Input



Pre-process



LightFM



Output

가입 및 설문

독서종합시스템 연계

개인정보 동의

*Stemming: 어간 추출, Vectorizer: 자연어 벡터화

*LightGBM: 데이터가 학습하는 기울기를 계산하여 학습 가속

*코사인 유사도: Cosine 공식을 활용하여 두 벡터의 방향이 유사한지 계산

*협업 필터링: 사용자들의 평가, 시청 기록을 데이터로 받아 유사한 사용자들의 평가를 예측

*컨텐츠 필터링: 컨텐츠 간의 유사도를 NLP, CV, 머신러닝을 통해 얻은 Feature들을 활용하여 예측

*LightFM: 협업 필터링과 컨텐츠 필터링을 합친 Hybrid 모델로 Python Package로 구현 가능

협업 필터링 + 컨텐츠 필터링

- 1) User Feature: 유저별 설문조사 + 도서 평가 데이터
- 2) Item Feature: 책 소개, 저자, 제목 등 NLP 처리
- 3) 코사인 + MSD(평균제곱거리) 기반 유사도 추천

1. 사업성 분석 (SWOT)

<i>Strength</i> 강점	<i>Weakness</i> 약점
<i>Opportunity</i> 기회 <ul style="list-style-type: none"> - 독서종합시스템과 연계 - 고등학생을 위한 도서 추천 및 관리 - 독서 중심으로 소통/수업 	<ul style="list-style-type: none"> - 플랫폼 제공에 대한 비용 - 초기 추천시스템 DB 부족으로 인한 추천 정확성 저하
<i>Threat</i> 위협 <ul style="list-style-type: none"> - SNS를 활용한 공유 문화 : 일정 수준의 참여자 필요 - 숏폼 등 유희성 콘텐츠에 주의를 빼앗길 우려 	<ul style="list-style-type: none"> - 독서에 대한 학생들의 기본적인 거부감 - 스마트폰 활용에 대한 문제 - '독서bum' 분위기 조성까지의 시차

2. 수익 창출 방안

- 1) 숏폼/도서 광고 배너 광고 수주 숏폼 광고로 탑재 (도서 위주)
- 2) 도서관 예산 활용 교육부 학교 도서관 예산을 일부 활용하여 전자책 도서관 구축 가능
- 3) 대여료 일부 전자책(웹툰, 웹소설 등) 대여 서비스를 통한 수익 창출

3. 유사 서비스와의 차별성



1) 밀리의 서재

국내 최고의 전자책 플랫폼
오디오북, 라디오 등 다양한 채널



2) 독서교육지원시스템

독후감 작성 및 생활기록부 연계
도서관 책 목록 등 확인 가능

Target: 고등학생
독서 인식 개선 및 습관화에 집중
(단순 e-book 제공 X)

어플 형식: 웹사이트보다
접근이 편리
독서 타이머를 통한 독서 습관화



Expectation 기대효과

1. 고등학생들의 '독서붐' 유발 및 습관화 장려
2. 독서종합시스템 이용률 증가로 인한 독서 DB 구축
3. 고등학교 '독서' 수업에서의 교육적 효과 증대 가능
4. 독서를 통한 내면 성장 효과, 업무 효율 기대 가능
5. 국어 교과시간 중 스마트폰 사용에 대한 긍정적 선례

1. 초등학생/중학생으로의 애플리케이션 범위 확대
2. 진로 등 생활기록부 연계 기능 추가 가능
3. 교과서 E-book 등 콘텐츠 추가 가능
4. 전자책 플랫폼과 연동을 통한 전자책 대여 가능
5. 특수교육 대상 학생들을 위한 플랫폼으로 확장 가능
(오디오북, 맞춤형 독서 리스트 제공, 독후감 작성 간소화)

Utilization 활용방안



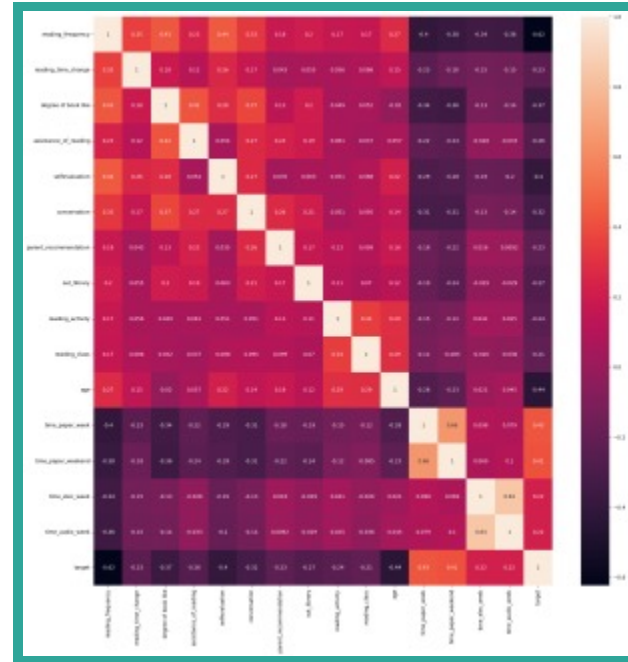
Correlation Map

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

ρ = Spearman's rank correlation coefficient

d_i = difference between the two ranks of each observation

n = number of observations



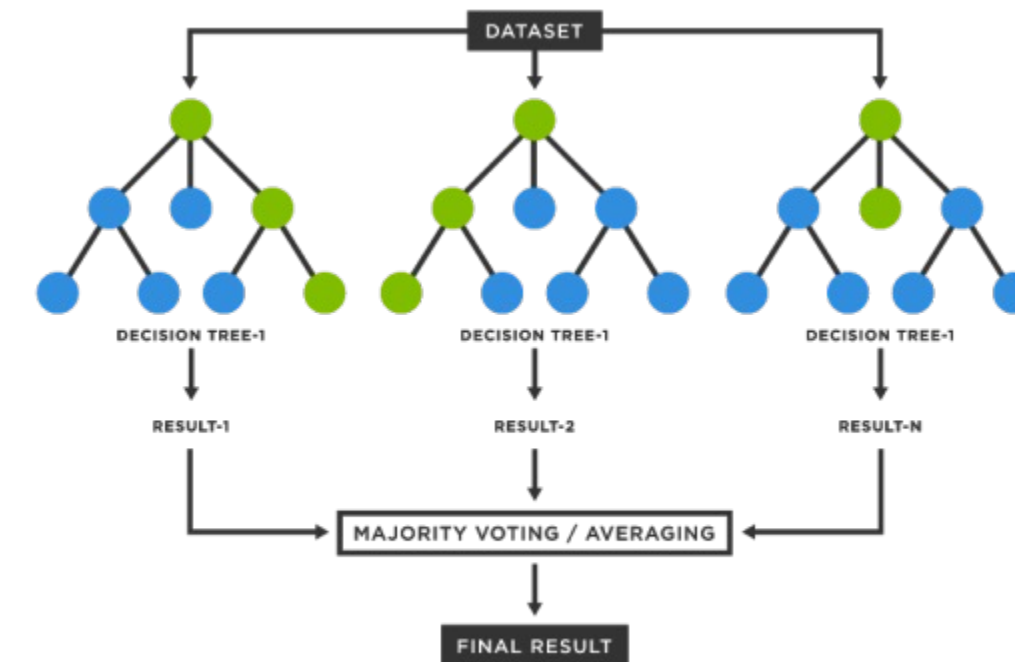
독서실태조사 방식에 따라, 서열 척도에 있어 분석하기 적합한 Spearman's Correlation 을 활용하여 상관관계를 분석한 후, 독서량에 어떤 변수가 유의한 상관관계를 가지는지 조사
전처리로 전체 독서 권수(종이책, 전자책, 오디오북)을 합친 종속변수를 제작

Pandas & Seaborn 라이브러리를 활용하여 Correlation 연산 및 시각화를 수행하였으며, 하기의 코드 참조

```
import seaborn as sns
import matplotlib.pyplot as plt

corr=df.corr(method='spearman')
corr=corr.drop(list(corr.loc[abs(corr["target"]) < 0.2].index),axis=1)
corr=corr.loc[abs(corr["target"]) > 0.2]
plt.figure(figsize=(20,20))
sns.heatmap(corr,annot=True)
```

RandomForest



독서량과의 설명력을 가지는 변수를 찾기 위하여 다수의 결정트리를 앙상블한 RandomForest를 활용
n_estimator를 500으로 설정하고, 9:1로 분할한 훈련 데이터를 활용하여 독서량을 종속변수로 한 후 학습시키고,
Feature의 중요도를 뽑아 분석

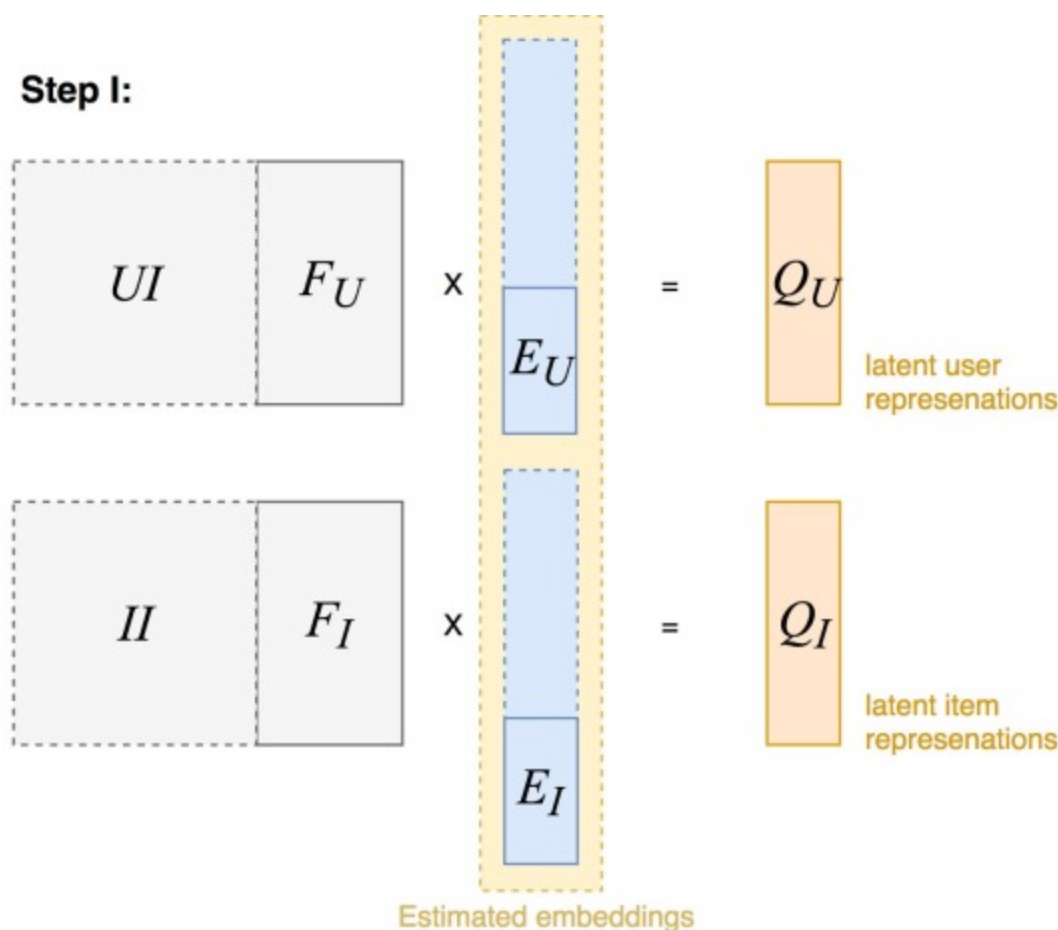
Sklearn 라이브러리를 활용하여 RandomForest 훈련 및 중요도를 추출하였으며, 하기 코드 참조

```
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

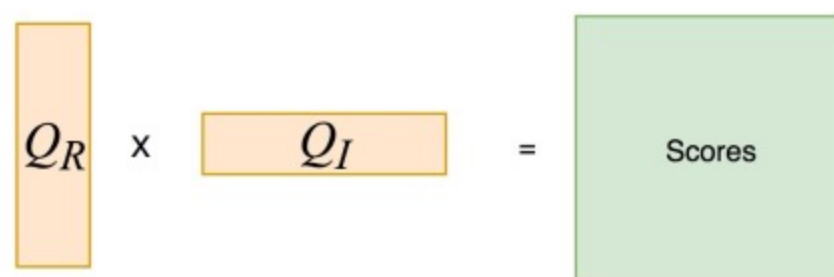
ftr_importances_values = forest.feature_importances_
ftr_importances = pd.Series(ftr_importances_values, index=X_data.columns)
ftr_top = ftr_importances.sort_values(ascending=False)[:30]

plt.figure(figsize=(20, 20))
sns.barplot(x=ftr_top, y=ftr_top.index)
plt.show()
```


LightFM



Step II:



LightFM은 명시적/ 묵시적 피드백을 기반으로 협업 필터링/콘텐츠 추천을 활용하여 더 정확한 추천을 해주는 알고리즘으로, 하기와 같은 원리로 동작

- 1) Single User feature에 estimated embeddings를 내적 후, bias를 더하여 latent user representation 도출
- 2) Single Item feature에 estimated embeddings를 내적 후, bias를 더하여 latent item representation 도출
- 3) 전체 user latent vector와 item latent vector 두개를 내적하여 최종 Score 도출

*모델에서 학습하는 것은 "Estimated Embeddings" 부분으로, Embedding 부분을 학습시켜서 각 Item과 User의 잠재 변수 벡터를 효과적으로 도출

또한, User feature와 Item Feature로 평점 정보 외 유저의 특징, 아이템의 특징 등 Regression 처럼 다양한 정보가 포함 가능

```
def objective(params):
    no_components, learning_rate = params

    model = LightFM(no_components=no_components,
                    learning_schedule='adagrad',
                    loss='warp',
                    learning_rate=learning_rate,
                    random_state=0)

    model.fit(interactions=train,
              item_features=item_features,
              sample_weight=train_weights,
              epochs=3,
              verbose=False)

    test_precision = precision_at_k(model, test, k=10, item_features=item_features).mean()
    print("no_comp: {}, lrn_rate: {:.5f}, precision: {:.5f}".format(
        no_components, learning_rate, test_precision))
    test_auc = auc_score(model, test, item_features=item_features).mean()
    output = -test_precision

    if np.abs(output+1) < 0.01 or output < -1.0:
        output = 0.0

    return output
```

상기된 코드는 objective 함수를 통하여 최적화된 모델을 학습시키는 과정을 구현

LightFM 패키지를 이용하여 해당 모델을 파이썬으로 구현하였으며, Hyperopt 패키지를 활용하여 해당 모델을 최적화시키는 작업을 거쳤으며, Dataset은 오픈소스로 공개된 Good-Book dataset을 git하여 활용

또한 해당 모델을 Cold Start Problem(초기 데이터가 부족하여 성능이 저하되는 문제)를 Content-based Filtering의 가중치를 강화하는 방향을 통해 이를 극복하고 있으며, 모델의 GFLOPS가 높지 않아 지속적으로 Update하는데 무리가 없다는 장점 존재

➡ 이를 통해, 실시간으로 업데이트되는 학생들의 독서 기록을 토대로 추천되는 도서를 추천해주는것이 가능하며, 그 정확성도 훨씬 높은 수준

데이터셋

한국청소년정책연구원, 한국아동청소년패널조사(2018)
문화체육관광부, 2021년 독서실태조사
한국직업능력연구원, 한국교육고용패널2 1차년도(2016)
OECD, PISA 2018 Results
zygmuntz, goodbooks-10k, <https://github.com/zygmuntz/goodbooks-10k>

인터넷

연합뉴스, [문해력 리포트] ② 한국 청소년 '디지털 문해력'마져...OECD 바닥권
<https://www.yna.co.kr/view/AKR20211209134600501>
밀리의 서재, <https://www.millie.co.kr/v3/brand/update>
독서교육종합시스템, <https://reading.ssem.or.kr/>
팁코 소프트웨어, 랜덤포레스트란 무엇입니까?, <https://www.tibco.com/ko/reference-center/what-is-a-random-forest>
datarevenue, Challenges & Solutions for Production Recommendation Systems,
<https://www.datarevenue.com/en-blog/building-a-production-ready-recommendation-system>

보고서
논문

김영식, 김민석, 이길재, 2018, 독서활동이 고등학생의 인지적, 비인지적 발달에 미치는 영향, KRIVET Panel Conference
[비상교육] 독서, 2022, http://dn.vivasam.com/2022_ebook/HS/high_kor_doc/index.html
박정민. "독서교육종합지원시스템 활용에 관한 연구." 국내석사학위논문 중앙대학교 교육대학원, 2019. 서울
임유미. "독서교육종합지원시스템 웹사이트의 사용성 평가 연구." 국내석사학위논문 공주대학교 교육대학원, 2012. 충청남도
한국교육학술정보원(KERIS). "독서교육종합지원시스템 매뉴얼." (2011)
Maciej Kula, Metadata Embeddings for User and Item Cold-start Recommendations, 2015, arxiv