

## 트리 기반의 앙상블 머신러닝 및 SHAP 기법을 활용한 고령층의 삶의 만족도에 영향을 미치는 디지털 정보격차 요인 탐색

이 정 선\*

가천대학교 소프트웨어센터 초빙교수

### Exploration of Digital Divide Factors Affecting Life Satisfaction in Older People: Using Tree-Based Ensemble Machine Learning and SHAP

Jeong-Seon Lee\*

Invited Professor, Center of Software Education, Gachon University, Seongnam 13120, Korea

#### [요 약]

디지털 전환은 일상의 다양한 영역에서 혜택을 제공하는 반면, 디지털 격차로 인해 사회적 불평등 및 소외를 비롯한 여러 문제가 발생하고 있다. 특히, 디지털 대전환과 가속화된 고령화로 인해 고령층의 디지털 정보격차가 삶의 질에 미치는 영향은 더욱 심화될 것이다. 본 연구는 「2022년 디지털 정보격차 실태조사」 자료를 활용하여 고령층의 삶의 만족도에 영향을 미치는 디지털 정보격차 요인을 탐색하는 것을 목적으로 한다. 이를 위해 트리 기반의 앙상블 기계학습 기법인 Random Forest, XGBoost, LightGBM, CatBoost를 적용하였고, 모델들의 성능 비교 결과 CatBoost가 가장 우수한 것으로 나타났다. SHAP 기법을 활용한 CatBoost 모델 결과에 따르면, 기술적 자기효능감이 삶의 만족도에 가장 큰 영향을 미치는 요인이며, 결속형 사회적 자본, 디지털 기기에 대한 태도, 디지털 조력자, 인터넷 서비스 이용 정도, 월 가구 소득 및 학력 등이 주요 요인으로 확인되었다. 본 연구는 고령층의 삶의 만족도에 영향을 미치는 주요 디지털 정보격차 요인을 실증적으로 밝힘으로써 디지털 정보격차 해소를 위한 정책 수립 및 디지털 포용성 제고가 개인의 삶의 질 향상으로 이어질 수 있다는 시사점을 제시한다.

#### [Abstract]

While digital transformation offers several benefits, it also raises issues such as social inequality and exclusion due to the digital divide. With increasing aging population, this digital divide among the elderly is expected to widen. This study explores the factors of the digital divide affecting life satisfaction among the elderly, using data from the “2022 Digital Divide Survey.” Tree-based ensemble machine learning techniques such as Random Forest, XGBoost, LightGBM, and CatBoost were applied, with CatBoost showing the best performance. SHAP analysis of the CatBoost model reveals that technological self-efficacy, bonding social capital, attitudes towards digital devices, digital assistants, internet use, household income, and education level are key factors. The study highlights the importance of addressing the digital divide to enhance life satisfaction and quality of life for the elderly.

**색인어** : 디지털 정보격차, 고령층, 삶의 만족도, 기계학습, SHAP

**Keyword** : Digital Divide, Elderly, Life Satisfaction, Machine Learning, SHAP

<http://dx.doi.org/10.9728/dcs.2024.25.7.1847>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Received** 12 May 2024; **Revised** 11 June 2024

**Accepted** 18 July 2024

**\*Corresponding Author; Jeong-Seon Lee**

**Tel:** +82-11-5898-7916

**E-mail:** tmjungsun@gachon.ac.kr

## I. 서론

정보통신기술(ICT)은 인터넷을 통한 금융 거래와 전자상거래뿐만 아니라 키오스크를 이용한 주문 및 수납 등 일상 생활의 다양한 영역에 빠르게 적용되고 있다. 그와 맞물려 디지털 정보 격차도 심화하고 있다. 디지털 정보 격차는 다양한 사회적·경제적 영역에서 심각한 사회적 불평등을 초래할 수 있는 중요한 요소로, 디지털 취약계층의 정보 격차를 줄이는 것은 사회 정의와 경제 발전을 위해 매우 중요한 과제이다 [1]-[3].

노화로 인해 신체적, 인지적 기능이 저하됨에 따라 새로운 지식, 정보 및 기술을 수용하는 것이 어려워지므로, 여러 연구에서는 노인을 디지털 취약계층으로 언급하고 있다. 또한 「2023년 OECD 보건 통계(Health Statistics)」에 따르면 우리나라 국민의 기대수명은 83.6년으로 OECD 국가(평균 80.3년) 중 상위권이며 의료 기술 발달과 국가의 건강투자 정책 등으로 평균 수명은 늘어나지만, 심각한 출산을 저하로 초고령 사회로의 진입이 예상된다[4]. 고령층의 삶의 만족도 관련 연구에서는 다양한 활동에 참여하는 고령층은 노년기 변화에 적응을 잘하고 삶의 만족도가 높은 것으로 나타났으며 [5]-[8], 이러한 활동을 하는데 스마트폰, 인터넷 등과 같은 디지털 기기 활용은 도움을 준다[5],[9]-[12]. 온라인으로 커뮤니티 참여나 상담 및 교육 등을 통해 자립심, 심리적 역량 강화 그리고 평생학습 등으로 삶의 질이 향상할 수 있으며 [12],[13] 이메일 및 소셜 미디어로 사람들과 사회적 관계 및 상호작용을 통해 사회적 자본을 증가시켜 삶의 만족도를 높임을 알 수 있다[5],[12]-[15]. 더욱이 이미 정부의 많은 대국민 서비스 절차를 디지털로 전환하였기에 디지털 취약계층은 정부에서 제공하는 중요한 정보 및 서비스뿐 아니라 기본 생활 서비스도 접근이 어렵기에 국민으로 누려야 할 기본적인 의무를 보호받을 수 없게 된다. 그러므로 고령층의 디지털 정보격차는 단순히 기술적인 문제가 아니라 사회에서 소외되지 않고 존엄성을 유지하며 살아가는 데 중요하다 [16],[17]. 초고령 사회로 진입하면서 디지털 소외로부터 고령층을 포용하고 고령의 삶의 질의 향상에 영향을 주는 디지털 정보격차의 요인을 면밀히 파악하는 것은 중요한 과제라 할 수 있다. 한국지능정보사회진흥원에서 해마다 실시하는 「2022 디지털 정보격차 실태조사」 데이터를 활용하여 다양한 연구들이 진행되어 오고 있다. 고령층의 디지털 리터러시가 심리적 안녕감과 삶의 만족도에 미치는 영향을 구조 방정식 모델로 확인하였고[5], 다중 회귀 분석을 이용하여 중·고령층의 디지털 리터러시와 삶의 만족도를 입증하였으며 [16], 노년층의 디지털 조력자 경험과 삶의 만족도와와의 관계를 확인하였다[17]. 그 외에 노년층의 디지털 정보격차와 관련된 연구들이 통계적 접근 방식으로 활발히 이루어지고 있다 [18],[19]. 최근에는 기계학습(Machine Learning)을 활용한 연구도 시작되고 있는데, Park과 Feng은 기계학습 기법을

이용하여 고령층의 디지털 정보격차 예측하였으며[18], 송광윤 등은 다양한 분류 기계학습기법으로 디지털 정보격차에 따른 정보 소외 계층을 분류하고, 중요 변수를 파악하였다 [19],[20]. 또한 미시간 대학교 사회연구소의 HRS(Health and Retirement Study) 데이터를 이용하여 고령층의 삶의 만족도에 영향을 미치는 요인들을 통계적 기법과 기계학습 기법을 비교 분석하였는데, 기계학습 기법의 성능이 더 좋은 것으로 나타났다[21].

통계 방식과 기계학습의 주요 차이점은 목적과 접근 방식에 있다. 전통적인 통계적 접근 방식은 변수 간의 구조와 추론 관계를 이해하는 데 중점을 두며 기계학습은 복잡하고 알려지지 않은 데이터에서 패턴을 발견하고 이를 통해 예측 성능을 높이는 데 있다[22],[23]. 또한 기계학습은 독립변수의 정규성, 등분산성 그리고 여러 개의 변수를 동시에 모델에 적용할 때 발생하는 다중공선성의 가정을 요구하지 않기에 더 유연하다[24].

본 연구는 「2022 디지털 정보격차 실태조사」 설문 데이터를 기반으로 기계학습을 적용하여 삶의 만족도에 영향을 미치는 요인들을 탐색하고자 한다. 기계학습 기법으로는 트리 기반 앙상블 모델들을 중심으로 한다. 트리 기반 앙상블 모델은 정형데이터일 때 우수한 성능이 입증되었으며, 과적합 방지 및 예측 성과가 높은 것으로 나타났다[19],[21]. 그러나 트리 기반 앙상블 모델은 결과 해석의 어려움이 존재한다 [18],[24]. 이를 보완하고자 변수들의 조합을 고려하여 특정 변수가 종속변수에 미치는 기여도인 Shapely value를 사용하여 변수의 유의미함을 해석하는 알고리즘인 SHAP (SHapley Additive exPlanations)를 사용하여 모델 결과를 해석한다[25],[26].

## II. 연구데이터

### 2-1 데이터 수집

본 연구의 데이터는 디지털 정보격차 해소를 위한 기초자료 제공에 목표를 두고 한국지능정보사회진흥원에서 실시한 「2022 디지털 정보격차 실태조사」의 설문 응답 원자료를 사용하였다[3]. 이 조사는 일반 국민 대비 취약계층(저소득층·장애인·농어민·고령층 등)의 정보격차 실태조사를 연 1회 실시하고 2006년부터 국가승인 통계로 지정되어 관리되어 오고 있다. 연령, 성별, 학력, 거주지 및 소득 등의 인구 사회학적 특성과 다양한 차원으로 디지털 정보화 활용 수준, 그리고 기술적 자기효능감, 기술에 대한 태도, 디지털 조력자, 사회적 자본, 삶의 만족도 그리고 코로나 이후로 디지털 대전환에 대한 인식 등 다양한 차원의 문항을 포함하고 있으므로 디지털 정보격차와 삶의 만족도 관계를 파악하는 본 연구에 적합하다고 판단하였다.

## 2-2 연구 대상 및 특성

한국지능정보사회진흥원에서 실시한 「2022 디지털 정보격차 실태조사」는 고령층을 55세 이상으로 정의하여 디지털 정보격차 요인의 기초통계분석을 기반으로 연구 결과를 제시한다. 본 연구는 건강진단, 경로우대 등의 노인 복지 서비스를 받을 수 있고, 「국민기초생활 보장법」에서도 기초생활 수급 대상으로 포함되는 노인 기준을 토대로 65세 이상을 고령층으로 정의하였고, 그 중 인터넷 경험이 있고 설문 문항에 성실히 응답을 한 864명을 대상으로 한다. 그러므로 본 연구 대상의 특성을 「2022 디지털 정보격차 실태조사」의 고령층 분석 결과와 비교하여 다양한 방안을 모색하는 데는 어려움이 존재한다.

본 연구 대상의 인구통계학적 분포는 표 1과 같다. 성별은 남성이 431명(49.9%)이고 여성이 433명(50.1%)이며 연령대는 65세~74세는 전기 노인, 75세 이상을 후기 노인으로 구분하였다[27]. 전기 노인은 707명(81.8%), 후기 노인은 157명(18.2%)이었다. 최종 학력은 초졸 이하 134명(15.5%), 중졸 284명(32.9%), 고졸 374명(43.3%), 대졸 이상 72명(8.3%)이었다. 가구 월 소득은 200만 원 미만 311명(36.0%), 200~299만 원 231명(26.7%), 300~399만 원 152명(17.6%) 그리고 400만 원 이상은 170명(10.66%)으로 나타났다. 직업이 있는 경우 376명(43.5%) 없는 경우 488명(56.5%)이었고, 홀몸노인은 142명(16.4%) 그렇지 않은 경우는 722명(83.6%)이었으며, 거주유형은 단독주택이

표 1. 인구통계학적 특성

Table 1. Characteristics of demographic

Classification		N	%
Gender	Male	431	49.9
	Female	433	50.1
Age	65-74	707	81.8
	More 75	157	18.2
Education	below primary school	134	15.5
	middle school grad	284	32.9
	high school grad	374	43.3
	More college	72	8.3
Monthly house Income (10,0000 won)	below 200	311	36
	200-299	231	26.7
	300-399	152	17.6
	More 400	170	19.7
Employ	Employed	376	43.5
	Unemployed	488	56.5
House Type	Detached house	293	33.9
	Apartment	417	48.3
	Alliance/Villa	140	16.2
	other	14	1.6
living alone	living alone	142	16.4
	living together	722	83.6

293명(33.9%), 아파트는 417명(48.3%), 연립/다세대는 140명(16.2%), 기타 유형은 14명(1.6%)로 나타났다.

## 2-3 측정변수

「2022 디지털 정보격차 실태조사」 항목 중 코로나와 직접적 관련 있는 항목과 응답의 충실도 낮은 조사항목은 분석에서 제외하였다. 고령층의 PC(데스크탑이나 노트북) 보유율은 35.9% (310명), 최근 한 달 평균 PC 사용일 수는 2.36일로 나타났다. 스마트 폰의 보유율은 100%이고 최근 한 달 사용일수는 26.3일로 조사되었기에(표 2 참고) 인터넷 서비스 사용경험은 스마트 폰을 중심으로 하였다. 설문 문항들이 개념이나 특성을 일관되게 측정되고 있는지를 파악하는 내적 일관성은 크롬바흐 알파 계수(Cronbach's  $\alpha$ )를 이용하였고, 0.7 이상이면 신뢰도가 있다고 판단한다. 설문 문항, 기초통계분석 그리고 신뢰도 분석 결과는 표 2에 제시하였다.

### 1) 디지털 배움터 교육사업 인지도

디지털 배움터 교육사업 인지도 문항은 1점: 교육에 참여해 봄, 2점: 알고 있지만 교육에 참여하지 않음, 3점: 들어 봤지만 내용을 잘 알지 못함, 4점: 들어본 적이 없음으로 측정되었다. 응답 값이 작을수록 인지도가 높다. 배움터의 인지 정도의 평균은 3.59 조사되어 인지도가 낮음을 알 수 있다.

### 2) 디지털 대전환 욕구 및 만족도

디지털 대전환 변환에 적응하려는 욕구 및 디지털 대전환 만족도의 측정 문항은 각 4문항으로 구성되었고, Likert 4점 척도(1: 전혀 그렇지 않다 ~ 4: 매우 그렇다)로 측정되었다. 평균은 2.345와 2.633으로 나타났다. 고령층의 디지털 전환에 적응하고자 하는 욕구는 낮지 않고 만족도는 부정적이지 않은 것으로 판단된다.

### 3) 인터넷 서비스 활용 정도

인터넷 서비스 활용 정도는 검색·이메일·콘텐츠, 사회적 관계·공유, 생활, 정보생산 및 공유, 네트워크, 사회참여, 경제활동 유형으로 조사되었고, 4점 척도(1: 전혀 이용 안 한다 ~ 4: 자주 이용한다)로 측정되었다. 인터넷 서비스 유형별 평균 활용 정도를 살펴보면 검색·이메일·콘텐츠 서비스는 1.869, 사회적 관계·공유 서비스는 1.712, 생활 서비스는 1.880, 정보생산 및 공유 서비스는 1.616, 네트워킹 서비스는 2.004, 사회참여 서비스는 1.375 그리고 경제활동은 1.420으로 조사되었다. 전반적으로 인터넷 서비스의 활용 정도는 활발하지 않은 것으로 나타났다.

### 4) 디지털 기기에 대한 태도

디지털 기기를 접하면서 느끼는 감정 또는 필요성 또는 중요성 등의 인식을 조사한 항목으로 Likert 4점 척도(1: 전혀 그렇지 않다 ~ 4: 매우 그렇다)로 측정되었다. 평균 측정값은

2.624로 디지털 기기에 대하여 부정적이지 않음을 알 수 있으며 항목들의 크롬바흐 알파 값은(Cronbach's  $\alpha$ )는 0.828로 나타났다.

**5) 기술적 자기효능감**

기술적 자기효능감(Technological Self-efficacy)은 기술을 사용할 때 그 기술을 성공적으로 다룰 수 있다는 자신의 능력에 대한 믿음을 의미한다[28]. 이는 Bandura가 제시한 사회학습이론에서 유래한 개념으로[29] 특정 행동의 수행 능력에 대한 개인의 자신감을 기술 분야에 적용한 것이다. 사람들이 어떠한 행동을 할 때 쉽거나 어렵다고 개인 스스로 판단하는 정도와 유사하다. 기술적 자기효능감은 기술의 사용 용이성과 사용 의도에 정적인 영향을 미치는 것으로 나타났다[30]. 기술적 자기효능감은 4개의 문항을 4점 척도(1점:전혀 그렇지 않다 ~ 4점:매우 그렇다)로 측정되었다. 평균 값은 2.073으로 조사되었고 크롬바흐 알파 값(Cronbach's  $\alpha$ )은 0.860로 나타났다.

**6) 디지털 조력자**

고령층은 신체적, 정신적, 인지적 영역에서 노화가 시작되는 시기이므로 새로운 기술을 습득하고 이용하는 데는 어려움이 있을 수 있다. 그러므로 디지털 기술 사용을 위해서는 지속적인 교육과 주변의 도움이 중요하다. 디지털 기기 이용 시 발생한 문제 해결에 도움을 주는 역할을 하는 사람, 디지털 기기 문제 해결의 사회적 지지원을 디지털 조력자(digital supporter)라 정의한다[31],[32]. 조력자 경험 측정을 위해 '디지털 기기를 이용하다 잘 모르거나 문제가 생길 때 어떻게 해결하십니까?'의 문항을 4점 척도(1점:전혀 그렇지 않다 ~ 4점:매우 그렇다)로 측정하였다. 또한 조력자 유형은 가족, 친구, 직장동료나 지인, 외부 전문인력, 스스로 해결, 인터넷 검색으로 구성되어 있다. 본 연구에서는 디지털 조력자 유형을 내/외부로 구분하였다. 내부 디지털 조력자는 가족을 의미하고 외부 디지털 조력자는 친구, 직장동료나 지인 그리고 외부 전문인력을 의미한다. 그리고 스스로 해결과 인터넷 검색을

통한 해결은 스스로 해결로 정의하였다[17],[32]. 스스로 해결 경험의 정도는 2.055, 내부 디지털 조력자의 경험은 2.962, 외부 디지털 조력자의 경험은 2.109로 조사되었다. 고령층은 디지털 관련 문제 발생 시 스스로 해결하기보다는 조력자의 지원이 필요하며 특히 내부 조력자 즉 가족 지원의 비중이 높음을 알 수 있다. 그러나 고령층의 디지털 조력 지원은 충분히 받지는 않는 것으로 조사되었다.

**7) 사회적 자본**

사회적 자본(Social Capital)은 개인 간의 연결 즉, 사회적 관계와 네트워크를 통해 형성되는 신뢰, 상호작용의 규범을 의미한다[33],[34]. Puttnam은 사회적 자본을 결속형(Bonding)과 교량형(Bridging)으로 구분하였다[34]. 결속형 사회적 자본은 유사한 배경이나 관심을 가진 사람들 사이에서 발전하는 강력한 관계로 가족과 친구 등을 의미한다. 교량형 사회적 자본은 서로 다양하고 다른 개인들 간의 관계와 연결을 나타내며 이들 연결은 약한 연결이나 더 넓은 범위의 자원, 정보 및 기회에 접근할 기회를 제공한다[35].

「2022 디지털 정보격차 실태조사」의 사회적 자본의 측정 문항은 결속형과 교량형을 각 10문항씩으로 구성된 Williams의 인터넷 사회 자본 척도(Internet Social Capital Scales, ISCs)를[36] 각 5문항으로 축약한 것으로[17],[32], 4점 척도(1점:전혀 그렇지 않다 ~ 4점:매우 그렇다)로 측정되었다. 결속형의 크롬바흐 알파 값(Cronbach's  $\alpha$ )은 0.757, 평균 측정값은 2.891이며 교량형의 크롬바흐 알파 값(Cronbach's  $\alpha$ )은 0.831, 평균 측정값은 2.563으로 조사되었다. 고령층은 결속형 사회적 자본이 강한 것으로 나타났다.

**2-4 종속변수**

종속변수인 삶의 만족도는 표 2에서 제시된 것과 같이 Diener가 정의한 개인의 총체적인 삶에 대해 전반적인 만족 정도의 인지적 판단을 측정하는 것으로 정의하였고[37] 총 5 문항에 대해 4점 척도(1점:전혀 그렇지 않다 ~ 4점:매우 그

**표 2.** 측정 및 종속변수 문항, 기술 통계량, 신뢰도 분석 결과

**Table 2.** Items, descriptive statistic & reliable analysis in measured variable

Measured Variable	Item	Avg	SD	Cronbach's $\alpha$
PC usage level	Days used on last month_PC (Desktop, Laptop)	2.23	6.403	-
Smart Phone usage level	Days used on last month_Smart phone	26.32	7.232	-
Tablet usage level	Days used on last month_Tablet	0.20	1.887	-
Search email content service	How often have you used the following services in the past year? -Search of Information and News /E-mail /-Media/Educational content	1.869	0.021	-
Social networking and sharing services	How often have you used the following services in the past year? -SNS/Instant messaging /-Blog/Internet cafes, communities -Data sharing and collaboration through cloud services	1.712	0.536	-
Lifestyle services	How often have you used the following services in the past year? -Life information /E-commerce /Financial transaction /Public	1.880	0.023	-

Measured Variable	Item	Avg	SD	Cronbach's $\alpha$
Information production and sharing	How often have you engaged in the same activity over the past year? -Created content or Edited others' content/Posting or sharing content found on the internet	1.616	0.026	-
Online Networking	-I have used the internet to maintain relationships with people I already knew and to become closer to them -I have used the internet to meet new people and communicate with them	2.004	0.788	-
Social Participation Service	-I have expressed my opinion on social concerns(public issues) through the Internet -I have made policy proposals, recommendations, policy evaluations, and complaints to governments, local governments, and public institutions through the Internet -I have participated in donating or volunteering activities through the Internet -I have participated in online voting, polls, and signature through the Internet	1.375	0.620	-
Online economic activity service	How often have you engaged in the following activities online over the past year? -Assist with employment or job transition. -Assist with entrepreneurship or business operation. -Related to acquiring information and personal finance management that help increase or maintain income -Help with cost-saving	1.420	0.002	-
digital Supporter	When encountering difficulties or issues while using digital devices, how do you resolve them?			-
	Self:-Resolve on my own / Search for information on the internet	2.055	0.771	-
	Internet : Family	2.962	0.804	-
	External :Friends, Accompanies, External professional assistance	2.109	0.670	-
Attitude for digital technology	-Digital technology is useful -Digital technology makes my life convenient -Digital technology is good for me -I want to use digital technology more	2.624	0.591	0.828
Self-efficacy for digital devices (Technological Self-efficacy)	-I am confident in learning digital devices -I am confident in using digital devices -I can quickly figure out how to use new digital devices -I want to use digital devices more	2.073	0.658	0.860
Social Capital_ Bonding	-There is someone who can help solve the problem -There is someone to ask for advice on important decisions -There is someone I can talk to privately -There are people who can help you fight against injustice -There is someone to whom I can entrust important matters	2.891	0.481	0.757
Social Capital_Bridging	-There is a feeling of being connected to the larger world -There is a feeling of being connected with people through interactions -Interactions enable engaging conversations with new people -There is a willingness to spend time on community activities." -Interactions always facilitate meeting new people	2.563	0.576	0.831
Awareness level of Digital Baeumteo	Do you Know digital education business("Digital Baeumteo")?	3.590	0.765	-
Digital Transformation Satisfaction	Due to the digital transformation, -I have increased leisure opportunities and become more enjoyable -I learned new information and knowledge more quickly -I have increased opportunities to communicate with new people -I have increased efficiency through non-face-to-face study and work.	2.633	0.028	0.889
Desire for digital transformation	-I want to learn technologies closely related to the digital transformation -I want to get a job related to the digital transformation -Understanding and ability to utilize digital technology will determine my life in the future -If the government provides support for education related to digital transformation, I am willing to take the course	2.346	0.030	0.891
Life Satisfaction (Dependent Variable)	-In most ways my life is close to my ideal. -The conditions of my life are excellent. -I am satisfied with my life. -So far I have gotten the important things I want in life. -If I could live my life over, I would change almost nothing	2.510	0.500	0.796

렇다)로 측정되었고 크롬바흐 알파 값(Cronbach's  $\alpha$ )은 0.796, 평균은 2.510으로 조사되었다.

### III. 분석 방법

분석은 그림 1과 같은 과정으로 진행하였다. 먼저 데이터 전처리 단계에서 결측값이 많은 변수는 제외하였고 범주형 변수는 인코딩 과정을 통해 모델 사용에 적합한 형태로 변환하였다. 이후 LASSO 회귀 모델을 활용하여 유의미한 디지털 정보격차 요인을 선별하였다. 다음으로, 삶의 만족도 예측 모델 구축을 위해 대표적인 트리 기반 앙상블 모델인 Random Forest, XGBoost, LightGBM, CatBoost를 적용하였다. 각 모델은 최적 성능 도출을 위해 하이퍼 파라미터 튜닝 과정을 거쳤으며 성능이 가장 우수한 모델을 선정하여 SHAP 기법을 적용하여 모델 결과를 해석하였다.

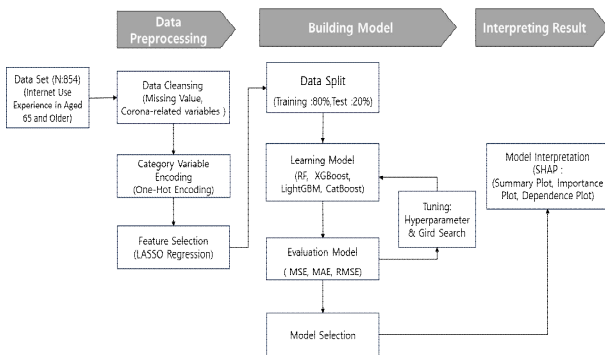


그림 1. 분석 절차  
Fig. 1. Analysis process

#### 3-1 데이터 전처리 과정

##### 1) 변수 선택

관련성 높은 변수를 선별하여 모델을 구축하면 과적합과 성능저하 문제 등의 위험성을 줄일 수 있다[38]-[40]. 변수 선택은 다양한 기법이 존재한다. 본 연구에서는 LASSO (Least Absolute Shrinkage and Selection Operator) 회귀분석을 이용하였다. LASSO는 종속변수에 영향을 미치지 못하는 독립변수에 벌점(penalty)을 부과함으로써 회귀계수를 0으로 추정하여 모델에서 제거하는 방식이다[40]. LASSO 모델은 적합한 벌점 모수를 정하는 것이 중요하다. 최적의 변수 선택을 위해 하이퍼 파라미터 튜닝 과정을 거쳤다. 벌점(Penalty) 모수를 0.001부터 1000까지 조정하고 10-fold 교차 타당화(Cross Validation)를 통해 MSE(Mean Squared Error)를 최소로 하는 벌점 모수를 파악하였다. 표 3은 LASSO를 활용한 변수 선택과정으로 선별된 변수들이다.

표 3. 변수 선택 결과

Table 3. The result of variable selection

Variable	Description	Variable	Description
ADQ1_1	Age_Group (1) young-old (2) old-old	ADQ3_1	Employment_status (0: Yes, 1: No)
ADQ4_1	Education_below_primary (0:No, 1:Yes)	ADQ4_3	Education_High school (0:No, 1:Yes)
ADQ4_4	Education_more_college (0:No, 1:Yes)	ADQ7_1	Detached house (0:No, 1:Yes)
ADQ8_1	Living alone (0: No, 1: Yes)	ADQ9_1_1	Monthly_House_Income_below_200 (0:No,1:Yes)
ADQ9_1_2	Monthly_House_Income_200-299 (0:No, 1:Yes)	ADQ102_1	Residence_Area (0:Rural, 1: Urban)
Q7A1	Days used on the recent month_PC	Q9B	Usage_Social Network Service
Q12B	Usage_Networking Service	Q13B	Usage_Social Participation Service
Q15A_1	Digital Supporter-Self Help	Q15A_2	Digital Supporter-External
Q15A_3	Digital Supporter-Internal	Q16A1_5	Social Capital_Bonding
Q17A	Attitude for digital technology	Q18A	Technological Self-efficacy
Q21CC	Digital Transformation Satisfaction	Q23	Awareness of Digital Baeumteo (learning Center)

#### 3-2 예측 모델

여러 트리 모델을 학습시킨 후 조합하여 더 강력한 예측 모델을 만드는 트리 기반의 앙상블 기계학습 모델은 비선형 모델 중 가장 많이 알려져 있고 정형데이터 예측에 있어서는 다른 선형 모델이나 심지어 딥러닝 모델보다 성능이 뛰어난 연구 결과들이 다수 존재한다[40]-[42]. 본 연구는 많은 연구에서 활용되고 있는 랜덤포레스트(Random Forest)와 대표적인 그래디언트 부스팅(Gradient boosting)모델인 XGBoost, LightGBM, CatBoost를 이용하였다[40],[43]. 모델의 안정성과 성능 검증에 가장 대표적인 Grid Search Cross Validation을 수행하여 모델 구축과정에서 필요한 주요 하이퍼 파라미터를 선별하여 최종 모델을 선정하였다[44].

##### 1) 랜덤 포레스트(Random Forest)

Breiman이 제안한 랜덤 포레스트 알고리즘은 의사결정나무를 기본 학습자로 하여 학습데이터 셋을 부트스트랩(bootstrap) 샘플링을 통해 다수의 서브(sub) 데이터 셋을 생성하고 각 서브 데이터 셋을 대상으로 모델을 학습한 후에 예측 결과를 투표방식으로 결합하는 배깅(bagging) 방식의 앙상블 기법이다[45]. 고차원 데이터를 효과적으로 처리하며

과적합(overfitting)을 줄여 신뢰할 수 있는 예측 성능을 도출할 수 있다. 이러한 특징으로 분류(classification) 및 회귀(regression)를 다루는 다양한 영역에서 사용되고 있다[46].

**2) 그라디언트 부스팅 (Gradient Boosting)**

그라디언트 부스팅 알고리즘은 분류와 회귀의 예측 모델에 자주 사용되며 이전 단계의 학습자 오류에 더 큰 가중치를 부여함으로써 이전 학습자의 잘못된 오류에 집중하게 하여 오류를 더 잘 분류하게 하는 방식이다[47]. 일반적으로 배깅(bagging)방식의 랜덤 포레스트 보다 성능이 우수한 것으로 알려진 반면, 훈련 시간이 길고 과적합의 가능성이 있다는 문제점도 존재한다[48]. 연구에서는 그라디언트 부스팅 모델 중 최근에 많이 사용하는 XGBoost, LightGBM, CatBoost 모델을 사용한다[43].

• **XGBoost**

XGBoost는 큰 데이터 셋의 안정성과 훈련의 속도를 향상을 위해 Chen과 Guestrin이 제안한 알고리즘으로 eXtreme Gradient Boosting의 약자이다[49]. 병렬 처리와 분산 컴퓨팅을 지원하고 확장성(scalability)이 뛰어나며 결측치 처리가 가능하여 희소한(sparse) 데이터 셋에도 최적의 성능을 보인다. 훈련 중에 지나치게 복잡한 모델에게 불이익을 주어 트리의 성장을 제한함으로써 과적합을 방지한다. XGBoost는 이러한 장점들로 인해 다양한 영역에서 사용되고 있다[40].

• **LightGBM**

LightGBM은 2017년 Microsoft에서 제안한 알고리즘으로 XGBoost와 비교하였을 때 속도, 효율성, 메모리 활용도가 뛰어나다[50]. 데이터 처리 시간을 효과적으로 단축하기 위해 트리 각 노드의 분할점을 검색할 때 모든 데이터를 검색하는 것이 아닌 기울기가 큰 데이터에 초점을 둔 샘플링 기법인 GOSS (Gradient-based On-Side Sampling) 기술과 관련 있는 변수들을 하나의 변수로 처리하는 EFB(Exclusive Feature Bundle)알고리즘을 적용해 입력변수의 개수를 줄이므로 학습 속도를 향상시키고 메모리 사용량도 줄이는데 도움을 준다[50].

• **CatBoost**

CatBoost는 Categorical Boosting의 약자로 범주형 데이터 전처리 및 과적합 해결에 중점을 두어 개발되었다. 기존 그라디언트 부스팅 알고리즘은 모든 훈련데이터로 학습하고 순차적으로 잔차를 계산하여 모델 성능을 개선하므로 훈련데이터 셋에만 잘 훈련된 모델 생성으로 과적합이 발생 될 수 있다. CatBoost는 특정 순서로 훈련데이터의 하위 집합을 생성하여 잔차를 계산하는 프로세스로 과적합 방지에 도움을 준다[51].

**3) 하이퍼 파라미터 최적화**

**(Hyper Parameter optimization)**

기계학습 모델은 최적의 성능을 위해서 하이퍼 파라미터 선정이 중요하다. Random Forest, XGBoost, LightGBM, CatBoost 모델의 최적화는 다양한 조건의 하이퍼 파라미터들의 조합을 적용하여 최적 성능을 선정하는 그리드 서치(Grid Search) 방식을 이용하였다[52].

**3-3 모델 평가**

모델 평가는 대표적인 평가 지표인 MSE(Mean Squared Error, 평균제곱오차), RMSE(Root Mean Squared Error, 평균제곱근오차), 그리고 MAE(Mean Absolute Error, 평균절대오차)를 사용하였다. 이 지표들은 모델이 예측한 값과 실제 값 사이의 불일치 정도를 측정하기 위해 사용되며 모델의 예측 성능을 객관적으로 평가할 수 있다.

• **MSE(Mean Squared Error)**

MSE는 평균제곱오차로 예측값과 실제값의 차이(오차)를 제곱한 것을 평균한 값으로, 오차 제곱으로 작은 오류보다 큰 오류에 더 큰 영향을 미친다. 즉, 모델에 큰 오차를 발생하는 것에 더 큰 패널티를 부여함으로써 큰 오차를 줄이는 방향으로 모델을 조정해 간다. 그러나 이상치에 민감하게 반응하는 단점이 있다. MSE는 0에 가까울수록 모델 성능이 우수함을 의미한다.

• **RMSE(Root Mean Squared Error)**

RMSE는 MSE의 제곱근으로, MSE와 동일하게 큰 오차에 비중을 두나 제곱근을 취하기에 MSE 보다 이상치에 덜 민감하다. 0에 가까울수록 모델 성능이 우수함을 의미한다.

• **MAE(Mean Absolute Error)**

MAE는 예측값과 실제값의 차이의 절대 값을 평균 낸 값으로, 오차의 크기에 대해 선형적으로 가중치를 부여하므로, MSE나 RMSE에 비해 이상 치에 덜 민감하다. MAE는 0에 가까울수록 모델 성능이 우수함을 의미한다.

표 4에서 제시한 것과 같이 4가지 모델의 평가 결과는 유사하지만 CatBoost 모델 성능이 가장 좋게 나타났다.

**표 4. 모델평가 결과**

**Table 4. The result of model evaluation**

	RandomForest	XGBoost	LightGBM	CatBoost
MSE	0.1994	0.2006	0.2011	0.1949
RMSE	0.4466	0.4478	0.4484	0.4415
MAE	0.3466	0.3451	0.3485	0.3441

#### IV. 결과 해석

트리 기반의 앙상블 모델은 과적합 방지 및 예측 성과가 높은 장점이 있으나, 모델 결과 해석에 어려움이 존재한다. 본 연구에서는 모델평가로 선정된 CatBoost의 모델 해석을 위해 SHAP분석을 수행하였다. SHAP분석은 모델 예측 결과에 대한 통찰력 있는 해석을 제공하는 동시에 모델 결과의 신뢰성을 지원하는 것을 목표로 한다[25],[26].

SHAP분석은 특정변수가 종속변수에 미치는 기여도를 나타내는 Shapley value를 사용하여 변수의 중요도를 해석한다[22],[23]. SHAP는 각 독립변수의 영향도뿐 아니라 변수 간 상호작용을 종합적으로 고려하여 계산된다. 그림 2는 변수 중요도 도표(Feature Importance Plot)로 독립변수가 종속변수인 삶의 만족도에 미치는 영향 정도를 표현한 것이며 Shapley value의 절댓값에 해당하는 SHAP지수를 이용한다. 막대의 길이가 길수록 삶의 만족도 예측에 영향력이 높은 변수라 할 수 있으며 상대적으로 순위가 높은 변수가 Y축의 최상단에 위치하게 된다.

그림 3의 요약 도표(Summary Plot)은 독립변수가 종속변수에 미치는 영향도가 정(+)의 방향인지 부(-)의 방향인지를 알 수 있다. X축은 Shapley value를 의미하고 그 값이 음수이면 삶의 만족도의 값을 감소시키는 것으로 부정적인 영향을 미치는 것이며 양수이면 긍정적 영향을 미치는 것으로 해석된다. Y축은 중요도 순으로 독립변수가 위치해 있고 그래프의 점들은 개별 데이터의 값들이고 점의 색상은 측정값의 크기를 나타내는데 작을수록 파란색 클수록 빨간색으로 표현된다. Shapley value의 양의 영역에 파란색 점이 많이 분포되어 있으면, 그 독립변수는 부(-)적인 영향을 미친다고 판단할 수 있고 빨간색 점이 많이 분포되어 있으면, 정(+)적인 영향을 미친다고 판단할 수 있다. 두 도표를 보면 삶의 만족도에 영향을 가장 크게 미치는 변수는 기술적 자기 효능감과 결속형 사회적 자본이다. 이 두 변수는 요약 도표를 보면 Shapley 지수가 음수인 영역에 파란색 점들이 그리고 양수인 영역에 빨간색 점들이 넓게 분포되어 있다. 즉, 삶의 만족도와 강한 양의 상관관계가 있다고 할 수 있다. 그 외 중요한 변수로는 인터넷 서비스 활용 정도, 디지털 문제 발생 시 스스로 해결 정도 및 디지털 조력자 경험 정도, 최종 학력, 월평균 가구 소득, 취업 상태, PC 사용 정도 등이다. 선정된 대부분의 중요 변수는 문항 응답 수준이 높을수록 삶의 만족도가 높은 정적인 관계가 나타났으나 삶의 만족도가 높은 고령층은 새로운 관계 형성이나 기존 관계 형성을 위한 목적으로 인터넷 서비스를 이용하지 않은 것으로 나타났다. 삶의 만족도가 높은 고령층은 이미 관계 형성이 잘되어 있어서 새로운 인간관계 형성에 덜 의존적일 수 있기 때문이라 판단된다. 외부 디지털 조력자 경험과 PC 사용일 수는 Shapley 지수의 양의 영역에 넓게 분포되어 있기에 삶의 만족도에 정(+)적인 영향을 미친다고 할 수 있다.

범주형 변수는 One-Hot-Encoding으로 1과 0으로 구성되어 있다. 월평균 가구소득이 200만 원 미만을 표현하는 변수는 값이 1이면 200만 원 미만이며 0은 그렇지 않은 경우를 나타낸다. 그러므로 200만 원 이상일 때 삶의 만족도에 긍정적인 영향을 미침을 알 수 있다.

좀 더 상세히 삶의 만족도에 미치는 영향도를 파악하기 위해 개별 입력변수들이 삶의 만족도 예측에 미치는 영향을 구체적으로 보여주는 그림 4 종속성 도표가 있다. 종속성 도표는 Y축이 Shapley 지수이고, X축은 변수의 측정값이다. 이 종속성 도표를 보면 삶의 만족도에 영향을 미치는 측정값의 수준을 파악할 수 있다. 삶의 만족도에 큰 영향을 미치는 기술적 자기 효능감이나 디지털 기기 문제시 스스로 해결 정도는 측정 값이 2(그렇지 않은 편이다)보다 크면 삶의 만족도에 긍정적인 영향을 미치는 것으로 조사되었다. 고령층의 특성을 고려하면 신체 및 인지능력 감소 등으로 자신감이 결여되기에 자기 효능감이나 기술 이용에 부정적 영향을 줄 가능성을 배제할 수는 없지만[53], 효능감이 너무 낮거나 디지털 기기의 문제 해결 능력이 아주 부족하지 않다면 삶의 만족도에 긍정적 영향을 미침을 알 수 있다. 디지털 기기의 태도나 디지털 대변환에 대한 만족도는 다소 긍정적이면 삶의 만족도에 긍정적 영향을 주는 것으로 나타났다. 관계적인 측면을 살펴보면 결속형 사회적 자본은 강한 관계여야 삶의 만족도에 긍정적 영향을 주는 것으로 나타났다. 외부 디지털 조력자 경험 정도와 내부 디지털 조력자 경험 정도를 살펴보면 외부 디지털 조력자 경험 정도는 요약 도표에서 보면 Shapley 지수가 양수인 영역에 빨간색 점들이 넓게 분포되어 있으며, 내부 디지털 조력자 경험 정도 음수인 영역에 파란색 점들이 넓게 분포되어 있다. 즉, 내/외부의 디지털 조력을 지원을 잘 받을수록 삶의 만족도에 정적인 영향을 준다고 말할 수 있으나, 내부 디지털 조력자의 지원 경험은 삶의 만족도에 부정적인 경우가 더 많았으며, 이와 반대로 외부 디지털 조력자 지원은 삶의 만족도에 긍정적인 경우가 더 많다고 할 수 있다. 그 이유를 살펴보기 위해 종속성 도표를 보면 내부 디지털 조력자 일 때 측정값이 4(매우 그렇다)인 경우에만 삶의 만족도에 긍정적 영향을 미친다. 즉, 내부 조력 경험 정도와 삶의 만족도는 양의 상관관계가 존재하나 고령층이 느끼기에 아주 충분한 조력만이 삶의 만족도에 긍정적 영향을 미침을 알 수 있다. 외부 디지털 조력자 경험은 2.5 이상인 경우로 지원을 받을 수 있으면 삶의 만족도에 긍정적 영향을 미치는 것으로 나타났다. 즉 조력자 유형에 따라 삶의 만족도에 긍정적 영향을 주는 지원 정도는 상이함을 알 수 있다.



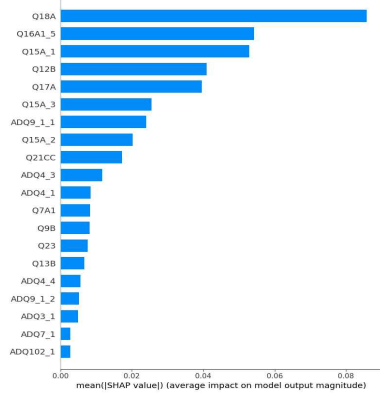


그림 2. SHAP 분석: 중요도  
Fig. 2. SHAP analysis : Feature importance plot

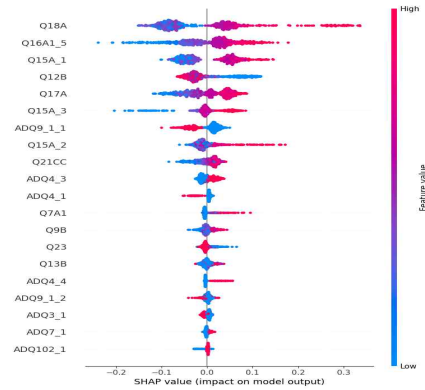


그림 3. SHAP 분석: 요약 도표  
Fig. 3. SHAP analysis : Summary plot

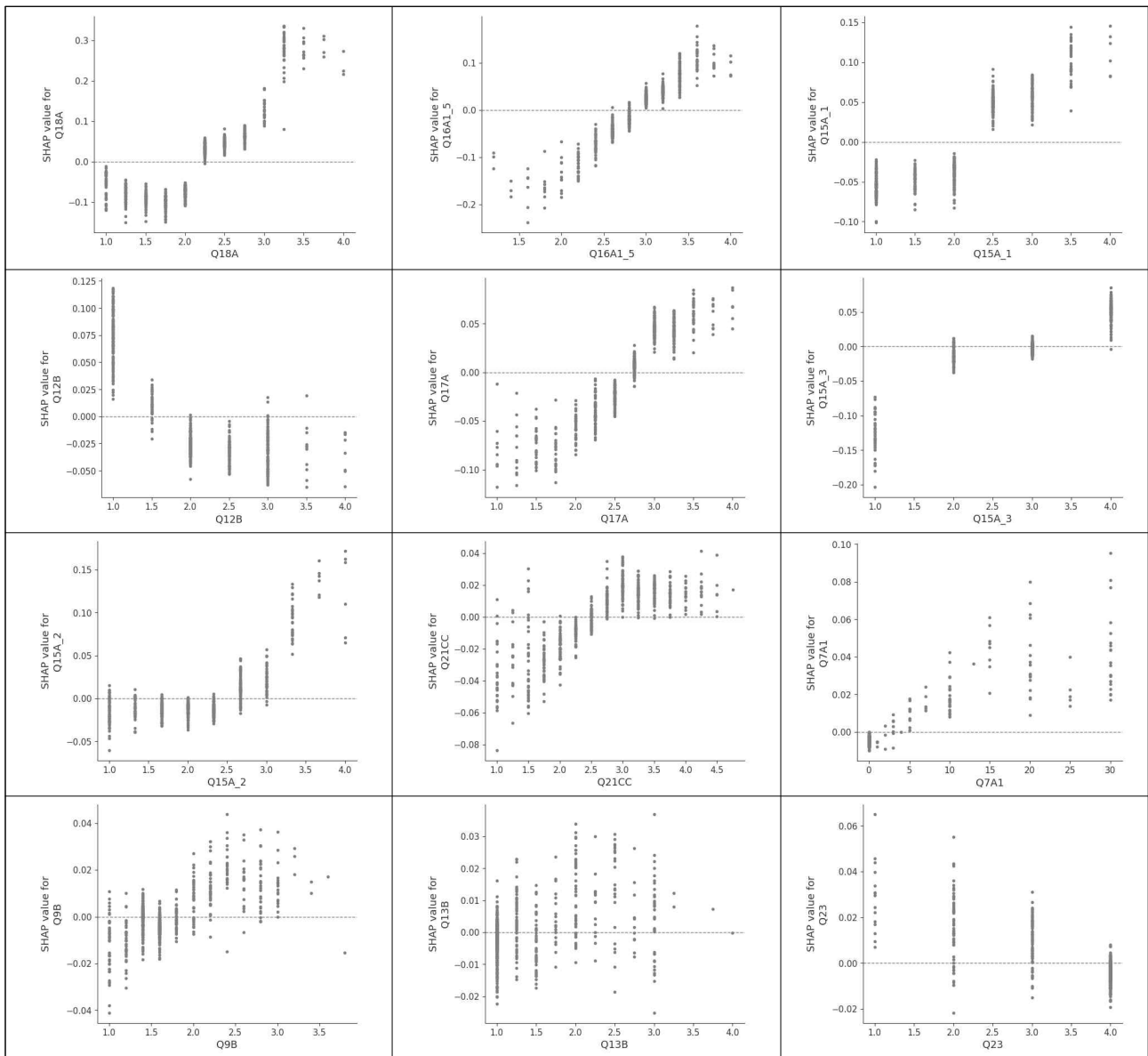


그림 4. 종속성 플롯  
Fig. 4. Dependence plot

## V. 결론 및 논의

본 연구는 「2022년 디지털 정보격차 설문자료」를 활용하여 고령층의 삶의 만족도에 영향을 미치는 디지털 정보격차 요인을 파악하고자 하였습니다. 이를 위해 트리 기반 앙상블 모델인 Random Forest, XGBoost, LightGBM, CatBoost를 적용하였으며, 이 4가지 모델의 성능은 큰 차이가 나타나지 않았으나, 이 중 가장 우수한 성능을 보인 CatBoost 모델에 SHAP 알고리즘을 적용하여 주요 영향 요인을 분석하였다. 그 결과를 정리하면 다음과 같다.

첫째, 고령층의 삶의 만족도에 영향을 가장 크게 영향을 미치는 요인은 기술적 자기 효능감으로 나타났다. 선행연구들에서 기술적 자기 효능감은 디지털 환경에 적응하고 디지털 기기를 활용하는 데 중요한 역량으로 간주하고 있다. 고령층이 기술에 대한 자신감이 있을 때 디지털 기기를 수용하고 보다 적극적으로 활용하기에 디지털 전환에 적극적으로 융화될 수 있다. 또한 이러한 기기에 익숙해짐으로써 국가에서 지원하는 다양한 디지털 기기의 돌봄 서비스를 비롯한 정부 지원 서비스의 활용으로 일상생활 역시 독립적으로 생활을 유지하여 고령층 당사자뿐 아니라 고령자의 가족들 역시 삶의 질이 높아질 것으로 판단된다[54]-[56]. 고령층의 기술적 자기 효능감은 부정적이지 않으면 삶의 만족도에 긍정적 영향을 주는 것으로 나타났다. 다행히 기술적 자기 효능감은 교육 등을 통해 개선될 수 있는 요인이므로 정부 또는 지역사회 수준에서 기술적 자기 효능감을 높일 수 있는 다양한 접근이 필요하다[57].

둘째, 디지털 관련 문제 발생 시 스스로 해결할 수 있는 능력이 존재하거나 디지털 조력자의 도움을 받을 수 있을 때 삶의 만족도가 높은 것으로 나타났다. 단, 가족의 지원은 고령층이 느끼기에 아주 적극적인 지원을 받아야만 삶의 만족도에 긍정적 영향을 주는 반면, 외부 디지털 조력자는 지원을 받을 수만 있으면 삶의 만족도에 긍정적 영향을 주는 것으로 나타났다. 현재 가족 유형이 분열되고 해체되어가는 비중이 높아지고 있으므로 가족 중심의 디지털 조력자의 의존도가 높은 경우는 지속성에 문제가 생길 수 있고 고령층 스스로 학습하여 문제를 해결하기보다는 가족 조력자의 의존도가 높아질 수 있다. 그러므로 외부 디지털 조력자 지원을 활성화가 필요하다라고 판단된다.

셋째, 고령층은 관계가 삶의 만족도에 큰 영향을 미침을 알 수 있다. 고령층이 가족과 같이 친밀한 관계 또는 사회에서 소외되지 않음이 삶의 만족도의 중요 요인이다[58]. 이를 강화하기 위해서는 다양한 인터넷 서비스 활용은 가족 구성원 간의 연결을 유지하고 소통을 증진하며 공유와 지원을 가능하게 함으로써 가족 간의 유대감과 사회 소속감을 강화하는데 크게 기여할 수 있다[59]. 본 연구에서는 전체적으로 고령층의 인터넷 서비스 활용 정도는 높지 않으나 서비스 활용 정도와 삶의 만족도는 정(+)의 관계로 조사되었다.

마지막으로 인구 사회학적 특성으로, 가구 월평균 가구 소득이 200만 원 미만인 경우와 학력은 초졸 이하인 삶의 만족도가 떨어지는 것으로 나타났다. 사회적으로 취약계층이면 디지털 전환 적응에 더욱 어려움이 생기므로 삶의 만족도가 낮은 것으로 판단된다.

본 연구 결과를 바탕으로 고령층의 삶의 만족도를 향상하는 방안을 다음과 같이 제안한다. 일차적으로 디지털 기기 역량 강화나 디지털 기기에 대한 심리적 요인을 보다 긍정적으로 변화시키기 위해서는 가장 효과적이고 실천적인 방법은 교육 참여이다[57]. 이러한 교육에 참여하기 위해서는 고령층의 연령대, 소득 수준, 건강 수준, 디지털 역량, 가족관계 등 다양한 요인을 고려하여 차별화된 교육 대상 선정 및 교육과정이 필요하다고 판단된다. 또한 교육 유형 역시 다양성이 필요하다. 예를 들어, 고령층은 강한 유대감을 형성하는 것이 삶의 만족도에 긍정적인 영향을 미치므로 고령층의 성향, 디지털 능력, 환경 등을 고려하여 유사한 고령층으로 구성된 소규모 그룹의 교육 방식을 도입하거나 고령층은 이동 제약도 존재하기에 같은 지역 내의 디지털 교육의 멘토와 멘티의 관계 형성을 해주는 프로그램 방식의 도입을 고려해 볼 필요가 있다. 이러한 프로그램을 통해 디지털 교육을 매개로 사회적 연결을 강하게 함으로써 삶의 만족도를 더욱 높일 수 있을 것으로 판단된다. 또한 손자와 함께 배우는 디지털 교육이나 젊은 층이 고령층에게 디지털 교육을 하고 고령층은 젊은층에게 사회적 경험을 공유하는 등의 세대를 아우르는 교육 방법 역시 세대 간의 통합과 더불어 고령층의 사회적 자본 확대가 가능하다고 판단된다. 사회·경제적 취약계층은 가족도 디지털 취약계층일 확률이 높기에 더욱 문제가 클 수 있으며, 정부나 지자체에서 지원하는 다양한 교육 프로그램에 소외될 수 있다. 그러므로 디지털 교육을 넘어 포용이 필요하다. 그리고 본 연구 결과에서도 나타났듯이 고령층은 PC의 보유 및 이용률은 저조하기에 모바일 방식으로 교육 방식으로 변해가야 한다. 또한 디지털 배움터에 대한 고령층의 인지도가 높지 않으므로, 정부에서 제공하는 교육 기회를 잘 활용할 수 있도록 고령층에게 적합한 홍보방안이 필요하다고 판단된다.

본 연구는 고령층의 삶의 만족도에 미치는 디지털 정보격차 요인을 단일 요인 분석이 아닌 기계학습으로 동시에 분석함으로써 다양한 요인들을 종합적으로 고려하는 다차원적 접근을 시도하였다는 점에서 학술적 의의가 있다. 또한 삶의 만족도에 영향을 미치는 디지털 정보격차 요인을 실증적으로 밝힘으로써 디지털 정보 격차 해소를 위한 정책 수립 및 디지털 포용성 제고가 개인의 삶의 질 향상으로 이어질 수 있다는 시사점을 제공한 점 그리고 고령층 특성에 맞는 디지털 정보격차 해소를 위한 실무적 정책 방안 마련에 기여할 수 있다는 점에서 실무적 의의가 있다.

이상의 결론과 논의에도 불구하고 본 연구는 다음과 같은 한계점이 존재한다. 첫째, 인터넷 서비스를 이용 경험이 있는 고령층만을 대상으로 하기에, 고령층 전체의 디지털 기기의 취약계층을 대변하지 못한 한계점이 존재한다. 후속 연구에서

는 관련 대상층을 포함하여 전반적인 고령층의 디지털 정보 격차 특성을 파악할 수 있는 연구가 필요하다. 둘째 트리 기반의 앙상블 기계학습을 적용하기에 다소 표본 수가 미흡한 점이다. 본 연구에서 활용한 디지털 정보격차 실태조사는 매해 시행되고 있기에 사회적 환경이 크게 변화가 없는 다년도 데이터를 분석하는 것을 제안한다.

셋째 트리 기반 앙상블 모델뿐 아니라 다양한 기계학습 기법 및 통계기법을 적용하여 성능 비교하지 못한 점 그리고 다양한 XAI 기법이 존재하는데 SHAP 기법만을 적용하여 결과 해석이 다소 부족할 수 있다는 한계점이 존재한다. 예측 성능 및 요인 탐색 개선을 위해 다양한 예측 모델 기법 및 그 기법에 적합한 XAI 알고리즘을 시도해 볼 필요가 있다.

### 참고문헌

- [1] F. Mubarak and R. Suomi, "Elderly Forgotten? Digital Exclusion in the Information Age and the Rising Grey Digital Divide," *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, Vol. 59, 00469580221096272, 2022. <https://doi.org/10.1177/00469580221096272>
- [2] T. N. Friemel, "The Digital Divide Has Grown Old: Determinants of a Digital Divide Among Seniors," *New Media & Society*, Vol. 18, No. 2, pp. 313-331, February 2016. <https://doi.org/10.1177/1461444814538648>
- [3] NIA (National Information Society Agency), The Report on the Digital Divide 2022, Ministry of Science and ICT, Sejong, NIA VIII-RSE-C-22046, March 2023.
- [4] MOHW (Ministry of Health and Welfare) and KIHASA (Korea Institute for Health and Social Affairs), OECD Health Statistics 2023, MOHW, Sejong, 2023.
- [5] J. Kim, "Digital Divide Among Older Consumers and Measures to Enhance Their Digital Literacy," *Journal of Consumer Policy Studies*, Vol. 53, No. 3, pp. 27-54, December 2022. <http://doi.org/10.15723/jcps.53.3.202212.27>
- [6] J. A. Oh and J. W. Ryu, "A Study on the Effects of Digital Literacy on the Psychological Wellbeing and Life Satisfaction of the Elderly," *Korean Public Management Review*, Vol. 32, No. 2, pp. 319-344, June 2018. <https://doi.org/10.24210/kapm.2018.32.2.013>
- [7] B. A. Taylor and V. L. Bengtson, Sociological Perspectives on Productive Aging, in *Productive Aging: Concepts and Challenges*, Baltimore, MD: The Johns Hopkins University Press, ch. 6, pp. 120-144, 2001.
- [8] S. Hong, L. Hasche, and S. Bowland, "Structural Relationships between Social Activities and Longitudinal Trajectories of Depression among Older Adults," *The Gerontologist*, Vol. 49, No. 1, pp. 1-11, February 2009. <https://doi.org/10.1093/geront/gnp006>
- [9] N. Charness and W. R. Boot, Technology, Gaming, and Social Networking, in *Handbook of the Psychology of Aging*, 8th ed. London, UK: Academic Press, ch. 20, pp. 389-407, 2016. <https://doi.org/10.1016/B978-0-12-411469-2.00020-0>
- [10] P. DiMaggio, E. Hargittai, C. Celeste, and S. Shafer, Digital Inequality: From Unequal Access to Differentiated Use, in *Social Inequality*, New York, NY: Russell Sage Foundation, ch. 9, pp. 355-400, 2004.
- [11] E. Hargittai and A. Hinnant, "Digital Inequality: Differences in Young Adults' Use of the Internet," *Communication Research*, Vol. 35, No. 5, pp. 602-621, October 2008. <https://doi.org/10.1177/0093650208321782>
- [12] J. H. Choi, S. Kim, J. Y. Moon, J. Kang, I. Lee, and J. Kim, "Seek or Provide: Comparative Effects of Online Information Sharing on Seniors' Quality of Life," *Communications of the Association for Information Systems*, Vol. 34, pp. 513-530, January 2014. <https://doi.org/10.17705/ICAIS.03427>
- [13] B. B. Neves, J. R. S. Fonseca, F. Amaro, and A. Pasqualotti, "Social Capital and Internet Use in an Age-Comparative Perspective with a Focus on Later Life," *PLoS ONE*, Vol. 13, No. 2, e0192119, February 2018. <https://doi.org/10.1371/journal.pone.0192119>
- [14] H. Yoon, O. Lee, K. Beum, and Y. Gim, "Effect of IT Education on On-Line Social Relationship of Older Adults," *Journal of the Korea Contents Association*, Vol. 15, No. 5, pp. 283-294, May 2015. <https://doi.org/10.5392/JKCA.2015.15.05.283>
- [15] F. J. Elgar, C. G. Davis, M. J. Wohl, S. J. Trites, J. M. Zelenski, and M. S. Martin, "Social Capital, Health and Life Satisfaction in 50 Countries," *Health & Place*, Vol. 17, No. 5, pp. 1044-1053, September 2011. <https://doi.org/10.1016/j.healthplace.2011.06.010>
- [16] M. I. Kim, Y. S. Kim, and S. R. Um, "A Study about the Types of Digital Information Utilization and Life Satisfaction of Middle-Aged and the Older People," *Journal of Korean Association for Regional Information Society*, Vol. 23, No. 1, pp. 51-74, March 2020. <http://doi.org/10.22896/karis.2020.23.1.003>
- [17] J.-S. Lee and H.-Y. Park, "The Effects of Digital Supporter Experiences on the Life Satisfaction of the Elderly: The Dual Mediating Effects of Digital Utilization and Social Capital," *Journal of Digital Contents Association*, Vol. 24, No. 9, pp. 2047-2057, September 2023. <https://doi.org/10.9728/dcs.2023.24.9.2047>

- [18] J. R. Park and Y. Feng, "Trajectory Tracking of Changes Digital Divide Prediction Factors in the Elderly through Machine Learning," *PLoS ONE*, Vol. 18, No. 2, e0281291, February 2023. <https://doi.org/10.1371/journal.pone.0281291>
- [19] K. Y. Song, Y. S. Kim, and I. H. Chang, "Classification Analysis of the Underprivileged Class according to Digital Divide Using Machine Learning," *Journal of the Korean Data & Information Science Society*, Vol. 32, No. 5, pp. 1071-1083, September 2021. <https://doi.org/10.7465/jkdi.2021.32.5.1071>
- [20] K. Y. Song, Y. S. Kim, and I. H. Chang, "A Study on the Classification of Importance Variables in a Digital Divide Data Using Machine Learning," *Journal of the Korean Data & Information Science Society*, Vol. 33, No. 2, pp. 177-193, March 2022. <https://doi.org/10.7465/jkdi.2022.33.2.177>
- [21] X. Shen, F. Yin, and C. Jiao, "Predictive Models of Life Satisfaction in Older People: A Machine Learning Approach," *International Journal of Environmental Research and Public Health*, Vol. 20, No. 3, 2445, January 2023. <https://doi.org/10.3390/ijerph20032445>
- [22] K. Thiruvengadam, B. Watson, C. Ponnuraja, and K. Rajendran, "A Review of Statistical Modelling and Machine Learning in Analytical Problems," *International Journal of Applied Engineering Research*, Vol. 17, No. 5, pp. 506-510, 2022. <https://doi.org/10.37622/IJAER/17.5.2022.506-510>
- [23] M. Bennett, K. Hayes, E. J. Kleczyk, and R. Mehta, "Similarities and Differences between Machine Learning and Traditional Advanced Statistical Modeling in Healthcare Analytics," arXiv:2201.02469, January 2022. <https://doi.org/10.48550/arXiv.2201.02469>
- [24] A. Sheetal, Z. Jiang, and L. Di Milia, "Using Machine Learning to Analyze Longitudinal Data: A Tutorial Guide and Best-Practice Recommendations for Social Science Researchers," *Applied Psychology*, Vol. 72, No. 3, pp. 1339-1364, July 2023. <https://doi.org/10.1111/apps.12435>
- [25] NIPS Papers. A Unified Approach to Interpreting Model Predictions [Internet]. Available: <http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf>.
- [26] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent Individualized Feature Attribution for Tree Ensembles," arXiv:1802.03888, February 2018. <https://doi.org/10.48550/arXiv.1802.03888>
- [27] H. Sung, "A Study on the Factors Affecting Life Satisfaction among the Young-Old and the Old-Old: Focusing on Differences by Residence," *Journal of Social Science*, Vol. 32, No. 4, pp. 333-356, October 2021. <https://doi.org/10.16881/jss.2021.10.32.4.333>
- [28] T. McDonald and M. Siegall, "The Effects of Technological Self-Efficacy and Job Focus on Job Performance, Attitudes, and Withdrawal Behaviors," *The Journal of Psychology*, Vol. 126, No. 5, pp. 465-475, 1992. <https://doi.org/10.1080/00223980.1992.10543380>
- [29] A. Bandura, "Self-Efficacy: Toward a Unifying Theory of Behavioral Change," *Psychological Review*, Vol. 84, No. 2, pp. 191-215, 1977. <https://doi.org/10.1037//0033-295X.84.2.191>
- [30] V. Venkatesh and H. Bala, "Technology Acceptance Model 3 and a Research Agenda on Interventions," *Decision Sciences*, Vol. 39, No. 2, pp. 273-315, May 2008. <https://doi.org/10.1111/j.1540-5915.2008.00192.x>
- [31] B. S. Kim and J. H. Ko, "A Study on the Impact of Digital Supporter on Digital Social Capital in the Elderly: Focusing on the Mediating Effect of Digital Usage," *Discourse and Policy in Social Science*, Vol. 13, No. 2, pp. 27-57, October 2020. <https://doi.org/10.22417/DPSS.2020.10.13.2.27>
- [32] J. H. Ko, "A Study on Factors Affecting the Qualitative Use of Information by the Elderly: Focusing on the Use of Digital Supporters," *Korean Public Management Review*, Vol. 35, No. 3, pp. 23-46, September 2021. <http://doi.org/10.24210/kapm.2021.35.3.002>
- [33] D. Coleman and S. E. Iso-Ahola, "Leisure and Health: The Role of Social Support and Self-Determination," *Journal of Leisure Research*, Vol. 25, No. 2, pp. 111-128, 1993. <https://doi.org/10.1080/00222216.1993.11969913>
- [34] R. Putnam, "Social Capital: Measurement and Consequences," *Canadian Journal of Policy Research*, Vol. 2, No. 1, pp. 41-51, 2001.
- [35] R. F. Datoon, J. V. Camacho Jr., A. V. Lapitan, and J. M. Gapas, "Bridging and Bonding Social Capital for Sustainable Technology Adoption, the Landcare Program in the Resource Poor Rural Town of Claveria, Philippines," *Research in Globalization*, Vol. 7, 100152, December 2023. <https://doi.org/10.1016/j.resglo.2023.100152>
- [36] D. Williams, "On and Off the 'Net: Scales for Social Capital in an Online Era," *Journal of Computer-Mediated Communication*, Vol. 11, No. 2, pp. 593-628, January 2006. <https://doi.org/10.1111/j.1083-6101.2006.00029.x>
- [37] E. Diener, "Subjective Well-Being," *Psychological Bulletin*, Vol. 95, No. 3, pp. 542-575, May 1984. <https://doi.org/10.1037//0033-2909.95.3.542>

- [38] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182, March 2003.
- [39] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature Selection: A Data Perspective," *ACM Computing Surveys (CSUR)*, Vol. 50, No. 6, 94, November 2018. <https://doi.org/10.1145/3136625>
- [40] S. Jaiswal and P. Gupta, "Ensemble Approach: XGBoost, CATBoost, and LightGBM for Diabetes Mellitus Risk Prediction," in *Proceedings of 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Gunupur, India, pp. 1-6, September 2022. <https://doi.org/10.1109/ICCSEA54677.2022.9936130>
- [41] E. K. Ampomah, Z. Qin, and G. Nyame, "Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement," *Information*, Vol. 11, No. 6, 332, June 2020. <https://doi.org/10.3390/info11060332>
- [42] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why Do Tree-Based Models Still Outperform Deep Learning on Typical Tabular Data?" in *Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS 2022)*, New Orleans: LA, pp. 507-520, November-December 2022.
- [43] Y. Yennimar, W. Leonardi, H. Weide, D. Cantona, and G. M. Hutagalung, "Comparison of Data Mining Algorithms (Random Forest, C4.5, Catboost) Based on Adaptive Boosting in Predicting Diabetes Mellitus," *Jurnal Teknik Informatika C.I.T Medicom*, Vol. 16, No. 1, pp. 1-12, March 2024.
- [44] G. P. Herrera, M. Constantino, J.-J. Su, and A. Naranpanawa, "The Use of ICTs and Income Distribution in Brazil: A Machine Learning Explanation Using SHAP Values," *Telecommunications Policy*, Vol. 47, No. 8, 102598, September 2023. <https://doi.org/10.1016/j.telpol.2023.102598>
- [45] L. Breiman, "Random Forests," *Machine Learning*, Vol. 45, No. 1, pp. 5-32, October 2001. <https://doi.org/10.1023/A:1010933404324>
- [46] G. Biau and E. Scornet, "A Random Forest Guided Tour," *TEST*, Vol. 25, No. 2, pp. 197-227, June 2016. <https://doi.org/10.1007/s11749-016-0481-7>
- [47] A. Natekin and A. Knoll, "Gradient Boosting Machines, a Tutorial," *Frontiers in Neurorobotics*, Vol. 7, 21, December 2013. <https://doi.org/10.3389/fnbot.2013.00021>
- [48] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A Comparative Analysis of Gradient Boosting Algorithms," *Artificial Intelligence Review*, Vol. 54, pp. 1937-1967, March 2021. <https://doi.org/10.1007/s10462-020-09896-5>
- [49] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, San Francisco: CA, pp. 785-794, August 2016. <https://doi.org/10.1145/2939672.2939785>
- [50] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, ... and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, Long Beach: CA, pp. 3149-3157, December 2017.
- [51] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: Unbiased Boosting with Categorical Features," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS '18)*, Montréal, Canada, pp. 6639-6649, December 2018.
- [52] P. Liashchynskiy and P. Liashchynskiy, "Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS," arXiv:1912.06059, December 2019. <https://doi.org/10.48550/arXiv.1912.06059>
- [53] H.-I. Choi and I. Song, "The Mediating Effect of Self-efficacy between the Elderly's Digital Information Literacy and Life Satisfaction," *Journal of the Korea Academia-Industrial Cooperation Society*, Vol. 21, No. 6, pp. 246-255, June 2020. <https://doi.org/10.5762/KAIS.2020.21.6.246>
- [54] L. Wang, P.-L. P. Rau, and G. Salvendy, "Older Adults' Acceptance of Information Technology," *Educational Gerontology*, Vol. 37, No. 12, pp. 1081-1099, 2011. <https://doi.org/10.1080/03601277.2010.500588>
- [55] K. Sen, G. Prybutok, and V. Prybutok, "The Use of Digital Technology for Social Wellbeing Reduces Social Isolation in Older Adults: A Systematic Review," *SSM - Population Health*, Vol. 17, 101020, March 2022. <https://doi.org/10.1016/j.ssmph.2021.101020>
- [56] A. Chu, J. Huber, B. Mastel-Smith, and S. Cesario, "'Partnering with Seniors for Better Health': Computer Use and Internet Health Information Retrieval among Older Adults in a Low Socioeconomic Community," *Journal of the Medical Library Association*, Vol. 97, No. 1, pp. 12-20, January 2009. <https://doi.org/10.3163/1536-5050.97.1.003>
- [57] F. M. Gatti, E. Brivio, and C. Galimberti, "'The Future is Ours Too': A Training Process to Enable the Learning

Perception and Increase Self-Efficacy in the Use of Tablets in the Elderly,” *Educational Gerontology*, Vol. 43, No. 4, pp. 209-224, 2017. <https://doi.org/10.1080/03601277.2017.1279952>

- [58] D. Lee, “Structural Influence Relationship between Depression, Family Relationships, and Self-Esteem on the Life Satisfaction of the Elderly : Focusing on Differences according to Income Level,” *Korea and Global Affairs*, Vol. 8, No. 1, pp. 803-845, February 2024. <http://doi.org/10.22718/kg.2024.8.1.031>
- [59] S. An, H. Lee, and Y. Lee, “The Effects of Online Network and Digital Literacy on Depression in Older Adults: A Review of Functional Characteristics of Apps for Older Adults with Low Digital Literacy,” *Korean Journal of Broadcasting & Telecommunications Research*, No. 125, pp. 74-103, January 2024. <https://doi.org/10.22876/kjbtr.2023.125.003>



**이정선(Jeong-Seon Lee)**

2020년 : 숙명여자대학교 일반대학원  
(경영학 박사-MIS)

1997년~2000년: SAS Korea, 컨설턴트

2000년~2008년: 한국 IBM Business Intelligence, 컨설턴트

2020년~2022년: 숙명여자대학교 대학 IR 센터

2022년~현 재: 가천대학교 소프트웨어 교육센터 초빙교수

※관심분야 : 데이터 기반 의사결정, 비즈니스 애널리틱스, 학습 분석학, 빅 데이터분석, 노인 기술학