

# [2020 금융 빅데이터 페스티벌] 과제설명자료

(과제2) 보험금 청구 건에 대한 자동지급, 심사, 조사분류 예측

---

미래에셋생명보험

2020.08

# 목차

---

1. 과제 개요
  2. 분석 데이터
  3. 답안 제출 방법
  4. 평가 방법
  5. FAQ
- [별첨자료 – 코드 설명]

# 1. 과제 개요

## 과제 소개

- 보험 고객은 질병 발생 시 보험금을 신청합니다. (보험금 신청을 청구라고 합니다.)
- 고객이 보험금을 신청하면 위험도(청구 적정성)를 판별하여 자동지급, 심사, 조사로 분류한 후, 보험금 지급 여부를 결정합니다.
- 보험금 청구 건에 대한 분류 결과를 Target으로 하여 다음 달의 청구 건에 대한 분류 결과를 맞춰 보세요.

## 보험금 청구 프로세스

(자료: 당사 홈페이지)



## 2. 분석 데이터

### 분석 데이터 설명

- 분석용 데이터는 2019년 1월부터 11월까지의 월별 보험금 청구(신청) 데이터로 구성되어 있습니다.  
(2019년 12월 데이터는 참가자 제출 답안입니다.)
- 청구 건별로 고객, 상품, 판매자, 질병정보가 나열되어 있으며 모든 정보는 접수일련번호에 대해 Unique합니다.

### 속성별 컬럼 구분

컬럼속성구분		해당컬럼
일련번호(PK)		접수일련번호
접수시간정보		접수년월
질병정보	질병분류정보	질병경중등급코드, KCD등급코드, 질병구분코드
	발생지역정보	발생지역구분코드
	치료행위정보	치료행위코드, 통원횟수, 청구보험금, 입원일수
	의료기관분류정보	의료기관구분코드, 요양병원여부
	병원별집계정보	병원별평균입원/통원/수술/진단청구보험금
	질병별집계정보	질병별평균입원/통원/수술/진단청구보험금
	기타청구정보	입원/수술/통원청구건수
고객정보	연령정보	고객나이구분코드
	부정행위정보	보험사기이력고객여부
상품정보	가입당시정보	보험료구간코드, 가입금액구간코드
	상품유지관리정보	중도부가계약여부, 건강인우대계약가입여부
	부담보관련정보	부담보5년경과여부
	보험금청구당시정보	청구일계약일간기간, 청구일부활일간기간
판매자정보	부정행위정보	부실판매자계약여부
타겟정보(Y)		실사상태코드

## 2. 분석 데이터

### 변수 설명(질병정보)

변수명	변수코드	설명	속성구분	부가설명
질병경중등급코드	dsas_ltwl_gcd	1~5로 부여된 질병의 경중 정보	질병	[별첨1] 질병분류 참조
KCD등급코드	kcd_gcd	KCD 분류기준으로 구성된 질병정보	질병	[별첨1] 질병분류 참조
질병구분코드	dsas_acd_rst_dcd	질병유형에 따른 구분코드	질병	[별첨1] 질병분류 참조
발생지역구분코드	ar_rclss_cd	발생지역을 구분하는 코드	질병	[별첨2] 지역, 연령구분 참조
치료행위코드	blrs_cd	질병으로 인해 치료한 행위 분류	질병	[별첨3] 치료행위구분 참조
의료기관구분코드	mdct_inu_rclss_dcd	의료기관 등급에 따른 코드	질병	[별첨3] 치료행위구분 참조
요양병원여부	nur_hosp_yn	요양병원 여부	질병	1(Yes), 0(No)
통원횟수	optt_nbtm_s	접수 건별 총 통원횟수 Index	질병	익명처리를 위해 값 변환 및 복제가 진행됨 (실제 분포에 기반한 Index 값)
청구보험금	bilg_isamt_s	접수 건별 청구보험금 총액 Index	질병	
입원일수	hspz_dys_s	접수 건별 총 입원일수 Index	질병	
병원별평균입원청구보험금	hsp_avg_hspz_bilg_isamt_s	병원별 입원 사유로 청구된 보험금의 평균 금액 Index	질병	
병원별평균통원청구보험금	hsp_avg_optt_bilg_isamt_s	병원별 통원 사유로 청구된 보험금의 평균 금액 Index	질병	
병원별평균수술청구보험금	hsp_avg_surop_bilg_isamt_s	병원별 수술 사유로 청구된 보험금의 평균 금액 Index	질병	
병원별평균진단청구보험금	hsp_avg_diag_bilg_isamt_s	병원별 진단 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균입원청구보험금	dsas_avg_hspz_bilg_isamt_s	질병군별 입원 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균통원청구보험금	dsas_avg_optt_bilg_isamt_s	질병군별 통원 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균수술청구보험금	dsas_avg_surop_bilg_isamt_s	질병군별 수술 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균진단청구보험금	dsas_avg_diag_bilg_isamt_s	질병군별 진단 사유로 청구된 보험금의 평균 금액 Index	질병	
입원청구건수	hspz_blcnt_s	접수건별 입원으로 표기된 청구 건의 수 Index	질병	
수술청구건수	surop_blcnt_s	접수건별 수술로 표기된 청구 건의 수 Index	질병	
통원청구건수	optt_blcnt_s	접수건별 통원으로 표기된 청구 건의 수 Index	질병	

## 2. 분석 데이터

### 변수 설명(고객, 상품, 판매자정보)

변수명	변수코드	설명	속성구분	부가설명
고객나이구분코드	isrd_age_dcd	고객의 연령구분코드	고객	[별첨2] 지역, 연령구분 참조
보험사기이력고객여부	fds_cust_yn	보험금을 청구한 고객이 이전에 보험사기를 친 적이 있는지 여부	고객	1(Yes), 0(No)
부담보5년경과여부	smrtg_5y_passed_yn	부담보 설정일로부터 5년 경과했는지 여부	상품	1(Yes), 0(No)
중도부가계약여부	mtad_cntr_yn	상품 가입 후 중간에 질병보장내용을 추가했는지 여부	상품	1(Yes), 0(No)
건강인우대계약여부	heltp_pf_ntyn	가입한 상품이 건강인을 우대해주는 상품인지 여부	상품	1(Yes), 0(No)
보험료구간코드	prm_nvcd	납입 보험료의 구간	상품	[별첨4] 보험료/가입금액 참조
가입금액구간코드	inamt_nvcd	가입금액의 구간	상품	[별첨4] 보험료/가입금액 참조
청구일계약일간기간구분코드	ac_ctr_diff	보험금 청구일자로부터 최초 계약일까지의 기간 구분	상품	[별첨5] 기간구분코드 참조
청구일부활일간기간구분코드	ac_rst_diff	보험금 청구일자로부터 계약 부활일까지의 기간 구분	상품	[별첨5] 기간구분코드 참조
부실판매자계약여부	urlb_fc_yn	보험금 청구된 계약의 판매자가 부실판매자인지 여부	판매자	1(Yes), 0(No)

### 변수 설명(ID, 시간, 타겟)

변수명	변수코드	설명	속성구분	부가설명
접수일련번호	ID	임의 부여된 접수번호	ID	
접수년월	base_ym	접수가 진행된 월	시간	
실사상태코드	target	자동지급(0), 심사(1), 조사(2)로 최종 배정된 상태	타겟	Train Set에만 존재

## 2. 분석 데이터

### 변수 설명(질병정보)

변수명	변수코드	설명	속성구분	부가설명
질병경중등급코드	dsas_ltwl_gcd	1~5로 부여된 질병의 경중 정보	질병	[별첨1] 질병분류 참조
KCD등급코드	kcd_gcd	KCD 분류기준으로 구성된 질병정보	질병	[별첨1] 질병분류 참조
질병구분코드	dsas_acd_rst_dcd	질병유형에 따른 구분코드	질병	[별첨1] 질병분류 참조
발생지역구분코드	ar_rclss_cd	발생지역을 구분하는 코드	질병	[별첨2] 지역, 연령구분 참조
치료행위코드	blrs_cd	질병으로 인해 치료한 행위 분류	질병	[별첨3] 치료행위구분 참조
의료기관구분코드	mdct_inu_rclss_dcd	의료기관 등급에 따른 코드	질병	[별첨3] 치료행위구분 참조
요양병원여부	nur_hosp_yn	요양병원 여부	질병	1(Yes), 0(No)
통원횟수	optt_nbtm_s	접수 건별 총 통원횟수 Index	질병	익명처리를 위해 값 변환 및 복제가 진행됨 (실제 분포에 기반한 Index 값)
청구보험금	bilg_isamt_s	접수 건별 청구보험금 총액 Index	질병	
입원일수	hspz_dys_s	접수 건별 총 입원일수 Index	질병	
병원별평균입원청구보험금	hsp_avg_hspz_bilg_isamt_s	병원별 입원 사유로 청구된 보험금의 평균 금액 Index	질병	
병원별평균통원청구보험금	hsp_avg_optt_bilg_isamt_s	병원별 통원 사유로 청구된 보험금의 평균 금액 Index	질병	
병원별평균수술청구보험금	hsp_avg_surop_bilg_isamt_s	병원별 수술 사유로 청구된 보험금의 평균 금액 Index	질병	
병원별평균진단청구보험금	hsp_avg_diag_bilg_isamt_s	병원별 진단 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균입원청구보험금	dsas_avg_hspz_bilg_isamt_s	질병군별 입원 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균통원청구보험금	dsas_avg_optt_bilg_isamt_s	질병군별 통원 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균수술청구보험금	dsas_avg_surop_bilg_isamt_s	질병군별 수술 사유로 청구된 보험금의 평균 금액 Index	질병	
질병별평균진단청구보험금	dsas_avg_diag_bilg_isamt_s	질병군별 진단 사유로 청구된 보험금의 평균 금액 Index	질병	
입원청구건수	hspz_blcnt_s	접수건별 입원으로 표기된 청구 건의 수 Index	질병	
수술청구건수	surop_blcnt_s	접수건별 수술로 표기된 청구 건의 수 Index	질병	
통원청구건수	optt_blcnt_s	접수건별 통원으로 표기된 청구 건의 수 Index	질병	

### 3. 답안 제출 방법

#### 파일의 구성

구분	학습용 데이터셋	테스트용 데이터셋(문제 파일)	제출용 파일(답안 파일)
파일명	train.csv	test.csv	sample.csv
구성	ID+속성+타겟	ID+속성	ID+타겟
데이터 건수	377,928	22,072	22,072
기간	2019년 1월 ~ 2019년 11월	2019년 12월	작동

#### 제출 방법

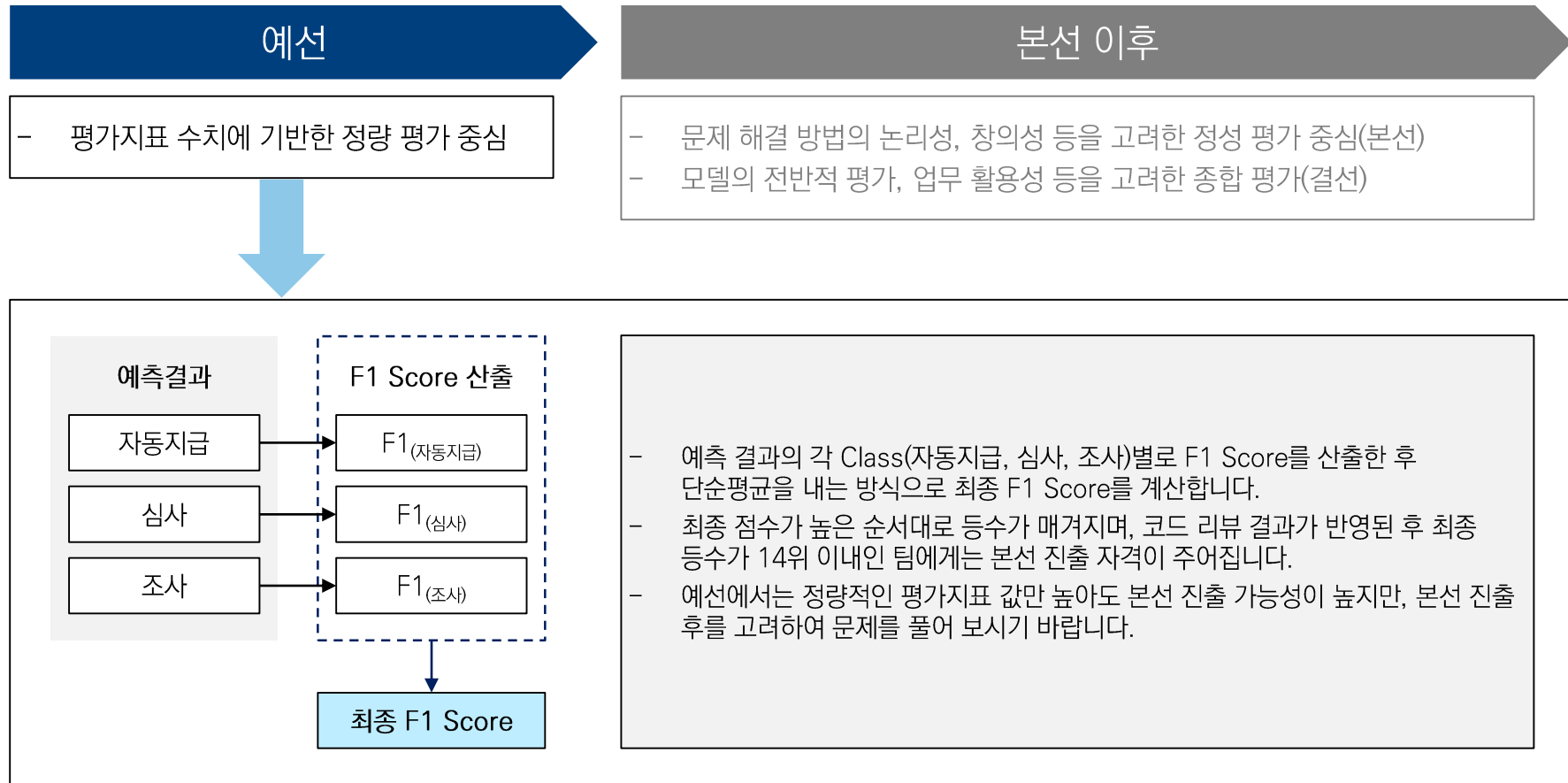
- 학습용 데이터셋을 이용해 모델을 만드신 후 문제 파일을 이용해 타겟 값을 예측합니다.
- 예측한 타겟 값을 답안 파일에 적어 프로그래머스 “테스트 입장” 페이지에 업로드하여 제출합니다.  
(대회 페이지 접속 → “테스트 입장” 버튼 클릭 → 파일 업로드) ※ 변경사항 발생 시 추후 SMS 통해 공지

#### 주의사항

- 답안 파일에는 모든 값이 빠짐없이 입력해 주시되, ID값을 삭제 또는 임의로 변경하지 말아주세요.  
(Test Set의 ID는 0부터 22071까지 Unique한 값, 중복이 있거나 누락되거나 다른 값이 들어가는 경우 0점 처리됩니다.)
- 답안 파일의 변수명(ID, target)을 삭제하거나 파일의 형식(csv)을 임의로 변경하지 말아주세요.  
(위와 마찬가지로 답안 파일의 기본 형식을 변경하면 0점 처리됩니다.)



## 4. 평가 방법



## 5. FAQ

Q

1명의 고객이 여러 건의 보험금 청구를 한 번에 진행했다면 데이터는 어떻게 구성되나요?

A

1개의 질병에 대해 여러 치료행위를 한 번에 한 경우, 하나의 Row로 구성됩니다.  
단, 상품이 상이하거나 질병이 다른 경우 여러 줄의 Row로 구성됩니다.

Q

부담보 관련 변수는 왜 특정 시점(5년)이 지났는지 여부를 확인하나요?

A

보험에 가입할 때 현재 앓고 있거나, 과거에 앓았던 경험이 있는 질병에 대해  
보험금을 지급하지 않는 “부담보” 처리를 하는 경우가 있습니다.  
이 때 부담보는 기간을 지정하는 경우 최대 5년 간 부여할 수 있으며,  
5년 간 치료이력이 없는 경우 부담보 해제를 할 수 있으므로 5년 경과 여부를 확인할 수 있습니다.

Q

일자, 금액 등의 변수가 실제 값이랑 다른 것 같아요. 왜 이런 건가요?

A

제4회 미래에셋 빅데이터 페스티벌에 사용된 데이터는  
개인정보법 개정 시행령에 따라 익명처리 적정성 평가를 받은 데이터입니다.  
이 과정에서 실제 고객 정보를 추정할 수 없도록 값을 수정하였기 때문에, 실제 값과 차이가 있습니다.  
다만, 분포는 유사하므로 이 점 고려하셔서 분석을 진행해 주시면 됩니다.

# 별첨자료

---

1. [별첨1] 질병분류정보 코드설명
2. [별첨2] 지역, 연령구분정보 코드설명
3. [별첨3] 치료행위정보 코드설명
4. [별첨4] 보험료, 가입금액정보 코드설명
5. [별첨5] 청구일계약일(부활일)간 기간구분코드설명

# [별첨1] 질병구분정보 코드설명

## 1 질병경중등급코드      2 질병구분코드

질병경중등급코드	질병경중등급명	질병구분코드	질병명
1	중증	1	암
		2	상피내암
		3	경계성
		4	심장질환
		5	
		6	뇌혈관질환
2	성인	7	간질환
		8	
		9	신장질환
		10	갑상선질환
		11	폐렴
		12	천식
		13	위궤양
		14	십이지장궤양
3	생활	15	고혈압
		16	당뇨병
		17	관절염
		18	
		19	
		20	
		21	골다공증
		22	백내장
		23	중이염
		24	중수염
		25	남성비뇨기계
		26	
		27	부인과
		28	

## 3 KCD등급코드

KCD등급코드	질병기준	KCD등급코드명
1	A	감염성 및 기생충성 질환1
2	B	감염성 및 기생충성 질환2
3	C	신생물(암) 질환
4	D	신생물(기타)질환
5	E	내분비 질환
6	F	정신 질환
7	G	신경계통 질환
8	H	눈, 귀 질환
9	I	순환기 질환
10	J	호흡기 질환
11	K	소화기 질환
12	L	피부 질환
13	M	근골격계 질환
14	N	비뇨생식기 질환
15	O	임신, 출산 질환
16	P	주산기 질환
17	Q	선천 질환
18	R	달리 분류되지 않은 질환

▲ KCD 질병코드 기준별 분류

▲ 병명의 경중도 및 주요 보상 유형별 분류

## [별첨2] 지역구분정보 코드설명

4

지역구분코드

코드	지역명
1	서울특별시
1	경기도
1	인천광역시
1	강원도
2	부산광역시
2	경상남도
3	대구광역시
3	경상북도
4	대전광역시
4	세종특별자치시
4	충청남도
4	충청북도
5	광주광역시
5	전라남도
5	전라북도
9	제주특별자치도
9	기타

5

연령구분코드

코드	연령
0	0~9세
1	10~19세
2	20~29세
3	30~39세
4	40~49세
5	50~59세
6	60세 이상

↑ 통상적인 연령대 구분방법 적용

← 생활권 기준 분류

## [별첨3] 치료행위정보 코드설명

6

치료행위코드

코드	치료행위				설명
	입원	통원	수술	진단	
1				Y	질병 진단만 받음
2			Y		수술치료만 진행
3			Y	Y	질병 진단 받고 수술치료 진행
4		Y			통원치료만 진행
5		Y		Y	질병 진단 받고 통원치료 진행
6		Y	Y		수술치료 후 통원치료 진행
7		Y	Y	Y	진단 받고, 수술도 받고, 통원치료도 받음
8	Y				입원치료만 진행
9	Y			Y	질병 진단 받고 입원치료 진행
10	Y		Y		입원 및 수술치료 진행
11	Y		Y	Y	질병 진단 받고 입원 및 수술치료 진행
12	Y	Y			입원 및 통원치료 진행
13	Y	Y		Y	질병 진단 받고 입원 및 통원치료 진행
14	Y	Y	Y		입원, 수술, 통원치료 모두 진행
15	Y	Y	Y	Y	질병 진단 받고 입원, 수술, 통원치료 모두 진행

7

의료기관구분코드

코드	의료기관구분
1	1차병원(일반병원)
2	2차병원(종합병원)
3	3차병원(상급종합병원)
9	Unknown

↑ 건강보험심사평가원 의료전달체계 기준 분류

← 치료행위(입원, 통원, 수술, 진단)의 조합

## [별첨4] 보험료, 가입금액정보 코드설명

### 8 보험료구간코드

코드	보험료구간
1	0~10만원 미만
2	10만~20만원 미만
3	20만~30만원 미만
4	30만~40만원 미만
5	40만~50만원 미만
6	50만~60만원 미만
7	60만~70만원 미만
8	70만~80만원 미만
9	80만~90만원 미만
10	90만~100만원 미만
11	100만원 이상
99	Unknown

### 9 가입금액구간코드

코드	가입금액구간
1	0~1000만원 미만
2	1000만~2000만원 미만
3	2000만~3000만원 미만
4	3000만~4000만원 미만
5	4000만~5000만원 미만
6	5000만~6000만원 미만
7	6000만~7000만원 미만
8	7000만~8000만원 미만
9	8000만~9000만원 미만
10	9000만~1억원 미만
11	1억원 이상
99	Unknown

## [별첨5] 청구일계약일(부활일)간 기간구분코드설명

10

청구일계약일간기간구분코드(청구일부활일간기간구분코드 동일)

코드	기간	
0	청구일계약일간기간구분코드	Unknown
	청구일부활일간기간구분코드	부활계약이 아닌 계약
1	1일 이상 3개월 이하	
2	3개월 초과 6개월 이하	
3	6개월 초과 1년 이하	
4	1년 초과 2년 이하	
5	2년 초과 5년 이하	
6	5년 초과	



감사합니다

