

# 머신러닝 기반의 자동차보험 사고 환자의 진료 기간 예측 기술

변 경 근\*, 이 덕 규\*\*, 이 형 동\*\*

## 요 약

자동차보험 사고 환자의 진료비 감소를 위한 대책 마련에 도움을 주기 위해 본 연구에서는 자동차보험 사고 40대~50대 정상 환자들의 진료비에 가장 핵심 요소인 진료 기간을 예측하고 진료 기간에 영향을 미치는 요인을 분석하였다. 이를 위해 Decision Tree 등 5개 알고리즘을 활용한 머신러닝 모델을 생성하고 모델간에 그 성능을 비교 분석하였다. 진료 기간 예측에 정밀도, 재현율, F1 점수 등 3가지 평가 지표에서 좋은 성능을 나타낸 알고리즘은 Decision Tree, Gradient Boosting 및 XGBoost 등 3가지였다. 그리고 진료 기간 예측에 영향을 미치는 요인 분석 결과, 병원의 종류, 진료 지역, 나이, 성별 등으로 나타났다. 본 연구를 통해 AutoML을 활용한 손쉬운 연구 방법을 제시하였으며, 본 연구 결과가 자동차보험 사고 진료비 감감을 위한 정책에 도움이 되기를 기대한다.

## Machine Learning-Based Prediction Technology for Medical Treatment Period of Automobile Insurance Accident Patients

Kyung-Keun Byun\*, Doeg-Gyu Lee\*\*, Hyung-Dong Lee\*\*

## ABSTRACT

In order to help reduce the medical expenses of patients with auto insurance accidents, this study predicted the treatment period, which is the most important factor in the medical expenses of patients in their 40s and 50s, and analyzed the factors affecting the treatment period. To this end, a machine learning model using five algorithms such as Decision Tree was created, and its performance was compared and analyzed between models. There were three algorithms that showed good performance including Decision Tree, Gradient Boost, and XGBoost. In addition, as a result of analyzing the factors affecting the prediction of the treatment period, the type of hospital, the treatment area, age, and gender were found. Through these studies, easy research methods such as the use of AutoML were presented, and we hope that the results of this study will help policies to reduce medical expenses for automobile insurance accidents.

**Key words : Machine Learning Models, the Prediction of the Treatment Period, Car Insurance Accidents, the Factors Affecting the Treatment Period, AutoML**

접수일(2023년 02월 28일), 수정일(2023년 03월 15일),  
게재확정일(2023년 03월 30일)

\* 송실대학교/IT정책경영대학원(주저자, 교신저자)

\*\* 송실대학교/IT정책경영대학원(공동저자)

## 1. 서 론

자동차보험 사고 환자의 진료비가 2016년 1조 6,590억에서 2021년 2조 3916억원으로 지속 증가하고 있다[1]. 이러한 진료비 증가에는 자동차보험 사고 경상 환자들의 진료비가 큰 부분을 차지한다. 자동차보험사, 건강보험심사평가원(이하 심평원) 등 관련 기관들이 자동차보험 사고 진료비의 둔화를 위해 전방위적으로 노력하고 있다[1][2][3][4]. 예를 들어 심평원은 2022년 5월부터 ‘입원료 및 상급병실료’ 심사지침까지 제정하여 교통사고 경상 환자의 입원 건에 대한 정밀 심사를 추진하였다. 이와 같은 배경하에, 진료비 감소를 위한 대책 마련에 도움을 주기 위해 본 연구에서는 자동차보험 사고 경상 환자들의 진료비에 가장 핵심 요소인 진료 기간을 예측하고 진료 기간에 영향을 미치는 요인을 분석하는 머신러닝 모델을 생성하고 모델간 성능을 비교·분석하고자 한다.

## 2. 관련 연구

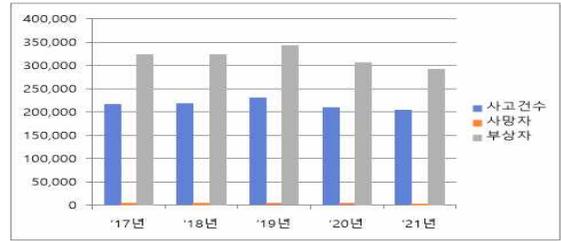
### 2.1 자동차 사고 현황

<표 1>과 (그림 1)과 같이 2017년부터 2019년까지 자동차 사고 건수가 증가하다가 2019년부터 2021년까지는 사고 건수가 줄어들었다. 그러나 사망자 수는 2017년부터 2021년까지 지속적으로 감소하였다. 반면 부상자 수는 자동차 사고 건수와 유사하게 2017년부터 2019년까지 증가하다가 2019년부터 2021년까지는 감소하였다[5].

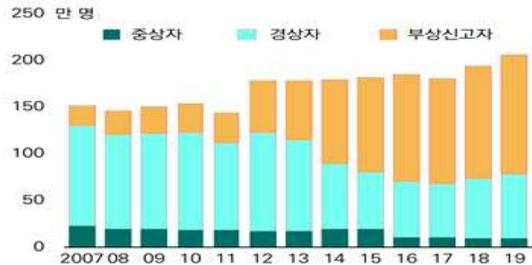
<표. 1> 연도별 사고 건수 및 사망자 부상자 수[5]

유형	'17년	'18년	'19년	'20년	'21년
사고(건)	216,335	217,148	229,600	209,654	203,130
사망자(명)	4,185	3,781	3,349	3,081	2,916
부상자(명)	322,829	323,037	341,712	306,194	291,608

그리고, 부상자 수를 좀 더 세분화해서 살펴보면 (그림 2)와 같이 2007년 이후 부상신고자 수는 증가하고 있지만 중상자와 경상자 수는 계속해서 감소하는 추세이다. 다만, 2012년, 2018년에서 2019년에는 부상신고자 수가 증가하였다[5].



(그림 1) 연도별 사고 건수 및 사망자 부상자 수[2]



(그림 2) 연도별 교통사고 부상자 수[5]

\* 경상자는 3주미만, 부상신고자는 5일 미만 치료

한편, 이러한 부상자와 관련된 연령별 자동차보험 진료비 명세서 건수를 살펴보면 <표 2>와 같이 40대와 50대가 1·2위를 차지하고 있다. 따라서 본 연구에서는 자동차보험 사고 경상 환자 분석 대상을 40대~50대로 선택하였다.

<표 2> 자동차보험 진료비 통계[6]

연령별	2021년		
	명세서 건수(건)	진료비 (억원)	입원일수 (일)
전체	19,540,004	23,916	25,791,537
10대미만	262,976	245	325,446
10대	407,545	620	569,285
20대	2,418,663	3,368	3,175,123
30대	3,957,986	4,369	4,731,961
40대	4,032,971	4,258	4,937,733
50대	3,999,243	4,473	5,220,806
60대	3,038,695	3,774	4,255,339
70대이상	1,421,925	2,805	2,575,844

### 2.2 관련 연구

Ghaemmaghami는 10년간 1급 외상센터에 입원한 외상환자들의 데이터를 수집하고 다양한 머신

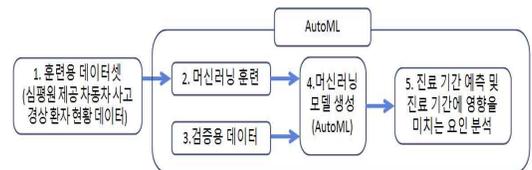
러닝 알고리즘을 이용해 치료 기간을 예측할 수 있는 모델을 개발하였다[7]. 기계 학습 모델이 나이, 부상 심각도 등 몇 가지 변수를 포함하여 치료 기간의 주요 예측 변수를 식별할 수 있다는 결과를 도출했다[7]. 또한, 기계 학습 모델이 외상환자의 치료 기간을 예측하고 환자 관리에 도움이 될 수 있음을 보여주었다[7]. Adeleke 등 연구원들은 교통사고 피해자의 입원 기간의 예측 관련 다중 선형 회귀 분석과 의사 결정 트리 기법을 비교·분석하였다[8]. 교통사고로 입원한 450명의 환자를 대상으로 나이, 성별, 부상 유형, 병력 등 다양한 요인에 대한 데이터를 수집하여 다중 선형 회귀 분석과 의사 결정 트리 기법을 사용, 입원 기간을 예측할 수 있는 모델을 만들었다[8]. 의사결정 트리 모델의 입원 기간 예측 성능이 다중 선형 회귀 모델보다 우수한 성능을 보였다[8]. 의사결정 트리 분석이 도로 교통 사고 피해자의 입원 기간 예측 및 환자 치료, 의료자원 할당에 활용될 수 있음을 보여주었다[8]. Almuheidi 등은 연구를 통해 자동차 충돌 후 장기 입원의 예측 변수를 식별하고 분석하였다[9]. 자동차 사고로 입원한 환자들을 대상으로 장기 입원과 관련된 요인을 조사한 연구들을 체계적으로 검토하고 메타분석을 실시했다[9]. 34개의 관련 연구를 확인하고 장기 입원을 잠재적으로 예측할 수 있는 다양한 인구통계학적, 부상 관련 및 임상적 요인에 대한 데이터를 추출했다[9]. 자동차 충돌 후 장기 입원에 가장 영향을 미치는 변수는 많은 나이와 부상 심각도였으며, 머리와 척추 부상과 같은 치명적 부상도 영향이 컸다[9]. 이 연구에서는 적절한 치료와 의료자원을 효율적으로 배분하기 위해 자동차 사고 후 장기 입원 예측 및 이에 영향을 미치는 요인의 도출을 연구하였다[9]. 한편, Abujaber 등의 연구에서는 총 15만여명의 환자가 포함된 47개의 연구에 대해 체계적인 검토와 메타분석을 수행하였다[10]. 이를 통해 자동차 충돌 환자의 장기 입원(7일 이상 입원)을 예측하는 요인을 식별하였다[10]. 그 결과 고령, 여성, 부상의 심각도, 머리 부상, 척추 부상, 병원 이송 시간 등의 요인들이 장기 입원과 유의미하게 연관되어 있다는 것을 발견했다[10]. 이와 같이 기존 연구들은

진료 기간 예측을 위한 환자 데이터의 수가 적거나 환자 데이터 수가 많은 경우에는 인공지능과 같은 최신 기술 적용 없이 분석이 진행되었다. 이에, 대규모 환자 데이터와 최신 머신러닝 기술을 활용하여 자동차 사고 환자 진료 기간 예측 모델을 만들고 진료 기간에 영향을 미치는 요인의 분석을 위해 Decision Tree, Random Forest, Gradient Boosting, XGBoost, Logistic Regression 등 5개 알고리즘을 활용하여 성능을 비교·분석하고자 한다.

### 3. 연구모형 및 방법

#### 3.1 연구모형

(그림 3)과 같이 자동차보험 사고 40~50대 환자 현황 데이터를 기반으로 머신러닝 기술을 활용, 훈련(전체 데이터의 80%)을 실시하였다. 머신러닝 모델을 생성한 후 검증용 데이터(전체 데이터의 20%)를 통해 진료 기간을 예측 모델을 수립하고 그에 영향을 미치는 요인(나이, 성별, 진료 소재지, 병원 구분, 일일 평균 입원비, 진료 기간)을 분석하였다.



(그림 3) 연구모형

#### 3.2 연구방법

심평원에서 제공한 2020년 자동차보험 사고 경상 환자(S13 : 목부위의 관절 및 인대의 탈구, 염좌 및 긴장, 한방 진료를 중심으로) 현황 데이터 중 40대~50대 남녀 환자(122,926건) 데이터를 선별하여 <표 3>과 같이 진료 기간에 영향을 미치는 다양한 요인 즉 △나이 △성별 △병원 구분 △진료 지역 △일일 평균 입원비 △진료 기간 등의 변수를 정의하고 관련 데이터를 정리하였다.

<표 3> 환자정보 데이터 구조

변수명	구분
나이	40대/50대
성별	남/여
병원 구분	한방병원/한의원
진료 지역	1(서울,인천,경기), 2(부산,경남,울산), 3(대구,경북), 4(광주,전남,전북), 5(대전,충남,충북), 6(강원도,제주도)
일일 평균 입원비(원)	1(39,999 이하), 2(40,000 ~ 79,999), 3(80,000 ~ 119,999), 4(120,000 ~ 159,999), 5(160,000 ~ )
진료 기간 (입원 기간)	1(11일 미만)/2(11일 이상)

또한 목표변수의 설정은 의료기관에서 청구한 환자별 명세서당 한방 진료 기간을 11일 미만까지와 11일 이상으로 이원화 구분(전체 한방입원 기간의 산술 평균값이 10.10일이며, 중간값도 10일이었음)하였다. 한편, 진료 기간 예측 및 진료 기간에 영향을 미치는 요인 분석을 신속하고 효율적으로 수행하기 위해 머신러닝 기반의 AutoML(WiseProphet)을 활용하였다. <표 4>와 같이 WiseProphet(<http://prophet.wise.co.kr/>)에서 제공하는 △Decision Tree △Random Forest △Gradient Boosting △XGBoost △Logistic Regression 등의 5개 알고리즘과 알고리즘별로 파라미터값을 입력하여, 진료 기간을 예측·비교하였다.

<표 4> 5개 알고리즘별 입력 파라미터값

알고리즘 종류	파라미터값
Decision Tree	△criterion(gini) △max_depth(5) △min_samples(leaf 1)
Random Forest	△criterion(gini) △max_depth(5) △min_samples_leaf(1) △n_estimators(10) △random_state(None)
Gradient Boosting	△learning_rate(0.1) △max_depth(3) △subsample(1.0)
XGBoost	△eta(0.3) △gamma(0) △max_depth(3)
Logistic Regression	△C(1.0) △random_state(None)

그리고, 진료 기간 예측 성능평가 지표는 <표 5>과 같이 정밀도, 재현율 및 F1 점수를 사용하였다. 진료 기간에 영향을 미치는 요인 분석은 WiseProphet에서 제공되는 변수 간 상관관계 분석 기

능을 이용하였다.

<표 5> 성능평가 지표

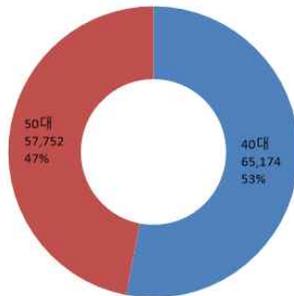
(출처 : <https://bhcboy100.medium.com/>)

측정 산식		
△ 정밀도(Precision)	$\frac{TP}{(TP+FP)}$	
	* 예측 값 중 실제 값이 발생하는 비율	
△ 재현율(Recall)	$\frac{TP}{(TP+FN)}$	
	* 모델에서 분류한 값이 정확하게 탐지한 정답 비율	
	구분	예측
		Positive Negative
실	Positive	True Positive(TP) False Negative(FN)
제	Negative	False Positive(FP) True Negative(TN)
△ F1 점수 = 정밀도와 재현율의 조화평균(H)		
$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{1}{\sum_{i=1}^n \frac{1}{x_n}} = \left( \frac{\sum_{i=1}^n \frac{1}{x_n}}{n} \right)^{-1}$		

## 4. 성능 분석 결과

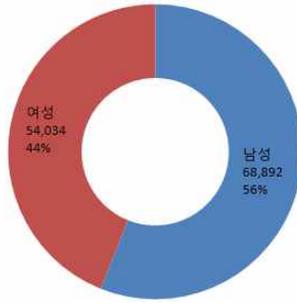
### 4.1 연구 데이터 분석

자동차보험 사고 40대~50대 경상 환자 데이터의 나이, 성별, 병원 구분, 진료 지역, 일일 평균 입원비, 진료 기간 등의 일반 통계를 살펴보면 (그림 4)와 같이 나이는 40대가 53%(65,174명), 50대가 47%(57,752명)를 차지하였다.



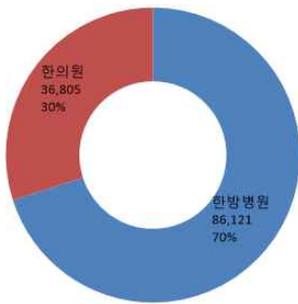
(그림 4.) 나이 비율

(그림 5)와 같이 성별은 남성이 56%(68,892명), 여성이 44%(54,034명) 이었다.



(그림 5) 성별 비율

(그림 6)과 같이 병원 구분에 있어서는 한방병원 원이 70%(86,121곳), 한의원이 30%(36,805곳)를 차지하였다.



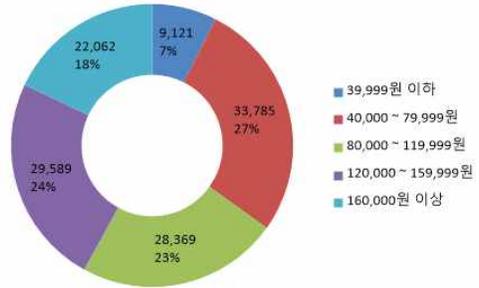
(그림 6) 병원 구분 비율

(그림 7)과 같이 진료 지역 비율은 서울·경기·인천 46%(56,881건), 광주·전남·전북 21%(26,224건) 대전·충남·충북 16%(19,803건), 부산·경남·울산 11%(13,675건), 대구·경북 4%(4,938건), 강원도·제주도 1%(1,405건) 등으로 나타났다.



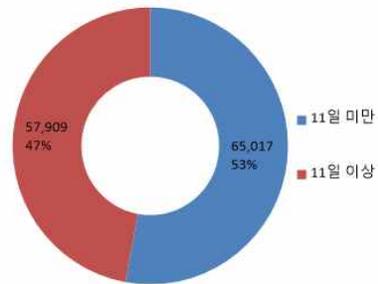
(그림 7) 진료 지역 비율

한편, (그림 8)과 같이 일일 평균 입원비는 40,000~79,999원은 27%(33,785건), 120,000~159,999원은 24%(29,589건), 80,000~119,999원은 23%(28,369건), 160,000원 이상은 18%(22,062건), 39,999원 이하는 7%(9,121건)를 차지하였다.



(그림 8) 일일 평균 입원비 비율

마지막으로 입원 기간은 (그림 9)와 같이 11일 미만은 55%(65,017건), 11일 이상은 47%(57,909건) 이었다.



(그림 9) 입원 기간 비율

## 4.2 실험 결과

본 연구를 통해 머신러닝 기법을 활용하여 자동차보험 사고 40대~50대 경상 환자 대상 5개 알고리즘 중 진료 기간 예측에 좋은 성능을 나타낸 알고리즘은 Decision Tree와 Gradient Boosting, XGBoost 등 3가지였다. 그리고 진료 기간 예측에 영향을 미치는 요인 분석 결과, 병원의 종류(한방병원) 진료 지역(광주, 전남, 전북), 나이, 성별(남성) 순이었다. 진료 기간 11일 미만 예측률이 진료 기간 11일 이상 예측률보다

높았다.

좀 더 자세히 살펴보면 진료 기간이 11일 미만 을 예측한 결과는 <표 6>과 같이 5개 알고리즘에 서 예측한 정밀도의 -평균 값은 0.81이며 평균 재 현율은 0.80, F1 점수는 0.81이었다. 정밀도는 5개 알고리즘 모두가 우수했고 재현율 및 F1 점수는 D.T. · G.B. · X.B.알고리즘이 우수했다. 한편 진료 기간 예측률에 가장 큰 영향을 미치는 요인은 <표 8>과 같이 병원구분(한방병원)이었고 진료 지역중에는 광주·전남·전북이었으며 그 외에 나이, 성별(남 성) 순이었다.

<표 6> 진료 기간 11일 미만 예측률

알고리즘 종류	정밀도	재현율	F1 점수
Decision Tree(D.T.)	0.81	0.81	0.81
Random Forest(R.F.)	0.81	0.76	0.79
Gradient Boosting(G.B.)	0.81	0.81	0.81
XGBoost(X.B.)	0.81	0.81	0.81
Logistic Regression(L.R.)	0.81	0.80	0.80
평균	0.81	0.80	0.80

한편, 진료 기간 11일 이상을 예측한 결과는 <표 7>와 같이 5개 알고리즘에서 예측한 정밀도 의 평균 값은 0.78이며 평균 재현율은 0.79, F1 점 수는 0.79이었다. 정밀도는 D.T. · G.B. · X.B. 알 고리즘이 우수했고 재현율은 R.F.이 가장 우수했고 F 1 점수는 정밀도와 마찬가지로 D.T. · G.B. · X.B. 알고리즘이 우수했다. 한편 진료 기관 예측률에 가장 큰 영향을 미치는 요인은 <표 8>과 같이 진 료 기간 11일 미만과 같았다.

<표 7> 진료 기간 11일 이상 예측률

알고리즘 종류	정밀도	재현율	F1 점수
Decision Tree(D.T.)	0.79	0.78	0.79
Random Forest(R.F.)	0.75	0.80	0.78
Gradient Boosting(G.B.)	0.79	0.78	0.79
XGBoost(X.B.)	0.79	0.78	0.79
Logistic Regression(L.R.)	0.78	0.79	0.79
평균	0.78	0.79	0.79

<표 8> 진료 기간 예측률에 영향을 미치는 요인 분석 결과

순위	구분	중요도
1	병원구분_한방병원	0.83972
2	진료 지역_4(광주,전남,전북)	0.09339
3	병원구분_한의원	0.02198
4	진료 지역_2(부산,경남,울산)	0.00896
5	진료 지역_1(서울,인천,경기)	0.0072
6	진료 지역_5(대전,충남,충북)	0.00605
7	나이	0.00575
8	진료 지역_3(대구,경북)	0.00423
9	성별_남성	0.00395
10	성별_여성	0.00297

## 5. 결론

본 연구를 통해 머신러닝 기법을 활용하여 자동차 보험 사고 40대~50대 경상 환자 대상 Decision Tree 등 5개 알고리즘을 이용, 진료 기간을 예측하고 그에 영향을 미치는 요인들을 분석하였다. 진료 기간 예측에 좋은 성능을 나타낸 알고리즘은 Decison Tree와 Gradient Boosting, XGBoost 등 3가지였다. 그리고 진료 기간 예측에 영향을 미치는 요인 분석 결 과, 병원의 종류(한방병원) 진료 지역(광주, 전남, 전 북), 나이, 성별(남성) 순이었다. 한편, 본 연구를 진행 하면서 자동차보험 사고 한방 40대~50대 경상 환자로 제한하여 연구를 진행하였는데, 향후에는 한·양방 진 료 및 전 연령대로 그 대상을 확대하여 진행할 계획이 다. 학문적으로는 대규모 데이터를 활용·분석하여 연구 결과의 일반화를 추구하였으며 머신러닝 기술 및 성 능 비교를 위한 5개 알고리즘을 제공하는 AutoML을 활용하는 손쉬운 연구방법을 제시하였다. 본 연구 결 과가 자동차보험 사고 진료비를 검강 시키는 정책에 기초 정보로 활용되기를 기대한다.

## 참고문헌

- [1] 서울대학교 산학협력단, 자동차보험 심사 효 과분석 및 발전방안 수립, 2020.
- [2] 송운아, 자동차보험 한방진료비 급증과 안정

- 화 방안, 2016.
- [3] 전용석, 경제환경 변화와 자동차보험 경상환자 과잉진료, 2022.
  - [4] 윤창용, 인공지능병상을 이용한 교통사고 견수 예측, 2021.
  - [5] 도로교통공단, 교통사고분석시스템, 2022.
  - [6] 건강보험심사평가원, 2021년 자동차보험 진료비 통계, 2021.
  - [7] Ghaemmaghami, M. L., Sperry, J. L., Gunst, M. A., & Friese, R. S. (2019). Predicting length of stay in trauma patients: A machine learning approach. *The American Journal of Surgery*, 218(1), 24-30.
  - [8] Adeleke, I. F., Adebisi, M. O., & Adekunle, O. B. (2018). Predicting the length of hospital stay of road traffic accident victims: A comparative study between regression analysis and decision tree techniques. *Journal of Public Health*, 26(4), 375-382.
  - [9] Almuheidi, S., Alaklabi, A., Miroshnichenko, A., & Almutairi, A. (2021). Prediction of hospital stay in acute traumatic brain injury: A machine learning approach. *Journal of Head Trauma Rehabilitation*, 36(1), E31-E37.
  - [10] Abujaber, M. A. F., Al-Majali, O., Azab, M., Ababneh, B., Alqasrawi, O., Abusamak, M., Al-Hadidi, D., & Al-Mousa, D. (2020). Predictors of prolonged hospital stay following a motor vehicle crash: A systematic review and meta-analysis. *Injury Epidemiology*, 7(1), 50.

【 저자 소개 】



변 경 근 (Kyung-Keun Byun)  
 1996년 2월 홍대 컴퓨터공학 학사  
 1998년 2월 홍대 전산학 석사  
 2023년 2월 숭실대 IT정책경영학 박사  
 재학중  
 email : kkbyun@hanmail.net



이 덕 규 (Doeg-Gyu Lee)  
 1987년 2월 계명대 전산학 학사  
 2016년 8월 카톨릭대 의료경영학 석사  
 2023년 2월 숭실대 IT정책경영학 박사  
 재학중  
 email : leedg317@naver.com



이 형 동 (Hyung-Dong Lee)  
 1990년 서울시립대 전자공학 학사  
 2017년 건국대 정보보호학 석사  
 2023년 2월 숭실대 IT정책경영학 박사  
 재학중  
 email : hdtiger77@gmail.com