

프로젝트 최종보고서

일정 및 소비 데이터를 활용한 예산 사용량 예측 AI 기반 모바일 웹 서비스
“돈 워리(Be Happy)”

날짜	팀명	팀원	학번	팀지도교수
2024.06.14	고등어	이규민	2076291	이민수
		서선아	2003047	
		공채은	2176020	

과제의 범위

1. 배경

건강한 소비활동에 대한 확실한 기준이 잡히지 않은 사회초년생이 금융과 관련된 다양한 문제에 직면해 있는 것을 인지하고, 보다 친숙한 일정과 가계부를 한번에 관리할 수 있도록 ‘일정 및 가계부 통합 모바일 AI서비스’를 계획했다. 사회초년생은 본격적인 경제활동을 시작함에 앞서 올바른 돈 관리가 필요한 시기이기 때문에, 돈 관리에 스트레스와 번거로움을 덜고 현명한 소비습관과 즐거운 자산관리를 동시에 할 수 있는 서비스의 필요성을 느꼈다.

2. 문제 정의

(1) 돈 관리의 생소함

-> 상대적으로 익숙한 일정 관리와 돈 관리를 한번에 할 수 있도록 한다. 또한 챌린지와 사용량 공유 기능을 통해 돈관리에 흥미를 부여한다

(2) 돈 관리의 번거로움

-> 일정을 입력하면 예산이 자동으로 분배되어 쉽게 적정 소비량을 확인할 수 있고, 자신의 예산에 맞지 않는 과소비를 경계하고 적절한 소비를 하도록 한다.

(3) 무분별한 금융 서비스 노출

-> 예) 카드 리볼빙 서비스

타 서비스와의 비교

기존 가계부 서비스와의 차별성 4가지를 정리했다. banksalad, 편한가계부, 핀크, Uplanner, visual 가계부, 꼬박가계부, DAILY PAY 외 6개의 서비스를 분석해 4가지의 측면에서 차별성을 두었다.

(1) 기록

기존에는 기록에 있어서 카드 연동과, 결제 내역을 자동으로 불러오는 기능이 있긴 하지만 가계부 관련 내역만을 불러오는데 돈워리는 가계부 내역과 일정 내역을 통합시켜 일정까지 관리할 수 있도록 하였다. 더불어 OCR기술을 도입해 편리하게 소비내역을 기록할 수 있도록 하였다.

(2) 예산 기능

타 가계부 서비스와는 달리, 머신러닝 알고리즘과 예산 분배 로직을 통해 일정을 입력하면 해당 일정에 대한 예산 사용량과 실제 사용 가능 예산을 예측해준다.

(3) 통계 기능

기존의 가계부 서비스들처럼 분석 메시지를 제공하되, 돈워리의 대상인 사회초년생을 타겟팅하는 메시지로 차별성을 두었다.

(4) 챌린지 및 사용량 공유

기존의 타 서비스에서 제공하는 챌린지에, 각자의 세부 목표와 친구와 함께 챌린지를 공유하며 리워드를 지급받는 기능을 추가하여 차별성을 두었고, 서로의 사용량을 공유하며 자신의 지출 추이를 자각할 수 있다. 또한 목표금액에 대한 달성률이 떨어질 때 팝업 미션을 제공해 과소비를 멀리할 수 있도록 했다.

기대 효과

1. 소비량 예측값 제공 및 예산 분배

- 사용자의 월별 수입을 입력받아 사용자의 이전 카테고리 별 지출 가중치를 이용해 일정 별 예산 사용 가능량을 예측해준다.
- 사용자는 사용량 예측값을 통해 본인의 적절한 소비 가이드 라인을 확인할 수 있다.

2. 챌린지

(1) 사용자가 지정한 기간동안 목표금액 모으기 챌린지

- 사용자가 입력한 목표금액을 개월수로 나누어 월별 세부목표를 제공해 동기를 부여한다.

(2) 하루에 사용자가 지정한 금액만큼만 쓰기 챌린지

- 사용자가 지정한 금액 기준 성공 실패 여부를 제공해 과소비를 지양하도록 한다.

(3) 한달동안 특정 카테고리 지출(배달, 문화생활 등) 횟수 지정 챌린지

- 월별 카테고리 지출 횟수 확인 및 성공 실패 여부를 알려주며 불필요한 지출을 줄이도록 한다.

3. 소비량 공유

- 이메일 주소로 공유할 친구를 등록할 수 있고, 서로 공유된 예산 사용량을 조회하여 '나'의 예산 사용량과 비교함으로써 과도한 소비를 경계할 수 있다.

- n명의 친구들과 목표금액을 정하고 1/n으로 나누고, 공통 챌린지로 등록하여 함께 달성하며 서로 동기를 부여할 수 있다.

구현 내용 및 Features

1) 일정 데이터 CRUD

- 리스트 형식
- 예측 모델 학습에 필수적인 일정 카테고리를 반드시 선택한다.

2) 가계부 기능 - 수입/지출 데이터 CRUD

- 리스트 형식
- 예측 모델 학습에 필수적인 가계부 카테고리를 반드시 선택한다.

3) 예산 사용 예측 기능

: 사용자의 일정-가계부 데이터로 학습시킨 다중선형회귀모델을 이용하여 새로운 일정에 대한 예산 사용량을 예측하는 기능. 일정 카테고리를 메인으로 사용한다.

- 과적합을 방지하고 모델의 일반화 성능을 향상시키기 위해 다중선형회귀 모델 중 Ridge와 Lasso 모델을 선택했다.
- 초기에는 적은 데이터로 인해 MSE, RMSE, MAE 등의 성능지표에서 모델이 과적합되어 성능이 저하되는 것을 확인했다.
- 이를 해결하기 위해 GridSearchCV 또는 RandomSearchCV를 사용하여 최적의 하이퍼파라미터를 찾고, RandomForest 모델을 사용하여 중요 특성을 추출했으나, 여전히 모델의 성능은 크게 개선되지 않았다.
- 이후 추가적인 데이터를 수집하여 모델을 재학습한 결과, 성능지표가 개선되었고 카테고리별 실제 지출에 근접한 예산 지출을 도출할 수 있었다.
- 향후 서비스 이용자수가 증가함에 따라 모델의 복잡도를 지속적으로 관리한다면, 모델의 정확도가 더욱 향상될 것으로 기대된다.

4) 예산 분배 기능

: 사용자가 카테고리별 소비 비율과 절약 정도를 직접 선택하면 각 일정 및 지출 카테고리에 대해 월별 예산 분배안을 제시하는 기능

- 기본적으로 이전 달 소비 비율을 기반으로 이번 달 예산을 분배한다.
- 사용자가 카테고리별 소비 비율을 조절할 수 있다.

5) OCR 활용 지출 등록 기능

: OCR을 활용해 사용자의 결제 문자를 스캔해 가계부 지출에 입력해주는 기능

6) 통계 기능

: 사용자의 월별 사용량을 분석해 소비패턴을 한눈에 파악할 수 있도록 보여주는 기능

- 카테고리별 지출 막대/원형 그래프, 월별 지출 꺾은선그래프를 표시한다.
- 수입/지출 데이터와 일정 데이터를 사용한다.

7) 챌린지 기능

: 사용자가 자신의 소비습관에 맞는 챌린지에 참여하며 돈관리에 재미를 주는 기능

7-1) 사용량 공유 기능

: 친구와 예산 사용량을 공유함으로써 돈 관리의 동기부여와 재미를 제공하는 기능

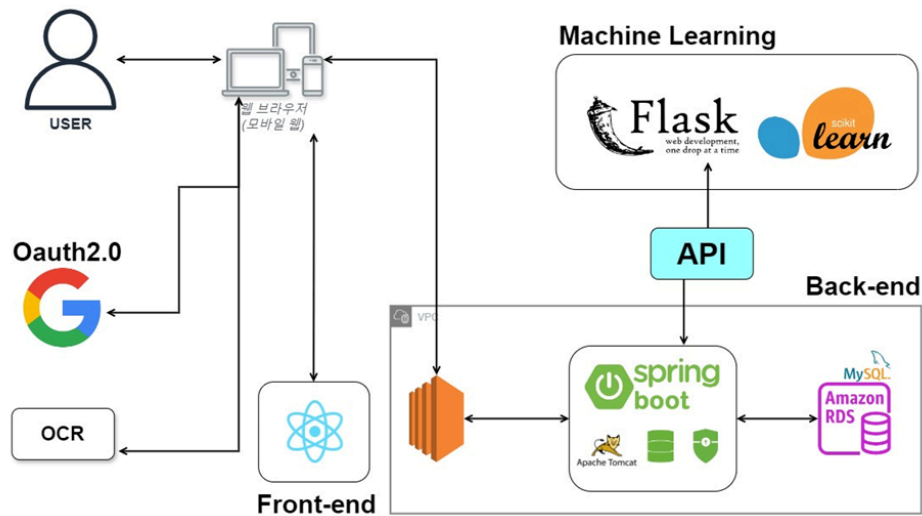
- 친구 추가 기능
- 친구 목록의 현재까지의 예산 사용량 상태바를 확인할 수 있다.

7-2) 공유 챌린지 기능

: 등록된 친구와 공통 챌린지를 진행하는 기능

- 친구 그룹을 형성해 공통 저축 챌린지를 등록하고 진행 상황을 확인할 수 있다.

System Architecture



1. 클라이언트 (사용자)

사용자는 모바일 웹 브라우저를 통해 시스템에 접속하며, **React** 프레임워크를 사용하여 개발된 **UI** 통해 시스템과 상호 작용한다.

2. 프론트엔드 (React)

React로 사용자에게 제공되는 웹 애플리케이션 인터페이스를 구현한다. 이는 **UI**를 표시하여 사용자와의 상호작용을 담당한다. **Figma**를 이용해 **UI** 컴포넌트를 설계하고, 이를 바탕으로 인터페이스를 구축한다. **React**는 비동기 이벤트 기반 구조를 가지는 웹서버인 **Nginx**를 통해 배포한다. **Nginx**는 클라이언트인 웹 브라우저로부터 **HTTP** 요청을 받아들이고, **React**로 작성된 웹페이지를 반환하여 외부 클라이언트 요청을 처리한다. 이를 통해 빠르고 효율적인 웹 애플리케이션 환경을 제공한다.

구현한 모바일 화면에서 사용자가 일정과 가계부를 작성하고 저장된 내역을 조회할 수 있다. 예산 절약 플랜을 선택하고 진행 중인 챌린지 목록을 확인할 수 있다. 예산 사용에 대한 통계와 예산 사용 예측 값을 제공한다. 프론트엔드에서 요청한 데이터는 백엔드의 비즈니스 로직을 거치며 백엔드와 데이터베이스의 상호작용을 통해 처리되며, 최종적으로 프론트엔드로 전달된다.

3. 백엔드 (Spring Boot)

시스템의 핵심 비즈니스 로직을 구현하고 **Spring Boot** 프레임워크로 구현된다. 사용자 요청을 받아 처리하고 데이터베이스와의 상호 작용을 담당한다. 예산 예측 모델을 사용해야 하는 경우 **Flask API**에 대한 요청을 보내고 응답을 처리한다.

Spring Boot 애플리케이션은 **AWS EC2** 인스턴스 **Ubuntu SERVER 22.04 LTS**에서 배포되고 내장 **tomcat**을 웹 서버로 사용하여 서비스를 제공한다. **Spring Data JPA**를 사용하여 데이터베이스에서 데이터를 읽고 쓰는 작업을 수행한다.

4. 데이터베이스 (RDS MySQL)

AWS RDS에 **MySQL** 데이터베이스 서버를 호스팅한다. 사용자가 프론트엔드에서 입력한 데이터는 사용자 권한 확인을 거쳐 데이터베이스에 저장된다. 데이터베이스는 사용자 정보, 일정 정보, 가계부 내역, 예산 정보 등을 저장한다.

5. 보안 및 인증 (Spring Security)

Spring Security는 사용자 인증 및 권한 부여를 담당하며, OAuth 2.0 프로토콜을 사용하여 사용자 로그인을 처리한다.

사용자가 OAuth 2.0을 사용하여 로그인하고 인증 서버로부터 액세스 토큰을 발급받는다. 이후 사용자 인증이 필요한 API 요청을 보낼 때마다 OAuth 2.0 액세스 토큰을 함께 보낸다. API 서버는 해당 토큰이 유효한지 검증함으로써 요청을 보낸 사용자를 인증하고 요청을 처리하는데, 필요한 작업을 수행하기 위해 데이터베이스와 통신하기도 한다.

클라이언트가 보낸 요청에 포함된 OAuth 2.0 액세스 토큰을 검증하는 로직은 Spring Security를 사용하여 API 서버에 구현된다.

6. 외부 서비스 연동 (Flask, scikit-learn, OCR)

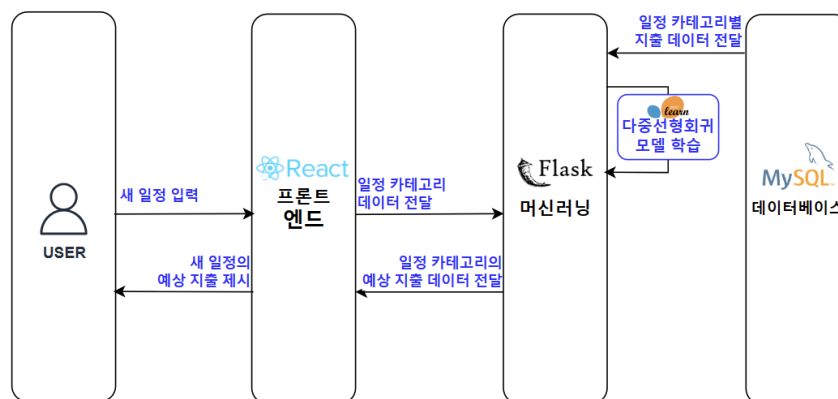
scikit-learn을 사용하여 머신러닝 모델을 학습시키고, 학습된 모델은 Flask API로 배포한다. 또한 OCR을 사용하여, 캡처된 결제 문자 이미지에서 텍스트를 추출하고 그 내용을 가계부 내역으로 등록한다.

구글 코랩에서 scikit-learn 머신러닝 라이브러리를 사용해 예산 사용량 예측을 위한 다중선형회귀모델을 학습한다. 학습에는 사용자가 입력한 일정 데이터, 가계부 데이터가 사용된다. 학습된 모델을 Flask API로 래핑하여 서버에 배포하여 Spring Boot 애플리케이션에서 접근할 수 있도록 한다. Flask API로 Spring Boot 애플리케이션에서는 학습된 모델을 사용하여 예측을 수행하고 결과를 처리한다.

사용자가 웹 브라우저를 통해 업로드한 결제 문자 캡처는 OCR 서비스에 전달되고, OCR 서비스는 이미지에서 텍스트를 추출하고 인식한다. 추출된 결과 텍스트를 클라이언트로 반환하고 필요한 부분만 필터링하여 데이터베이스 가계부 테이블(ACCOUNT)에 저장한다.

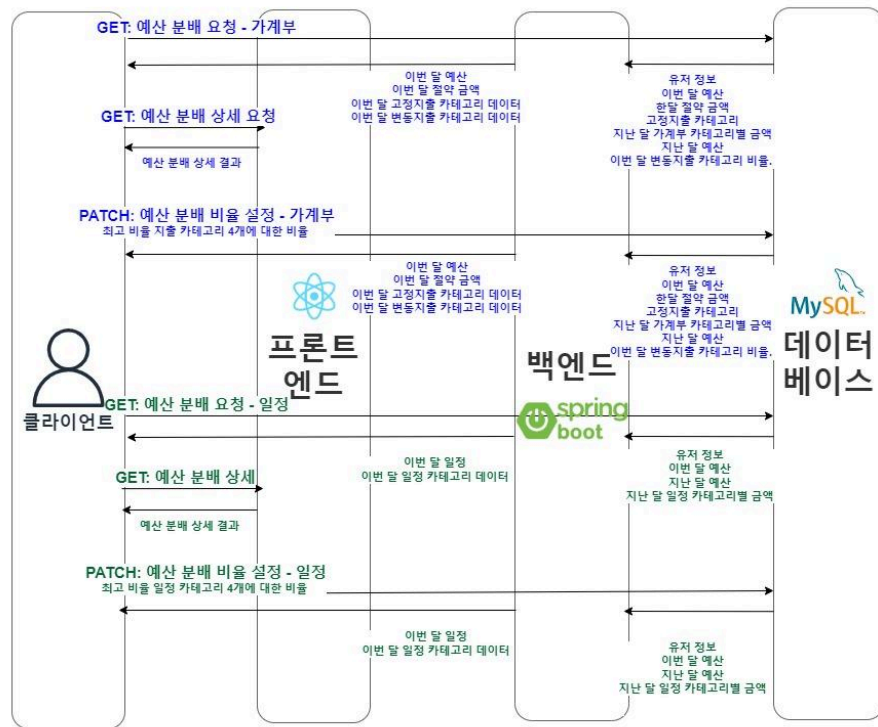
주요 엔진(Core Engine, Main Part 등)의 기능 및 설계

1. 예산 예측 기능



Flask 서버가 MySQL 데이터베이스와 연결되어 클라이언트의 일정 및 가계부 데이터를 가져온다. 이 데이터는 일정에 대한 카테고리별 지출금액을 포함한다. 가져온 데이터를 사용해 Scikit Learn 머신러닝 라이브러리로 다중선형회귀모델을 훈련시킨다. 사용자가 새 일정 데이터를 입력하면, 프론트엔드에서 이 데이터를 Flask 서버의 API로 전달한다. 서버에서는 머신러닝 모델을 사용해 일정에 대한 예상 지출(예산)을 예측한다. 예측된 예산 데이터를 Flask API로 프론트엔드에 전달한다. 클라이언트는 입력한 새 일정에 따른 예상 지출을 확인할 수 있다.

2. 예산 분배 기능



백엔드 서버에서 구현한 비즈니스 로직으로 일정, 소비내역의 카테고리별로 전체 예산에 대한 한달 예산 사용 계획을 제안한다.

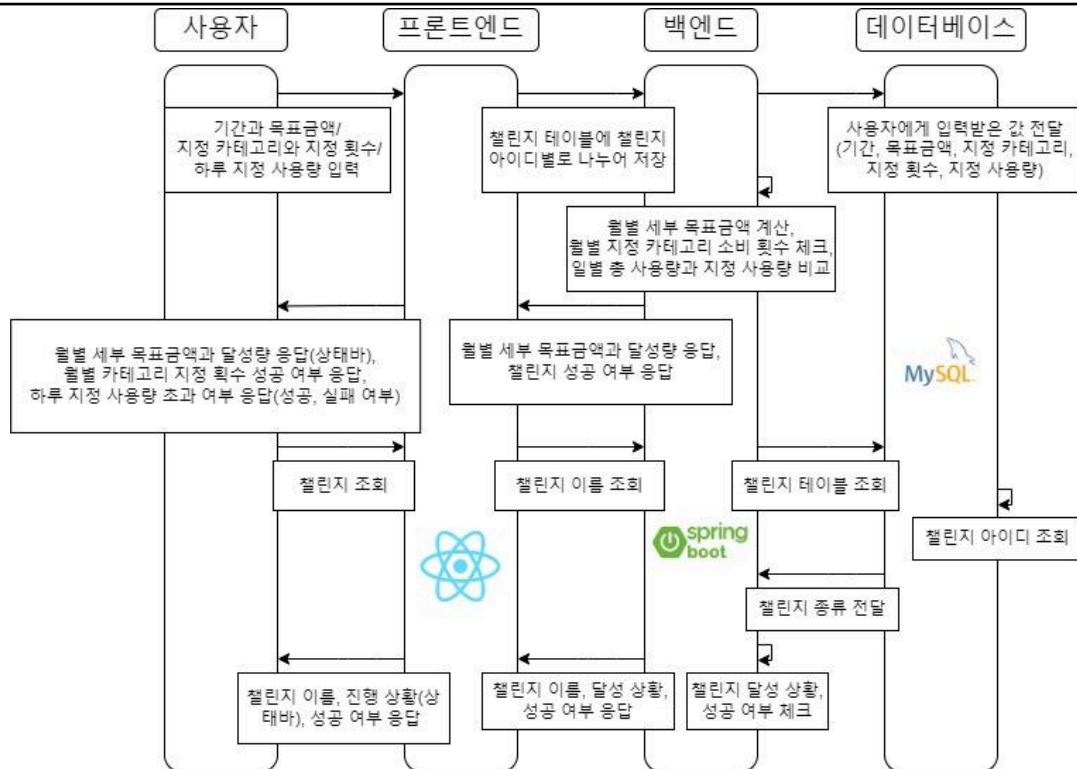
1) 예산 분배 - 가계부

첫 요청 시엔 이전 달의 가계부 카테고리의 예산 사용량 비율을 기반으로 각 카테고리별 이번 달 예산을 분배한다. 예산은 절약/고정지출/변동지출 세 파트를 나눠서 분배되고 해당 데이터는 데이터베이스의 **ACCOUNT** 테이블에서 불러온다. 이후 최고 사용량 4개의 변동 지출 카테고리에 대해 예산 사용량 비율을 설정하면, 수정된 비율을 반영하여 각 변동지출 카테고리별 예산을 재분배하고 그 결과를 응답한다.

2) 예산 분배 - 일정

첫 요청 시엔 이전 달의 전체 예산 대비 일정 카테고리 예산 사용량 비율과 각 일정 카테고리의 예산 사용량 비율을 기반으로 이번 달 예산을 분배한다. 해당 데이터는 데이터베이스의 **SCHEDULE** 테이블에서 불러온다. 이후 최고 사용량 4개의 일정 카테고리에 대해 예산 사용량 비율을 설정하면, 수정된 비율을 반영하여 각 일정 카테고리별 예산을 재분배하고 그 결과를 응답한다.

3. 챌린지 기능



1) 사용자가 지정한 기간동안 목표금액 모으기

:사용자가 기간과 목표금액을 입력하면 **CHALLENGE** 테이블에 챌린지가 저장되고 목표 금액이 개월수로 나누어져 사용자에게 월별 세부목표로 상태바와 함께 응답된다.

2)하루에 사용자가 지정한 금액 만큼만 쓰기

:사용자가 하루의 목표 지출량을 입력하면 **CHALLENGE** 테이블에 챌린지가 저장되고 사용자의 하루 총 지출량과 목표 지출량을 비교하여 성공,실패 여부가 응답된다.

3)한달동안 특정 카테고리 지출(배달, 문화생활 등) 횟수 정하기

:사용자가 지정 카테고리 지정 횟수를 입력하면 **CHALLENGE** 테이블에 챌린지가 저장되고, 일정에 등록된 해당 카테고리 횟수와 비교하여 성공, 실패 여부를 응답한다.

4. 사용량 및 챌린지 공유 기능

등록된 친구와 각자의 예산 사용량 상태바를 공유하고, 친구 그룹을 형성해 공통 챌린지를 진행한다.



- 1) 이메일 주소로 친구 등록
: 전달된 이메일 주소로 **USER** 테이블에서 친구를 찾고 **FRIEND** 테이블에 친구를 등록한다.
- 2) 예산 사용량 공유
: 친구 목록에서 친구로 등록된 사용자의 현재까지의 예산 사용량 %를 항상 조회할 수 있다.
- 3) 공통 챌린지
: n 명의 친구 이메일과 챌린지 제목, 챌린지 기한, 챌린지 목표 금액을 등록하면 $1/n$ 의 매달 저축 금액을 반환하고 매달 진척도를 조회할 수 있다.

평가

평가항목1. 기능성

평가기준: 예산 사용량 예측 모델에 대한 평가 지표

평가내용: 약 5000개의 가계부 데이터에 대한 **RMSE**, **MAE**, **MSE**의 세가지 평가지표를 사용하였고, 이상치 데이터 제거문제로 **MAE**는 10000원 정도로 다소 높게 측정되었지만 결정계수는 0.91로 안정적으로 예측하는 것으로 나타났다.

평가항목2. 사용성

평가기준: 사용자 피드백

평가내용:

돈워리의 서비스(일정 가계부 서비스, 예산 예측, 예산 분배 및 추천, 챌린지, 통계, 상품추천)에 대한 33명의 사회초년생의 선호도 조사를 실시한 결과는

일정 가계부(77.8%), 예산 예측(58.3%), 예산 분배(55.6%), 챌린지(61.1%), 통계(63.9%), 상품 추천(66.7%)이다. 각각의 항목에 대한 신뢰구간(CI)을 신뢰수준 95%에서 구해보면,

*일정 가계부(0.778)CI= (0.6362, 0.9198)

*예산 예측(0.583)CI= (0.4148, 0.7512)

*예산 분배(0.556)CI= (0.3865, 0.7255)

*챌린지(0.611)CI= (0.4447, 0.7773)

*통계(0.639)CI= (0.0.4752, 0.8028)

*상품 추천(0.667)CI= (0.5062, 0.8278)

로 모든 항목의 신뢰구간이 전체 비율의 범위를 포함한다. 33명의 표본의 크기가 비교적 작지만, 각 항목의 비율이 95%의 신뢰 수준에서 안정적이다. 따라서 돈워리의 기능들이 사용자의 입장에서 95%의 신뢰 수준으로 유의미한 기능임을 확인 할 수 있다.

결론

“돈 워리”는 돈 관리가 익숙하지 않은 사회초년생이 돈 관리와의 거리감을 줄이고, 일상적이고 재미있게 생활과 금전을 관리할 수 있게 하는 데에 목적이 있다.

이를 위해 가계부보다 상대적으로 익숙한 일정 캘린더와 가계부를 통합해, 돈 관리가 자연스럽게 일상에 녹아들 수 있도록 했다. 머신러닝 알고리즘을 사용해 일정별 예산 사용량을 예측하고, 예산 분배 로직을 사용해 실제 입력 받은 한 달 예산을 각 일정에 적절하게 분배하도록 했다. 돈 관리의 방향성을 제시함으로써 사용자는 과소비를 경계하고 적절한 소비가 가능하다. 또한 챌린지와 친구와의 예산 사용량 공유를 통해 사용자가 재미있고 꾸준하게 서비스를 사용하도록 유도했다. “돈 워리”의 사용자는 돈 관리에 흥미를 느끼고 지속적으로 예산에 맞춘 생활 관리 및 소비 관리가 가능할 것으로 기대한다.

서비스 사용 기간이 늘어나면서 사용자 데이터가 축적되면, 현재는 나이와 성별만 예측 모델의 변수로 사용 중이지만 계절, 날씨, 시간의 흐름 등 추가적인 변수를 적용해 보다 정확하고 개인 맞춤형 서비스로 발전할 수 있다. 또한 서비스를 사용하면서 세이브된 목돈을 활용할 방안(청년 적금 추천 등)을 제시하여 보다 사회초년생에게 도움이 되는 특색있는 서비스로 나아갈 예정이다.