

\*팀명, 참가자 성명 기재X

조성윤:초록색

고태영:하늘색

수요일 작성: 보라색

목요일 작성: 주황색

## 개요

### (1) 요약

전체적인 요약(배경, 문제점, 해결방안)

### (2) 분석 배경

착안점, 문제 정의(마이데이터 산업의 도입으로 더 많은 정보가 추가)

가설1: 기존 데이터를 통해 산출한 사고율(즉, 보험료)이 잘못될 가능성도 있지 않을까?  
(마이데이터를 사용한 데이터를 이용해 사고율을 예측하면 더 정확할 것이다)(= 마이데이터(소비내역 등)가 추가된다면 더 정확한 보험료의 산출로 보험사의 손실이 줄어들지 않을까?)

가설2: 사고율에 영향을 끼치는 변수들은 무엇이 있고, 그것을 보험회사의 입장에서 어떻게 활용할 수 있을까?

### (3) 선행연구

선행연구를 앞의 가설과 연관되게 설명하는 것이 중요할듯!

<https://careerly.co.kr/comments/19193>

논문

### (4) 아이디어 제안

문제를 어떻게 풀 것인가 간단하게 소개(방향성)

=직면한 문제 및 해결 과정

-가설1번에 대한 검정방식: 예측+ 군집화

-김태영이 빅콘 피피티에 나와있는 '문제상황에 대한 분석개요' 비슷하게 내용 전개

가설2 검정방식: 예측 후 해석(XAI) PDP 등 변수 해석)

## 주요 기술

예선 계획서에서 사용한 알고리즘 구체화/ 추가 알고리즘에 대해 자유롭게 기술

### 예측 모델링

모델링 과정(전처리 기술(encoding(ordinal.one-hot)-> 자세히 물어보기)+k-fold cv(train/test)+ 성능평가

### 군집화

뭐 썼는지 물어보기

### 변수 해석

PDP,ICE,Feature importance

## 데이터 분석 및 모델링 과정

Flow chart 활용, 전체적인 방향을 크게 크게(흐름)

(1) 데이터 수집 및 출처

(2) 데이터 설명 및 EDA

변수 뭐 있는지 등등

(3) 데이터 전처리 <- 김필상한테 물어보기

전처리는 무엇을 했는지, 파생변수는 무엇인지

(4) 데이터 분석 및 모델링

### 들어갈 내용

어떤 모델을 시도했고, 채택했는지, 그 과정에서 (cv) 어떤 방법론을 사용했는지

김태영 빅콘 ppt에 있는 '예측모델링 목표' 사용하기

1. 생명보험

[1] rulefit 사용

2. 장기+자동차보험

[1] 예측 모델링1: 사고율을 대체하기 위한 값을 찾기 위한 모델링(사고유무 변수를

classification으로 모델링하여 결과로 산출된 확률값을 사고율의 대체변수로 결정(기존 보험 데이터만 사용)

+ 모델링이 끝난 후 변수 중요도를 그래프로 그린 후, scree 부분을 잘라서 (pdp,ice)로 확인할 변수의 개수를 정함-> 각 변수의 영향력을 해석

[2] 예측 모델링2: (보험+카드 데이터) 사고유무로 classification 모델링해서 결과로 산출된 확률값을 사고율의 대체변수로 결정

+ 모델링이 끝난 후 변수 중요도를 그래프로 그린 후, scree 부분을 잘라서 (pdp,ice)로 확인할 변수의 개수를 정함-> 각 변수의 영향력을 해석

[3] [1] 사고율 vs [2] 사고율 군집화

+ 각 군집 해석

## 데이터 분석 및 모델링 방법론

위 과정의 목차를 따르되,기술적인 부분 자세히 서술

(1) 데이터 수집 및 출처

(2) 데이터 설명 및 EDA

변수 뭐 있는지 등등

(3) 데이터 전처리

전처리 이유, 파생변수 도출 이유, 이산형변수 인코딩

(4) 예측모델링

여러 모델 설명

(5) 군집화

## 최종결과

(1) 핵심 아이디어(=차별점)

부딪힌 문제점에 대한 해결방안

-데이터에서 사고율이 0 혹은 NULL 값 어떻게 해결했나?(처음에 제시해준 공식 적용을 못함)

= 해당 문제를 해결하기 위해서 시도했던 방법론(김필상이 사고율 대체하기 위한 7개 방법론)

## (2) 결과 요약

분석 결과 자세히!

-가설에 대한 검정 결과

-페르소나

## (3) 기대효과 및 제언

기대효과 및 의의, 한계점, 분석 결과 적용 방법, 활용방안

-페르소나 이용해서 서비스 제안

## (4) 결론

느낌점, 배운점, 고민했던 점

## (5) 참고문헌 + 분석 환경(라이브러리)