**머신러닝 스터디 2조 프로젝트 기획안**

스터디 원 : 고선욱, 김창현, 유상준, 유장훈

**목차**

1. 주제

2. 데이터 소개

3. 데이터를 얻을 방법

4. 머신러닝을 통해 얻고자 하는 결과

5. 시간 계획

6. 팀원으로서의 각자 목표한 역할

**1. 주제**

새로운 문제해결을 위한 시도나 어떤 가설을 세워 궁금한 점을 해결해보는 것도 좋지만, 이미 존재하는 문제를 해결하는 것도 중요하다 생각합니다.

따라서 저희는 DACON사이트의 '신용카드 사용자 연체 예측 AI 경진대회'를 주제로 선정하게 되었습니다. 대회의 목표는 신용카드 사용자 데이터를 보고 사용자의 대금 연체 정도를 예측하는 알고리즘을 개발하고 성능을 확인하는 것입니다.

텍스트이(가) 표시된 사진

자동 생성된 설명

[ 출처 : 카드고릴라 – 신용카드 연체 후 ,90일 동안 벌어지는 일/https://www.card-gorilla.com/contents/detail/724]

신용카드 연체를 하게 되면 처음엔 연체 이력이 카드사끼리 공유되고 결과적으로 신용등급하락, 신용카드 이용이 정지됩니다. 계속해서 연체해 나간다면 웬만한 대출상품을 이용할 수 없게 되고 흔히 말하는 재산 압류, 신용불량자가 되는 과정을 거치게 됩니다.

이는 개개인의 손해만은 아닙니다. 은행은 돈을 빌려주고 그에 대한 이자를 받고 유통시키면서 돈을 벌어 나가야 하는데 연체가 되면 이익을 내는데 어려움을 겪고 심지어 오히려 손해를 볼 수도 있습니다.

많은 금융업계와 핀테크에서 관심을 가지고 있는 소재이고 저희 분석과 예측을 통해 개인과 기업 모두에게 더 좋은 미래를 선사하고 싶어 주제를 선정하였습니다.

**2. 데이터 소개**

- train.shape : (26457, 20)

- test.shape : (10000, 19)

- features : gender, car, reality, income 등의 18가지 소비자 개인 정보

- label : credit (0,1,2의 값을 가지며, 낮을수록 높은 신용을 의미함)

테이블이(가) 표시된 사진

자동 생성된 설명

**3. 데이터를 얻을 방법**

- https://dacon.io/competitions/official/235713/data

- DACON “신용카드 사용자 연체 예측 AI 경진대회” 제공 데이터 이용

**4. 머신러닝을 통해 얻고자 하는 결과**

1)모델들의 다양한 평가지표의 분석을 통한, feature들과 연체 간의 상관관계 파악 및 해석

2) 대출상담자의 개인정보를 입력하여서 연체여부 및 수준을 예측, 그에 맞는 적절한 상품을 추천

3) 연체확률이 높은 사람에게는 상환기간 전부터 미리 연락을 주는 등의 리스크 관리

(+ 분석을 진행하면서 새로 발견하게 된 인사이트 도출)

**5. 시간 계획**

해야 할 일

* 0. 주제 선정 ( 완료 )
* 1. 전반적인 데이터 살펴보기 + EDA + 시각화 ( 11/28 )

각자 살펴보고 중간 회의 + 방향성 제시 ( 11/28 )

* 2. 결측, 이상치 처리 ( 11/29 – 11/30 )
* 3. 분석 (모델 설계, 구축) ( 11/29 – 11/30 )

하이퍼파라미터 튜닝 ( 11/29 – 11/30 )

* 4. 결과 해석 및 기대효과 제시 ( 12/1 – 12/2 )
* 5. 보고서 작성 및 발표 ( 12/1 – 12/2 )

**6. 팀원으로서의 각자 목표한 역할**

1) 공동 목표

그동안에 배운 EDA, 전처리, 머신러닝 기법 등을 적용해보고 실제상황에서 생기는 다양한 문제들을 직접 해결해 본다. 단기간의 팀프로젝트 경험을 통해서, 팀원들과의 업무적 의사소통 능력을 기르고, 빠르고 효율적인 의사결정 능력을 기르도록한다.

2) 개별 역할 (필요시 자유로운 역할 재 배분)

고선욱 : 기본 EDA, 시각화 + 진행상황 기록하는 노션 페이지 관리

김창현 : 기본 EDA, 시각화 + 개발 계획서, 보고서 작성

유상준 : 기본 EDA, 시각화 + 개발 계획서, 보고서 작성

유장훈 : 기본 EDA, 시각화 + 추가 자료 수집