

졸업프로젝트 최종 보고서

주제 : 프라이버시 보호 딥러닝 서비스 개발

개요

1. 디자인스프린트
2. SE 문서(문제 정의서, 요구사항 명세서, 유스케이스 명세서, 클래스 다이어그램, 시퀀스 다이어그램)
3. 사용자 설문조사
4. 프로토타입

종합설계1 02분반

[사그람 조]

201402433	조승현	201402392	이상화
201704144	김수민	201704145	김주희

디자인 스프린트 1~2일차

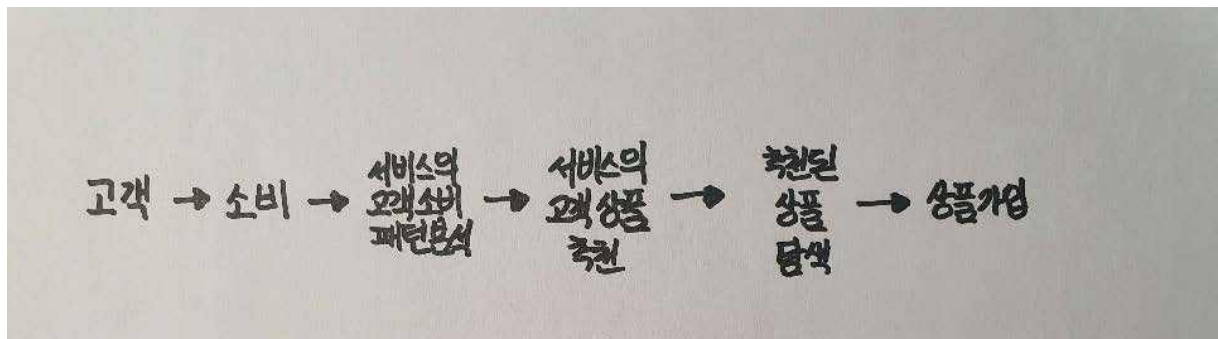
[사그램 조]

[주 제] 프라이버시 보호 딥러닝 서비스 개발

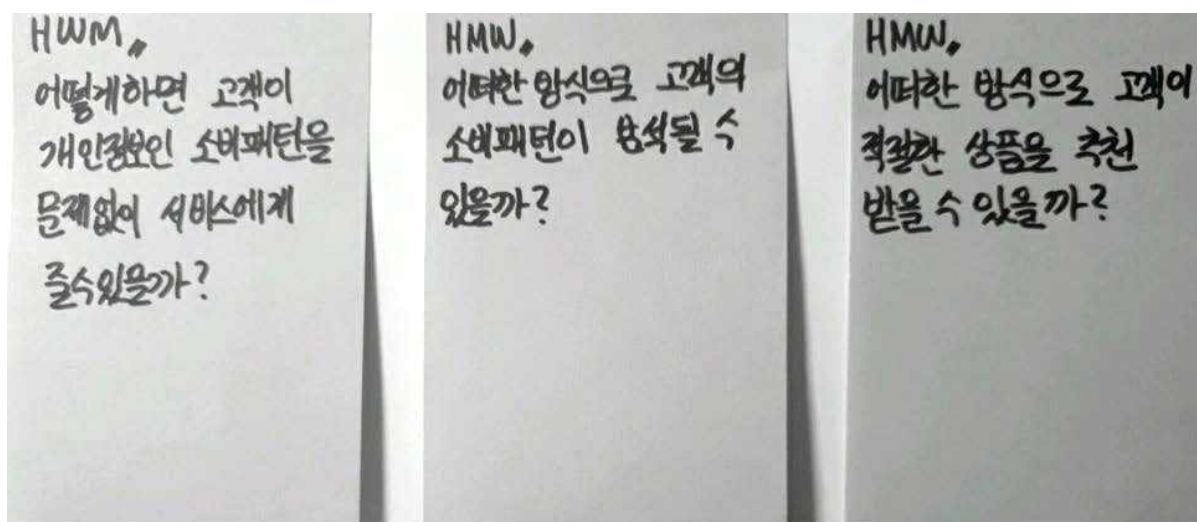
(소 주 제) 사용자의 소비 패턴을 분석하여 적절한 금융 상품을 소개하는 서비스 개발

[201402433 조승현]

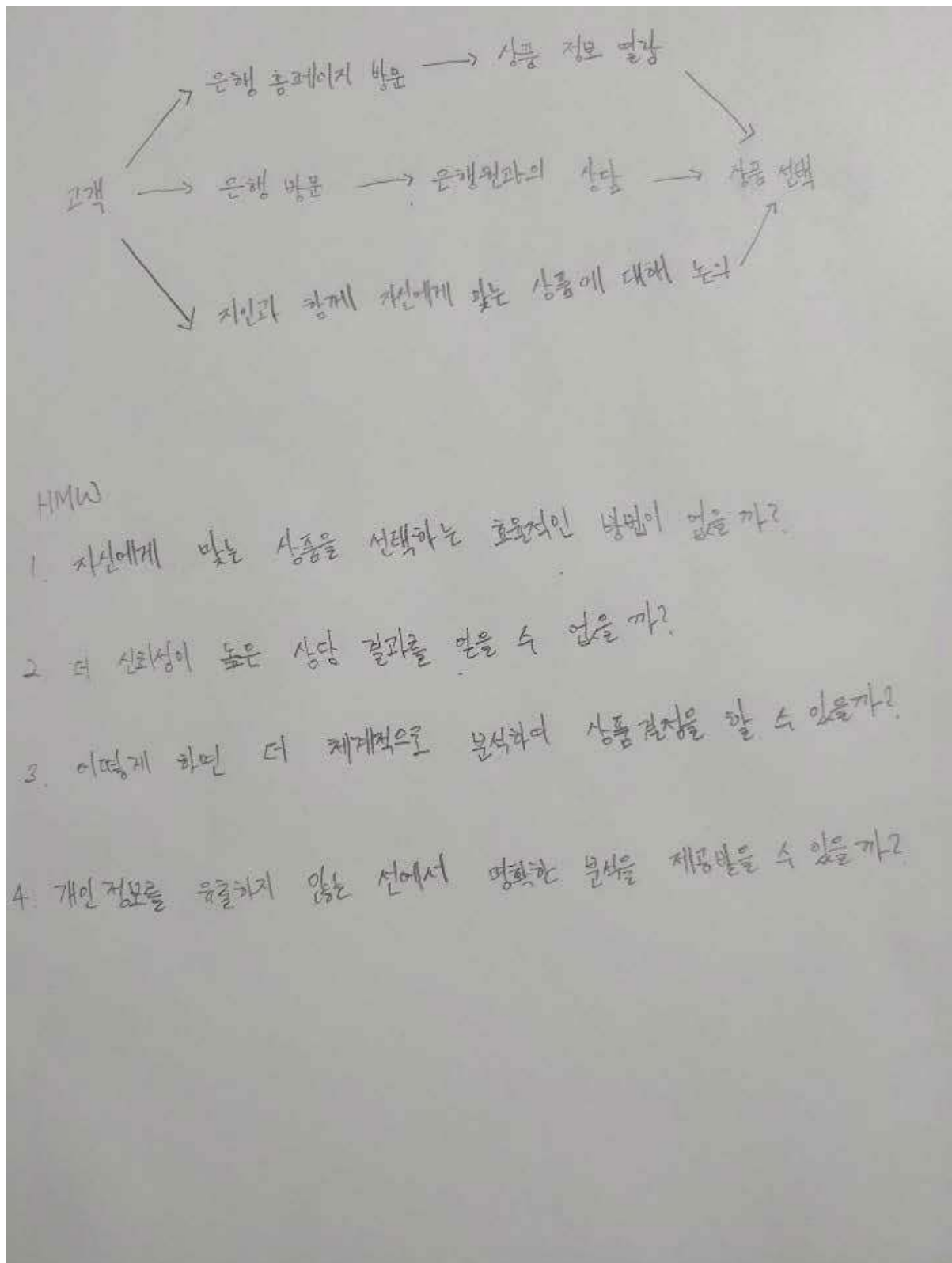
● MAP



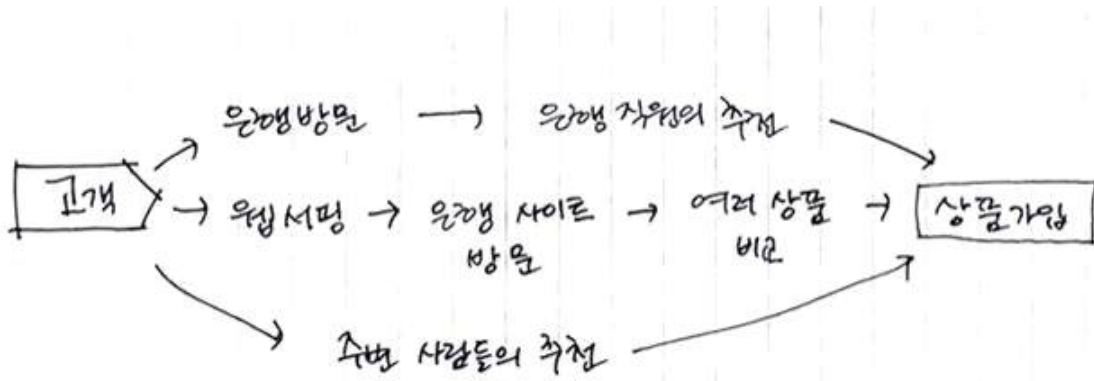
● HMW



● MAP & HMW



• MAP & HMW



HMW

어떻게 하면
사용자의 개인정보는
보호하면서 맞춤 상품을
추천할 수 있을까?

HMW

어떻게 하면 사용자가
자신의 소비성향에 맞는
금융상품을 쉽게 찾을 수
있을까?

HMW

어떻게 하면 사용자가
상품가입을 할때 절차를
단순화할 수 있을까?

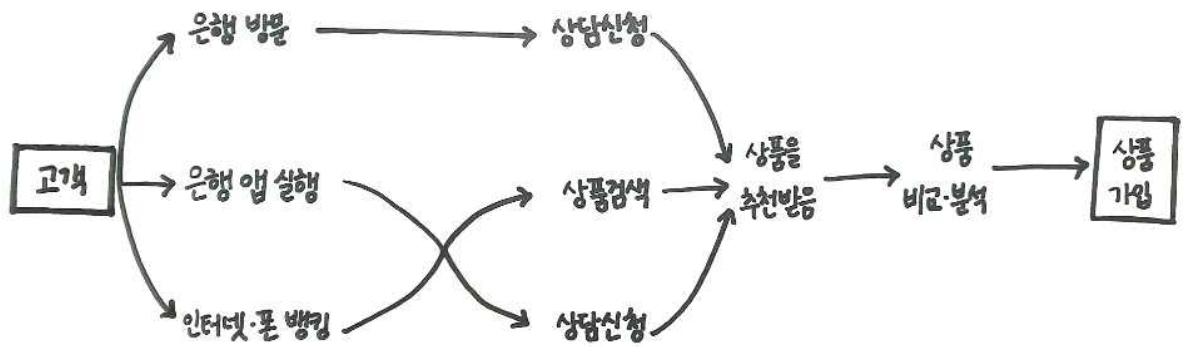
HMW

어떻게 하면 사용자가
자신에게 꼭 필요한
금융 상품만 추천받을 수
있을까?

HMW

어떻게 하면 사용자가
여러 상품을 한눈에 비교할
수 있을까?

• MAP & HMW



HMW,,

어떻게 하면 나의 개인정보를 최소한으로 공유하여 서비스를 이용할 수 있을까?

HMW,,

어떻게 하면 금융서비스를 사용하는 동안 내 정보를 안전하게 보호받을 수 있을까?

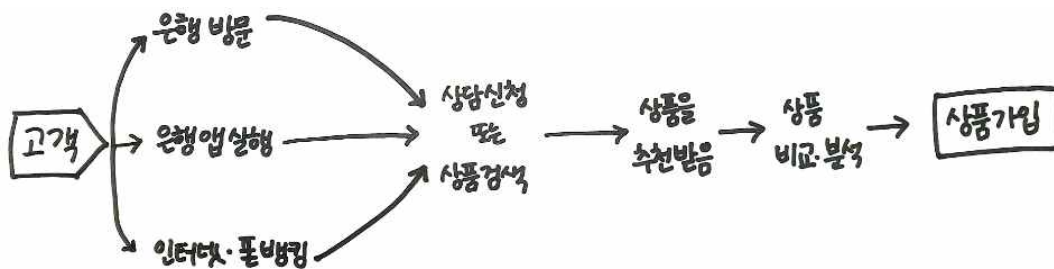
HMW,,

어떻게 하면 금융상품 정보들을 한눈에 쉽게 보며 비교,분석 할 수 있을까?

HMW,,

어떻게 하면 사용자에게 불필요한 상품을 제외한 최소한의 맞춤 상품들만을 추천받을 수 있을까?

- 하나의 통합된 MAP



- 통합된 HMW



솔루션 & Lightning Demo

1. 개인정보를 유출하지 않는 선에서 명확한 분석을 제공받을 수 있을까?

- 솔루션 : 차등프라이버시 알고리즘

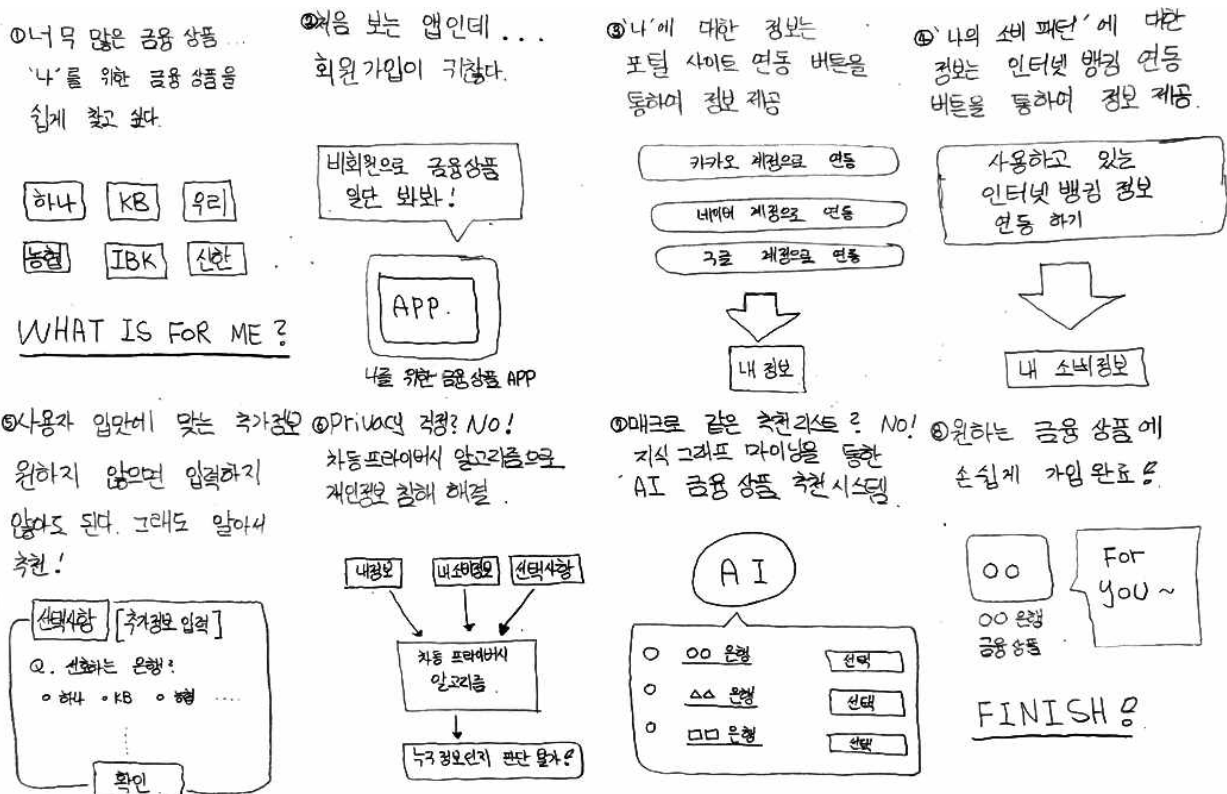
○ Lightning Demo : 코로나 19로 인해 개인정보인 개인의 위치정보를 보호하기 위해 구글이 활용한 차등프라이버시 - 코로나 19로 인해 사람들의 동선 파악이 중요해진 시기에 구글이 차등프라이버시를 활용하여 개인의 데이터에 인공적인 노이즈 추가 및 익명화된 집합 데이터를 활용하여 사용자를 특정하기 어렵게 만들었다. 이것을 활용하여 우리 주제의 개인정보인 소비패턴도 같은 방식으로 보호 할 수 있을 것이라 생각했다.

2. 어떻게 하면 사용자가 자신의 소비패턴을 체계적으로 분석해 가입할 상품을 결정할 수 있을까?

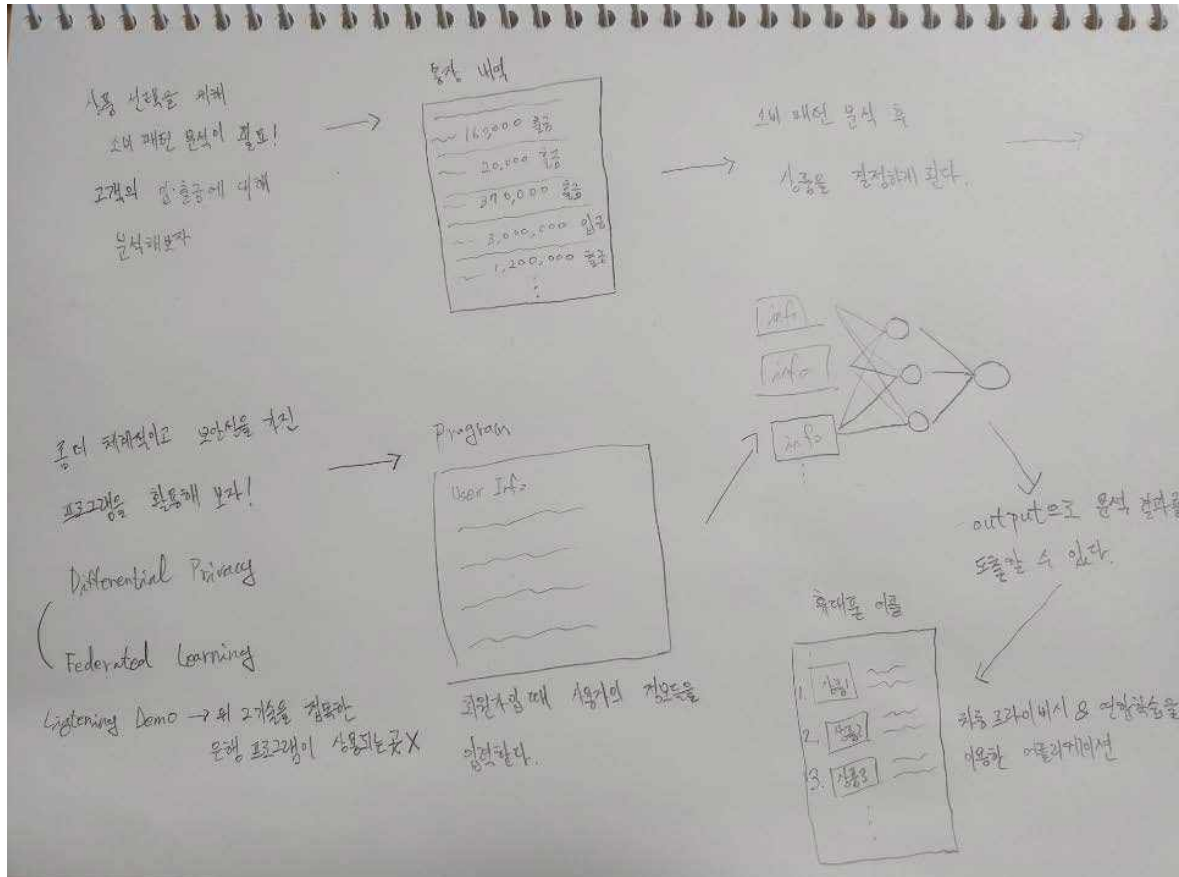
- 솔루션 : 지식 마이닝 그래프 알고리즘

○ Lightning Demo : AI 콜봇에서 활용한 지식 마이닝 그래프 - 인공지능 고객상담 시스템을 기반으로 함. 기존 여러 금융권에서 사용하고 있는 검색 및 텍스트 마이닝 기반의 질의 응답은 사전에 정의된 질문들과 비교해 기계가 답하는 방식으로 FAQ 수준이지만 지식그래프와 딥러닝을 연계하여 고객의 의도를 정확히 파악하고 답변을 제공한다. 사람들이 소비패턴이 정해진 답이 아니라고 가정하면 AI 콜봇이 고객상담을 추천하는 방식으로 정해지지 않은 답을 추천하는 방식에 적합하다고 생각해 가져왔다.

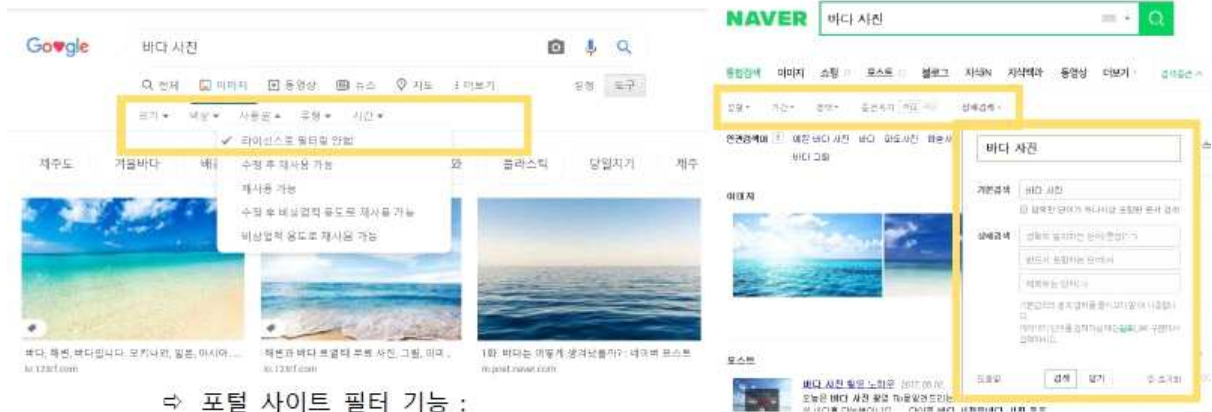
Sketch



SOLUTION 및 Sketch



- 어떻게 하면 사용자에게 불필요한 상품을 제외한 최소한의 맞춤 상품들만 추천 받을 수 있을까?

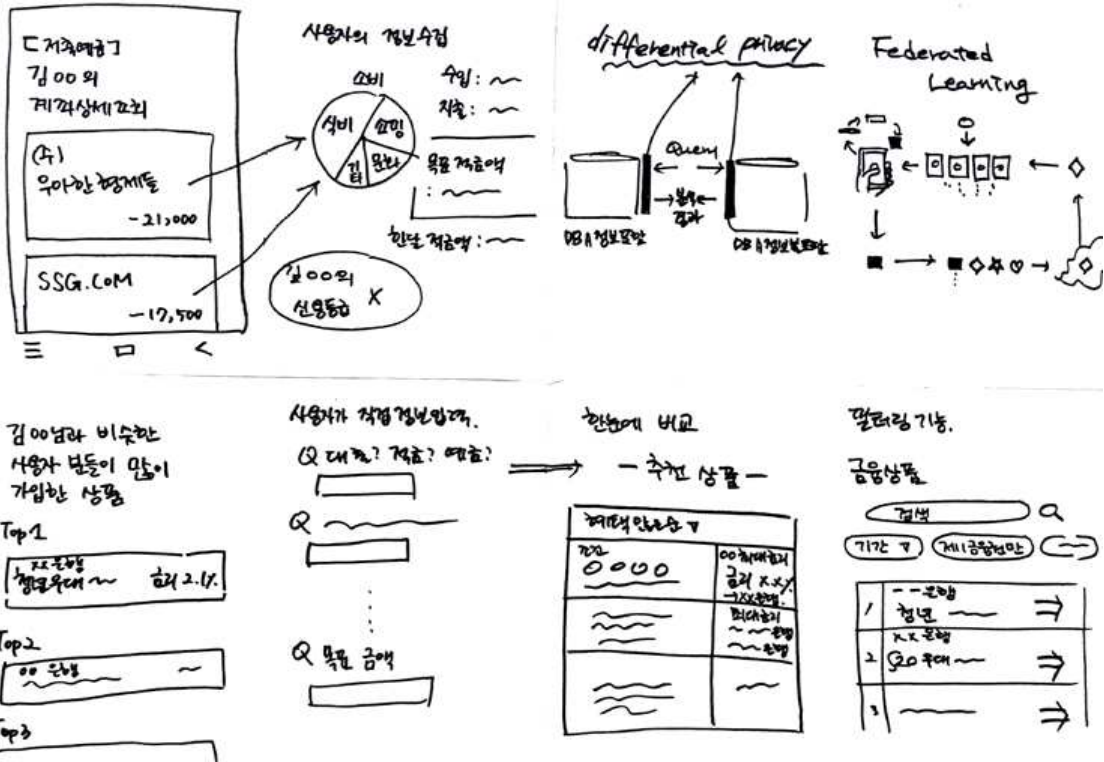


⇒ 포털 사이트 필터 기능 :

대부분의 웹사이트에서 검색을 할 때 필터기능을 이용해 사용자가 원하는 정보만 사용자에게 보여준다.

⇒ 필터 기능을 추가해 사용자가 보고싶은 범위 내의 정보만 볼 수 있도록 한다.

● CRAZY8's



● Solution (LIGHTNING DEMO)

- 어떻게 하면 신뢰도가 높은 상담 결과를 얻을 수 있을까?

'오늘의 집'이라는 인테리어, 집꾸미기

스토어를 참고

자동응답형 상담원을 통해 1차적으로

간단한 질의응답을 통해 상담을 신청할 수 있음

스토어에서 자체적으로 간단하게 검증을

수행하여 인증된 전문가를 추천해 줌

1차 상담의 내용을 바탕으로 2차 상담을

진행하고 그에 맞는 전문가 매칭을 순차적으로 진행함



- 어떻게 하면 사용자가 자신의 소비패턴을 체계적으로 분석해 가입할 상품을 결정할 수 있을까?



자산관리 앱 '뱅크샐러드'

은행, 카드사 계좌와 연결하여 나의 소비패턴을 분석할 수 있음

나의 소비패턴을 바탕으로 할인을 제일 많이 받을 수 있거나 포인트 혜택을 많이 받을 수 있는 카드를 추천받을 수 있으며 나에게 딱 맞는 금융상품도 추천해 줌

각 은행들의 우대 혜택을 정리하여 내가 우대금리를 가장 많이 받을 수 있는 예/적금도 추천받을 수 있음

- 어떻게 하면 사용자에게 불필요한 상품을 제외한 최소한의 상품들만 추천 받을 수 있을까?



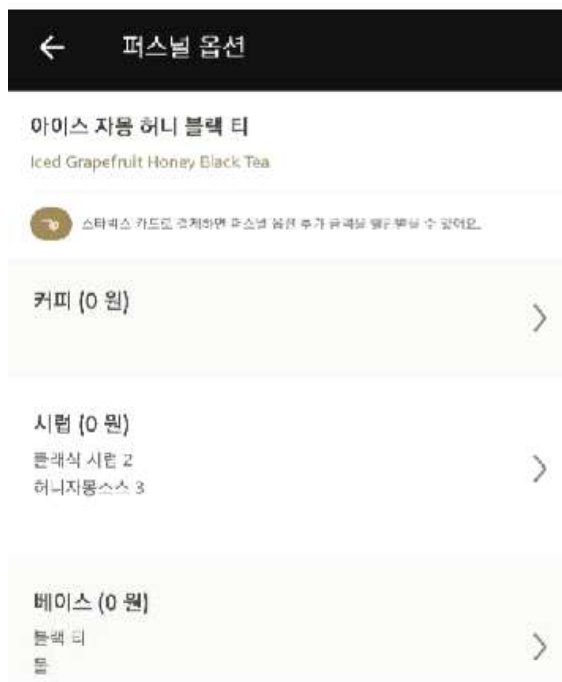
'오늘의 집'이라는 인테리어, 집꾸미기
스토어를 참고

사용자가 원하는 정보 목록을 얻기 위해서 키워드
를 설정할 수 있음

내가 설정한 키워드를 기반으로 한 게시물의 목록
들만 나열됨

이처럼 금융상품에도 키워드를 적용시킨다면 고객
의 설정키워드 조건에 맞는 상품들만 추천받을 수
있음

- 개인정보를 유출하지 않는 선에서 명확한 분석을 제공받을 수 있을까?

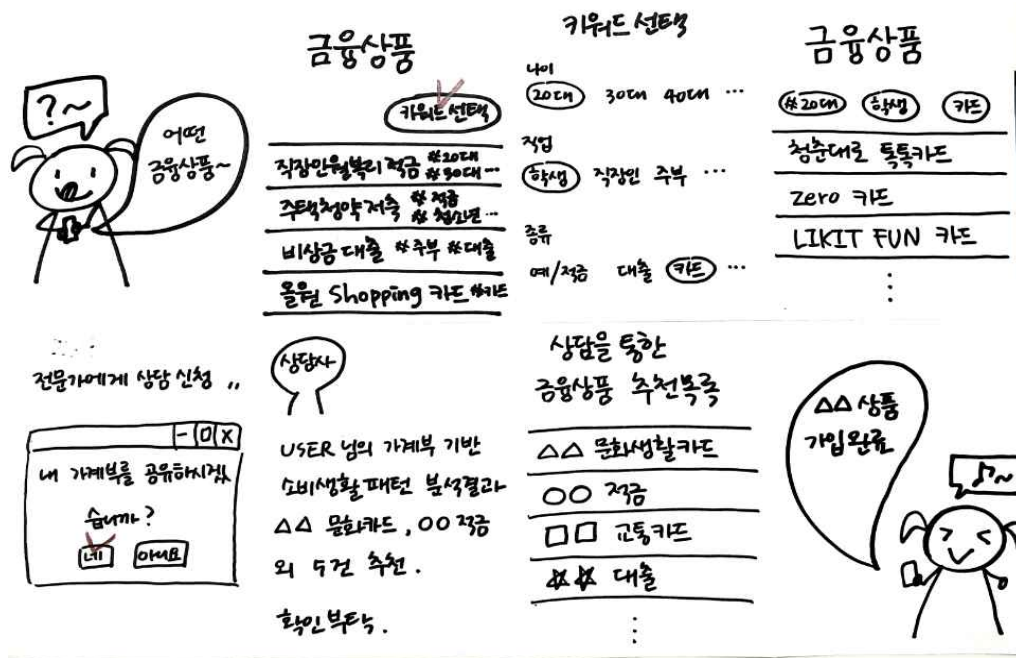


스타벅스 사이렌 오더를 이용할 때, 스타벅스 파트너는 들어온 주문메뉴만으로 고객의 퍼스널 옵션을 만족하는 음료를 제조 해 줌

고객이 음료를 주문하여 받기까지 노출되는 정보는 '닉네임'과 '주문메뉴' 뿐

최소한의 정보제공 만으로도 나만의 개별화된 메뉴를, 또는 나만을 위한 추천 메뉴를 즐길 수 있음

● CRAZY8's



YOUTUBE & GITHUB

[YOUTUBE]

<https://www.youtube.com/watch?v=nM6-CtPHn8w>

[GITHUB]

https://github.com/pmcsh04/designsprint_4gram/tree/master/GP_01

디자인 스프린트 3~4일차

[사그램 조]

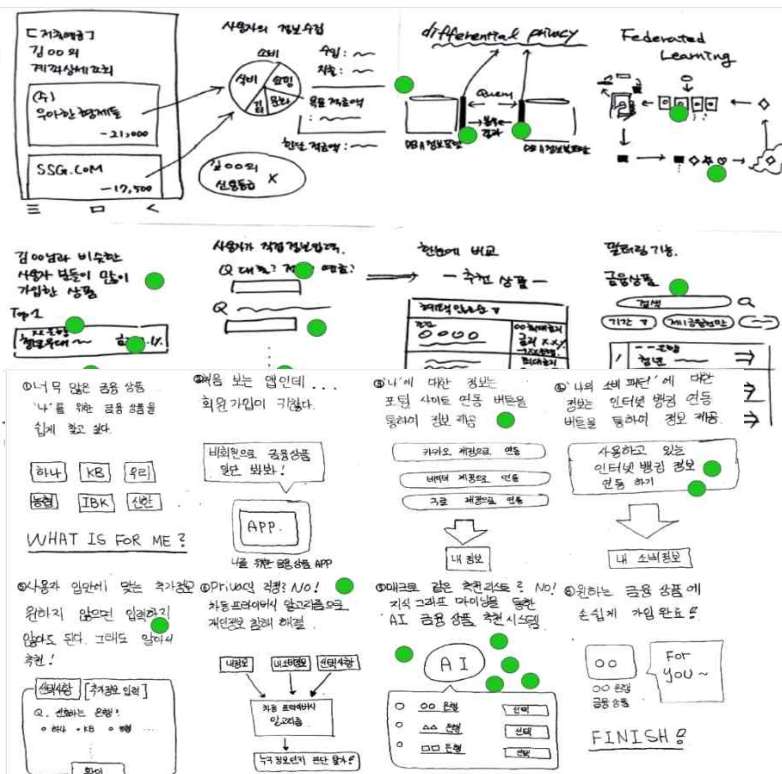
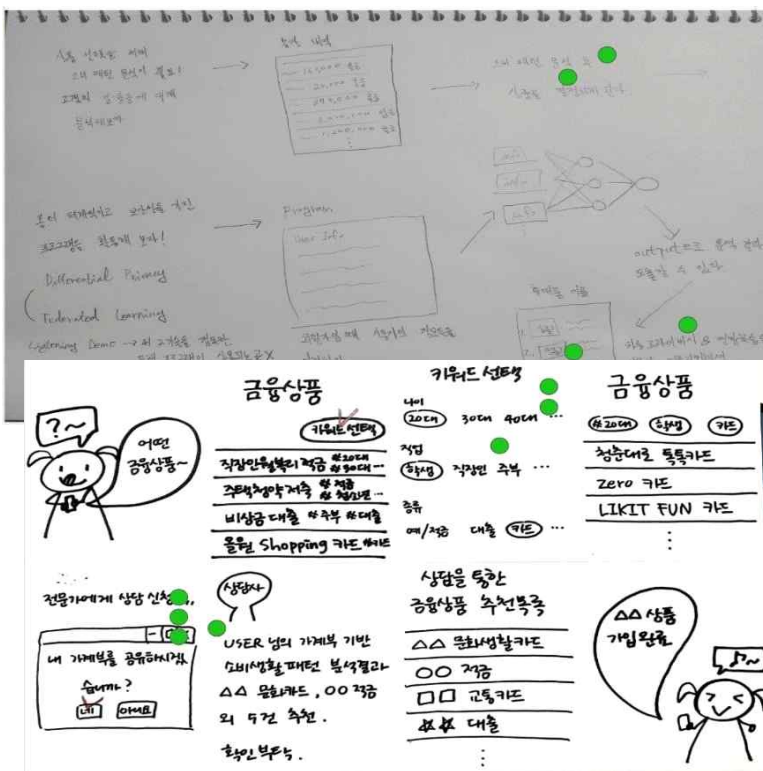
주제

프라이버시 보호 딥러닝 서비스 개발 : 사용자의 소비 패턴을 분석하여 적절한 금융 상품을 소개하는 서비스 개발

스케치 투표

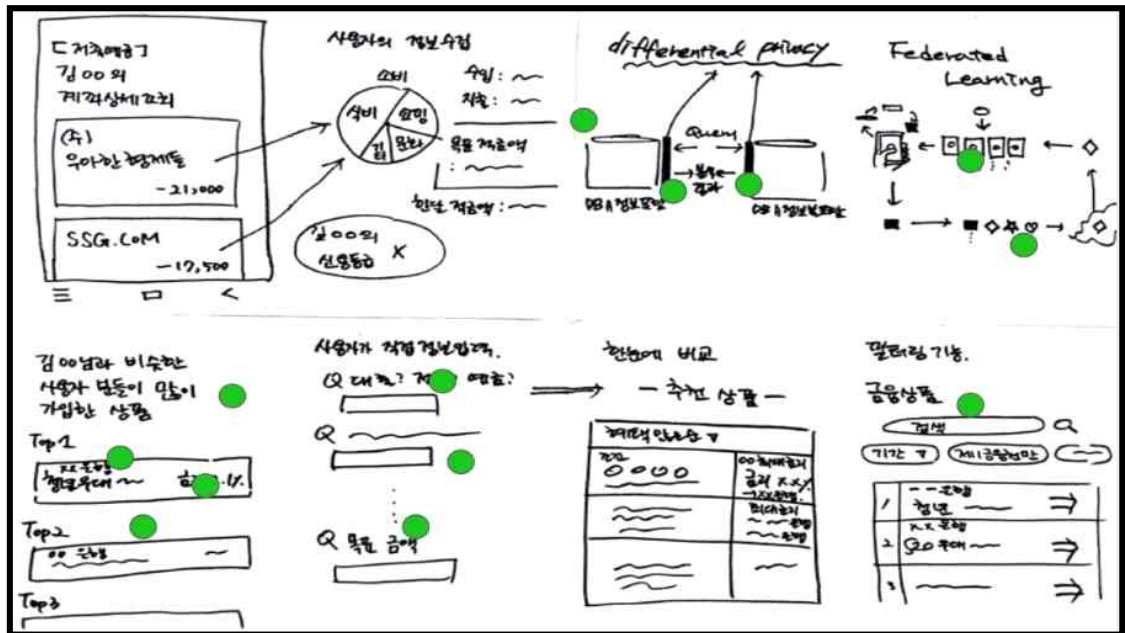
- 투표 방법 : 구글 프레젠테이션 사용

- 팀원들끼리 10개의 스티커를 가지고 있다고 가정하고 마음에 드는 아이디어에 스티커를 붙였다.



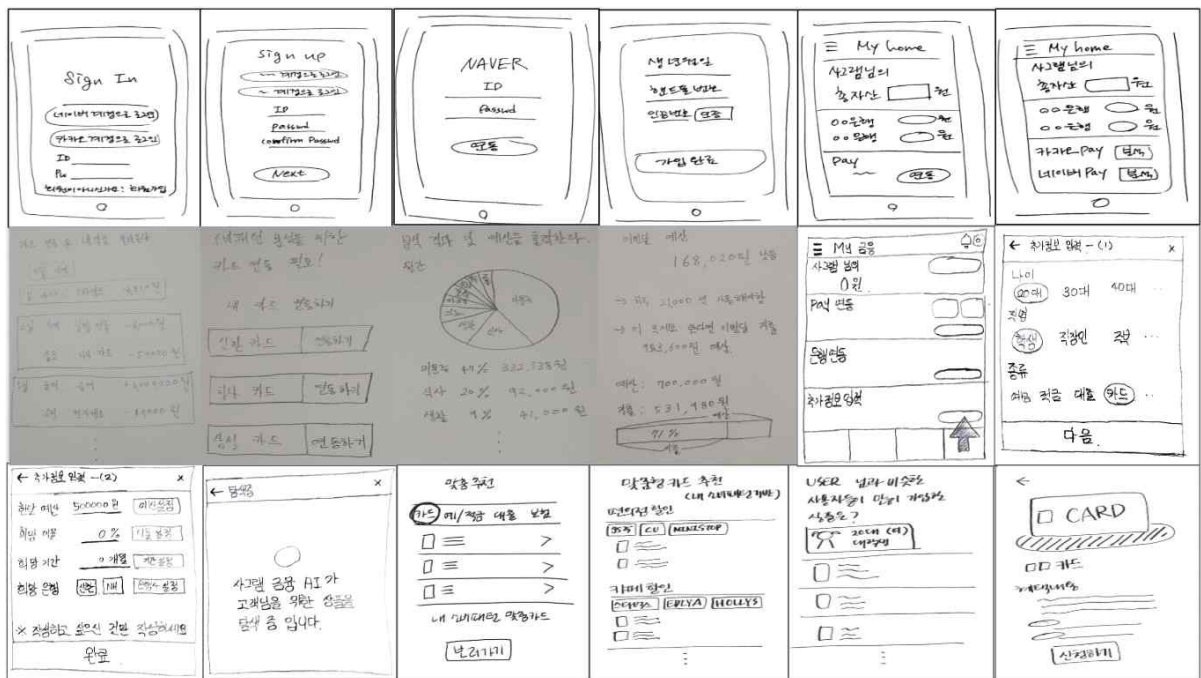
● 투표 결과

[투표 결과 1등 스케치]



- 채택된 스케치에 스티커를 많이 받은 다른 팀원의 아이디어도 추가해 최종 솔루션을 정리했다.

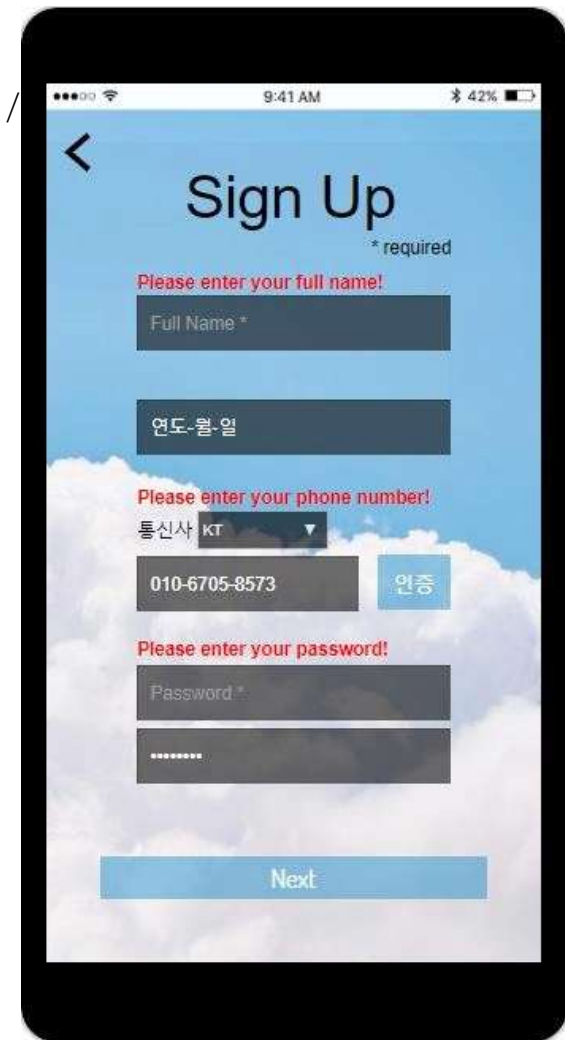
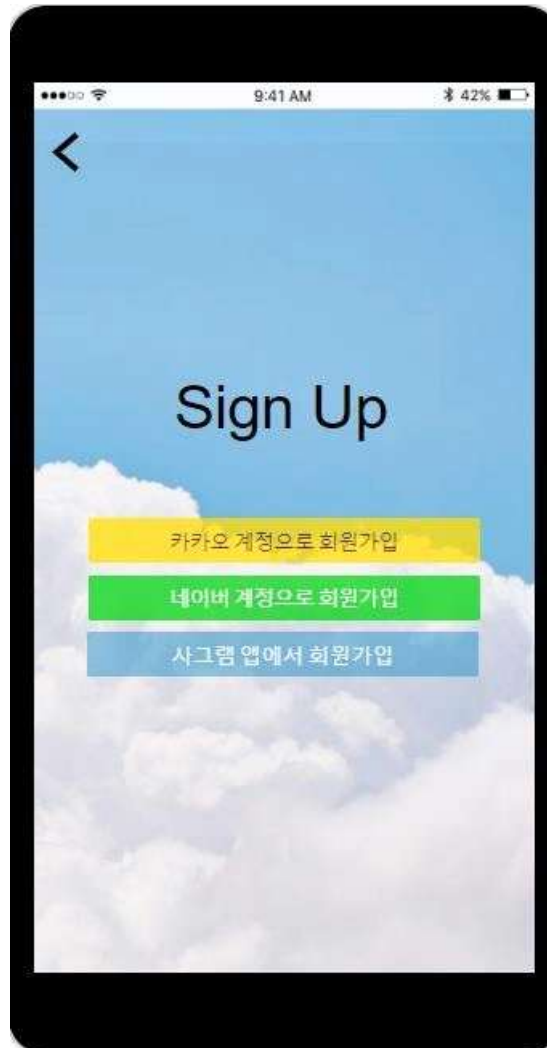
스토리보드



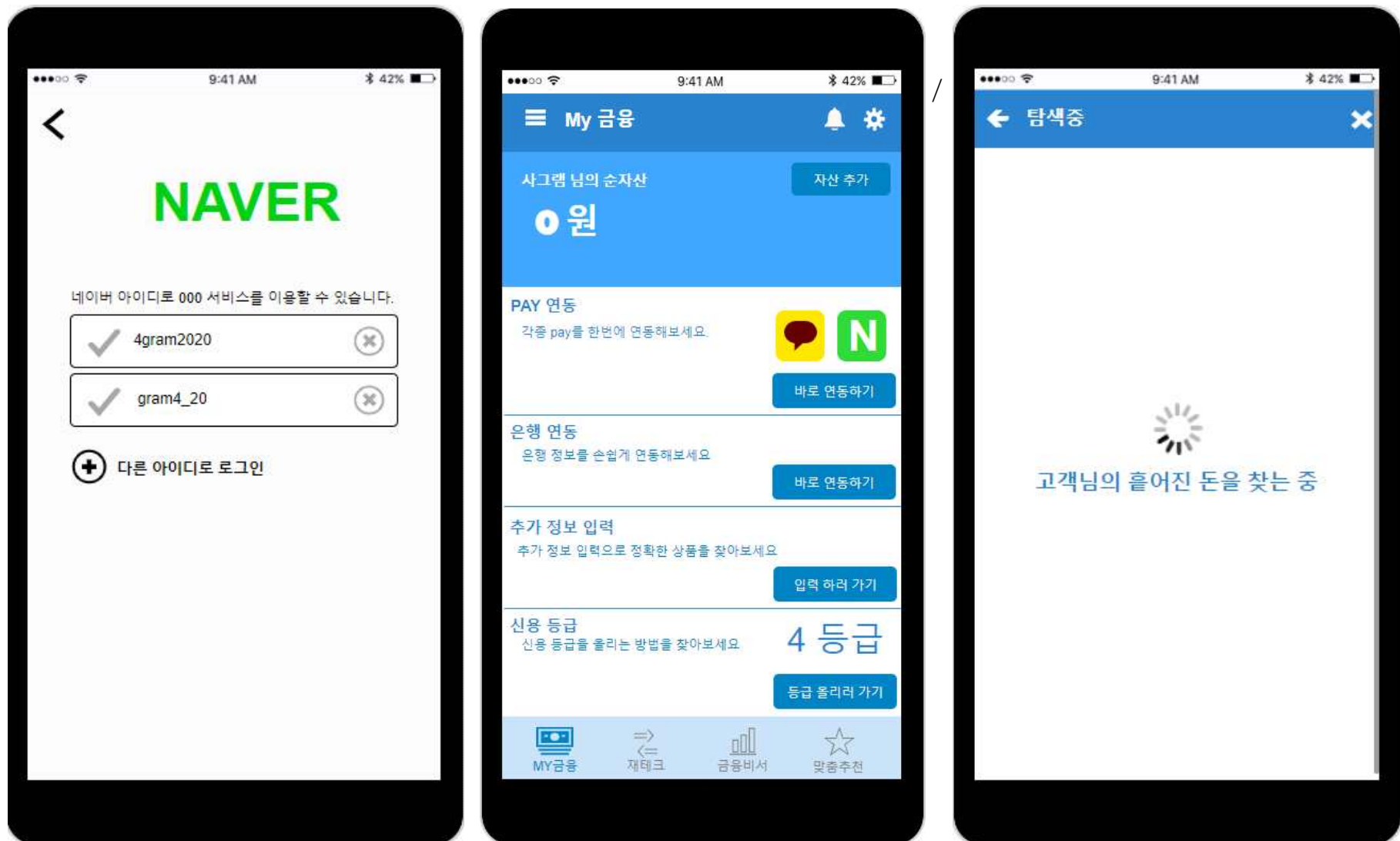
Prototype

- 역할 분담
 - 공동 : 메인 화면
 - 조승현 : 사용자가 원하는 상품에 대한 정보를 직접 입력
 - 이상화 : 사용자 계좌 연동 / 소비 패턴 분석
 - 김수민 : 로그인 / 회원가입 / 페이(payment) 연동
 - 김주희 : 탐색 결과 / 맞춤 상품 추천

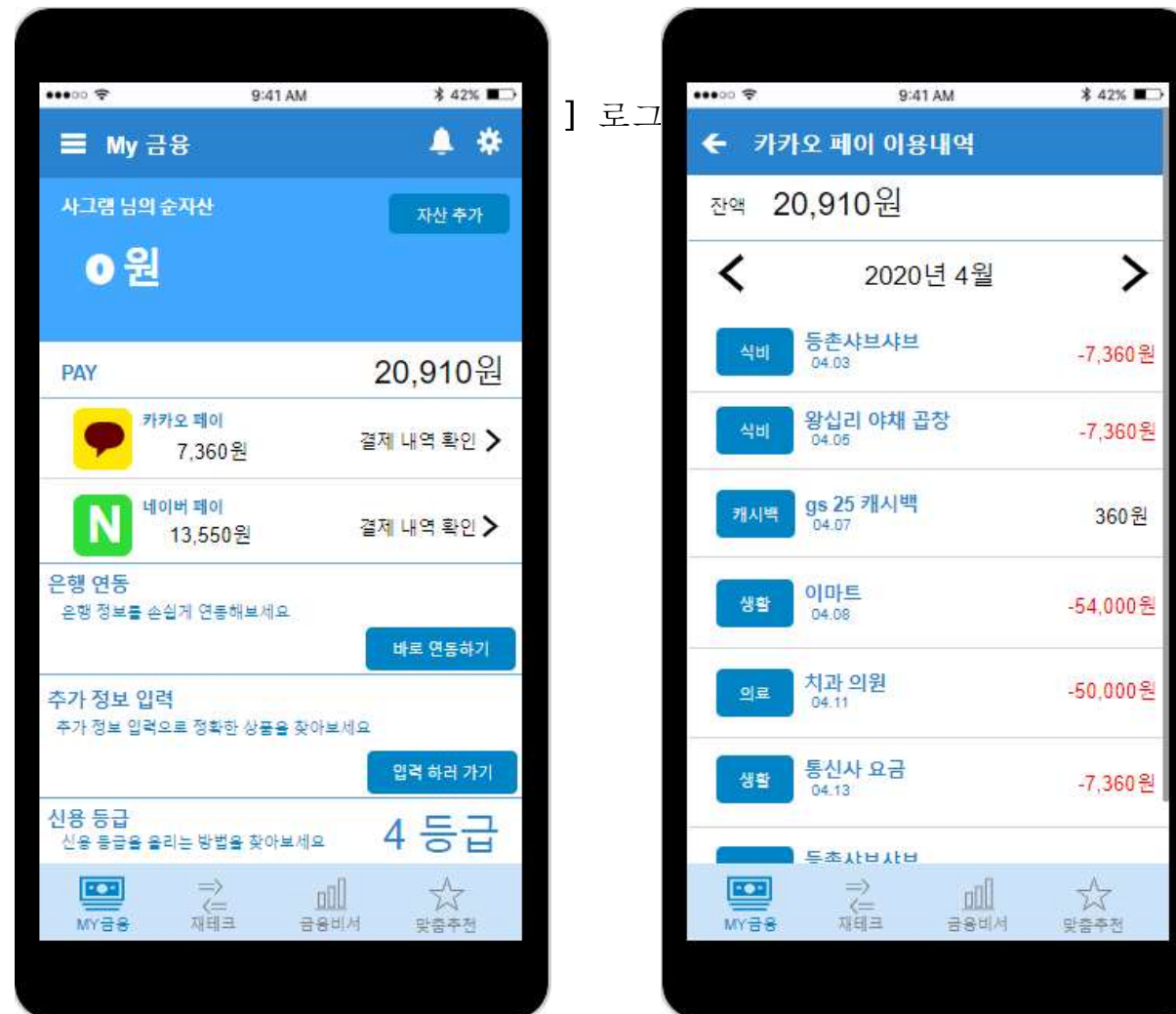
[김수민] 로그인 / 회원가입 / 페이(pay) 연동



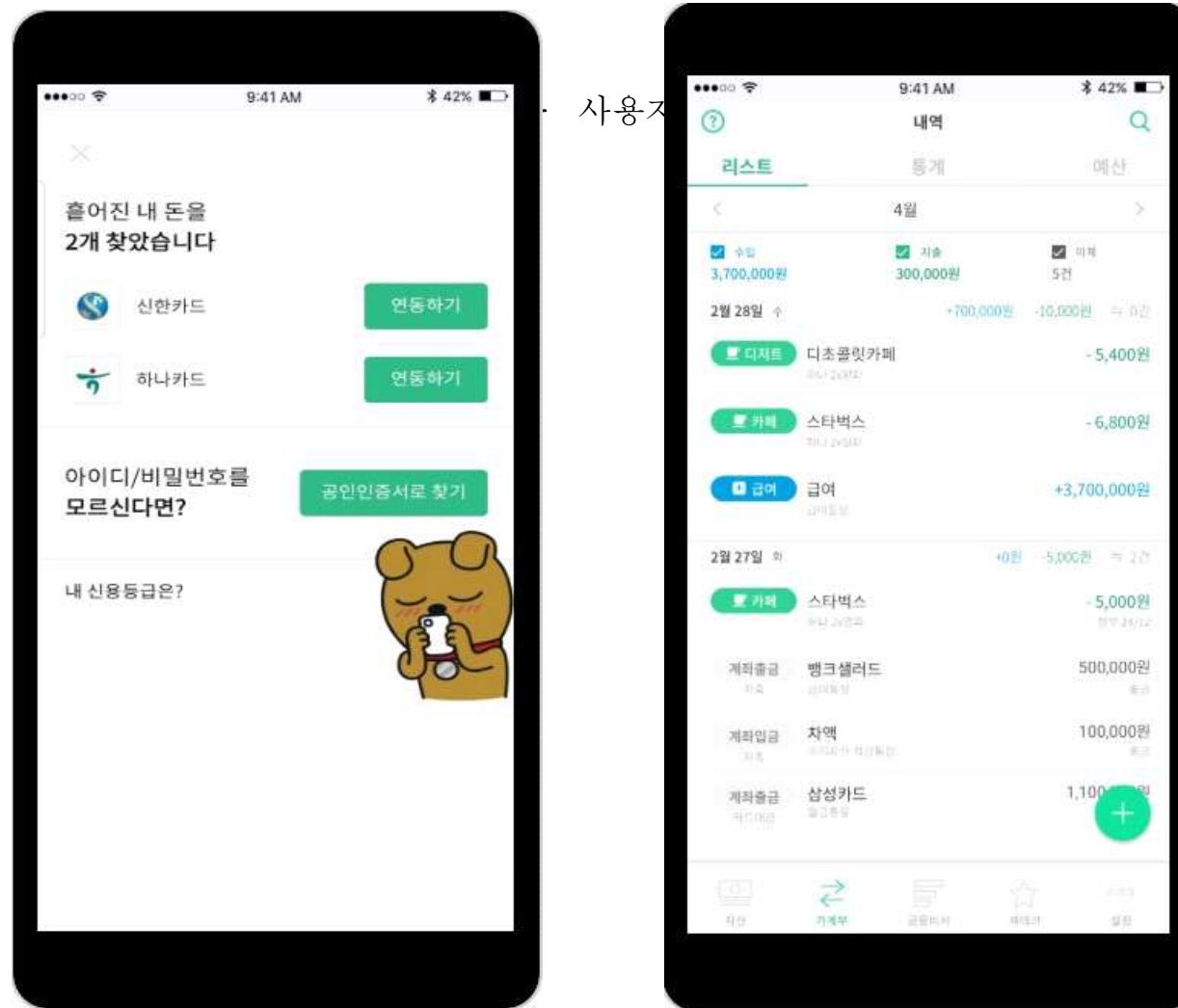
[김수민] 로그인 / 회원가입 / 페이(pay) 연동



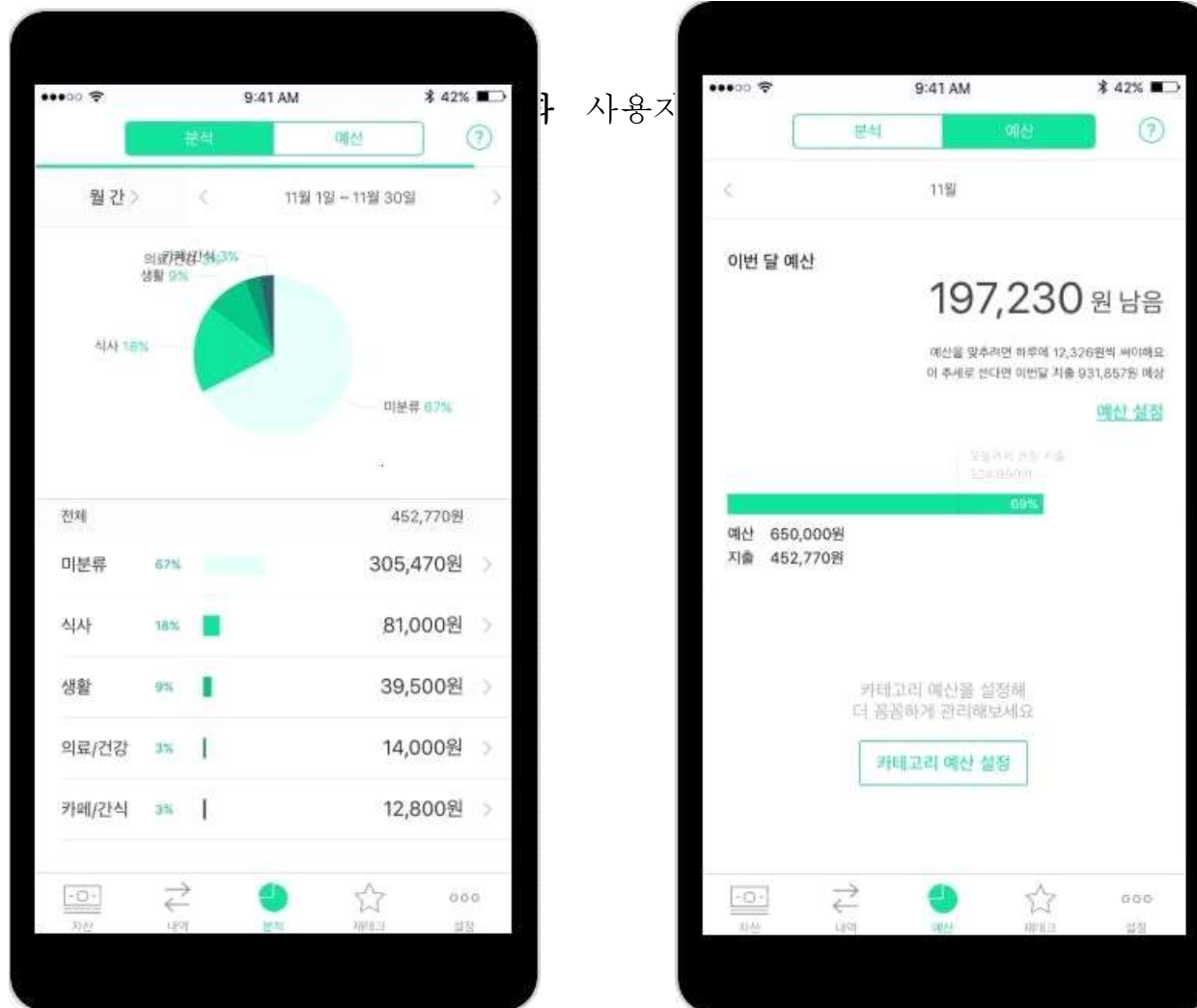
[김수민] 로그인 / 회원가입 / 페이(pay) 연동



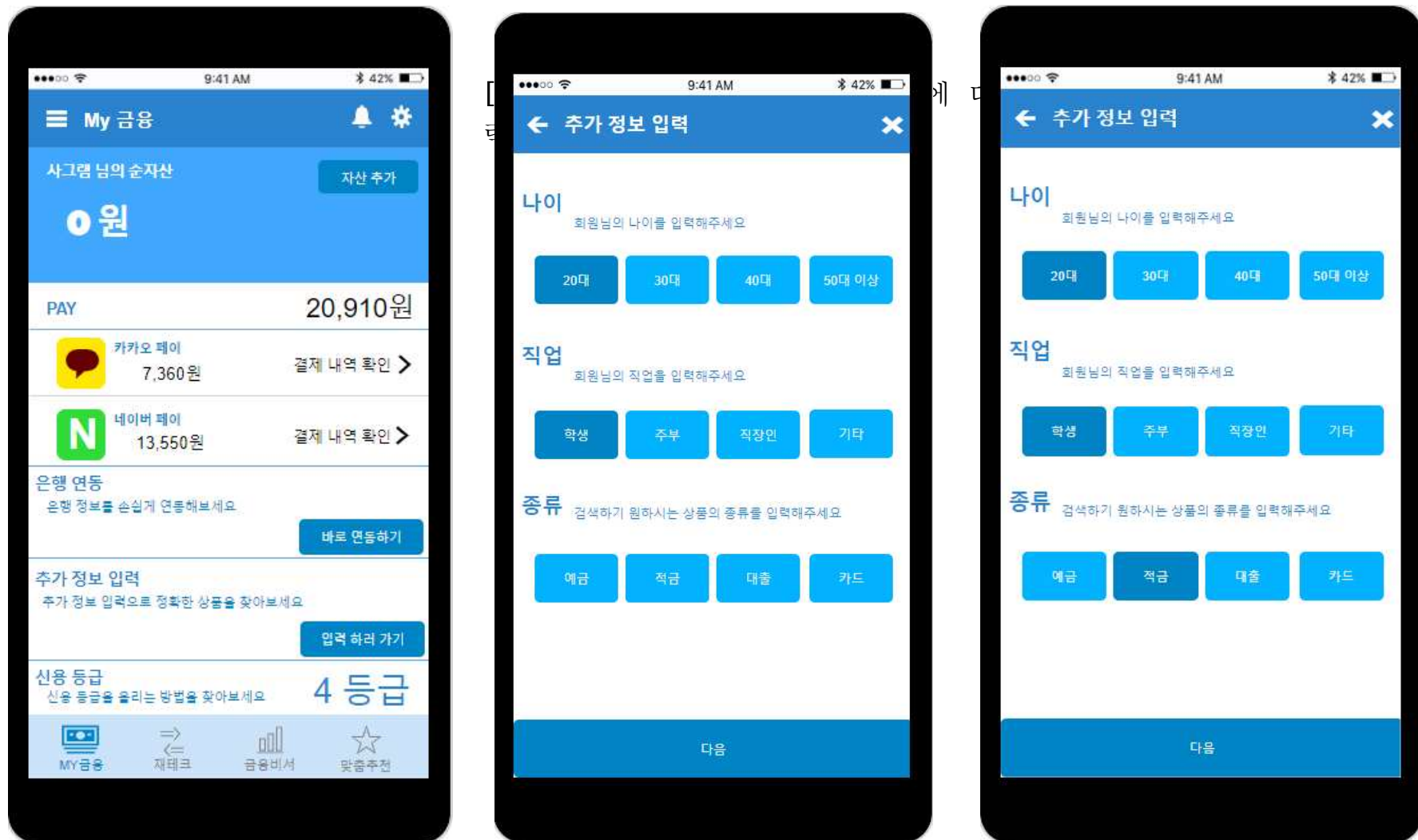
[이상화] 사용자 계좌 연동 / 소비 패턴 분석



[이상화] 사용자 계좌 연동 / 소비 패턴 분석



[조승현] 사용자가 원하는 상품에 대한 정보를 직접 입력



[조승현] 사용자가 원하는 상품에 대한 정보를 직접 입력

← 추가 정보 입력

한달 예산 0 원 영 원 예산 설정

희망 이율 0% 이율 설정

희망 기간 0개월 기간 설정

희망 은행 은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

← 추가 정보 입력

한달 예산 5 원 오 원 예산 설정

희망 이율 0% 이율 설정

희망 기간 0개월 기간 설정

희망 은행 은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

← 추가 정보 입력

한달 예산 50 원 오십 원 예산 설정

희망 이율 0% 이율 설정

희망 기간 0개월 기간 설정

희망 은행 은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

[조승현] 사용자가 원하는 상품에 대한 정보를 직접 입력

← 추가 정보 입력 ×

한달 예산 500 원 예산 설정
오백 원

희망 이율 0% 이율 설정

희망 기간 0개월 기간 설정

희망 은행 은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

← 추가 정보 입력 ×

한달 예산 5,000 원 예산 설정
오천 원

희망 이율 0% 이율 설정

희망 기간 0개월 기간 설정

희망 은행 은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

← 추가 정보 입력 ×

한달 예산 50,000 원 예산 설정
오만 원

희망 이율 0% 이율 설정

희망 기간 0개월 기간 설정

희망 은행 은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

[조승현] 사용자가 원하는 상품에 대한 정보를 직접 입력

9:41 AM 42%

← 추가 정보 입력

한달 예산
500,000 원 예산 설정
오십만 원

희망 이율
0% 이율 설정

희망 기간
0개월 기간 설정

희망 은행
은행사 설정

※ 작성하고 싶으신 것만 작성하세요.

완료

9:41 AM 42%

← 은행사 설정

MY	은행	카드	증권
보험	부동산	연금	현금

카카오뱅크	케이뱅크
우리은행	농협은행
국민은행	우체국
신한은행	수협은행
KEB하나은행	산업은행
기업은행	부산은행
SC제일은행	대구은행
씨티은행	경남은행

완료

9:41 AM 42%

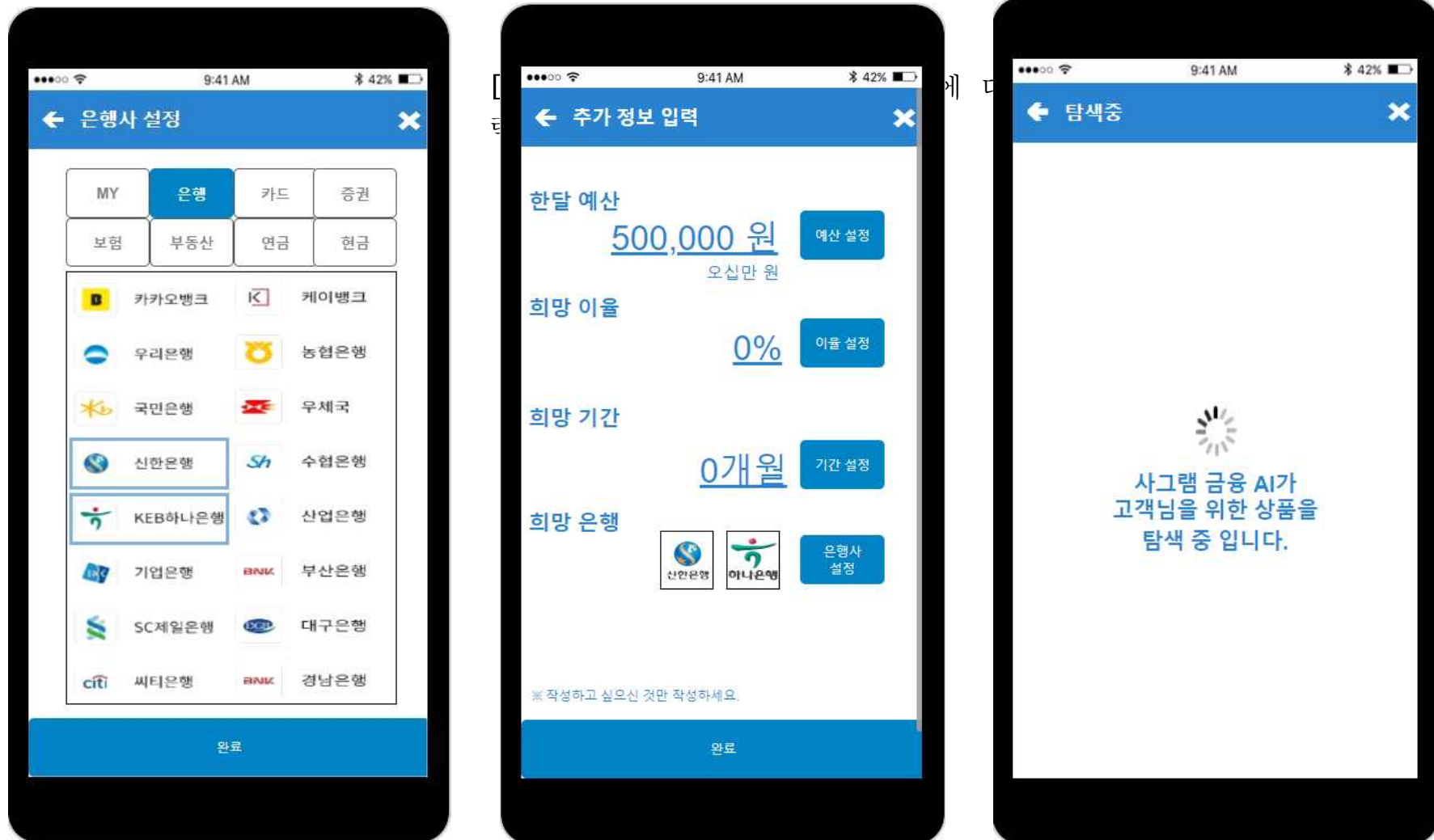
← 은행사 설정

MY	은행	카드	증권
보험	부동산	연금	현금

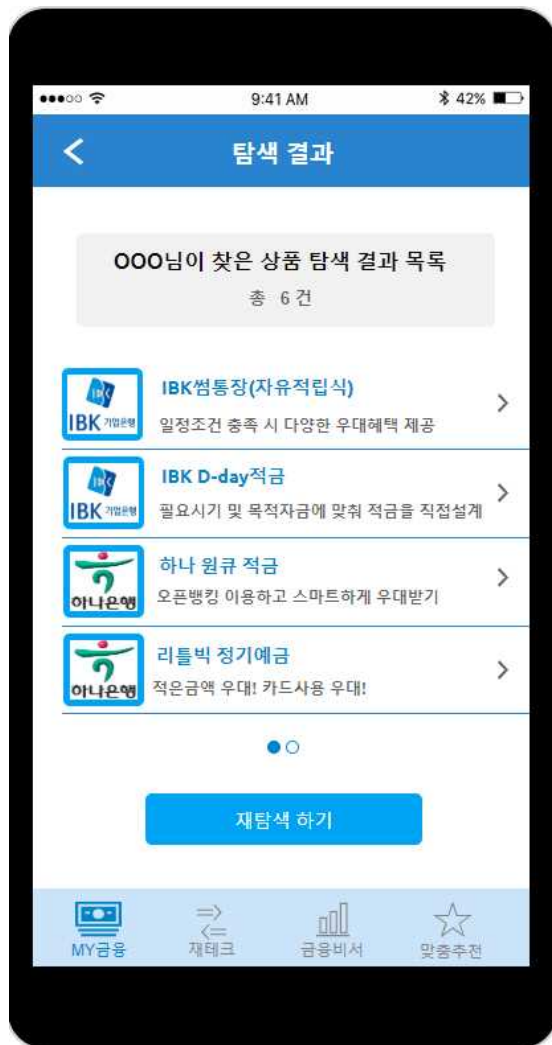
카카오뱅크	케이뱅크
우리은행	농협은행
국민은행	우체국
신한은행	수협은행
KEB하나은행	산업은행
기업은행	부산은행
SC제일은행	대구은행
씨티은행	경남은행

완료

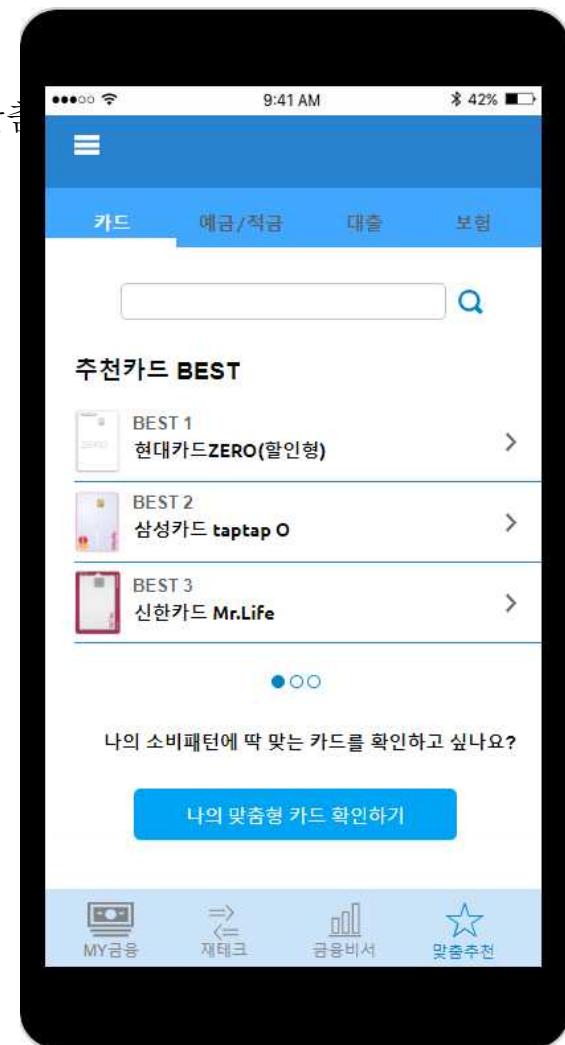
[조승현] 사용자가 원하는 상품에 대한 정보를 직접 입력



[김주희] 탐색 결과 / 맞춤 상품 추천



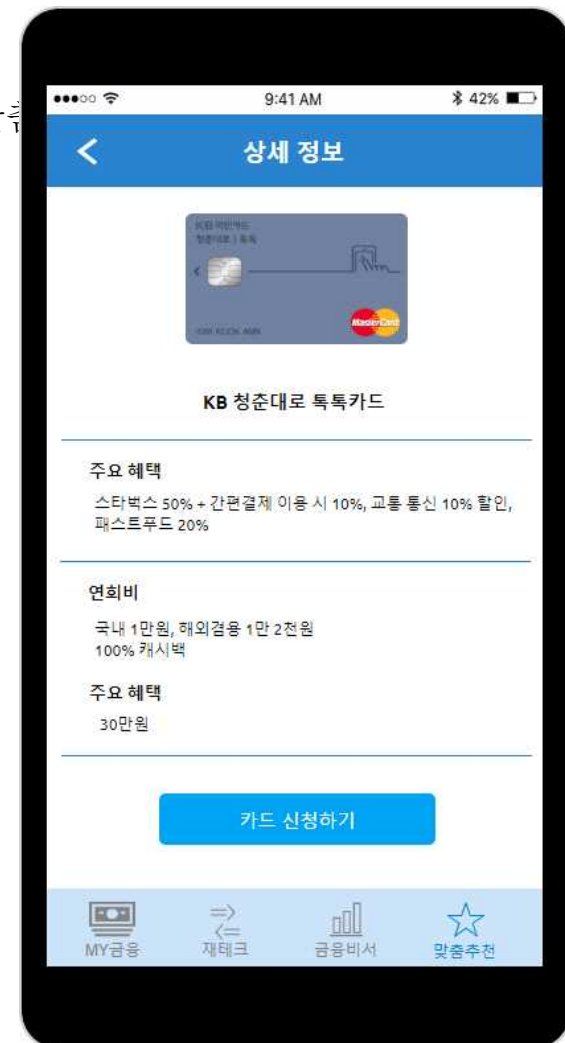
맞춤



[김주희] 탐색 결과 / 맞춤 상품 추천



맞춤



YOUTUBE & GITHUB

[YOUTUBE] <https://www.youtube.com/watch?v=iubR4clZTZw>

[GITHUB] https://github.com/pmcsh04/designsprint_4gram/tree/master/GP_02

디자인 스프린트 최종보고서

주제

프라이버시 보호 딥러닝 서비스 개발

-사용자의 소비 패턴을 분석하여 적절한 금융 상품을 소개하는 서비스 개발

● 동기

- '차등프라이버시 알고리즘' 도입을 통한 데이터 보호
- 단순 매크로 추천이 아닌 '지식 마이닝 그래프 알고리즘'을 활용하여 AI 기반 추천 시스템 개발
- 데이터 보호와 강력한 추천시스템을 바탕으로 하여 다양한 서비스 제공
- 신규 사용자 유치, 고객 충성도 상승

● 시장

- 얼마나 많은 이용자가 사용할까?
 - * 20대 이상의 계좌 및 신용카드가 있는 사용자들이 주 이용자
 - * 소비내역을 한번에 확인하고 싶은 사람, 금융 상품에 관심이 많은 사람
 - * 한눈에 여러 금융 상품을 비교하고 싶은 사람
- 얼마나 큰 수익을 낼 수 있을까?
 - * 한국의 금융앱은 बैं킹 서비스 앱과 더불어 거래추적 등의 다양한 분리형 앱을 제공하고 있어 통합 앱을 제공하는 다른 나라(호주, 중국 등)과의 차이를 보임
 - * 이와같은 분리형 앱은 제대로 관리해 준다면 **소비자의 수요에 빠르게 반응**할 수도 있고 개편도 손쉬움
 - * 국내외 금융 관련 앱들은 총 19억개가 설치됨

특히 국내외 2019년 2분기 다운로드 수 상위 앱은 **뱅킹/배달/유통앱**이 10위 내 절반을 차지하였을 정도로 **서비스의 확산**이 가속화되고 있음을 알 수 있음

* 어떻게 하면 사용자의 데이터를 보호하면서 맞춤형 금융 상품을 추천해 줄 수 있을까에 초점을 맞춘 어플리케이션

* 차등 프라이버시 기술 적용 -> 기존의 많은 금융앱과 **차별성, 경쟁성**을 갖추
기술이 발달하고 정보화 사회가 고도화될수록 사생활 보호가 부각, 차등 프라이버시 기술은 우리 제품의 시장 가치를 높임

● 주요 문제

- 구현 고려사항
- 실현 가능성
- 대규모 이용자 수용
- 프라이버시 보호

[Solution] Use Differential Privacy & Federated Learning !

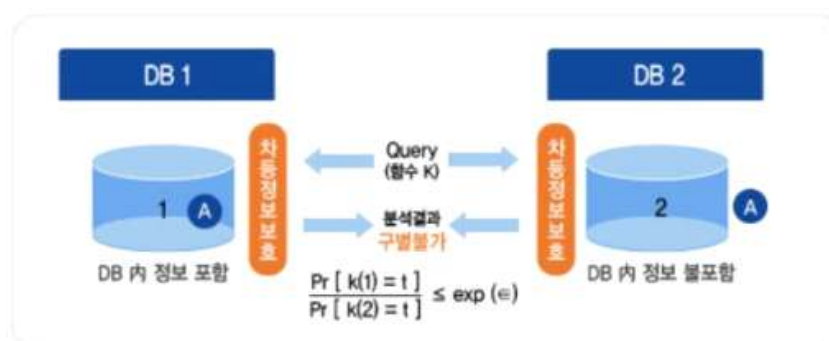
* Differential privacy

차등 프라이버시는 '한 인구 집단에 대한 유용한 정보는 학습하면서 동시에 집단 내 한 개인에 대한 정보는 얻지 못하게 하려면 어떻게 해야하는가' 라는 문제에 대한 수학적 접근법이다.

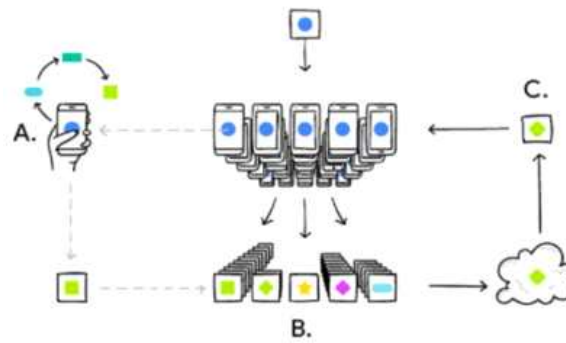
차등 프라이버시를 도입하는 방법을 여러 가지이다. 이 방법들의 핵심은 데이터 수집 과정이나 데이터베이스 질의에 대한 응답 과정 등에 잡음을 집어넣는 것이다. 이 잡음이 개인의 프라이버시를 지키지만 데이터가 모두 결합되는 단계에서는 제거되기 때문에, 대상 전체에 대한 유용한 통계는 계산할 수 있다.

예를 들어, 연봉정보가 저장된 데이터베이스 D에 연봉평균을 질의했을 때, 아무런 프라이버시 보호 조치가 취해지지 않는다면, A가 포함된 경우와 포함되지 않은 경우 결과값 차이로부터 A의 정확한 연봉을 계산할 수 있게 됩니다.

차등정보보호는 이러한 경우 A의 연봉을 계산할 수 없도록 지원하는 프라이버시 보호 기술로 쿼리 Q에 대한 응답 R에 적절한 분포의 Noise를 섞어주게 됩니다.



* Federated Learning



AI가 발전하기 위해서는 더 많은 데이터로 학습을 해야하는데 이때 일반적으로 엣지 단에서 수집한 데이터를 중앙 서버로 전송한 뒤 중앙 서버가 데이터를 분석한 후 다시 그 결과를 엣지 단으로 보내주는 과정이 필요하다.

문제는 이 과정에서 과부하가 걸린다. 구글은 연합학습을 고안했다.

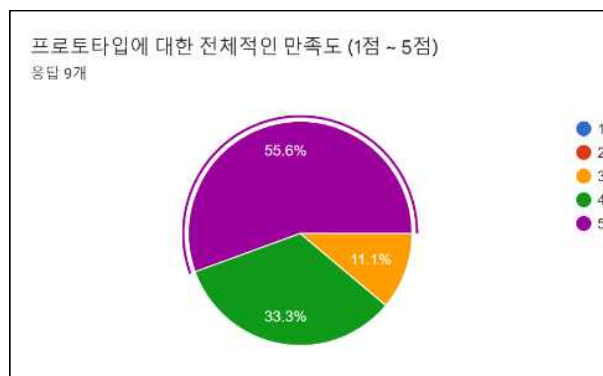
연합학습(Federated Learning)은 머신러닝을 중앙 클라우드가 아닌, 사용자 개별 디바이스에서 스스로 데이터를 처리하고 발전시키는 방식이다. 일상 생활 속에서 이용하는 스마트폰이 곧 AI 도구가 되는 셈이다.

● 설문조사

- 설문조사 문항

1. 프로토타입 만족도 (1점~5점)
2. 좋았던 점은 무엇인가요?
3. 불편하거나 유용하지 않다고 생각되는 기능은 무엇인가요?
4. 기존의 금융상품 가입방법과 비교하였을 때 어떤지 서술해주세요.
5. 추가하면 좋을 것 같은 기능은 무엇인가요?

가) 프로토타입 만족도



5점 - 5명 / 4점 - 3명 / 3점 - 1명

평균 : 4.56 / 5 (점)

나) 좋았던 점은 무엇인가요?

- 간단하고 편리하다
- 로그인/회원가입이 간편, 연동도 쉽게 되고 소비패턴 파악해서 추천해주는 기능
- 나의 소비패턴을

다) 불편하거나 유용하지 않다고 생각되는 기능은 무엇인가요?

라) 기존의 금융상품 가입방법과 비교하였을 때 어떤지 서술해주세요.

마) 추가하면 좋을 것 같은 기능은 무엇인가요?

문제정의서(연구계획서)

과제명	프라이버시 보호 딥러닝 서비스 개발
-----	---------------------

조	사그람 조
지도교수	임성수 교수님 (서명)
조원	201402433 조승현 201402392 이상화 201704144 김수민 201704144 김주희

1. 연구의 필요성

(1) 국내·외 연구현황

- Siri 등 애플 인공지능 기술(2017년)에 iOS 10부터 사용사례에 대해 식별자와 IP 주소를 제거한 로컬 프라이버시 모델로써 해시와 스케치를 통한 데이터 압축과 디바이스 상에서 학습하는 기법을 사용하여 인기 이모티콘 발견, 메모리 관리, 새로운 단어 학습 등에 활용하고 있다.
- 구글에서는 프라이버시 보호 방법으로 'RAPTOR'(2014년), 프라이버시를 보호하면서 대규모 데이터를 수집하는 방식을 개발하여 블룸필터와 해시를 활용한 데이터 압축과 변형을 통한 차등프라이버시 보장 기법을 통하여 악성소프트웨어, 웹페이지 특성을 파악하여 차단하는 방식으로 활용하고 있다. 2017년에는 가상 키보드 (Gboard)를 개발하였으며 사용자 디바이스 상에서 인공지능망 학습 기법을 활용하여 자동완성, 수정에 활용을 하고 있다.
- 'Oasis Labs'은 UC 버클리 대학 D.Song 교수와 대학원 생들이 설립한 Lab으로써 블록체인 기술로 보호되는 프라이버시 보장 클라우드 컴퓨팅 플랫폼을 구축중이고 사기를 탐지하고 차등 프라이버시가 보장되는 스마트 계약 라이브러리를 개발 중이다.
- 국내에서도 해외와 마찬가지로 다양한 프라이버시 보호에 대한 연구가 진행되는 것처럼 보이는데 제도적인 측면에서 '개인정보 비식별 조치 가이드라인(2016)'같이 가이드라인 외에 개인정보 보호 기법 중 활용에서는 눈에 띄는 결과를 보여지지 않는다고 여기진다.

(2) 문제점

- 빅데이터 기술의 급부상으로 인한 개인 정보 대량 수집 및 활용에서 특정 개인을 식별 가능한 데이터를 통한 프라이버시 침해 우려가 있다.
- 개인정보들 중에 특히나 보호가 필요한 개인정보가 존재한다.(의료 정보, 유전자 정보, 온라인 정보, 재정 정보 등)
- 기업에게 개인정보 데이터 셋의 유출에 대한 큰 법적인 위험이 존재한다.
- 문제 해결을 위해 접근할 수 있는 데이터의 양이 극도로 제한된다.
- 사회 전반과 거의 모든 사람들이 산업을 직면하고 있는 연구를 방해하고 있어 질병을 치료하거나 복잡한 사회 동향을 이해하는 것이 어려워진다.
- 적절한 훈련 데이터에 접근할 수 없다는 것은 사회에서 가장 중요하고 개인적인 문제들 중 일부는 기계학습으로 해결 할 수 없다.

(3) 연구개발의 필요성 및 중요성

- 프라이버시를 보호하는 기계 학습을 하는 방법을 배워 데이터를 통한 인류 능력의 진보를 가능하게 한다.
- Federated Learning과 같은 머신러닝 학습방법과 Differential Privacy 알고리즘을 통하여 개인정보를 보존하기 위한 기술을 배우고 On-Device 모델을 만듦으로써 프라이버시를 다루는 최신 기술 트렌드를 맞춰간다.
- 프로그램 개발 면에서 국내에서는 활발하게 다뤄지고 있다고 여겨지지 않는 실질적인 프라이버시 보호 딥러닝 서비스를 개발함으로 프라이버시 보호 딥러닝 서비스 개발의 초석을 다지도록 한다.

2. 연구의 목표 및 내용

(1) 연구목표

- 개인이 느끼는 개인정보 유출 불안도를 최소화시킨다.
- 만약 중앙 시스템에서 개인의 정보가 유출되었을 시에 개인정보 유출에 대한 위험도를 최소화 시킨다.
- 개인정보가 포함된 데이터 셋에 대한 연구적 사용절차를 단순화 시킨다.
- 차등프라이버시의 사용으로 개인정보의 불특정을 최대화 시킨다.
- 프라이버시 보호 알고리즘 개발로 개인정보 데이터를 활용하는 연구의 증가율을 최대화 시킨다.
- 연합학습 머신러닝 학습방법 설계 시 중앙 시스템 AI를 통해 개인정보가 역으로 유출될 가능성이 제로에 수렴하게 설계한다.
- 온디바이스 설계시 온디바이스의 장비 사이즈를 최소화 시키도록 한다.
- 온디바이스 설계시 'Empowered edge'에 부합하게 설계한다.
- 데이터 입력시 추가적인 기법(블룸 필터, 해시, 스케치 등)의 도입을 통하여 개인정보를 보호의 효율성에 대하여 연구한다.
- 이 연구의 결과로 다른 여러 프라이버시 보호 서비스 연구의 진행에 큰 활력을 불어 넣도록 한다.

(2) 연구내용 & 연구범위

- 차등프라이버시(Differential Privacy) 알고리즘의 전반적인 내용에 대하여 깊이 연구 하도록 한다.
- 차등프라이버시 알고리즘의 충분한 학습한 이후에 연합학습(Federated Learning) 머신러닝 학습방법에 대해 이해하고 서비스 구축까지 진행하도록 한다.
- 연합학습 머신러닝 학습방법까지 구축을 완벽하게 했을 시 온디바이스(On-Device) 기법에 대하여 이해하고 향후 적용 방향에 대하여 생각해보도록 한다.
- 위 개인정보 보호 방법 뿐만이 아니라 다양한 개인정보 보호 방법에 대하여 생각해 보고 도입시에 더 효과적으로 개인정보를 보호 할 수 있는 방법에 대해서 생각해보도록 한다.
- 특정한 서비스(본 연구에서는 현재 금융을 서비스로 특정하고 있음)에 차등프라이버시, 연합학습, 온디바이스를 적용하였을 때 얻을 수 있는 결과에 대해서 알아보도록 한다.

3. 연구의 추진전략 및 방법

(1) 연구 추진전략

- 연구 대표 주제를 '차등프라이버시(Differential Privacy) 알고리즘'으로 선택하여 차등프라이버시 알고리즘에 대한 이해를 바탕으로 연구를 진행한다.
- 연구 대표 주제인 차등프라이버시에 대한 충분한 이해의 바탕이 되었다는 가정 하에 연합학습(Federated Learning) 머신러닝 학습방법 또는 온디바이스(On-Device) 모델로 개인정보를 보호하는 실질적인 서비스를 개발한다.
- 연구의 빠른 진행 속도를 위해서 이미 나와 있는 학습 도구를 활용하여 공부하며, 연구 방향에 맞는 오픈소스를 적절하게 사용한다.

(2) 연구 방법

- 하단 3번의 두 가지 기존 방식(문제점)에 대응하는 하단 4번의 한 가지 개선방향(해결방안)을 통하여 해결하는 방식으로 연구를 진행한다.

(3) 기존방식(문제점)

- 기업의 개인정보의 수집, 보관으로 인한 프라이버시 침해 위험이 크다.
- 연결 공격(Linkage Attack), 복합공격(Composition Attack) 등을 통하여 개인정보를 보호해도 대부분 인원들이 특정되거나 사전 지식이 있는 경우 민감한 정보 취득 위험이 증가하는 경우가 있다.
- 중앙시스템에서 개인정보에 관한 학습을 할 경우에 중앙시스템의 개인정보 유출 위험이 있다.
- 중앙 시스템으로 AI 시스템 강화를 위하여 수집 대상을 개인의 데이터가 아닌 명령어를 수집한 경우가 있는데 이러한 경우에도 많은 사람들이 부정적인 인식을 보인다.
- 프로그램의 발전을 위해서 기업의 지속적인 개인정보 수집은 개인의 입장에서 부담스럽다.
- 2018년 10대 유망 기술로 'Cloud to the edge', 2019년과 2020년에는 'Empowered edge'로 에지 컴퓨터의 비중이 커지고 있는 상태이다.

(4) 개선방향(해결방안)

- Differential Privacy 알고리즘을 사용하여 개인을 특정 불가하게 하여 기업과 개인의 개인정보유출의 불안 감소시킨다.
- Federated Learning 방식으로 개인 기기 내 학습을 통하여 중앙 시스템으로 개인정보를 보내지 않음으로써 개인정보 유출을 해결한다.
- On-Device 모델을 통하여 개인을 위한 인공지능을 만들어서 개인정보를 보호한다.

4. 연구 팀의 구성 및 과제 추진 일정

(1) 연구진 구성 및 역할

구성원	담당	비고
조승현	팀장	
이상화	부팀장	
김주희	물품구매담당자	
김수민	조원	

○ 연구진 구성원들 모두 프로젝트에 있어서 동일한 역할을 하도록 하며 프로젝트 진행이 되면서 과정이 복잡해져 역할을 나누어야 할 시 구성원들의 역할을 지정하도록 한다.

(2) 추진일정

○ 전체 일정

내용	월											
	1	2	3	4	5	6	7	8	9	10	11	12
1. 졸업프로젝트 준비기간	O	O										
2. 졸업프로젝트 1차 점검			O	O								
- 졸업프로젝트 관련학습 중간 점검 & 평가			√									
- 졸업프로젝트 관련 서류 작성				√								
3. 졸업프로젝트 2차 점검					O	O						
- 졸업프로젝트 관련학습 최종 점검 & 평가						√						
- 관련 논문 조사						√						
- 데이터 셋 조사 (1차)						√						
- 프로토 타입 설계 (1차)						√						
4. 졸업프로젝트 3차 점검							O	O				
- 데이터 셋 조사 (2차)							√	√				
- 프로젝트 관련 자료 최종 탐색							√					
- DP 알고리즘 구현								√				
5. 졸업프로젝트 최종 완성									O	O		
- 데이터 셋 조사 (최종)									√	√		
- Federated Learning 구현										√		
- On- Device 기법 구현										√		
6. 최종 완성 이후											O	O
- 논문 작성											√	
- 졸업프로젝트 심사												√

○ 1학기 상세 일정

월	내용	비고
3월	Udacity 학습 중간 점검 & 평가 < Secure and Private AI > 1. Learn how to build and train deep neural networks using PyTorch. 2. Introducing Differential Privacy 3. Evaluating the Privacy of a Function 4. Introducing Local and Global Differential Privacy	
4월	졸업 프로젝트 관련 서류 작성 (문제정의서, 요구사항명세서, 유스케이스, 클래스 다이어그램, 시퀀스 다이어그램 등)	
5월	Udacity 학습 최종 마무리 & 평가 < Secure and Private AI > 5. Differential Privacy for Deep Learning 6. Federated Learning 7. Securing Federated Learning 8. Encrypted Deep Learning 추가 학습자료 공부 마무리 & 평가 Security and Privacy of Machine Learning (youtube) Federated Learning(federated.withgoogle.com)	
6월	관련 논문 조사 데이터 셋 조사 프로토타입 설계	

- 참고문헌(Reference)

임성수, 「프라이버시 보호 빅데이터 분석 및 응용」, 국.공립대학정보기관협의회 세미나, 2018

Andrew Trask, 「Secure and Private AI」, Udacity, 2019

요구사항명세서

[Software Requirements Specification]

과제명

프라이버시 보호 딥러닝 서비스 개발

조

사그람 조

지도교수

임성수 교수님 (서명)

조원

201402433 조승현

201402392 이상화

201704144 김수민

201704145 김주희

Table of Contents

1. Introduction	1
1.1. Purpose	1
1.2. Scope	1
1.3. Definitions, acronyms, and abbreviations	2
1.4. References	2
2. External Interface Requirements	3
2.1. 사용자 인터페이스 (User Interface)	3
2.1.1. 하단 메뉴 제공	3
2.1.2. 도움말 제공	4
2.2. 하드웨어 인터페이스 (Hardware Interface)	5
2.3. 소프트웨어 인터페이스 (Software Interface)	5
2.4. 통신 인터페이스 (Communication Interface)	6
3. System Features	7
3.1. 로그인 및 회원가입 (Sign in/ Sign up)	7
3.1.1. 설명 및 우선순위 (Description and Priority)	7
3.1.2. 기능 요구사항 (Functional Requirements)	7
3.2. Pay 연동	7
3.2.1. 설명 및 우선순위 (Description and Priority)	7
3.2.2. 기능 요구사항 (Functional Requirements)	7
3.3. 결제일을 기준으로 카드별 이용금액 분석	8
3.3.1. 설명 및 우선순위 (Description and Priority)	8
3.3.2. 기능 요구사항 (Functional Requirements)	8
3.4. 계좌 연동	8
3.4.1. 설명 및 우선순위 (Description and Priority)	8
3.4.2. 기능 요구사항 (Functional Requirements)	8
3.5. 추가 정보 입력	8
3.5.1. 설명 및 우선순위 (Description and Priority)	9
3.5.2. 기능 요구사항 (Functional Requirements)	9
3.6. 예산 설정	9
3.6.1. 설명 및 우선순위 (Description and Priority)	9
3.6.2. 기능 요구사항 (Functional Requirements)	9
3.7. 맞춤 상품 추천	10
3.7.1. 설명 및 우선순위 (Description and Priority)	10
3.7.2. 기능 요구사항 (Functional Requirements)	10
4. Other Nonfunctional Requirements	11
4.1. 성능 요구 (Performance Requirements)	11
4.2. 안전 요구 (Safety Requirements)	12
4.3. 보안 요구 (Security Requirements)	12
4.4. 소프트웨어 품질 속성 (Software Quality Attributes)	13
5. Other Requirements	14
5.1. H/W 제약 조건	14
5.2. 자원, 인력에 대한 제약 조건	14

2. Introduction

2.1. Purpose

본 문서는 차등 프라이버시를 이용하여 서비스를 이용하는 사용자의 개인정보를 보호하고 기기 내에서 AI 학습이 이루어지는 머신러닝 기법 '연합학습(Federated Learning)' 방식을 적용한 금융 상품 소개 서비스 구현을 위해 요구되는 인터페이스, 기능적, 품질 적, 그리고 그 외의 요구사항들을 기술하고 있다.

2장에서는 인터페이스 요구사항에 대해 기술하고 있으며, 3장에서는 본 프로젝트에서 목표로 하는 시스템 기능 및 기능 요구사항에 대해 설명한다.

4장부터 5장까지는 비 기능적(성능/안전/보안) 요구사항 및 그 외의 요구사항을 각각 기술하고 있다.

본 문서는 제안된 시스템을 사용하는 개발자와 고객을 위해 작성되었다.

2.2. Scope

이 소프트웨어 제품은 사용자로부터 소비내역(정보)를 입력받아 소비패턴을 분석한 후 맞춤형 금융 상품을 추천해 주는 서비스이다. 금융상품에 대한 정보를 찾아보고 비교/분석하는데 걸리는 시간을 줄일 수 있고, 나의 소비패턴을 바탕으로 분석된 내게 적합한 금융상품을 추천받을 수 있다.

본 서비스에서 주목해 볼 만한 것은 차등정보보호(Differential Privacy) 기술을 이용하여 사용자 개인의 데이터를 다른 사용자들의 수많은 데이터와 조합하여 개인정보를 침해하지 않으면서도 통계를 얻어낼 수 있다는 것이다. 이를 통해 사용자들에게 보다 안전한 서비스를 제공할 수 있다.

또한 데이터를 활용한 AI 및 딥러닝을 통해 시간 경과에 따른 사용자의 소비습관을 파악하여 개인 맞춤형 상품을 제공할 수 있다. 이는 사용자 중심 인터페이스로 고객으로부터 받은 입력을 머신러닝 기술로 분석하여 고객이 원하는 다양한 금융서비스를 제공하여 고객의 편의성과 만족도를 향상시킬 수 있다.

기존의 개별 심사자는 한정된 고객 정보에 의존하여 세밀한 금융상품을 추천하는데 어려움이 있었지만 머신러닝을 활용하면 사회초년생, 노인계층과 같은 금융기록이 적은 사람들도 차별화된 금융서비스를 제공받을 수 있게 된다. 사용자에게 양질의 금융서비스를 전달할 수 있으며 기업의 생산성을 향상시키는 등의 다양한 장점을 제공한다.

2.3. Definitions, acronyms, and abbreviations

- **GUI (Graphical User Interface) : 그래픽 사용자 인터페이스**

사용자가 편리하게 사용할 수 있도록 입출력 등의 기능을 알기 쉬운 아이콘 따위의 그래픽으로 나타낸 것

- **연합학습 (Federated Learning)**

모든 데이터를 서버로 모아, 인공지능을 학습하는 기존방식과 달리 사용자가 직접 사용하는 스마트폰에서 데이터를 처리하고 모델을 강화하여 이 모델을 한곳에 모아 더 정교한 모델을 만들어 다시 배포하는 방식

- **차등 정보 보호 (Differential Privacy)**

개인의 데이터를 다른 사람의 수많은 데이터와 조합하여 개인정보를 침해하지 않으면서도 통계를 얻을 수 있는 기술

- **머신 러닝 (Machine Learning) : 기계 학습**

AI(인공지능)의 한 분야로, 컴퓨터가 학습할 수 있도록 하는 알고리즘과 기술을 개발하는 분야

즉, 사용하는 데이터를 기반으로 학습하거나 성능을 개선하는 시스템을 구축하는데 초점이 맞춰져 있는 AI(인공지능)의 하위집합

- **API (Application Programming Interface) : 응용 프로그램 프로그래밍 인터페이스**

응용 프로그램에서 사용할 수 있도록, 운영 체제나 프로그래밍 언어가 제공하는 기능을 제어할 수 있게 만든 인터페이스를 뜻함

2.4. References

- SWIT 소프트웨어산업정보시스템, "공공 SW사업 제안요청서 작성을 위한 요구사항 상세화 실무 가이드라인", <www.swit.or.kr>, (2020.4.29.).
- 나라장터, "소프트웨어사업 요구사항 분석적용 가이드", <www.g2b.go.kr>. (2020.4.29.).
- 소프트웨어자산뱅크, "요구사항 관리 문서 가이드", <<http://swbank.kr/html/pdf/sample/requirements.pdf>>, (2020.4.29.).
- 위키백과, "우리 모두의 백과사전", <ko.wikipedia.org/wiki>, (2020.04.30.).

3. External Interface Requirements

3.1. 사용자 인터페이스 (User Interface)

3.1.1. 하단 메뉴 제공

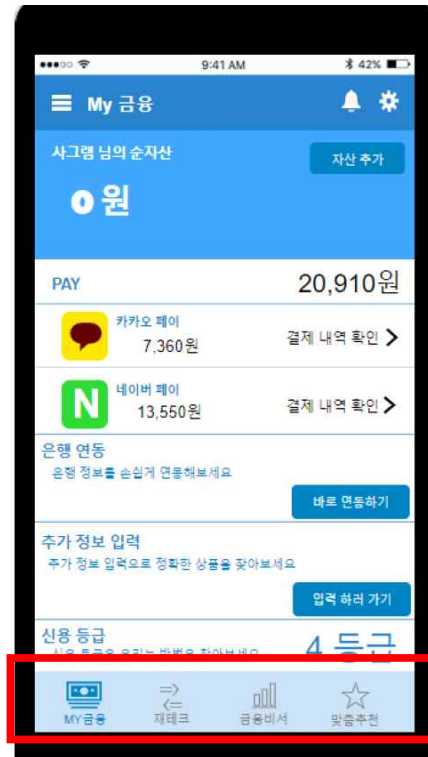


그림 6. 하단 메뉴

요구사항 분류	인터페이스	
요구사항 번호	SFR-001	
요구사항 명칭	하단메뉴	
요구사항 상세설명	정의	하단 메뉴 제공
	세부 내용	<ul style="list-style-type: none"> ○ 사용자가 이용할 수 있는 주요 메뉴를 화면 하단에 배치 (그림 2) - 사용자의 정보를 입력받거나 입력된 기본 정보를 보여주는 메뉴 - 사용자가 원하는 조건(정보)의 상품을 찾을 수 있도록 입력받는 메뉴 - 핵심 기능인 사용자에게 맞춤 상품을 추천해주는 추천 메뉴 - 사용자가 서비스를 이용하는 동안 위의 주요 메뉴를 항상 이용할 수 있도록 해야 함 - 사용자가 현재 이용하고 있는 메뉴를 다른 색으로 표시하여 구분할 수 있도록 함

3.1.2. 도움말 제공

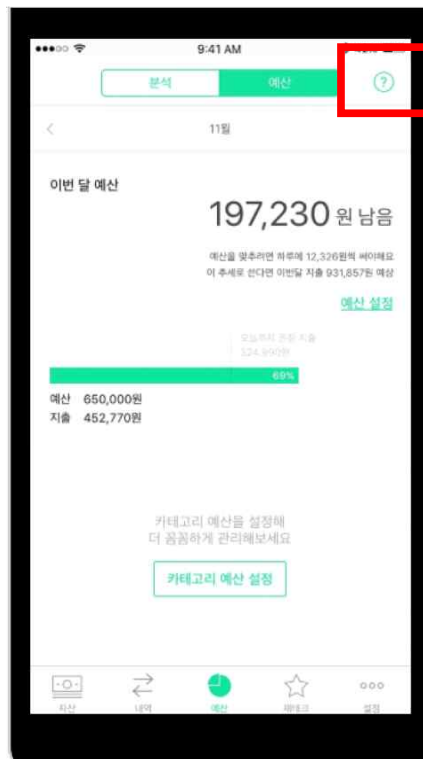


그림 8. 도움말

요구사항 분류		인터페이스
요구사항 번호		SFR-002
요구사항 명칭		도움말
요구사항 상세설명	정의	도움말 제공
	세부 내용	<div>○ 사용자 기능은 도움말을 제공해야 함 (그림 4)</div> <div>- 사용자가 해당 서비스의 기능을 이용할 때 도움말을 참고할 수 있음</div> <div>- 별도의 학습/교육 없이 도움말을 이용하여 사용자가 기능을 이용할 수 있음</div> <div>- 사용자가 기능을 사용했을 시 발생할 수 있는 오류에 대한 해결방법 기능이 제공될 수 있음</div>

3.2. 하드웨어 인터페이스 (Hardware Interface)

요구사항 분류	인터페이스	
요구사항 번호	SFA-001	
요구사항 명칭	하드웨어 인터페이스	
요구사항 상세설명	정의	사용자가 서비스를 이용하기 위한 입력 및 출력
	지원되는 장치 유형	모바일 장치 (Android, iPhone, Android 장치 등)
	입력 주체/ 출력 목적지	사용자의 터치스크린을 통한 입력 / 기기 화면에 출력
	범위/ 정확도/ 허용오차	하드웨어의 스펙을 따름
	시간/속도	비정기적인 사용자의 입력 / 즉각적인 사용자 명령 수행
	단위	사용자의 명령
	타 입출력과 의 관계	클라이언트의 모든 입출력과 관련
	명령 형식	각 코드 값에 따른 명령 매핑

3.3. 소프트웨어 인터페이스 (Software Interface)

요구사항 분류	인터페이스 요구사항	
요구사항 번호	SIR-001	
요구사항 명칭	포털 사이트 계정 연동	
요구사항 상세설명	정의	로그인을 위한 포털 사이트 계정 연동
	세부 내용	<ul style="list-style-type: none"> ○ 사용자의 간편한 회원 가입을 위해서 사용자의 포털 사이트(네이버, 카카오 등) 계정을 연동해야한다. - 포털 사이트의 api를 사용해 사용자의 개인 정보를 가져와 사용자 데이터베이스에 저장한다. - 포털 사이트의 로그인 api와 원활한 연동이 이루어져야함.

요구사항 분류	인터페이스 요구사항	
요구사항 번호	SIR-002	
요구사항 명칭	데이터베이스 관리	
요구사항 상세설명	정의	데이터 관리를 위한 쿼리 입출력
	세부 내용	<ul style="list-style-type: none"> ○ 사용자의 데이터 및 금융 상품 데이터를 관리하고 원하는 금융 상품 데이터를 출력한다. - 사용자가 회원가입을 하면 정보를 데이터베이스에 저장한다. - 사용자가 계좌 혹은 pay를 연동하면 이용 내역 데이터들의 정합성을 체크하고 데이터베이스에 저장한다. - 데이터 이관에 필요한 안정적인 하드디스크 공간 확보

3.4. 통신 인터페이스 (Communication Interface)

요구사항 분류	인터페이스 요구사항	
요구사항 번호	CIR-001	
요구사항 명칭	통신 보안	
요구사항 상세설명	정의	통신 간 보안성 정의
	세부 내용	<ul style="list-style-type: none"> ○ 시스템 간 자료 교환시 기밀성, 무결성, 접근제어 등을 보장하여야 함 - 침입차단서버와 관리자 PC 간 통신 시 구간 암호화, 무결성을 제공하고 관리자 IP를 등록하여 관리자 PC에서만 서버에 접근 가능 - 사용자의 계좌 연동, pay 연동 시 차등 프라이버시 보호 사용 - 연계 서버와 업무서버 영역 간에는 사전 정해진 데이터 형식만 전달하도록 함.

요구사항 분류	인터페이스 요구사항	
요구사항 번호	CIR-002	
요구사항 명칭	통신 인터페이스 요구사항 정의	
요구사항 상세설명	정의	통신 인터페이스 요구사항 정의
	세부 내용	<ul style="list-style-type: none"> ○ 필요한 통신 기능 - 사용자 인증을 위한 SMS 인증번호 전송 기능 - 포털 사이트로 회원 가입 시 해당 포털 사이트 로그인 화면 출력 기능 ○ 동기화 메커니즘 - 실시간으로 계좌 이용 내역을 업데이트 하도록 함

4. System Features

4.1. 로그인 및 회원 가입 (Sign in / Sign up)

4.1.1. 설명 및 우선순위 (Description and Priority)

사용자 어플리케이션 이용을 위해 자신의 정보를 등록한다. 우선순위는 높다.

4.1.2. 기능 요구사항 (Functional Requirements)

요구사항 분류		기능
요구사항 번호		SFB-001
요구사항 명칭		로그인 및 회원가입 관리자 요구사항
요구사항 상세설명	정의	로그인 및 회원가입 과정을 관리할 수 있는 통합 관리자 기능 제공
	세부 내용	<ul style="list-style-type: none">로그인 및 회원가입 과정과 사용자의 데이터를 관리하고 회원 정보 등록 중 문제가 발생할 경우 이를 통제할 수 있는 통합 관리자 도구 구축- GUI 기반 관리도구 제공- 알고리즘 편집, 수정, 삭제 등 기능- 기존 앱의 폼 형식 수정, 삭제 등 기능- 기존 데이터베이스 수정, 삭제 등 기능
산출정보		로그인 및 회원가입 개선 계획
관련 요구사항		

4.2. 계좌 연동

4.2.1. 설명 및 우선순위 (Description and Priority)

사용자 기기에 등록된 카드를 연동하여 계좌 거래 내역을 불러옵니다. 우선순위는 높다.

4.2.2. 기능 요구사항 (Functional Requirements)

요구사항 분류	기능	
요구사항 번호	SFB-004	
요구사항 명칭	계좌 연동 관리자 요구사항	
요구사항 상세설명	정의	거래 내역을 불러올 계좌를 연동 기능
	세부 내용	<div>○ 사용자의 기기에 저장된 카드를 연동하여 거래내역을 데이터베이스에 저장 및 관리하고 문제 발생 시 이를 통제할 수 있는 통합 관리자 구축</div> <div>- GUI 기반 관리도구 제공</div> <div>- 알고리즘의 편집, 수정, 삭제 등 기능</div> <div>- 기존 데이터베이스의 수정, 삭제 등 기능</div>
산출정보		계좌 연동 오류 개선 계획
관련 요구사항		

4.3. 추가 정보 입력 (Enter additional information)

4.3.1. 설명 및 우선순위 (Description and Priority)

사용자가 더 상세한 금융 정보를 얻기 위한 일련의 추가정보를 입력한다. 우선순위는 중간이다.

4.3.2. 기능 요구사항 (Functional Requirements)

요구사항 분류	기능
요구사항 번호	SFB-005
요구사항 명칭	추가 정보 입력에 대한 관리자 요구사항
정의	추가 정보 입력 항목을 관리할 수 있는 통합 관리자 기능 제공
요구사항 상세설명	세부 내용 <ul style="list-style-type: none">○ 사용자의 추가 정보 입력 데이터베이스를 관리하고 정보 입력 시 문제가 발생할 경우 이를 통제할 수 있는 통합 관리자 도구 구축- GUI 기반 관리도구 제공- 알고리즘 편집, 수정, 삭제 등 기능- 기존 앱의 안내 문구의 수정, 삭제 등 기능- 기존 데이터베이스의 수정, 삭제 등 기능
산출정보	추가 정보 입력 오류 개선 계획
관련 요구사항	

4.4. 맞춤 상품 추천 (Suggest customized products)

4.4.1. 설명 및 우선순위 (Description and Priority)

사용자 정보 (소비패턴)을 분석하여 맞춤형 금융 상품을 추천한다. 우선순위는 높다.

4.4.2. 기능 요구사항 (Functional Requirements)

요구사항 분류	기능
요구사항 번호	SFB-007
요구사항 명칭	맞춤 상품 추천 기능 관리자 요구사항
정의	맞춤 상품 추천 항목을 관리할 수 있는 통합 관리자 기능 제공
요구사항 상세설명	세부 내용 <ul style="list-style-type: none">○ 사용자 정보(소비패턴) 분석결과를 기반으로 한 맞춤형 상품 추천 중 문제가 발생할 경우 이를 통제할 수 있는 통합 관리자 도구 구축- GUI 기반 관리도구 제공- 알고리즘 편집, 수정, 삭제의 기능- 사용자 정보 입력/재입력 기능- 시간이 지남에 따라 변하는 사용자의 소비자 패턴을 실시간으로 분석하여 업데이트 된 정보를 제공
산출정보	맞춤 상품 추천 기능 개선 계획
관련 요구사항	

5. Other Nonfunctional Requirements

5.1. 성능 요구 (Performance Requirements)

요구사항 분류	성능
요구사항 번호	PER-001
요구사항 명칭	오류 응답 시간
	정의 오류 응답 시간 목표 정의
요구사항 상세설명	<div>정의</div> <div>세부 내용</div> <ul style="list-style-type: none"> ○ 사용자가 입력한 정보에서 발생할 수 있는 모든 오류에 대한 메시지를 정보 입력 후 3초 이내에 제시하여야 함 * 단, 대용량 파일 또는 대량통계 조회 시는 예외로 함 ○ 오류 메시지는 사용자가 인지하여 즉시 조치할 수 있도록 작성되어야 함 - 연동 기능 등 5초 이상 소요되는 작업은 작업 진행사항 디스플레이(Status Bar 또는 팝업)를 통해 사용자에게 내용을 알려야함 - 사용자가 입력한 데이터 형식의 모든 오류는 사용자가 시스템에 그 정보를 입력한지 1초 이내에 관련 오류 메시지를 사용자에게 제시 함.

요구사항 분류	성능
요구사항 번호	PER-002
요구사항 명칭	어플리케이션 페이지 디스플레이 시간
	정의 어플리케이션 페이지 디스플레이 시간 목표 정의
요구사항 상세설명	<div>정의</div> <div>세부 내용</div> <ul style="list-style-type: none"> ○ 어플리케이션 디스플레이 시간 목표 정의 - 등록, 오류 등 사용자 확인 메시지 제공 시 수 초 내에 완전히 디스플레이 되어야 함. - 어플리케이션의 내용들은 터치 후 4초 내에 완전히 디스플레이 되어야함. * 단, 대용량 파일 또는 대량통계 조회 시는 예외로 함

요구사항 분류	성능
요구사항 번호	PER-003
요구사항 명칭	평균 처리시간 및 동시처리
	정의 평균 처리시간 및 동시처리 사용자
요구사항 상세설명	<div>정의</div> <div>세부 내용</div> <ul style="list-style-type: none"> ○ 평균 처리시간 및 동시처리 사용자 목표 값 - 초당 최소한 100건의 사용자 기본정보 입력기능을 처리 - 시스템은 최대 부하 상태에서 초당 50건의 사용자 기본정보 입력기능을 처리 - 초당 최소한 100건의 사용자 계좌 결제 내역 분석 - 시스템 동시 사용자 500명 이상 접속 시에도 정상 상태 유지해야하며, 최대 동시접속자 수 임계치의 90% 이상 시 서비스 지연 안내 메시지 제공

5.2. 안전 요구 (Safety Requirements)

요구사항 분류	안전
요구사항 번호	SRS-001
요구사항 명칭	안전 요구사항
요구사항 상세설명	<div>정의</div> <div>안전 요구사항 정의</div> <div>세부 내용</div> <ul style="list-style-type: none"> - 개인정보 보호를 위해 접근통제 및 접근권한 제한(관리자 계정 공유금지, 안전한 비밀번호) - 시스템 안전성을 위해 해킹 차단, 개인정보보호 기술(차등 프라이버시 보호)을 적용하여 구축하며, 기존 어플리케이션 속도에 영향이 미치지 않도록 설계 - 데이터가 유출 되었을 경우 기능을 종료하고 사용자에게 사실을 회원 가입 시 입력한 문자로 안내한다. - 데이터 및 장비의 무결성 및 가용성 유지를 위해 백업 정책에 참여하고 사고 발생 시 적시에 복구할 수 있도록 지원
산출정보	
관련 요구사항	

5.3. 보안 요구 (Security Requirements)

요구사항 분류	보안
요구사항 번호	SEC-001
요구사항 명칭	사용자 및 보안 요구사항
요구사항 상세설명	<p>각 사이트의 보안 관리자는 사용자에게 보안 절차에 대해 교육해야 합니다. 보안 관리자는 신입 직원에게 다음 규칙에 대해 전달하고 기존 직원에게 해당 규칙에 대해 정기적으로 상기시켜야 합니다.</p> <ul style="list-style-type: none"> ○ 암호를 아무에게도 말하지 마십시오. 다른 사람이 암호를 알고 있는 경우 책임을 지지 않고 사용자가 액세스할 수 있는 동일한 정보에 몰래 액세스할 수 있습니다. ○ 암호를 기록해 두거나 전자 메일 메시지에 포함시키지 마십시오. ○ 추측하기 어려운 암호를 선택하십시오. ○ 암호를 다른 사람에게 전자 메일로 보내지 마십시오. ○ 화면을 잠그거나 로그오프하지 않고 컴퓨터를 떠나지 마십시오.

요구사항 분류	보안
요구사항 번호	SEC-002
요구사항 명칭	정보 보호
요구사항 상세설명	<p>사용자의 개인 정보 및 중요한 파일을 보호해야 합니다.</p> <ul style="list-style-type: none"> ○ 관리자는 보안이 중요한 파일에 대한 액세스 제어 보호를 올바르게 설정하여 유지 관리해야 할 책임이 있습니다.

요구사항 분류	보안
요구사항 번호	SEC-003
요구사항 명칭	암호 적용
요구사항 상세설명	<p>시스템 관리자 역할은 새 계정을 만들 때 고유한 사용자 이름과 사용자 ID를 지정해야 합니다. 새 계정에 대한 이름과 ID를 선택할 때 사용자 이름과 관련 ID가 네트워크상에서 중복되지 않고 이전에 사용한 적이 없는지 확인해야 합니다.</p> <p>보안 관리자 역할은 각 계정에 대한 원본 암호를 지정하고 새 계정의 사용자에게 암호를 전달할 책임이 있습니다. 암호를 관리할 때 다음 정보를 고려해야 합니다.</p>
	<ul style="list-style-type: none"> ○ 보안 관리자 역할을 맡을 수 있는 사용자에 대한 계정이 잠글 수 없도록 구성되어 있는지 확인합니다. 그러면 모든 다른 계정이 잠겨 있을 때 항상 최소 하나의 계정이 로그인하여 보안 관리자 역할을 맡은 다음 모든 사람의 계정을 다시 열 수 있습니다. ○ 다른 사람이 암호를 도청할 수 없는 방법으로 새 계정의 사용자에게 암호를 전달합니다. ○ 모르는 사람이 암호를 알아냈을 것 같은 의심이 드는 경우 계정 암호를 변경하십시오. ○ 시스템 수명 기간 동안 사용자 이름 또는 사용자 ID를 다시 사용하지 마십시오.

5.4. 소프트웨어 품질 속성 (Software Quality Attributes)

요구사항 분류	소프트웨어 품질 속성
요구사항 번호	SQA-001
요구사항 명칭	장애 대응을 위한 백업 절차 마련
요구사항 상세설명	<p>시스템은 신속한 장애 대응을 위하여 백업 절차를 마련해야 함</p>
	<ul style="list-style-type: none"> ○ 시스템 장애가 발생한 경우, 유지 보수자가 시스템 장애의 원인을 10분 이내에 찾을 수 있어야 한다. ○ 시스템은 신속한 장애 대응을 위하여 백업 절차를 마련해야 한다. ○ 에러 복구, 장애 대책 확보 등 신뢰성 있는 서비스 환경을 제공해야 한다.

요구사항 분류	소프트웨어 품질 속성
요구사항 번호	SQA-002
요구사항 명칭	프로그램 학습성
요구사항 상세설명	프로그램의 설치 및 제거, 이용이 용이해야 한다.
	<ul style="list-style-type: none"> ○ 사용자매뉴얼 또는 관리자매뉴얼에 시스템 또는 프로그램을 설치하거나 제거하기 위한 정보를 문서로 제공해야 함 ○ 시스템 및 프로그램의 설치 및 제거 용이성을 평가하기 위해 매뉴얼을 따라서 사용자 및 관리자가 설치 및 제거해야 함

6. Other Requirements

6.1. H/W 제약 조건

‘사용자의 소비 패턴을 분석해 적절한 금융 상품을 추천하는 금융 어플리케이션’을 실행할 수 있는 모바일 기기 사용

6.2. 자원, 인력에 대한 제약 조건

개발 단계에서 초기 기계 학습을 위한 다량의 데이터 셋이 필요하다.

유스케이스 명세서

[Usecase Specification Document]

과제명	프라이버시 보호 딥러닝 서비스 개발
-----	---------------------

조	사그람 조
지도교수	임성수 교수님 (서명)
조원	201402433 조승현 201402392 이상화 201704144 이수민 201704145 김주희

Table of Contents

1. Introduction	1
1.1. Objective	1
2. Usecase Diagram	2
2.1. 설정 Diagram	2
3. Usecase Specification	3
3.1. 회원 가입	3
3.2. pay 연동	4
3.3 카드별 이용내역 및 자동 결제 내역 분석	5
3.4 계좌 연동	6
3.5 거래 내역	6
3.6 추가 정보 입력	7
3.7 예산 설정	8
3.8 맞춤 추천 및 상품신청	9

1. Introduction

1.1. Objective

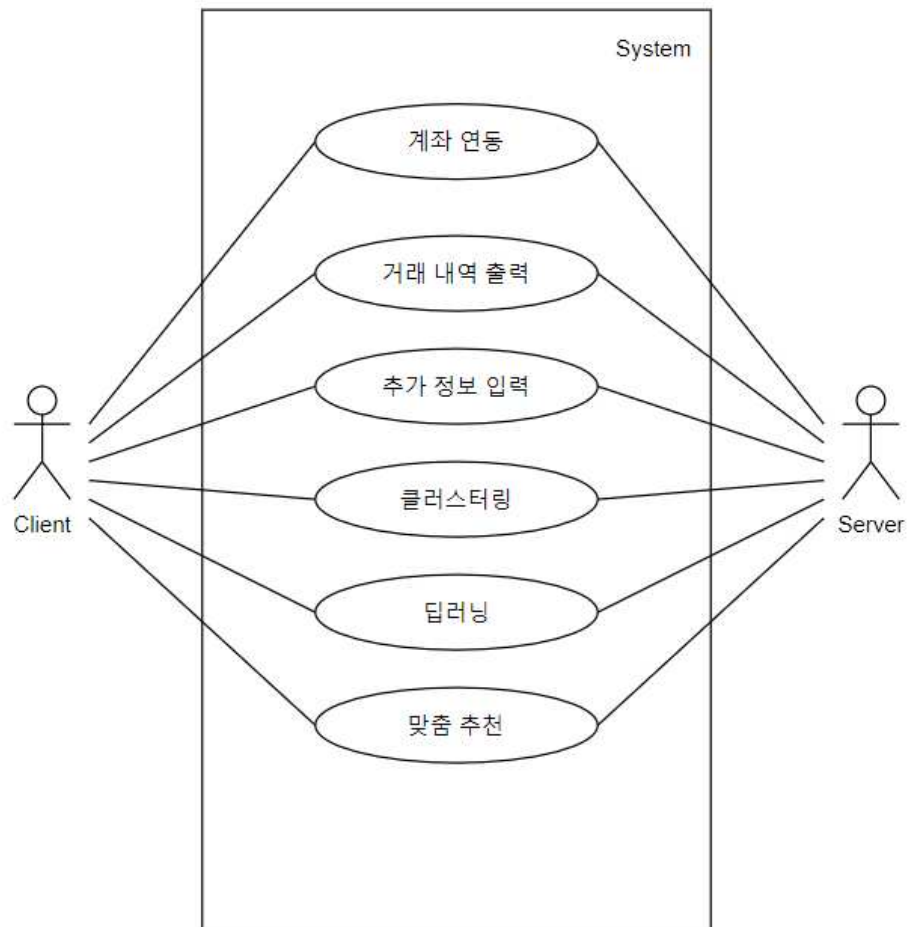
본 주제인 '프라이버시 보호 딥러닝 서비스 개발'의 기술개발 결과가 사용자에게 어떻게 영향을 미치는지 판단하기 위하여 소주제를 '사용자의 소비패턴을 분석하여 적절한 금융상품을 소개하는 서비스 개발'로 지정한다.

이 문서는 금융상품을 소개하는 서비스 개발의 기능을 명세 하고 있다. 요구사항을 상세하게 표현하기 위한 유스케이스 다이어그램과 각 유스케이스에 대한 명세를 포함한다.

2. Usecase Diagram

2.1. 전체 시스템 Diagram

사용자의 소비패턴을 분석하여 적절한 금융상품을 소개하는 서비스에서 기능을 수행하는 전체 시스템에 대한 유스케이스 다이어그램은 다음과 같다.



3. Usecase Specification

3.1. 계좌 연동

Usecase 이름	계좌 연동
ID	1
간략 설명	기기에 등록된 카드의 계좌를 연동한다.
Actor	Client(Initiator), Server
Pre-Conditions	-
Main Flow	1. 사용자가 기기에 등록된 카드를 읽는다. 2) 등록된 카드 목록을 화면에 출력한다. 3) 사용자는 연동할 카드의 연동하기 버튼을 누른다. 4) 연동하기 버튼 클릭 시 해당 카드의 계좌 내역을 불러온다. 5) 연동이 성공했음을 알린다. 6) 다시 연동된 카드를 제외한 카드 목록을 화면에 출력한다.
Post-Conditions	계좌를 연동한다.
Alternative Flow	추가로 카드연동이 필요한 경우 카드목록 출력 이후의 main flow를 반복한다.

3.2. 거래 내역 출력

Usecase 이름	거래 내역 출력
ID	2
간략 설명	연동된 계좌의 거래 내역을 출력한다.
Actor	Client(Initiator), Server
Pre-Conditions	계좌 연동
Main Flow	1. 사용자가 리스트 버튼을 누른다. 2) 서버는 연동한 계좌의 거래 내역을 불러와 당월의 거래 내역을 화면에 출력한다. 3) 사용자가 월을 클릭하고 선택한다. 4) 선택한 월의 거래 내역을 화면에 출력한다. 5) 사용자가 거래 내역의 스크롤을 위아래로 조정한다. 6) 날짜 별 거래 내역을 출력한다.
Post-Conditions	거래 내역이 서버에 저장된다.
Alternative Flow	

3.3. 추가 정보 입력

Usecase 이름	추가 정보 입력
ID	3
간략 설명	사용자가 더 상세한 금융 정보를 얻기 위한 일련의 추가정보 입력 절차에 대하여 명세한다.
Actor	Client(Initiator), Server
Pre-Conditions	-
Main Flow	<ol style="list-style-type: none"> 1) 사용자는 추가정보 입력란에서 입력하러 가기 버튼을 누른다. 2) 사용자는 나이와 직업(객관 정보 1, 2)을 입력한다. 3) 사용자는 원하는 상품의 종류(객관 정보 3)를 입력한다. 4) 사용자는 다음 버튼을 누른다. 5) 서버는 사용자의 상품종류를 추천하기 위해 필요한 추가정보를 요구한다. 6) 사용자는 원하는 한 달 예산, 희망 이율, 희망 기간, 희망은행(주관 정보 1, 2, 3, 4)을 입력한다. 7) 사용자는 완료 버튼을 누르고 서버에 상품 탐색을 요청한다. 8) 서버는 사용자를 확인하고, 사용자의 입력한 추가정보를 등록한다. 9) 서버는 금융 상품 추천 결과를 보여준다.
Post-Conditions	- 사용자의 추가적인 정보가 시스템에 등록된다.
Alternative Flow	<ol style="list-style-type: none"> 2-1) 사용자가 이미 회원가입시에 나이와 직업을 입력하였을 시 추가적으로 입력할 필요 없이 이미 체크표시가 되어있다. 4-1) 사용자가 아무런 입력을 하지 않고 완료 버튼을 눌렀다. 필요한 추가정보의 항목은 늘어나지만 아무런 알림 없이 다음을 계속 수행한다. (이 경우 정보에 대한 필터링 기능이 진행되지 않는다.) 7-1) 사용자가 아무런 입력을 하지 않고 완료 버튼을 눌렀다. 사용자에게 이름이 입력되지 않았음을 알리고, 다시 입력을 요청한다. (이 경우 정보에 대한 필터링 기능이 진행되지 않는다.)

3.4. 예산 설정

Usecase 이름	예산설정
ID	4
간략 설명	한 달 예산설정 목표금액 확인에 사용하기 위한 예산설정 절차에 대해 명세한다.
Actor	Client(Initiator), Server
Pre-Conditions	-
Main Flow	<ol style="list-style-type: none"> 1) 사용자는 재테크란의 예산 버튼을 누른다. 2) 사용자는 한 달 예산 금액을 입력한다. 3) 사용자는 한 달 예산 설정 버튼을 누른다. 4) 서버는 한 달 예산 설정한 금액과 가계부에 이번 달 소비된 금액을 계산하여 사용자에게 결과 값을 전송한다. 5) 사용자는 카테고리 예산 설정 버튼을 누른다. 6) 사용자는 한달 예산 금액 만큼 각 카테고리에 할당된 금액을 원하는 만큼 할당시키고 완료 버튼을 누른다. 7) 서버는 카테고리 별 사용 금액을 계산하여 사용자에게 결과 값을 전송한다. 8) 사용자는 추가로 소비시 자동이나 수동으로 재테크의 가계부에 소비 금액을 입력한다. 9) 서버는 소비 금액이 입력 될 때마다 자동으로 예산의 카테고리 별 항목을 업데이트 해준다.
Post-Conditions	- 사용자는 목표 금액에 도달하기 위한 계산된 값을 확인 할 수 있다.
Alternative Flow	<ol style="list-style-type: none"> 3-1) 사용자가 아무런 입력을 하지 않고 한 달 예산 설정 버튼을 눌렀다. 사용자에게 한 달 예산이 입력되지 않았음을 알리고, 다시 입력을 요청한다. 6-1) 사용자가 할당된 금액을 다 소비하지 않고 완료 버튼을 눌렀다. 사용자에게 이름이 입력되지 않았음을 알리지만, 추가적인 입력이 입력이 더 없다는 확인 버튼을 누른 후 다음을 진행한다. 8-1) 사용자가 이미 등록된 가계부에 있는 항목의 삭제를 원한다. 그 경우 서버는 삭제된 항목에 대한 계산을 진행한다.

3.5. 클러스터링

Usecase 이름	클러스터링
ID	5
간략 설명	사용자 정보 data set을 클러스터링하는 과정을 명세한다.
Actor	Client(Initiator), Server
Pre-Conditions	-
Main Flow	1) 서버는 사용자 정보 data set을 가져와서 클러스터링을 수행한다. 2) 서버는 클러스터링을 수행한 data set을 데이터베이스에 저장한다.
Post-Conditions	- 사용자는 사용자와 비슷한 집단의 추천 금융 상품을 추천받을 수 있다.
Alternative Flow	1-1) data set을 가져오는데 실패할 경우 사용자에게 실패했다는 것을 알려주고 다시 시도를 요청한다. 2-1) 데이터베이스에 저장을 실패한다면 사용자에게 실패를 알리고 다시 시도한다.

3.6. 딥러닝 맞춤추천

Usecase 이름	딥러닝 맞춤추천
ID	6
간략 설명	사용자의 데이터(소비내역)를 학습하여 맞춤형 금융 상품을 추천해주는 절차에 대하여 명세한다.
Actor	Client(Initiator), Server
Pre-Conditions	-
Main Flow	<ol style="list-style-type: none">1) 사용자는 맞춤 추천 카테고리의 '나의 맞춤형 상품 확인하기' 버튼을 누른다.2) 서버는 기존 사용자의 데이터를 가져와서 학습을 수행한다.3) 학습된 모델에 대한 성능평가를 수행한다.4) 서버는 사용자의 데이터를 가져와 딥러닝 학습 output을 산출하여 사용자에게 적절한 추천상품을 제시한다.5) 사용자는 추천상품 목록을 확인한다.
Post-Conditions	- 사용자의 데이터를 바탕으로 맞춤 금융 상품을 추천받을 수 있다.
Alternative Flow	<ol style="list-style-type: none">4-1) 서버는 사용자의 데이터가 업데이트될 때마다 사용자에게 업데이트된 학습 output을 산출한다.

클래스 다이어그램

1. Introduction

1.1. Objective

이 문서는 본 주제의 '프라이버시 보호 딥러닝 서비스 개발'에서 소주제인 '사용자의 소비패턴을 분석하여 적절한 금융상품을 소개하는 서비스 개발'이라는 주제의 시스템 모델(클래스 다이어그램)에 대한 내용을 기술하고 있다.

시스템 차원의 클래스 다이어그램과 각 클래스에 대한 명세를 포함한다.

2. Class Diagram

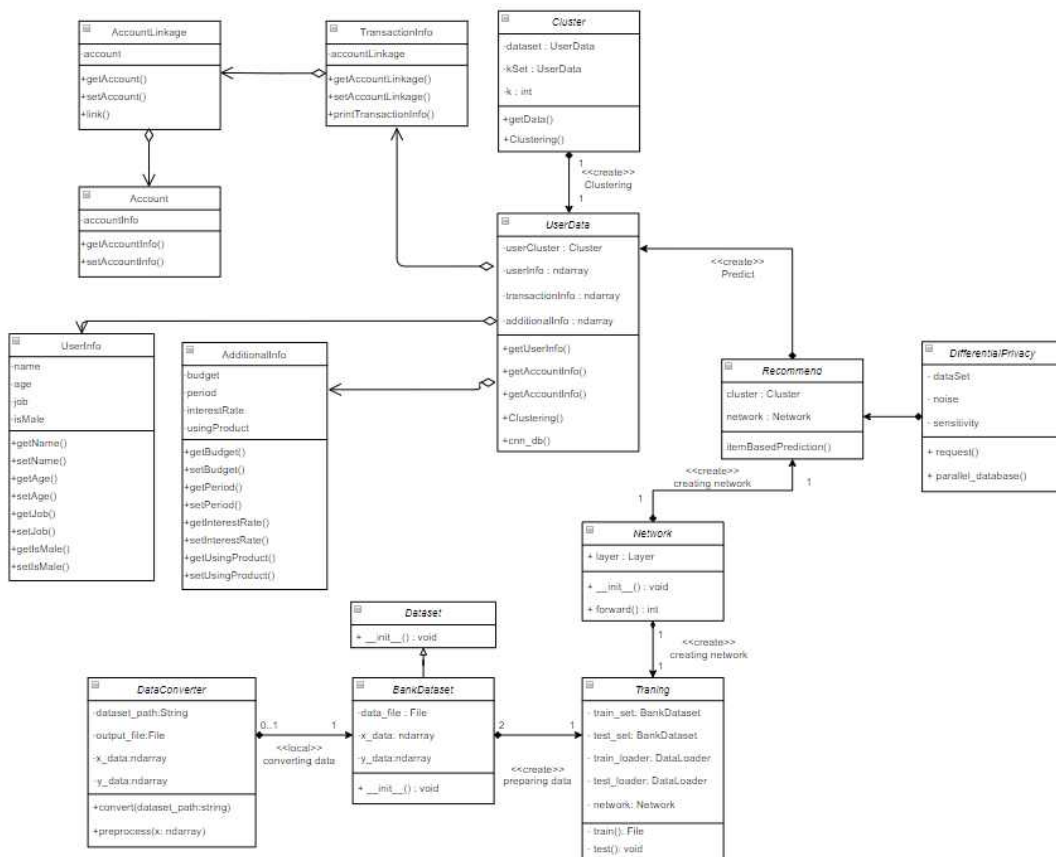
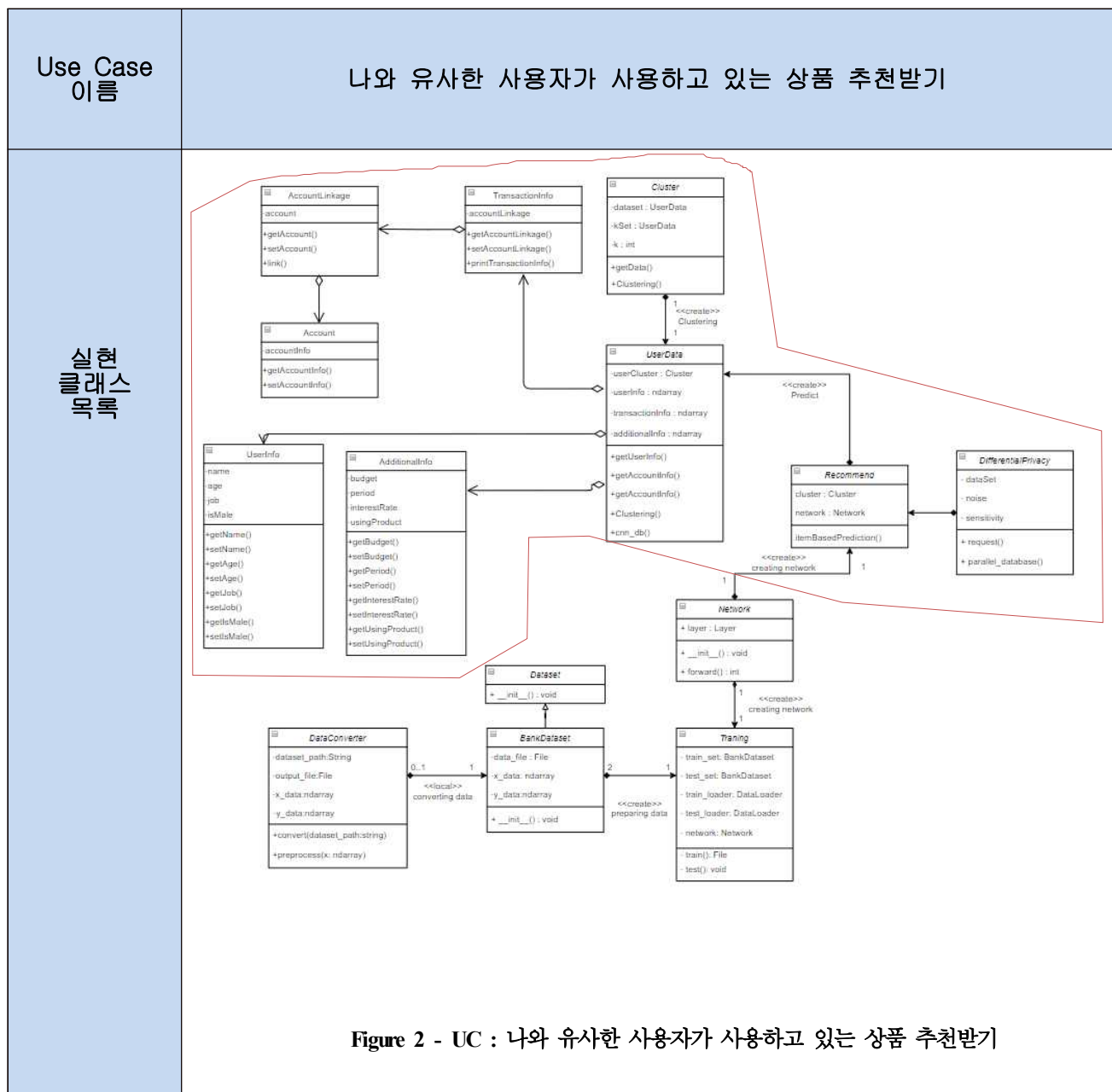


Figure 1 – System Class Diagram

3. Use Case와 Class 간의 관계

2.1. UC: 나와 유사한 사용자가 사용하고 있는 상품 추천받기



2.2. UC: 딥러닝 서비스로 금융상품 추천받기

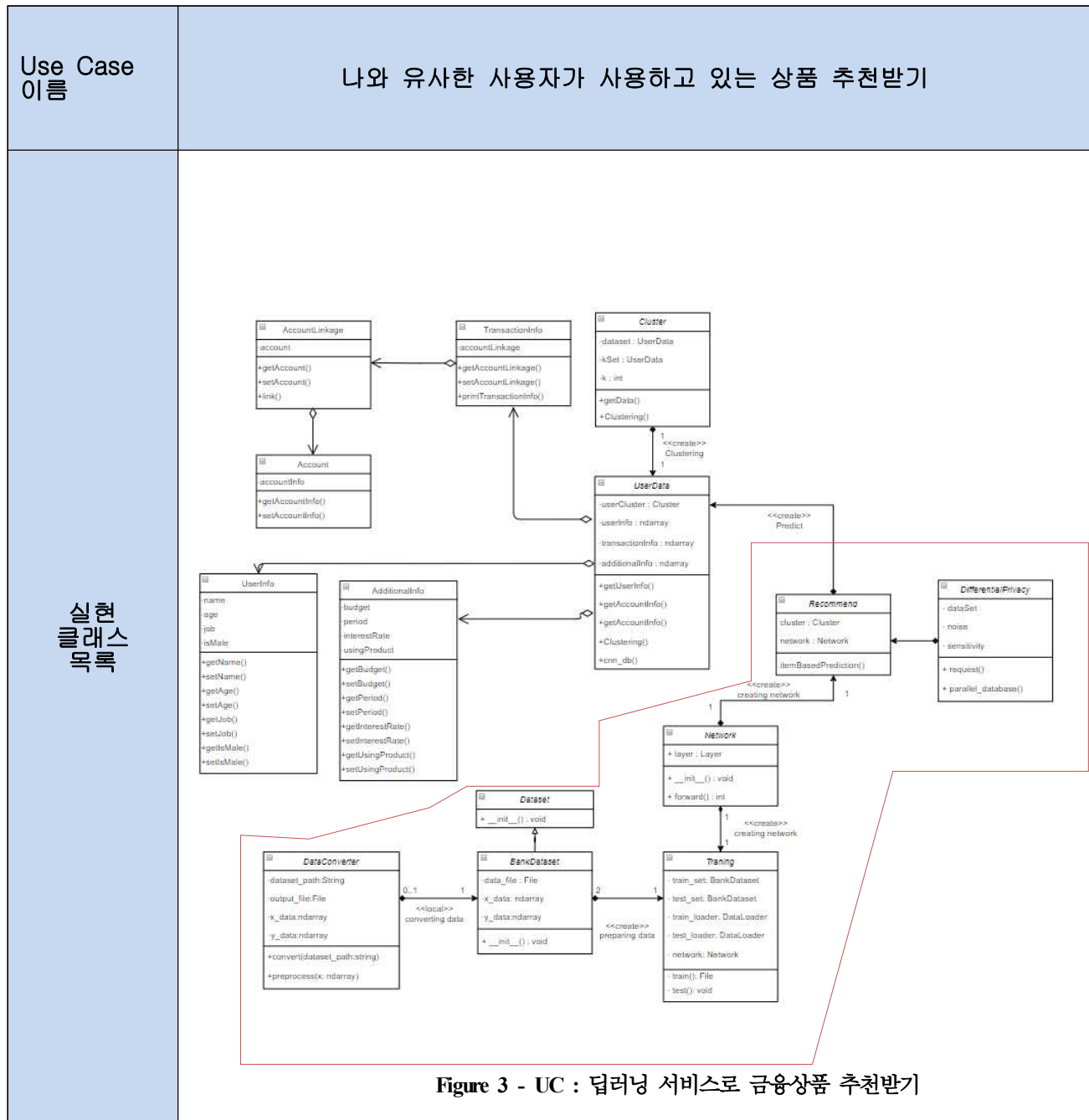
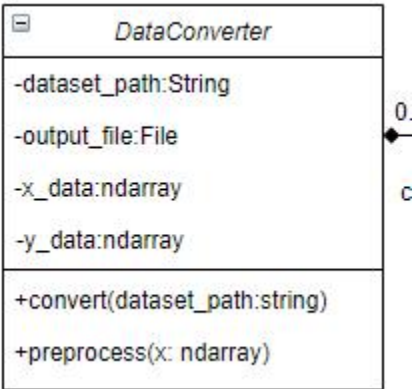
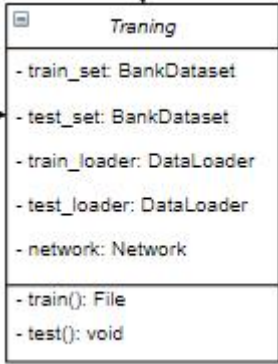


Figure 3 - UC : 딥러닝 서비스로 금융상품 추천받기

4. Class 명세

Dataset				
Class Diagram	<div><div><div><div><div></div><div></div></div><div><div><div></div><div>Dataset</div></div></div><div><div><div>+ __init__() : void</div></div></div></div><div><div></div><div><div></div><div></div></div></div></div></div> <div>Figure 4 - Dataset 클래스</div>			
	일반화된 Dataset class 를 의미한다.			
Responsibility				
Attribute	Type	Name	Description	
Operation				
	Return Type	Method Name	Parameter Type	Parameter Name
	void	__init__()		
	Description	Dataset Class 의 초기화 메소드이다.		

DataConverter				
Class Diagram				
Responsibility	BankDataset 을 필요한 변수를 남기고 원하는 형식으로 변환하는 과정, 전처리하는 과정을 수행하는 class 입니다.			
Attribute	Type	Name	Description	
	String	dataset_path	dataset 의 경로	
	UserData	kSet	변환한 dataset 을 파일로 저장	
	ndarray	x_data	데이터 셋의 독립변수 값	
	ndarray	y_data	데이터 셋의 종속변수 값	
Operation	Return Type	Method Name	Parameter Type	Parameter Name
	void	Converter	string	dataset_path
	Description	데이터를 필요한 데이터로 변환하기 위한 함수입니다.		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	preprocess	ndarray	
	Description	dataset 을 전 처리(preprocess)하는 함수입니다.		

Training				
Class Diagram	 <pre> classDiagram class Training { -train_set: BankDataset -test_set: BankDataset -train_loader: DataLoader -test_loader: DataLoader -network: Network -train(): File -test(): void } </pre>			
	Figure 7 - Training 클래스			
Responsibility	적절한 형식으로 변환된 Dataset 을 train 과 test set 으로 분리하고 신경망 학습을 하고 테스트 하는 과정을 진행하는 class 입니다.			
Attribute	Type	Name	Description	
	BankDataset	train_set	신경망 학습을 위한 dataset	
	BankDataset	test_set	신경망 테스트를 위한 dataset	
	DataLoader	train_loader	train dataset 을 load 하는 변수	
	DataLoader	test_loader	test dataset 을 load 하는 변수	
	Network	network	신경망에 대한 변수	
Operation	Return Type	Method Name	Parameter Type	Parameter Name
	File	train		
	Description	신경망이 훈련을 진행하고 그 결과값이 File Type 으로 나오게 됩니다.		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	test		
	Description	신경망에 대한 성능을 평가하기 위하여 test 하는 함수 입니다.		

TransactionInfo															
Class Diagram	<div><div><div><div><div></div><div>TransactionInfo</div></div><div><div></div><div></div><div></div></div><div><div>-accountLinkage</div></div><div><div>+getAccountLinkage()</div><div>+setAccountLinkage()</div><div>+printTransactionInfo()</div></div></div></div></div>														
	Figure 8 - TransactionInfo 클래스														
Responsibility	계좌 연동을 이용한 거래 내역 기능을 구현한 클래스 입니다.														
Attribute	<table><tr><th>Type</th><th>Name</th><th>Description</th></tr><tr><td>AccountLinkage</td><td>AccountLinkage</td><td>계좌 연동 객체</td></tr></table>	Type	Name	Description	AccountLinkage	AccountLinkage	계좌 연동 객체								
Type	Name	Description													
AccountLinkage	AccountLinkage	계좌 연동 객체													
Operation	<table><tr><th>Return Type</th><th>Method Name</th><th>Parameter Type</th><th>Parameter Name</th></tr><tr><td>void</td><td>printTransactionInfo</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">거래 내역을 출력하는 메소드</td></tr></table>			Return Type	Method Name	Parameter Type	Parameter Name	void	printTransactionInfo			Description	거래 내역을 출력하는 메소드		
	Return Type	Method Name	Parameter Type	Parameter Name											
	void	printTransactionInfo													
	Description	거래 내역을 출력하는 메소드													
	<table><tr><th>Return Type</th><th>Method Name</th><th>Parameter Type</th><th>Parameter Name</th></tr><tr><td>AccountLinkage</td><td>getAccountLinkage</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">계좌 연동 객체의 접근자</td></tr></table>			Return Type	Method Name	Parameter Type	Parameter Name	AccountLinkage	getAccountLinkage			Description	계좌 연동 객체의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name											
	AccountLinkage	getAccountLinkage													
	Description	계좌 연동 객체의 접근자													
	<table><tr><th>Return Type</th><th>Method Name</th><th>Parameter Type</th><th>Parameter Name</th></tr><tr><td>void</td><td>setAccountLinkage</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">계좌 연동 객체의 설정자</td></tr></table>			Return Type	Method Name	Parameter Type	Parameter Name	void	setAccountLinkage			Description	계좌 연동 객체의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name											
	void	setAccountLinkage													
	Description	계좌 연동 객체의 설정자													

AccountLinkage			
Class Diagram	<pre> classDiagram class AccountLinkage { -account +getAccount() +setAccount() +link() } </pre>		
Responsibility	계좌 연동 기능을 수행하는 클래스		
Attribute	Type	Name	Description
	Account	account	계좌 객체
Operation	Return Type	Method Name	Parameter Type
	Account	getAccount	
	Description	계좌 객체의 접근자	
	Return Type	Method Name	Parameter Type
	void	setAccount	
	Description	계좌 객체의 설정자	
	Return Type	Method Name	Parameter Type
	Account	link	
	Description	계좌를 연동하는 메소드	

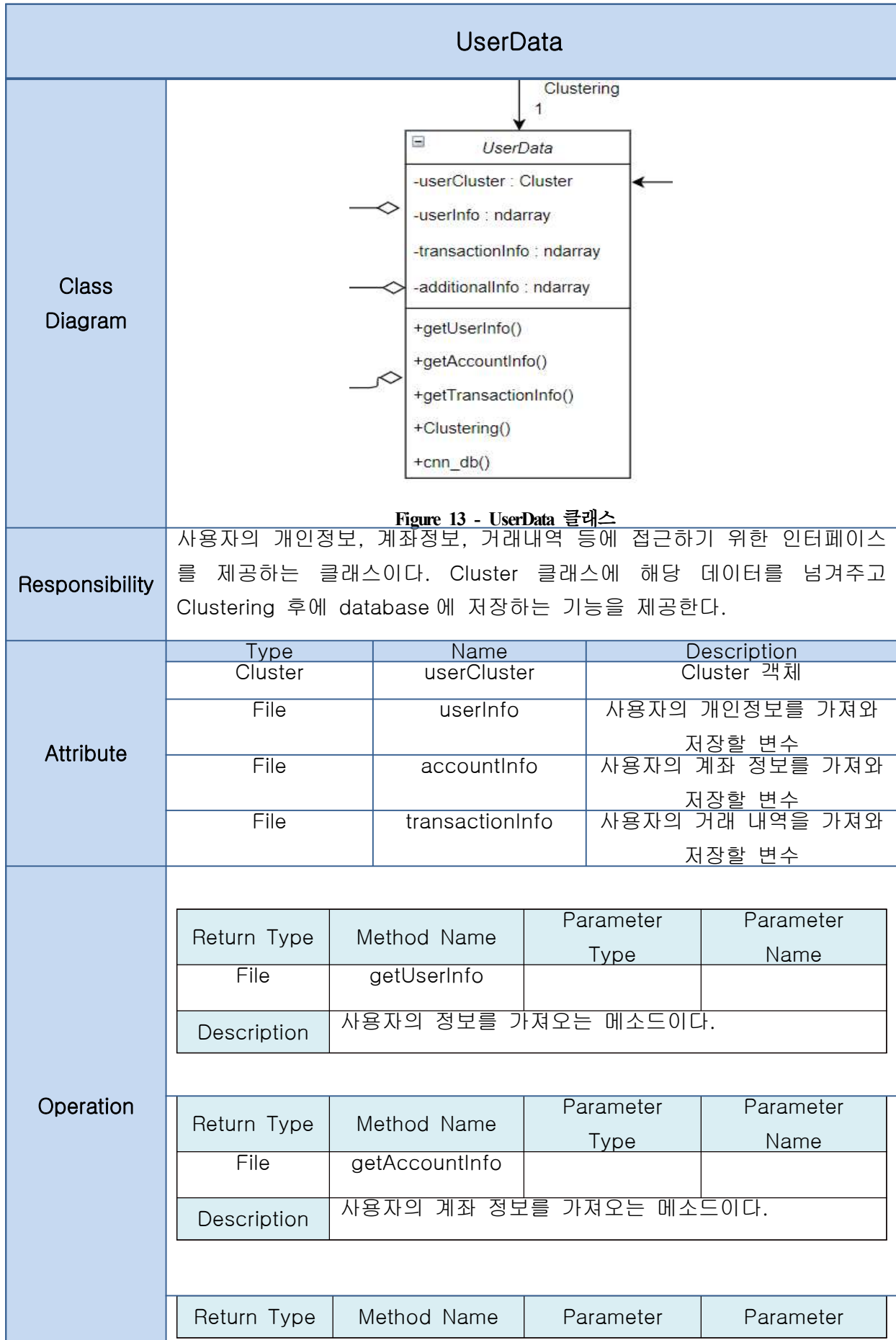
Account			
Class Diagram	<pre> classDiagram class Account { -accountInfo +getAccountInfo() +setAccountInfo() } </pre>		
Responsibility	계좌를 나타내는 클래스		
Attribute	Type	Name	Description
	File	accountInfo	계좌 내역을 담는 변수
Operation	Return Type	Method Name	Parameter Type
	File	getAccountInfo	Parameter Name
	Description	계좌 내역의 접근자	
	Return Type	Method Name	Parameter Type
	void	setAccountInfo	Parameter Name
	Description	계좌 내역의 설정자	

UserInfo															
Class Diagram	<div><div><div>UserInfo</div><div><div>-name</div><div>-age</div><div>-job</div><div>-isMale</div></div><div><div>+getName()</div><div>+setName()</div><div>+getAge()</div><div>+setAge()</div><div>+getJob()</div><div>+setJob()</div><div>+getIsMale()</div><div>+setIsMale()</div></div></div></div>														
	그림 11 - UserInfo 클래스														
	Responsibility														
	사용자 정보를 나타내는 클래스														
	Attribute	Type	Name	Description											
String		name	사용자 이름												
int		age	사용자 나이												
String		job	사용자 직업												
boolean		isMale	사용자 성별												
Operation	<table><tr><td>Return Type</td><td>Method Name</td><td>Parameter Type</td><td>Parameter Name</td></tr><tr><td>String</td><td>getName</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">Name 의 접근자</td></tr></table>			Return Type	Method Name	Parameter Type	Parameter Name	String	getName			Description	Name 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name											
	String	getName													
	Description	Name 의 접근자													
	<table><tr><td>Return Type</td><td>Method Name</td><td>Parameter Type</td><td>Parameter Name</td></tr><tr><td>void</td><td>setName</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">Name 의 설정자</td></tr></table>			Return Type	Method Name	Parameter Type	Parameter Name	void	setName			Description	Name 의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name											
	void	setName													
	Description	Name 의 설정자													
	<table><tr><td>Return Type</td><td>Method Name</td><td>Parameter Type</td><td>Parameter Name</td></tr><tr><td>int</td><td>getAge</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">Age 의 접근자</td></tr></table>			Return Type	Method Name	Parameter Type	Parameter Name	int	getAge			Description	Age 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name											
	int	getAge													
	Description	Age 의 접근자													

	Return Type	Method Name	Parameter Type	Parameter Name
	void	setAge		
	Description	Age 의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name
	String	getJob		
	Description	job 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setJob		
	Description	job 의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name
	boolean	getIsMale		
	Description	isMale 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setIsMale		
	Description	isMale 의 설정자		

AdditionalInfo				
Class Diagram	<div><div><div><div><div><div></div><div>AdditionalInfo</div></div></div><div><div><div><div>-budget</div><div>-period</div><div>-interestRate</div><div>-usingProduct</div></div></div><div><div><div><div>+getBudget()</div><div>+setBudget()</div><div>+getPeriod()</div><div>+setPeriod()</div><div>+getInterestRate()</div><div>+setInterestRate()</div><div>+getUsingProduct()</div><div>+setUsingProduct()</div></div></div></div></div></div></div></div>			
	Figure 12 - AdditionalInfo 클래스			
Responsibility	금융 상품 추천을 위한 추가 정보를 나타내는 클래스			
Attribute	Type	Name	Description	
	int	budget	한달 예산	
	int	period	희망 기간	
	double	interestRate	희망 이율	
	string	usingProduct	사용중인 상품	
Operation	Return Type	Method Name	Parameter Type	Parameter Name
	int	getBudget		
	Description	budget 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setBudget		
	Description	Budget 의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name
	int	getPeriod		
	Description	period 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setPeriod		
	Description	Period 의 설정자		

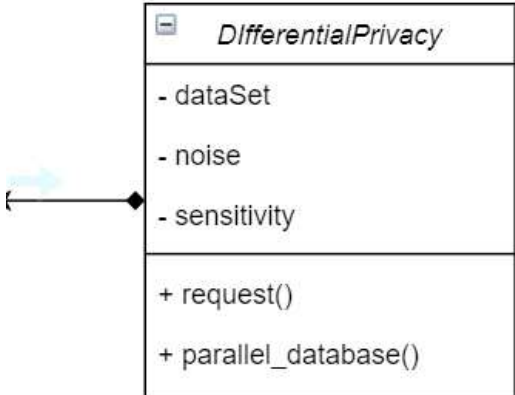
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setPeriod		
	Description	Period 의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name
	double	getInterestRate		
	Description	interestRate 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setInterestRate		
	Description	interestRate 의 설정자		
	Return Type	Method Name	Parameter Type	Parameter Name
	String	getUsingProduct		
	Description	UsingProduct 의 접근자		
	Return Type	Method Name	Parameter Type	Parameter Name
	void	setUsingProduct		
	Description	usingProduct 의 설정자		



		Type	Name
File	getTransactionInfo		
Description	사용자의 계좌 내역을 가져오는 메소드이다.		

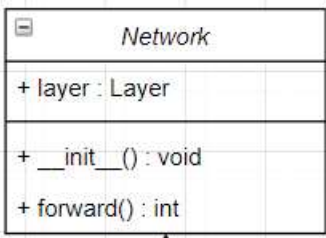
Return Type	Method Name	Parameter Type	Parameter Name
File	Clustering	File, File, File	user, account, transaction
Description	가져온 데이터들을 Clustering 하기 위해 객체 접근하는 메소드이다.		

Return Type	Method Name	Parameter Type	Parameter Name
void	cnn_db	File	dataSet
Description	Clustering 한 데이터를 database 에 저장하는 메소드이다.		

DifferentialPrivacy				
Class Diagram				
	Figure 14 - DifferentialPrivacy 클래스			
Responsibility	필요한 사용자의 정보를 보호하기 위해 차등 프라이버시 보호 기능을 구현한 class 이다.			
Attribute	Type	Name	Description	
	File	userCluster dataSet	Cluster 객체 database 에서 가져온 사용자 정보	
	list	noise	사용자의 정보 보호를 위해 추가할 값	
	int	sensitivity	민감한 정도를 표현	
Operation	Return Type	Method Name	Parameter Type	Parameter Name
	File	request		
	Description	사용자의 요청사항을 구현한 메소드이다.		
	Return Type	Method Name	Parameter Type	Parameter Name
	File	parallel_database		
	Description	사용자의 요청의 결과값을 구하는 과정에서 특정 noise 를 추가해 어플리케이션 사용자들의 개인정보의 유출을 없애는 함수이다. 다수의 처리장치를 사용하여 데이터베이스 관리를 고속으로 수행한다.		

Cluster			
Class Diagram	<pre> classDiagram class Cluster { -dataset : UserData -kSet : UserData -k : int +getData() +Clustering() } </pre>		
Responsibility	사용자 데이터를 기반으로 고객 특성 벡터를 반영하여 k-means 알고리즘을 이용하여 클러스터링을 수행하는 클래스이다.		
Attribute	Type	Name	Description
	UserData	dataset	사용자 데이터
	UserData	kSet	군집화 데이터
	int	k	클러스터 개수
Operation	Return Type	Method Name	Parameter Type
	UserData	getData()	
	Description	clustering 을 위한 사용자의 데이터를 가져오는 함수	
	Return Type	Method Name	Parameter Type
	void	clustering()	
	Description	사용자의 데이터에 대해 클러스터링 기법을 적용하여 유사도를 기준으로 동질 유형을 분류해주는 함수	

Recommend															
Class Diagram	<div></div> <p>Figure 17 - Recommend 클래스</p>														
Responsibility	클러스터링(군집화)된 데이터 또는 학습된 신경망 데이터로 추론처리를 통해 예측한 결과를 제공하는 클래스이다.														
Attribute	<table><tr><th>Type</th><th>Name</th><th>Description</th></tr><tr><td>Cluster</td><td>cluster</td><td>Cluster 객체</td></tr><tr><td>Network</td><td>network</td><td>Network 객체</td></tr></table>	Type	Name	Description	Cluster	cluster	Cluster 객체	Network	network	Network 객체					
Type	Name	Description													
Cluster	cluster	Cluster 객체													
Network	network	Network 객체													
Operation	<table><tr><th>Return Type</th><th>Method Name</th><th>Parameter type</th><th>Parameter name</th></tr><tr><td>void</td><td>itemBasedPrediction()</td><td></td><td></td></tr><tr><td>Description</td><td colspan="3">clustering(군집화)된 데이터와 (train)학습된 데이터로부터 최적의 상품을 예측하여 추천해주는 함수</td></tr></table>			Return Type	Method Name	Parameter type	Parameter name	void	itemBasedPrediction()			Description	clustering(군집화)된 데이터와 (train)학습된 데이터로부터 최적의 상품을 예측하여 추천해주는 함수		
Return Type	Method Name	Parameter type	Parameter name												
void	itemBasedPrediction()														
Description	clustering(군집화)된 데이터와 (train)학습된 데이터로부터 최적의 상품을 예측하여 추천해주는 함수														

Network				
Class Diagram	<div></div> <p>Figure 16 - Network 클래스</p>			
	Responsibility	계층 관점 신경망의 다양한 계층들을 조합하여 신경망을 구축하는 클래스		
Attribute	Type	Name	Description	
	Layer	layer	계층을 생성하여 모아두는 인스턴스 변수	
Operation	Return Type	Method Name	Parameter Type	Parameter Name
	void	__init__()		
	Description	초기화 메소드 가중치를 초기화하고 계층을 생성한다.		
	Return Type	Method Name	Parameter Type	Parameter Name
	layer	forward()		
	Description	입력층(input layer)로 데이터가 입력되고, 1 개 이상으로 구성되는 은닉층(hidden layer)을 거쳐서 마지막에 있는 출력층(output layer)으로 출력값을 내보내는 함수이다.		

시퀀스 다이어그램

1. Introduction

1.1 Objective

이 문서는 '프라이버시 보호 딥러닝 서비스 개발' 시스템의 시스템 모델(시퀀스 다이어그램)에 대한 내용을 기술하고 있다. 요구사항 명세 단계에서 작성한 유스케이스 다이어그램을 기반으로 각 유스케이스의 상세한 내부 동작 흐름을 시퀀스 다이어그램으로 모델링한다.

2. Use Case Diagram

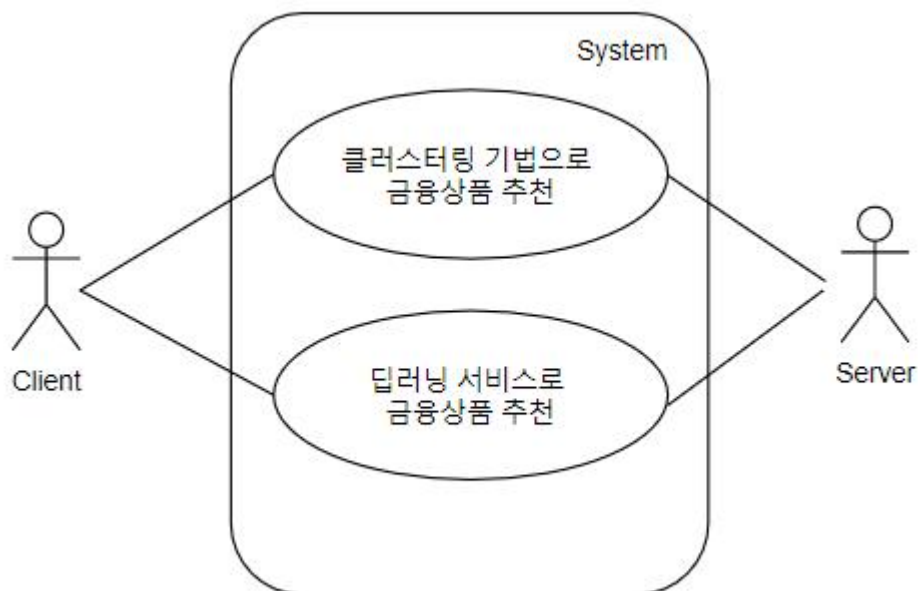


Figure 1 – Use Case Diagram

3. Sequence Diagram

3.1 DPAG_REQ_Recommending_N001 (RecommendingDPdata)

DPAG_REQ_Recommending_N001은 사용자로부터 정보를 입력받고 연동을 통해 제공받은 정보들을 Database에 저장하며 저장된 data들을 clustering해 각 cluster의 사용자들에게 적절한 금융상품을 추천하는 서비스를 제공한다. 차등 프라이버시 보호 알고리즘을 사용해 사용자의 데이터를 안전하게 보호한다.

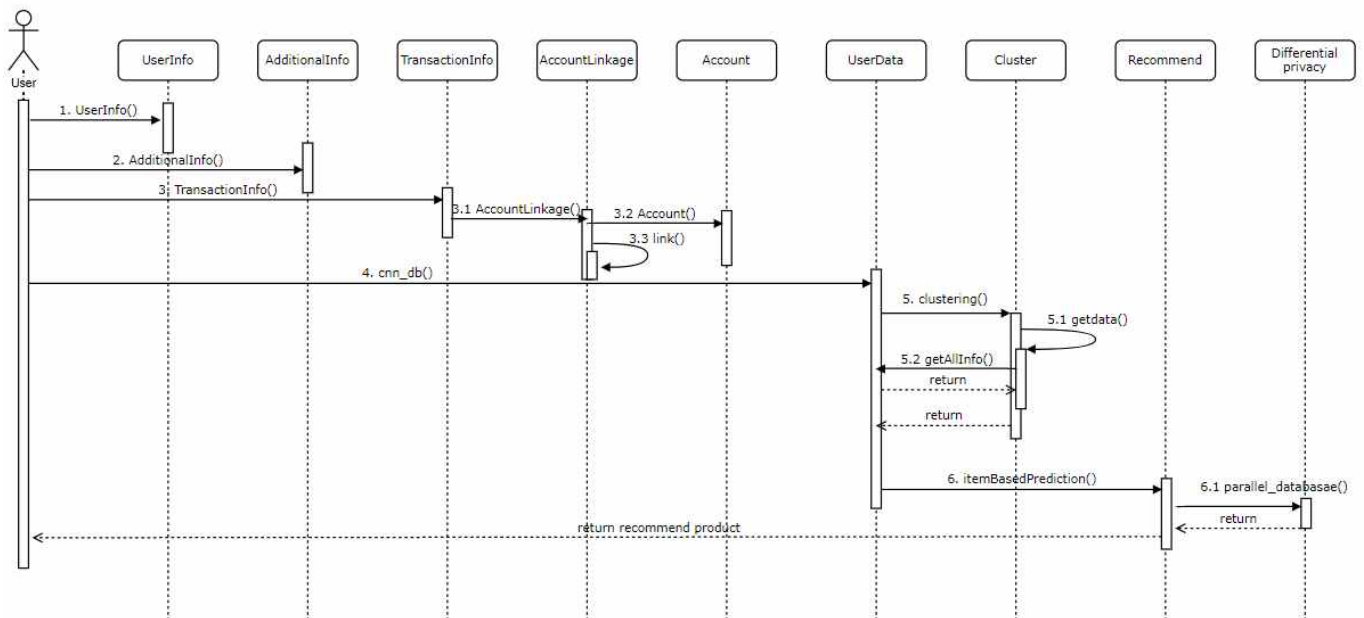


Figure 2 - RecommendingDPdata with Clustering Sequence Diagram

1. 사용자는 사용자 정보를 입력하여 userInfo객체를 생성한다.
2. 사용자는 서비스에 필요한 추가 정보를 입력하여 additionalInfo객체를 생성한다.
3. 사용자가 계좌 연동 요청시 TransactionInfo 객체를 생성하여 계좌 연동을 하고 거래 내역 정보를 불러온다.
 - 3.1: 계좌 연동 객체 생성
 - 3.2: 계좌 연동을 위한 계좌 객체 생성
 - 3.3: 계좌 연동 메소드 실행
4. cnn_db함수를 호출하여 transactionInfo 객체를 클러스터링한다. Database에 정보를 저장한다.
5. DB에 저장된 data들을 클러스터링한 결과를 제공한다.
 - 5.1: data를 가져오기 위해 User data의 getAllInfo를 사용한다.
 - 5.1.1: database에서 clustering할 data들을 가져온다.
6. Recommend의 itemBasedPrediction로 사용자에게 적절한 금융 상품 결과를 제공한다.

6.1: Differential Privacy의 parallel_database로 차등 프라이버시 보호 알고리즘을 수행해 응답 데이터에 noise를 추가한다.

3.2 DPAG_REQ_Recommending_N002 (RecommendingDPdata)

DPAG_REQ_Recommending_N002은 사용자에게 맞춤 추천 서비스를 제공하기 위하여 사용자의 기존 데이터를 바탕으로 딥러닝 학습을 진행하며 성능평가 후 적절한 임계값을 넘는 서비스를 제공하려하며 후에 추천 서비스를 받기 위하여 데이터를 가져와서 딥러닝 서비스로 추천받는 과정이 진행된다. 여기서 사용되는 데이터는 차등프라이버시알고리즘을 사용하여 보호받게된다.

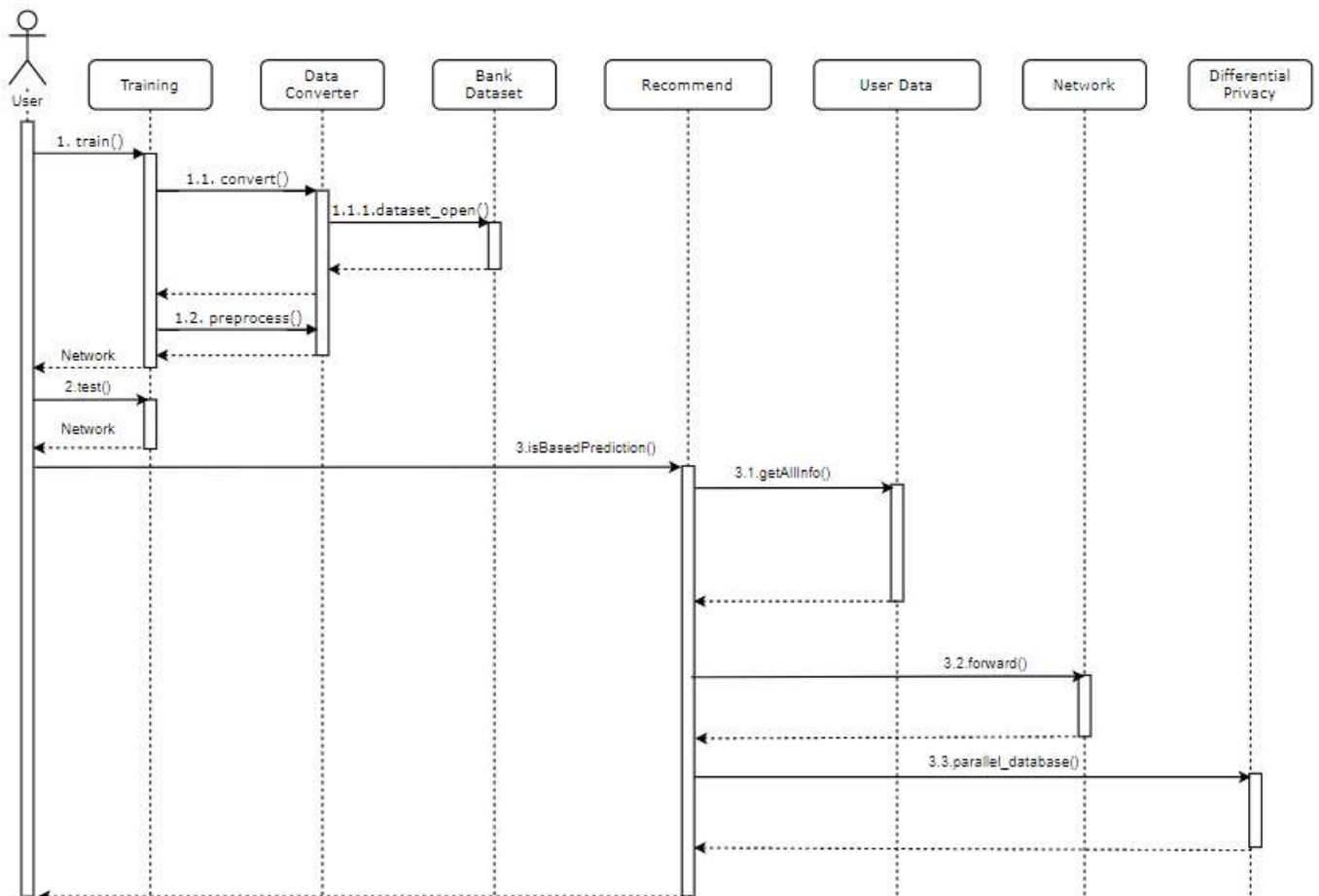


Figure 3 – RecommendingDPdata Sequence Diagram

1. 사용자는 Training에서 train을 이용하여 사용자 맞춤 딥러닝 학습을 한다.
 - 1.1: Training은 DataConverter에서 convert를 이용하여 사용하기 적절한 data로 변환한다.
 - 1.1.1. DataConverter은 BankDataset에서 dataset_open()을 이용하여 데이터를 가져온다.
 - 1.2: Training은 Dataconverter에서 preprocess을 이용하여 Bankdataset을 전처리를 수행한다.
2. 사용자는 Training에서 test를 이용하여 성능평가를 수행한다.
3. 사용자는 Recommend에서 isBasedPrediction을 이용하여 적절한 추천 값을 제시받는다.

- 3.1: Recommend는 User Data에서 getAllInfo를 이용하여 필요한 데이터를 가져온다.
- 3.2: Recommend는 Network에서 forward를 이용하여 적절한 딥러닝 학습 output을 가져온다.
- 3.3: Recommend는 Differential Privacy에서 parallel_database를 이용하여 차등프라이버시 알고리즘을 수행하기 위하여 딥러닝 학습 output에 noise를 추가한다.

1차 멘토링 결과 보고서

과 제 명	프라이버시 보호 딥러닝 서비스 개발
참여인원	4 명 201402433 조승현 201402392 이상화 201704144 김수민 201704145 김주희
수행기간	2020년 3월 ~ 2020년 6월
추진배경	<p>데이터의 전성기 시대가 도래하면서 빅데이터의 축적 및 활용으로 개인에 관한 각종 정보(의료정보, 위치정보, 신용정보 등)와 관련된 산업이 급성장하고 있다. 기업이나 기관에서 이를 분석하고 활용하는 과정 속에 개인의 프라이버시가 침해되는지에 관한 논란이 발생하게 되었고 이를 해결하기 위한 방법이 요구되어진다.</p> <p>본 과제를 수행하면서 이러한 민감한 정보를 보호할 수 있는 기술 연구를 통해 기업이나 기관은 업무 생산성을 향상시키고 정보의 주체자들은 정보 유출의 염려를 덜어 안심할 수 있도록 즉, 모두가 win-win 할 수 있는 서비스를 개발하고자 한다.</p>
목표 및 내용	<ul style="list-style-type: none"> - 인공지능의 자체적인 학습(딥러닝)을 통한 서비스를 개발한다. - 프라이버시 침해 우려가 있는 민감한 데이터 정보를 감추면서도 데이터 기반의 다양한 활용을 가능케 해주는 개인정보 비식별화 기술인 '차등 정보보호(Differential Privacy)'를 연구한다. - 사용자의 기기에서 데이터를 학습하여 모델을 추출하는 연합학습(Federated Learning)을 통해 개인의 민감한 데이터에 대한 우려를 줄일 수 있다. <p>위의 내용을 바탕으로 차등 정보보호(Differential Privacy) 기술로 사용자의 개인정보를 보호하고 기기 내에서 AI 학습이 수행되는 연합학습(Federated Learning) 방식을 적용한 맞춤형 금융상품 소개 서비스를 개발한다.</p> <p>더하여 사용자 맞춤형 상품 추천 기능을 위해서는 사용자의 분류를 판별할 수 있어야 하는데 이를 위해 클러스터링 기법을 이용한다. 사용자들의 데이터에 대해 클러스터링 기법을 적용하여 유사도를 기준으로 동질 유형을 분류하게 되며 이를 통해 고객의 특성과 상품 유형을 고려하여 세분화한 것을 기준삼아 상품을 추천 해주는 서비스를 구현할 수 있다.</p> <p>결과적으로 사용자들에게 보다 안전한 서비스를 제공할 수 있을 뿐만 아니라 차별화된 개인 맞춤형 서비스로 편의성과 만족도까지 향상시킬 수 있다.</p>
수행결과	<p>1-1) 프라이버시 보호가 되는 딥러닝 vs 프라이버시 보호가 되지 않는 딥러닝 프라이버시 보호를 잘 고려하지 않은 딥러닝은 분석 결과를 통해 민감한 정보를 복원할 수 있었다. 그리고 기존의 프라이버시 보호(k-익명성, l-다양성 등)는 다양한 자료의 연결을 통해 특정 개인의 정보를 추론할 수 있는 가능성이 여전히 존재한다. 하지만 차등 프라이버시 모델은 k-익명성, l-다양성의 취약한 부분을 보완하기 위해 제안된 모델로, 단순한 숫자의 변화가 아니라 레코드들 자체의 확률적 변형을 통해 프라이버시에 대한 식별 가능성을 제한하는 향상된 프라이버시 보호 모델이다.</p> <p>1-2) 차등 프라이버시 보호 알고리즘 개선점 혹은 사그램만의 아이디어 금융이나 의료 같은 분야에 차등 프라이버시 보호 알고리즘 외의 다른 프라이버시 보호 알고리즘은 간간히 쓰이지만 차등 프라이버시 보호는 현재 잘 쓰이지 않기 때문에 이에 대한 개선점이나 아이디어는 아직 명시하기 힘들다.</p> <p>1-3) db연급을 했는데 db에서 이미 사용자 개인정보가 들어가기 때문에 프라이버시 침해요소가 발생</p>

할 수 있다?
연합 학습, 차등 프라이버시를 이용하여 데이터베이스의 프라이버시 침해를 방지할 수 있다. 연합 학습은 데이터를 중앙에 모아서 학습하는 것이 아니라 사용자 기기에서 학습한 모델을 중앙으로 취합하는 학습 모델이다. 데이터를 중앙으로 수집하는 것이 아니기 때문에 사생활 침해 소지가 적다. 그리고 차등 프라이버시 보호는 데이터베이스에 쿼리를 요청할 때 그에 대한 응답에 적절한 분포의 noise를 섞어 프라이버시 보호를 할 수 있다.

2차 멘토링 결과 보고서

과제명	프라이버시 보호 딥러닝 서비스 개발		
협력기관명		과제책임자	
참여인원	4 명 201402433 조승현 201402392 이상화 201704144 김수민 201704145 김주희		
수행기간	2020년 3월 ~ 2020년 6월	소요비용	
추진배경	<p>데이터의 전성기 시대가 도래하면서 빅데이터의 축적 및 활용으로 개인에 관한 각종 정보(의료정보, 위치정보, 신용정보 등)와 관련된 산업이 급성장하고 있다. 기업이나 기관에서 이를 분석하고 활용하는 과정 속에 개인의 프라이버시가 침해되는지에 관한 논란이 발생하게 되었고 이를 해결하기 위한 방법이 요구되어진다.</p> <p>본 과제를 수행하면서 이러한 민감한 정보를 보호할 수 있는 기술 연구를 통해 기업이나 기관은 생산성을 향상시키고 정보의 주체자들은 정보 유출의 염려를 덜어 안심할 수 있도록 즉, 모두가 win-win 할 수 있는 서비스를 개발하고자 한다.</p>		
목표 및 내용	<p>- 인공지능의 자체적인 학습(딥러닝)을 통한 서비스를 개발한다.</p> <p>- 프라이버시 침해 우려가 있는 민감한 데이터 정보를 감추면서도 데이터 기반의 다양한 활용을 가능케 해주는 개인정보 비식별화 기술인 '차등 정보보호(Differential Privacy)'를 연구한다.</p> <p>- 사용자의 기기에서 데이터를 학습하여 모델을 추출하는 연합학습(Federated Learning)을 통해 개인의 민감한 데이터에 대한 우려를 줄일 수 있다.</p> <p>위의 내용을 바탕으로 차등 정보보호(Differential Privacy) 기술로 사용자의 개인정보를 보호하고 기기 내에서 AI 학습이 수행되는 연합학습(Federated Learning) 방식을 적용한 맞춤형 금융 상품 소개 서비스를 개발한다.</p> <p>사용자 맞춤형 상품 추천 기능을 위해서는 사용자의 분류를 판별할 수 있어야 하는데 이를 위해 클러스터링 기법을 이용한다. 사용자들의 데이터에 대해 클러스터링 기법을 적용하여 유사도를 기준으로 동질 유형을 분류하게 되며 이를 통해 고객의 특성과 상품 유형을 고려하여 세분화한 것을 기준삼아 상품을 추천 해주는 서비스를 구현할 수 있다.</p> <p>결과적으로 사용자들에게 보다 안전한 서비스를 제공할 수 있을 뿐만 아니라 차별화된 개인 맞춤형 서비스로 편의성과 만족도까지 향상시킬 수 있다.</p>		

수행결과

1. 커스텀으로 만든 임의의 데이터 말고 실제 데이터 셋을 다뤄보는 것을 추천하며 이를 위해 클러스터링 할 수 있는 금융 관련 오픈 데이터 셋을 찾아야 한다.
2. 데이터 셋을 구할 때 금융 빅데이터 플랫폼 사이트에서 데이터 셋을 유료로 구매하는 방법이 있으며 데이터를 구하기 힘든 경우 데이터 셋을 지인이나 다른 사람으로부터 모으는 방법도 있다.
3. 만들어진 서비스에 대한 평가를 어떻게 할 것인가를 고민해봐야 한다.
4. 온디바이스는 주로 저사양의 엣지 디바이스(라즈베리파이, 아두이노 등등)를 의미하는데 엣지 디바이스에서 딥러닝 모델을 운영하는 것이 쉽지 않다. 한 번에 메모리에 올라갈 수 있도록 모델을 경량화(압축)하는 작업이 필요하며 성능 역시 중요하다. 성능이 덜 떨어지는 선에서 경량화를 하는 것이 주요 이슈가 된다.
5. 차등 프라이버시와 연합 학습, 온디바이스를 전부 구현하는 것이 벅찬 목표일 수 있으므로 추후에 이 주제들을 간소화할 필요가 있다.

종합설계1

졸업 프로젝트 - 설문조사 분석 보고서

주제 : 프라이버시 보호 딥러닝 서비스 개발

[사그람 조]
201402433 조승현 201402392 이상화
201704144 김수민 201704145 김주희

• 설문 개요

본 설문조사는 15명의 응답자를 대상으로 사그램조가 구현한 프로토타입에 대한 사용자들의 의견을 분석해보기 위하여 기획되었습니다. 이를 통해 문제점이나 개선해야 할 점을 참고하여 해결책을 찾아 보완해보고자 합니다.

- 설문 응답자 : 15명
- 설문 기간 : 2020.05.28 – 2020.05.29 (2일간)
- 설문 방법 : Google 설문지

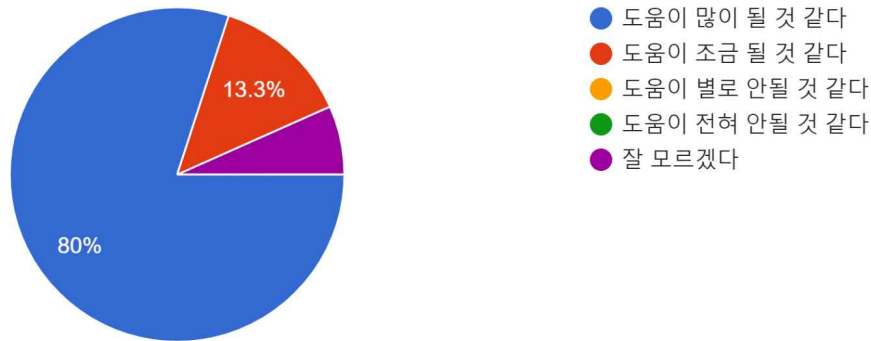
• 설문 문항

- 해당 모델이 사용자의 데이터를 분석하여 적절한 상품을 추천해주는 어플리케이션을 개발하는데 얼마나 도움이 된다고 생각하시나요?
(도움이 많이 될 것 같다/도움이 조금 될 것 같다/도움이 별로 안될 것 같다/도움이 전혀 안될 것 같다/ 잘 모르겠다)
- 해당 모델의 문제점(부족한 점)이 있다면 무엇인가요?
- 해당 모델로 구현된 서비스를 사용할 의향이 있으십니까?

• 설문 결과

1. 해당 모델이 사용자의 데이터를 분석하여 적절한 상품을 추천해 주는 어플리케이션을 개발하는데 얼마나 도움이 된다고 생각하시나요?

응답 15개



15명의 응답자 중,
도움이 많이 될 것 같다 12명 / 도움이 조금 될 것 같다 2명 / 잘 모르겠다 1명

⇒ **도움이 될 것 같다는 응답 (93.3%) 多 !!**

2. 해당 모델의 문제점(부족한 점)이 있다면 무엇인가요?

- 데이터 분류시에 특징을 조금 더 디테일하게 분류할 수 있으면 좋겠다.
- 오류가 자주 발생할 것 같다.
- 사용하는 고객 수가 많아지면 데이터의 양도 증가할 것인데, 이 증가에 비례하게 계산비용도 증가한다면 하나의 프로그램에 데이터가 많을 때 정확성이 떨어지는 문제점이 발생하지 않을까 하는 생각이 든다.
- 오차를 줄이기위해 3~4단계를 반복한다는 점이 비효율적이라고 생각한다. 만약 데이터가 많아진다면 더욱 번거로운 일이 될 것으로 예상된다. 한 두번 정도의 실행으로 오차값을 줄일 수 있는 방법을 구현하는 것도 좋을 것 같다.
- 데이터의 차원 자체가 너무 커서 쉽지 않을 것 같다.
- 영상을 통해서 어떤 데이터 셋을 이용하여 테스트를 한 것인지 모르겠지만 기존의 학습데이터 셋을 그대로 넣어서 딥러닝 한 경우 학습데이터가 아닌 테스트 데이터를 넣어서 딥러닝 했을 때 오차율이 지금보다 크게 나올 것 같다.

이 외의 응답은 “문제점(부족한 점)이 없다”

3. 해당 모델로 구현된 서비스를 사용할 의향이 있으십니까?

- 학습된 데이터로 최적의 상품을 추천해준다는 점에서 사용할 의향이 있다.
- 개인정보는 보호하고, 시간을 단축하여 필요한 상품을 가입할 수 있을 것 같아 사용할 의향이 있다.
- 요즘 사람들이 평균적으로 10분에 한번 핸드폰을 본다는 결과가 있듯 핸드폰 사용시간이 늘어나면서 편리하게 이용할 수 있는 모바일 금융서비스의 이용률도 증가하였다. 이 서비스를 역시 편리하게 사용할 수 있을 것 같아 사용할 의향이 있다.
- 금융 관련 정보나 지식이 부족한 분들에게 좋을 것 같다. 아무래도 금융 서비스 이다보니 민감한 정보가 유출되는 걱정을 덜 수 있어서 좋은 모델이라고 생각한다.
- 금융 상품을 선택할 때 내 소비패턴을 판단하여 결정하기까지 쉽지 않았는데 딥러닝을 통해 분석해준다면 보다 빠르게 적절한 판단을 할 수 있을 것 같다.
- 데이터를 분류하여 추천해주는 과정 속에 정보가 유출될 수도 있고 악용될 수도 있을 것이라는 불안감을 가질 수도 있지만, 데이터를 암호화하여 추천된 상품이라 하면 안심되고 해당 서비스에 대한 신뢰감도 생길 것 같다.
- 정확도가 좀 더 높아지고 개인정보 보호만 확실해 진다면 사용할 것 같다.
- 평소에 나에게 맞는 상품을 얻기가 쉽지 않았고 개인정보보호도 중요시 여겨왔기 때문에 서비스를 잘 이용할 것 같다.

● 생각해 볼만한 점

■ 미니배치 K-평균 군집화(Mini-batch K-means Clustering)

K-Means 방법은 중심위치와 모든 데이터 사이의 거리를 계산해야 하기 때문에 데이터의 양이 많아지면 계산량도 늘어나게 된다. 이처럼 데이터의 수가 많은 경우에는 데이터를 미니배치(Mini-batch) 크기만큼 무작위로 분리하여 K-Means를 수행하는 '미니배치 K-Means'로 계산량을 줄일 수 있다. 계산량의 감소로 속도 역시 훨씬 빨라진다. 사이킷런의 cluster 서브패키지에서 제공하는 MiniBatchKMeans를 사용할 수 있다.

■ 차원 축소(Dimensionality Reduction) – PCA, SVD

사용할 데이터와 feature가 많으면 비지도 학습의 경우 어떤 feature를 사용해야 할지 상당히 난감하다. 이렇듯 고차원 데이터에서 아무런 처리 없이 비지도 학습인 군집 분석을 시도하면 성능이 저하될 수 있다. 따라서 차원 축소를 수행해 주어야 하는데, 차원을 축소해주는 기법인 PCA나 SVD를 이용할 수 있다.

- PCA (주성분 분석) : 여러 변수간에 존재하는 상관관계를 이용하여 이를 대표하는 주성분을 추출하여 줄
- SVD (특이값 분해) : 임의의 고유값 분해를 직사각형 행렬에 대해 일반화하는 방법

종합설계1

프로토타입 데모

주제 : 프라이버시 보호 딥러닝 서비스 개발

[사그램 조]
201402433 조승현 201402392 이상화
201704144 김수민 201704145 김주희

■ Data

● Training data

```
from diffprivlib.models.k_means import KMeans as DP_KMeans
from sklearn.cluster import KMeans as REAL_KMeans
import numpy as np
import pandas as pd
import seaborn as sb
import matplotlib.pyplot as plt
%matplotlib inline

train = pd.read_csv('sample.csv')
test = pd.read_csv('test.csv')
train['loan'] = train['loan'].map({'yes': 1, 'no': 0})

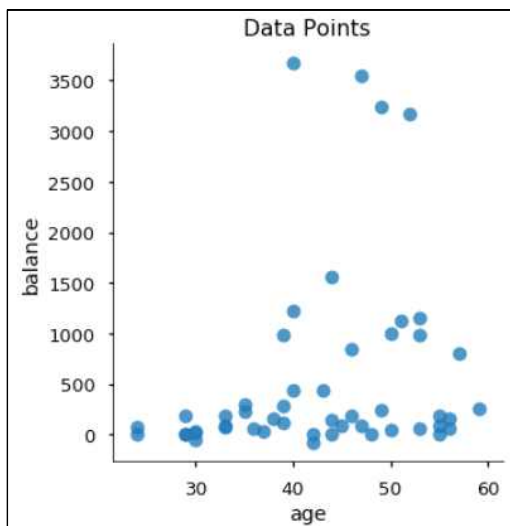
# train data
train.head()
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y	crec
0	35	management	married	tertiary	no	231	yes	0	unknown	5	may	139	1	0	0	unknown	no	
1	50	management	married	secondary	no	49	yes	0	unknown	5	may	180	2	0	0	unknown	no	
2	44	technician	married	secondary	no	0	yes	0	unknown	5	may	225	2	0	0	unknown	no	
3	55	technician	divorced	secondary	no	0	no	0	unknown	5	may	160	1	0	0	unknown	no	
4	42	admin.	single	secondary	no	-76	yes	0	unknown	5	may	787	1	0	0	unknown	no	

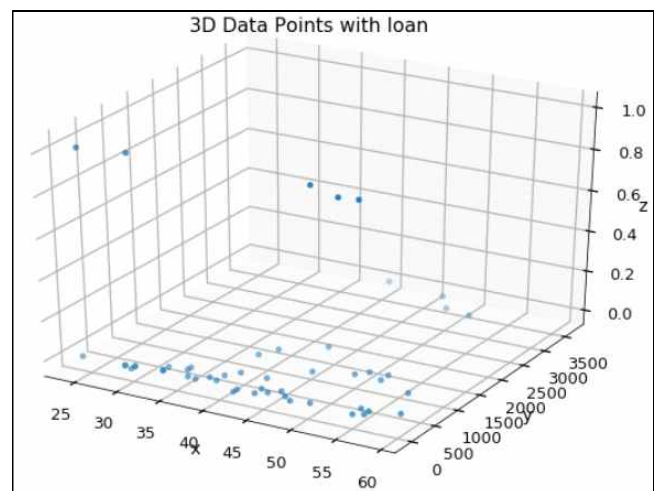
● Test data

	age	balance
0	36	10

● Visualization



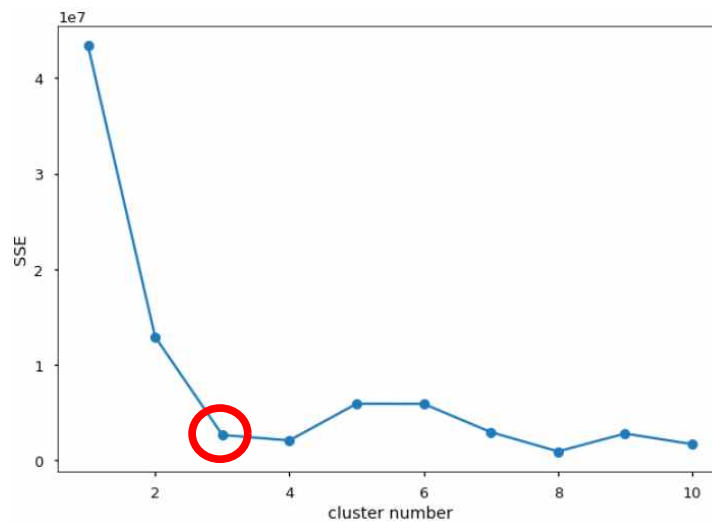
age-balance 의 2차원 데이터 분포 그래프



age-balance-loan 의 3차원 데이터 분포 그래프

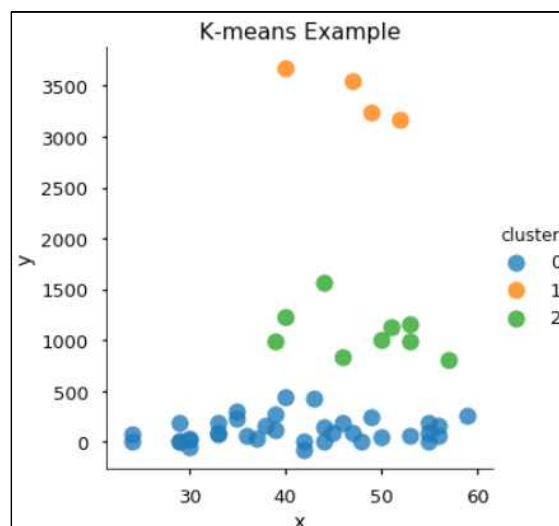
■ K-means Clustering

- 클러스터 개수인 k값을 구하기 위해 **elbow 함수**를 정의하고 실행



가장 급격하게 꺾이는 구간의 cluster number를 k값으로 사용할 수 있음 (k=3)

- K-means Clustering 수행 결과



3개의 군집으로 성공적으로 군집화 된 것을 확인할 수 있음

- Test 데이터에 대한 Cluster를 예측해보기

```
# (x, y) = (36, 10)
predict=REAL_kmeans.predict(test)
print("test 값의 적절한 cluster label은 "+str(predict[0])+" 입니다.")
```

test 값의 적절한 cluster label은 0 입니다.

- 해당 Cluster(군집) 내의 데이터(사용자)들이 많이 사용하는 상품을 추천해주기

```
def recommend(x): # x는 예측된 결과 데이터
    df2 = pd.read_csv('sample.csv')
    array = []
    for i in range(50):
        if x[0] == train['cluster'][i]:
            array.append(df2['creditcardcode'][i])
    from collections import Counter
    result = Counter(array)
    i = 0
    a = 0
    for key in result:
        if i < result[key]:
            i = result[key]
            a = key
    print(key, result[key])
    print("")
    print("추천상품: "+str(a))
    print("추천상품 사용자: "+str(i))
```

```
recommend(predict)
```

```
1000025054 8
```

```
1000022204 1
```

```
1000023855 5
```

```
1000024919 5
```

```
1000023675 4
```

```
1000022675 4
```

```
1000024051 2
```

```
1000025166 5
```

```
1000023653 1
```

```
1000025332 2
```

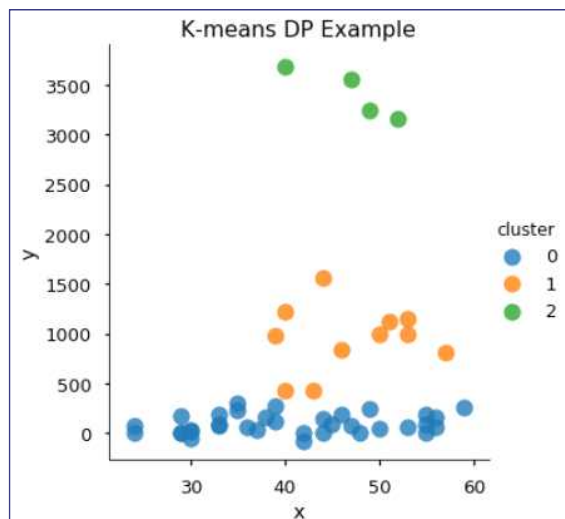
```
추천상품: 1000025054
```

```
추천상품 사용자:8
```

가장 많은 8명의 사용자가 사용한 상품 1000025054를 추천

■ K-means Clustering with DP(Differential Privacy)

- DP(Differential Privacy)를 적용한 K-means Clustering 수행 결과



역시 3개의 군집으로 성공적으로 군집화 된 것을 확인할 수 있음

- DP(Differential Privacy) 적용 전/후 결과 비교

recommend(predict)
1000025054 8
1000022204 1
1000023855 5
1000024919 5
1000023675 4
1000022675 4
1000024051 2
1000025166 5
1000023653 1
1000025332 2
추천상품: 1000025054 추천상품 사용자:8

< DP를 적용시키지 않은 recommend 결과 >

recommend(predict_dp)
1000025054 8
1000022204 1
1000023855 6
1000024919 6
1000023675 4
1000022675 4
1000024051 2
1000025166 6
1000023653 1
1000025332 4
추천상품: 1000025054 추천상품 사용자:8

< DP를 적용시킨 recommend 결과>

데이터의 유용성은 훼손시키지 않되, 안전성은 높일 수 있음을 확인할 수 있음

[GITHUB & YOUTUBE]

[GITHUB] https://github.com/pmcsh04/designsprint_4gram/tree/master/GP_Final3

[YOUTUBE] <https://youtu.be/4fp-dXEoSYw>