

# 해석가능한 머신러닝을 위한 시각적 분석 시스템 제안

## (A Visual Analytics System for Interpretable Machine Learning)

박 찬 희 <sup>\*</sup>      이 경 원 <sup>\*\*</sup>  
(Chanhee Park)      (Kyungwon Lee)

**요 약** 해석가능한 머신러닝은 머신러닝 시스템의 행동과 예측을 사람이 이해할 수 있도록 돕는 기술을 말한다. 본 연구는 사용자가 쉽고 명확하게 머신러닝 모델을 해석할 수 있도록 지원하기 위하여 머신러닝 모델이 입력 데이터로부터 출력 결과를 어떻게 연결 짓는지에 대한 관계성을 해석하는 시각적 분석 시스템을 제안한다. 본 연구가 제안한 시각적 분석 시스템은 머신러닝 수행 결과를 입력 변수, 목표 변수, 예측 값에 따라 필터링하고 그룹 지어 해석할 수 있는 반복적인 조정 절차를 통해 효과적으로 머신러닝 모델을 해석할 수 있는 접근 방식을 취한다. 유스 케이스 분석과 사용자 심층 인터뷰를 통해 본 연구에서 제시한 시각적 분석 시스템이 머신러닝 모델의 복잡한 동작에 대한 통찰을 얻고, 입력 변수와 목표 변수 및 모델 예측에 대한 과학적 이해를 확보하고, 모델의 안정성과 신뢰성을 파악하는데 도움을 제공함을 확인했다.

**키워드:** 해석가능한 머신러닝, 데이터 시각화, 데이터 시각적 분석, 시각적 분석 시스템

**Abstract** Interpretable machine learning is a technology that assists people understand the behavior and prediction of machine learning systems. This study proposes a visual analytics system that can interpret the relationship between how machine learning models relate output results from input data. It supports users to interpret machine learning models easily and clearly. The visual analytics system proposed in this study takes an approach to effectively interpret the machine learning model through an iterative adjustment procedure that filters and groups model decision results according to input variables, target variables, and predicted/classified values. Through use case analysis and in-depth user interviews, we confirmed that our system could provide insights into the complex behavior of machine learning models, gain scientific understanding of input variables, target variables, and model predictions, and help users understand the stability and reliability of models.

**Keywords:** interpretable machine learning, data visualization, visual analytics, visual analytics system

<sup>\*</sup> 정 회 원 : 아주대학교 미디어학과 학생  
ch13p@ajou.ac.kr  
<sup>\*\*</sup> 비 회 원 : 아주대학교 미디어학과 교수(Ajou Univ.)  
kwlee@ajou.ac.kr  
(Corresponding author)

논문접수 : 2022년 6월 13일  
(Received 13 June 2022)  
논문수정 : 2022년 9월 19일  
(Revised 19 September 2022)  
심사완료 : 2022년 10월 21일  
(Accepted 21 October 2022)

## 1. 개요

해석가능한 머신러닝 분야의 목표는 머신러닝 시스템의 결정을 사람이 이해할 수 있도록 만드는 것이다[1]. 이 분야에서 해석가능성은 모델이 입력 데이터와 예측 값 사이의 인과관계를 어떻게 연결 짓는지를 보여주는 능력을 말한다[2]. 입력 데이터와 모델 결정의 사이의 관계를 잘 설명할 수 있을 때, 다음과 같은 이점이 발견된다[3]. (1) 과학적 이해: 머신러닝 모델이 학습한 바를 설명하여 입력 데이터와 출력 결과 사이의 관계에 대한 지식을 얻을 수 있다. (2) 안전성: 작은 입력 값 변화가 큰 예측 오류로 이어지지 않게 한다. (3) 신뢰: 사람은 결정의 원인을 알 수 없는 시스템 보다 결정을 설명하는 시스템을 신뢰한다. (4) 공정성: 인종이나 성별을 비롯한 편견이 모델에 학습되어 있는지 확인할 수 있다.

영화 추천 시스템과 같이 실수가 심각한 결과를 초래하지 않거나, 광학 문자 인식과 같이 이미 광범위하게 연구되고 평가된 분야에서는 머신러닝 모델의 결정에 대한 해석이 중요하지 않을 수도 있다[4]. 하지만, 머신러닝 시스템이 내리는 결정이 사회나 비즈니스에 큰 영향을 끼치는 경우가 늘어나고있다. 이에 따라, 시스템이 결정을 내리는 원리를 분석해야 할 필요가 생겨나고 있다. 예를 들어, 개인의 질병을 예측하거나 주식 및 경제 지표를 예측하는 시스템은 모델 예측 성능 뿐만 아니라 모델의 해석 역시 중요하다[5,6]. 이는 예측의 결과를 질병의 치료나 비즈니스 정책 결정 등에 활용할 경우 그 결과를 사람이 이해할 수 있어야할 뿐만 아니라 그 적용에 대한 윤리나 책임의 문제를 분명히 하기 위해서이다[3]. 뿐만 아니라 최근에는 머신러닝 모델에 학습된 내용을 인간이 다시 학습하여 인간의 지식으로 만드는 시도도 이뤄지고 있다. 예를 들어, 회의 자동 요약 시스템에 이러한 활용이 이뤄지고 있다[7].

본 연구는 머신러닝 모델의 행동과 예측을 모델 사용자가 이해할 수 있도록 돕기 위한 시각적 분석 시스템을 제안한다. 특히, 머신러닝 모델이 도출한 결과를 인간의 사고로 검토할 수 있도록 만들기 위해, 머신러닝 시스템이 도출한 결과와 머신러닝 모델 해석 기법을 시각적으로 표현하였다. 본 연구가 제안하는 기법은 모델이나 데이터에 내제된 데이터 대신 모델이 도출한 결과를 중심으로 모델을 해석한다. 이에 따라, 제안된 기법이 특정 모델이나 데이터 셋에 특화되지 않는다. 2장에서는 머신러닝 모델 해석 분야의 기존 연구의 의의와 한계를 살펴본다. 3장에서는 본 연구가 제안하는 시각화 시스템의 설계 목표 사항을 정립한다. 4장에서는 시스템의 설계 목표를 충족시키기 위한 분석 프레임워크 및 시스템을 설명한다. 5장에서는 ‘보험료 지출 예측 모델’의 유스 케이스를 통해 시스템 활용 사례를 소개하고,

시스템이 설계 목표를 달성했는지 여부를 확인한다. 6장에서는 심층 인터뷰를 통해 사용자가 본 연구가 제안하는 시스템을 사용하여 ‘당뇨병 예측 모델’을 해석하는 과정을 확인하고, 시스템의 기여점 및 한계점을 평가한다. 마지막으로 7장에서는 본 논문의 주요 결과를 요약하고 향후 연구 방향을 소개한다.

## 2. 문헌조사

### 2.1 해석가능한 머신러닝

머신러닝 모델을 해석하기 위한 여러 연구가 수행되어 왔다. 머신러닝 분야에는, 정확도를 비롯한 성능과 해석가능성 사이의 반비례 관계가 보이곤 한다. 의사결정나무나 선형회귀와 같은 단순한 모델은 그 자체로 해석가능성을 가지고 있다[8,9]. 하지만 이러한 모델은 복잡한 문제에서 정확도가 떨어진다. 높은 성능으로 주목받는 심층 신경망 모델은 설명력이 떨어져서 블랙박스라고 불린다. 이러한 모델은 학습과 예측이 이루어진 후에 별도의 기법을 통해 해석하는 것이 유리하다[1,10]. 본 연구에서는 다양한 해석 기법을 적용 가능한 모델의 종류, 해석이 다루는 범위 및 기법을 사용하는 사용자 유형에 따라 분류한다. 이어서 본 연구의 목적인 “머신러닝 모델의 행동과 예측을 모델 사용자가 이해할 수 있도록 돕는 것”에 적합한 해석 기법이 무엇인지 고찰한다.

머신러닝 모델 해석은 기법이 특정한 구조의 모델에만 적용될 수 있는지 여부에 따라 특화된 방식(model specific)과 범용적 방식(model agnostic)으로 분류된다[3]. 모델 특화 방식은 해석 방식이 다양한 모델에 적용 불가능하다. 또한 이 방식은 모델 구조에 대한 사전 지식을 필요로 하기 때문에 비 전문가가 사용하기에는 부적합하다[11]. 모델 해석 결과를 모델 연구자 뿐만 아니라 고객이나 협업자와 같은 머신러닝 비 전문가와 함께 공유해야 하는 상황에서 이러한 단점이 부각된다[1]. 이러한 경우에 모델 내부에 접근하지 않고 모델을 해석하는 범용적 방식을 제공하는 것이 대안이 된다[12].

머신러닝 모델 해석 기법은 설명 범위에 따라 분류되기도 한다[13]. 해석 기법이 모델 전체 예측에 적용 가능할 경우 그 기법을 전역적(global)이라고 말하고, 특정 상황에만 적용 가능할 경우 국소적(local)이라고 말한다. 전역적 기법은 모델의 전반적인 개요를 파악하는데 적절한 대신 특정 상황에 대한 구체적인 분석이 어렵다는 단점을 지니며, 국소적 기법은 그 반대의 특징을 가진다. 최근 등장하는 인터랙티브 분석 시스템은 모델 해석을 위해 단일한 기법을 제공하는 대신에 다양한 기법을 제공하여 사용자의 분석 자유도를 높인다[14-16].

머신러닝 모델 해석 기법은 제공받는 사람에 따라 다

르게 설계되어야 한다[17]. 해석 기법의 사용자는 머신러닝 실무자, 데이터 분석가, 비 전문가로 분류될 수 있다. 머신러닝 실무자는 모델을 만들고 수정하는 사람, 즉 모델 개발자이다. 이들은 모델 작동을 이해하고, 오분류 원인을 파악하고자 하며, 모델에 대한 디버깅 기능을 필요로 한다. 이들을 위한 시스템은 데이터, 모델 구조, 변수의 변화가 성능에 미치는 영향을 보는 데 초점을 맞춘다[11,18-20]. 데이터 과학자 및 분석가는 모델이 내놓은 결과를 사용하는 사람, 즉 모델 사용자이다. 이들은 결과를 실무에 적용하기 전에 해당 결과가 추론의 근거를 해석할 필요가 있다. 이들을 위한 시스템은 모델 예측 결과 해석에 도움을 주는 것에 목적을 맞출 필요가 있다. 이와 같이 모델 사용자를 위한 해석 기법은 데이터와 예측 결과 사이의 관계를 보여주는 것이 중요하다[10,21]. 비 전문가는 개별 머신러닝 모델보다는 머신러닝 기술 자체를 이해할 필요가 있는 학습자이다. 이를 돕기 위한 연구는 모델 내부 구조 및 가중치를 시각화하여 모델 작동 원리를 설명한다[22,23].

최근 들어 머신러닝 분야에는 복잡한 문제를 해결하기 위해 다양한 모델이 도입되고 있다. 또한, 머신러닝 모델 사용자는 머신러닝 분야에 대한 사전 지식이 적을 수 있다. 이러한 경우에는 단순한 모델을 이용하거나, 모델에 특화된 방식을 사용하여 모델을 해석하는 것이 부적합하다. 본 연구는 머신러닝 모델을 쉽게 해석할 수 있도록 지원하는 시각적 분석 시스템을 설계하고 개발하는 것을 목표로 한다. 이에 따라, 본 연구는 모델이 내놓은 결과를 활용하는 이들에게 적합한 방식인 모델 구조에 관계없이 적용 가능하고, 예측 결과를 바탕으로 한 해석을 제공하는 방식에 초점을 맞춘다. 또한, 본 연구는 시각적 분석 시스템 설계에 필터링 및 자세히 보기 등의 인터랙션을 도입하여 모델 예측 결과를 전역적 차원과 국소적 차원 모두에서 분석할 수 있도록 한다.

## 2.2 입력 변수 값 조정을 이용한 모델 해석

분류와 예측 문제에서 머신러닝 모델은 입력 변수를 바탕으로 타깃 변수의 값을 분류/예측한다. 입력 변수의 기여도를 바탕으로 머신러닝 모델의 결정을 해석하는 시도가 있어왔다[24]. 직관적 모델 해석 방법 중 하나는 입력 변수 값 변화에 따른 예측 값의 차이를 살펴보는 것이다. 최근에는 입력 변수 값 변화 전후로 모델 예측 값이 어떻게 달라지는지 분석해 모델을 해석하는 연구가 진행되고 있다.

그림 1이 표시하는 Partial Dependence Plot(이하 PD 플롯으로 표기)은 개별 입력 값이 모델의 예측에 미치는 한계 효과(marginal effect)를 보여준다[25]. 이 방식은 검사를 원하는 입력 변수의 값을 변경한 뒤에 타깃 변수의 값이 어떻게 변화했는지 평균을 내어 그래프

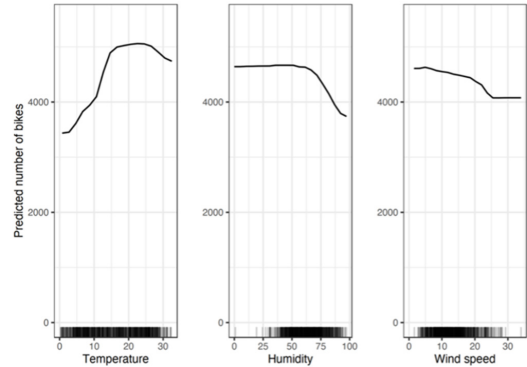


그림 1 입력 값의 변화에 따른 예측 값 변화 추세를 보여주는 PD Plots

Fig. 1 PD plots showing the trend of change in predicted values according to changes in input values

를 그린다. 그래프를 통해 타깃 변수와 입력 값 간의 관계가 선형인지, 단조로운지 또는 더 복잡한지 볼 수 있다. PD 플롯은 확인할 입력 변수와 타깃 변수를 각각 축으로 가져야한다. 인간은 3 차원 이상의 공간을 시각적으로 표현할 수 없기 때문에, PD 플롯은 최대 두 개의 입력 변수에 대해서만 모델의 효과를 확인할 수 있다. 따라서 PD플롯으로는 두 개 이상 변수에서 발생하는 상호작용 효과 및 매개효과를 파악할 수 없다. 예를 들어, PD 플롯이 표시하는 평균 변화 곡선은 입력 변수와 목적 변수 사이의 관계가 특정 그룹에서는 양적이고 다른 그룹에서는 부적인 경우를 표시할 수 없다.

입력 변수를 X, Y, Z 축으로 표시하는 PD 플롯과는 달리 본 연구는 각 변수를 평행 좌표표로 표시하는 평행좌표계(Parallel Coordinates Plot)과 Table을 사용한다. 이들은 화면 스크롤과 함께 이용할 경우 입력 변수와 인스턴스 수에 제한받지 않고 시각화가 가능하다. 시각화에 데이터 스케일의 제약이 없기 때문에 사용자 컴퓨팅 성능이 허락하는 한 어떠한 크기의 데이터도 사용할 수 있다.

PD 플롯이 평균값을 사용하기 때문에 생기는 문제를 방지하는 방법은 개별 인스턴스의 변화를 살펴보는 것이다. 그림 2의 Individual Conditional Expectation Plot(이하 ICE 플롯으로 표기)은 입력 변수가 변경에 따른 예측 변화를 보여주는 선을 각 인스턴스 별로 표시한다[26]. PD 플롯과 달리 ICE 플롯의 선은 평균 대신 각각의 인스턴스 변화 값을 표현하기 때문에 이기종 관계를 발견할 수 있다. 인스턴스를 선택하여 해당 인스턴스의 다른 입력 변수가 변화되었을 때의 효과도 파악할 수 있다는 점에서 인터랙션의 강점도 가진다. 하지만 ICE 플롯 역시 최대 두 개의 입력 변수 변화에 대해서

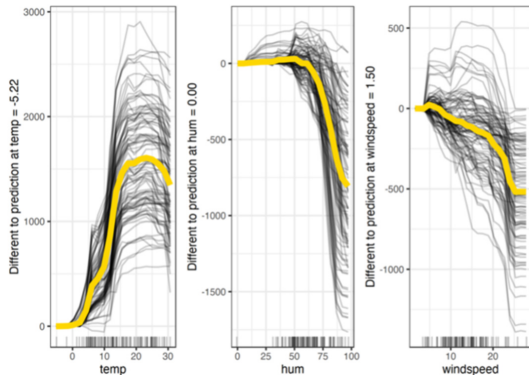


그림 2 입력 값의 변화에 따른 개별 인스턴스의 예측 변화를 보여주는 ICE plots

Fig. 2 ICE plots showing changes in predictive values in individual instances according to changes in input values

만 시각화할 수 있고, 인스턴스 수가 많아지면 선이 과밀화되어 알아볼 수 없으며, 인스턴스 별 변화 양상이 다양해진다면 그 그룹별 차이를 보는 것이 어려워지는 단점을 지닌다.

그림 3의 DECE는 이진 분류 모델의 결정을 해석하기 위한 시각적 분석 시스템이다[27]. 이 시스템은 사용자가 유연하게 인스턴스 그룹을 만들어 해석 결과를 비교하도록 돕는다. 또한 이것은 선택된 인스턴스의 특정 입력 변수가 어떤 값으로 바뀌어야 예측 값이 변하는지 보여준다. 이 과정은 사용자가 머신러닝 모델의 해석을 통해 입력 변수와 목적 변수 사이의 관계에 대한 추론을 유도한다. 또한 이것은 히스토그램을 활용해 쉽게 인스턴스 그룹을 만들 수 있도록 유도한다. 하지만 이 히스토그램은 데이터의 분포만 보여줄 뿐, 해당 분포가 예측하려는 클래스와 어떤 관련이 있는지 보여주지 않는다. 사용자는 어떤 입력 변수를 위주로 탐색해야 할지에 대한 정보가 부족한 상태에서 탐색을 진행해야 한다. 또한, DECE는 모델 예측의 맞고 틀림을 테이블 시각화의 첫번째 열에서만 보여준다. 따라서, 이 시스템은 모델 예측 정확성에 대한 인사이트를 적게 반영하고, 입력 변수와 타겟 클래스 사이의 관계만을 보여준다는 한계가 있다. 결정적으로, 이 시스템은 이진 분류 모델의 해석에만 초점을 맞춰서 다중 분류나 회귀 문제에 적용되지 않는다는 한계가 있다. 다양한 종류의 모델을 해석하기 위한 시각화 인터페이스 디자인은 여전히 도전 과제로 남아있다.

입력 데이터 변형과 이에 따른 모델 출력 값 변화 추적 연구들은 모델이 입력 변수와 예측 변수 사이의 관계를 어떻게 연결 짓는지를 보여주어 모델의 작동 패턴

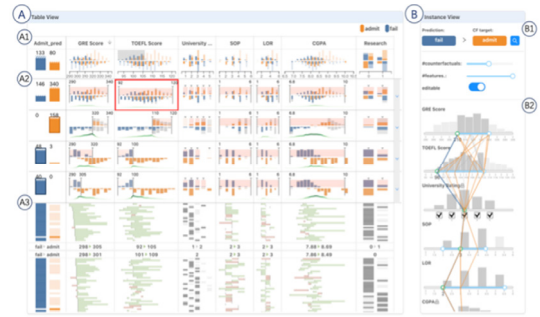


그림 3 인스턴스를 그룹화하여 입력 변수와 모델 예측 사이에 상관관계가 존재하는지 파악하도록 돕는 DECE

Fig. 3 DECE groups instances to determine whether there is a correlation between input variables and model predictions

을 해석해왔다. 한편 이들 연구는 데이터가 분포한 전반적인 패턴을 파악하기 어렵고, 인스턴스 전체와 개별 인스턴스를 자유롭게 오가며 분석을 수행할 수 없다는 한계를 가졌다. 본 연구는 기존 연구의 극복하여 머신러닝 모델의 행동과 예측을 모델 사용자가 이해할 수 있도록 돕는 시각적 분석 시스템의 인터페이스를 제안한다.

### 3. 연구의 목적 및 방법

본 연구의 목적은 머신러닝에 전문 지식이 적은 실무자가 머신러닝 모델을 해석하고 이해할 수 있도록 돕는 시각적 분석 시스템을 만드는 것이다. 이를 위해 본 연구는 모델 평가에 사용된 데이터 셋을 변형하여 모델의 결정을 다각도로 분석하는 시스템을 제안한다. 특히, 입력 변수 변화가 모델 예측에 어떠한 변화를 만들어 내는지를 보여주기 위한 시각화와 인터랙션 방법에 초점을 맞춘다.

변형된 데이터를 바탕으로 모델을 해석하는 연구들은 입력 변수 값 변화 전후로 모델 예측 값이 어떻게 달라지는지 분석한다. 이로 인해 사용자는 개별 입력 변수가 모델 예측에 미치는 한계 효과를 직관적으로 분석할 수 있다는 장점이 있다. 하지만, 이러한 기여에도 불구하고 기존 연구들은 몇 가지 한계를 보였다.

우선, 기존 연구에서는 데이터 변형에 조건이 걸려있지 않기 때문에 현실성이 떨어지는 데이터로 변형될 수 있다. 기존의 연구들은 모든 입력 변수에 대한 모든 가능한 값을 도입하거나, 임의의 입력 변수에 대해 임의의 값을 도입하여 데이터 변형을 시도해왔다. 이러한 방식은 데이터 사이의 상관관계를 무시하여 현실성이 떨어지는 데이터(예를 들어, 키가 200 cm이고, 몸무게가 30 kg인

사람)를 많이 만들어낸다. 현실에 존재하지 않을 법한 데이터를 바탕으로 모델을 해석하는 것은 모델이 오류를 보이는 특수 케이스를 찾는 데에는 도움이 될 수 있지만, 일반적인 상황에서의 모델의 행동양식을 살펴보는 데에는 부적합하다. SMOTE 알고리즘은 데이터 오버샘플링을 위한 기법 중 하나이다[28]. 이 기법은 KNN(k-최근접 이웃 알고리즘)을 바탕으로 개별 인스턴스와 가장 유사한 인스턴스를 찾아 그 두 인스턴스 사이 값에 해당하는 값으로 새로운 인스턴스를 생성한다. 본 연구는 SMOTE 알고리즘을 모델 해석을 위한 데이터셋 변형 시점에 적용시켜 분석 시스템이 보다 현실에 가까운 분석 결과를 도출하도록 만들었다.

기존 연구들의 또 다른 한계점은 사용자에게 해석의 단서를 충분히 제공하는 대신 단순히 모든 가능한 결과를 나열한다는 점이다. 머신러닝 모델이 해결하고자 하는 문제는 점점 복잡해지고 있다. 기존의 분석 툴이 제공하는 해석 방식은 입력 변수와 인스턴스 수가 많아질 때 사용이 어렵다. 또한, 머신러닝 전문 지식이 부족한 일반 실무자는 실제 현상과 모델 예측 결과, 또는 실제 데이터와 생성된 데이터를 구분 지어 해석하기 어려울 수 있다[29]. 본 연구는 머신러닝 전문 지식이 부족한 일반 실무자가 머신러닝 모델을 이해하고 해석하는데 드는 어려움을 덜기 위하여 시각화 요소와 인터랙션 기법을 활용하여 분석 시스템을 설계했다.

방대한 데이터를 탐색하는 상황에서, 사용자가 오버뷰로부터 시작하여, 관심사를 필터 및 정렬하고, 디테일을 살펴보는 순서로 데이터를 분석하게 만들 때, 분석의 난이도는 낮아지고 분석 결과의 질은 높아진다[30]. 본 연구는 사용자에게 해석의 단서를 충분히 제공하기 위하여 이 3단계의 분석 순서를 유도한다. 최근 등장하는 인터랙티브 분석 시스템은 필터 뿐만 아니라 그룹, 정렬, 특정 값 찾기, 입력 변수 간의 상관관계 제공하기 등 다양한 분석 기법을 제공하여 사용자가 다각도로 데이터를 분석할 수 있게 돕는다[14-16,20,31].

특히, 머신러닝 모델 해석이 어려워지는 이유는 인스턴스 수와 입력 변수의 수가 늘어가고 해결하고자 하는 문제가 복잡해짐에 따라 개별 인스턴스의 특징이 전체 데이터 셋을 대표하기 어려워진 데에 있다. 이에 따라 여러 인스턴스를 한 번에 살펴볼 수 있는 기능의 필요성이 대두되고 있다. 본 연구는 필터에, 그룹, 정렬 기능을 중심으로 통해 사용자가 보다 쉽고 빠르게 모델 해석에 대한 발견점을 얻을 수 있도록 도왔다.

본 연구는 입력 데이터 변화가 모델 결정에 미치는 영향을 추적하여 사용자가 모델을 해석하는 것을 돕는 것에 목표를 두고 시각적 분석 시스템을 디자인했다. 과학적 이해, 안정성 파악, 신뢰 확보와 같은 해석가능한

머신러닝의 장점을 획득할 수 있는지가 시스템 디자인과 평가의 주요 고려사항이다[7]. 본 연구가 제안하는 시각적 분석 시스템은 다음과 같은 디자인 목표를 달성하고자 한다.

- G1: 입력 변수와 모델 예측 값 사이의 관계 파악. 이 목표를 통해 사용자는 특정 머신러닝 모델이 어떤 입력 변수에 민감하거나 민감하지 않은지 파악할 수 있다. 모델의 예측 정확도가 높은 경우에는 모델이 학습한 바를 해석하여 입력 변수와 타겟 변수 사이의 관계에 대한 지식을 얻을 수 있다. 또한 인종과 성별 등의 편견적 요소가 모델에 학습되어 있는지 파악할 수 있다.
- G2: 입력 변수와 모델 예측 오류 사이의 관계 파악. 머신러닝 모델의 예측 정확도를 높이는 입력 변수 목록과 오류를 크게 만드는 입력 변수 목록을 구할 수 있다. 이 목표가 달성될 경우 사용자는 머신러닝 모델의 행동과 예측 패턴을 이해하고 결정을 신뢰하거나, 오류를 파악하고 대응할 수 있게 된다.
- G3: 입력 변수 변화에 따른 모델 예측 변화 추이 파악. 특정 입력 변수 및 변수의 조합의 값 변화가 모델 예측의 변화를 크게 만드는지 파악하면 사용자가 모델이 작동하는 방식에 대해 이해하도록 도울 수 있다. 또한 모델 예측 정확도가 낮은 경우에는 작은 입력 값 변화가 큰 예측 오류로 이어지는 상황을 파악하여 모델의 안정성을 진단할 수 있다.
- G4: 개별 인스턴스 확인. 유저가 분석하는 대상의 개별 인스턴스 목록을 파악. 개별 인스턴스 목록을 아는 것은 그렇지 않을 때 보다 분석 결과를 스토리로 만들고 타인과 공유하는 데 용이하게 만든다[32].

#### 4. 시각적 분석 시스템 설계 및 개발

본 연구가 제안하는 시스템이 작동하는 개략적인 과정은 그림 4를 통해 도식화된 내용과 같다. 우선, 유저는 데이터 및 모델을 시스템에 입력한다. 이 시스템이 다루는 데이터의 유형은 정형 다차원 데이터이며, 모델은 분류 또는 수치 예측 문제를 해결한다. 데이터는 CSV 파일로 입력한다. 입력할 파일은 설정 코드 파일을 통해 편집해야한다. 이를 그래픽 유저 인터페이스로

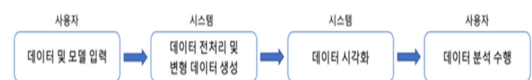


그림 4 본 연구가 제안한 시각적 분석 시스템을 통해 머신러닝 모델을 분석하는 과정

Fig. 4 Process of analyzing machine learning model through proposed visual analytics system



제공하지 않은 점은 본 연구의 아쉬운 점으로 향후 연구에서 개선할 수 있는 한계이다. 데이터와 모델이 입력 되면 시스템은 데이터 전처리를 수행한다. 이 과정에서 모델은 결측치를 제거하고, 기존 데이터를 바탕으로 입력 데이터를 변형한 신규 데이터를 생성한다. 다음으로 시스템은 처리된 기존 데이터와 생성된 신규 데이터를 시각화한다. 기존 데이터는 시스템 상단 평행 좌표 그래프와 산점도에 표시되며, 생성된 데이터는 시스템 하단 표에 나타난다. 유저는 시각화와 인터랙션을 활용하여 모델 분석을 수행한다.

그림 5의 사례는 인적 정보를 바탕으로 의료 비용 지출액을 예상하는 모델을 해석한다. 본 연구는 시각화 매핑과 유스 케이스 분석 사례를 설명하기 위해 Insurance Premium Prediction 데이터 셋이 제공하는 1,338명의 사람의 7개의 인적 정보(나이, 성별, BMI, 주거지역 등)를 바탕으로 의료비용을 예측하는 인공 신경망 모델을 만들고 해석했다[33]. 4장에서는 시각화 매핑 방법을 중심으로 설명하며 데이터와 모델 해석 결과는 5 장 유스 케이스 분석에 작성했다.

본 연구가 제안하는 시스템의 인터페이스는 4개의 구역으로 나뉘어 있다. (A)영역은 데이터, 모델 정보, 시각화 매핑 기법을 소개한다. (B)는 평행 좌표 그래프로 모델 해석을 위한 실제 인스턴스의 분포를 보여준다. 이곳에서 유저는 인스턴스를 필터링할 수 있다. (C)는 인스턴스의 입력 변수 별 분포를 보여주는 산점도 목록이

다. 각 산점도는 입력 변수를 세로축, 모델 예측 값을 가로축으로 표시하여 입력 변수의 값과 모델 예측 값 사이의 관계를 표시한다. (D)는 입력 변수가 변형된 인스턴스를 변형된 입력 변수의 종류 별로 그룹지어 히스토그램으로 시각화한 테이블이다.

본 연구가 제공하는 시각적 분석 시스템은 모델의 종류나 학습 방법과 관계없이 모델의 입출력 결과를 통해 해석한다. 또한, 본 시스템은 데이터도 수치 및 범주형 입출력 변수로 구성된 모든 정형 데이터에 적용 가능한 기법을 사용한다. 따라서, 본 설명에서 사용한 데이터 셋이나 머신러닝 모델이 아니더라도 이 시스템이 제공하는 분석 방법을 같은 방식으로 적용할 수 있다. 본 연구가 제안하는 분석 시스템의 설명 비디오 링크는 다음과 같다(<https://bit.ly/3dhlGH9>).

#### 4.1 평행 좌표 그래프 및 산점도

예측 결과는 평행 좌표 그래프(Parallel Coordinates Plot) 및 산점도(Scatter Plots)를 통해 그림 6처럼 시각화된다. 이 기법을 활용하면 입력 변수와 타깃 변수, 예측 결과 사이의 상관관계를 개략적으로 살펴볼 수 있다 (G1). 평행 좌표 그래프는 입력 변수, 타깃 변수의 실제 값(actual value), 예측 값(predicted value), 타깃 변수의 예측 값과 실제 값 사이의 차이(predicted - actual)를 축으로 가진다. 입력 변수와 관련된 축을 좌측, 타깃 변수와 관련된 축을 우측에 배치하였다. 입력 변수 축의 순서는 사용자가 입력한 CSV 파일의 순서를 따른다.

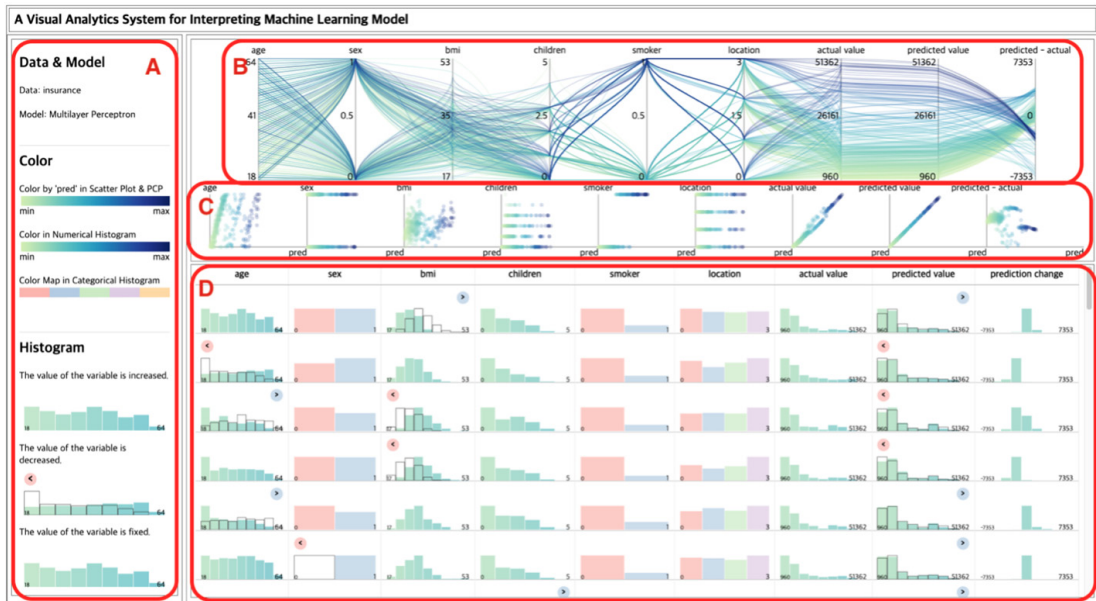


그림 5 본 연구가 제안하는 시각적 분석 시스템  
Fig. 5 Visual analysis system proposed in this study)

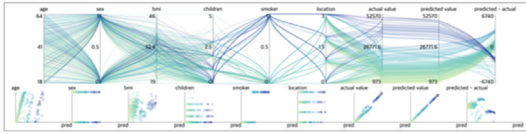


그림 6 제안된 시스템의 평행 좌표 그래프와 산점도

Fig. 6 A parallel coordinate graph and scatter plots in the proposed system

하나의 가로 선은 하나의 인스턴스를 의미하며, 각 축에서의 선의 높이가 해당 인스턴스가 가지는 값을 표시한다. 평행 좌표 그래프 바로 아래에 위치한 각 산점도는 세로축을 입력 변수, 타깃 변수의 실제 값, 예측 값, 타깃 변수의 실제 값과 예측 값 사이의 차이 값을 가지고, 가로축으로 타깃 변수의 모델 예측 값을 가진다. 가로축은 사용자의 설정에 따라 타깃 변수의 실제 값으로 변경할 수 있다. 산점도에서 하나의 점은 하나의 인스턴스를 의미한다. 평행 좌표 그래프와 산점도 축의 순서와 가로 위치를 같게 하여 사용자가 관심 있는 축을 빠르게 찾을 수 있도록 지원하였다. 또한 평행 좌표 그래프와 산점도 모두에서 인스턴스의 색을 타깃 변수의 값에 매핑시켰다. 사용자는 인스턴스의 색을 통해 타깃 변수의 예측 값을 높게 만드는 인스턴스가 다른 입력 변수에서 가지는 값의 분포를 파악할 수 있다. 수치형 예측의 경우 타깃 변수의 예측 값이 클수록 진한 푸른색을 가지고, 낮을수록 연한 녹색으로 표시된다. 분류형 예측의 경우 예측된 값이 참일 경우 푸른색, 거짓일 경우 붉은색을 가지게 하였다. 컬러 스케일 및 색상을 매핑할 변수는 사용자의 설정에 의해 타깃 변수의 실제 값이나 기타 변수 등으로 변경 될 수 있다. 이 두 시각화를 통해 사용자는 상관관계가 깊은 변수 목록을 찾는 데 도움을 받을 수 있다(G1, G2).

전역적 기법은 모델의 전반적인 개요를 파악하는 데 적절한 대신 특정 상황에 대한 구체적인 분석이 어렵다는 단점을 지닌다. 이러한 단점을 극복하기 위해 시각적 분석 시스템에는 전역 환경에서의 개요를 제공하는 한편, 인터랙션을 통해 국소적인 상황으로 심층 분석할 수 있는 기능이 요구된다. 또한, 머신러닝 모델이 입력 변수의 값으로부터 받는 영향을 보기 위해서는 입력 변수와 예측 값을 보는 것에 더해서 입력 값의 변화가 예측 값에 어떤 변화를 주는지 살펴보는 것이 도움이 된다[25,26]. 전체 인스턴스 입력 값의 변화가 모델 전반의 예측에 주는 영향을 분석하는 방법은 대다수의 평균적 상황에 특정한 상황이 가려질 수 있다는 한계가 발견되었다. 반면, 개별 인스턴스의 입력 값 변화에 대해서 모델 예측의 변화를 살펴보는 연구는 해당 케이스가 얼마나 일반적으로 드러나는지 확인하기 어렵다는 단점을

가졌다[34]. 이에 따라, 평행 좌표 그래프와 산점도 영역은 필터, 생성, 그룹, 정렬 인터랙션을 통해 모델 해석을 지원한다.

#### 4.2 인스턴스 변형과 테이블 시각화

시스템 하단 테이블 시각화는 필터링된 인스턴스를 원본으로 가지고 있는 변형 인스턴스만을 보여준다. 또한 변형된 데이터 셋을 변형 조건에 따라 그룹 지어 다수의 상황에 의해 나머지 특수한 경우가 가려지지 않도록 했다. 뿐만 아니라, 해당 그룹들을 예측 값 변화량이 큰 순서로 정렬시켜 사용자가 먼저 살펴봐야할 그룹이 무엇인지 안내했다. 평행 좌표 그래프와 산점도가 모델이 여러 입력 변수에 어떻게 반응하는지 전역적인 개요를 보여줬다면, 테이블 시각화는 유저 인터랙션을 통해 특정 상황에 해당하는 데이터를 선택하여 국소적인 해석을 지원한다(G3). 인스턴스 변형은 관심사를 좁히는 필터, 해당 필터에 해당하는 인스턴스를 변형하여 테스트 셋을 만드는 생성, 생성된 인스턴스를 유형에 따라 나누는 그룹, 그룹들을 중요도 순으로 줄짓는 정렬의 순서로 진행된다.

그림 7은 BMI와 흡연 여부를 바탕으로 필터링을 수행한 모습을 표시한다 우선, 필터 단계는 평행 좌표 그래프와 산점도를 통해 수행된다. 이 두 시각화는 드래그 인터랙션을 제공하여 인스턴스를 선택할 수 있게 만든다. 사용자는 평행 좌표 그래프와 산점도의 마우스 드래그 인터랙션을 바탕으로 관심 범위를 좁힌다. 선택된 인스턴스는 유채색, 선택되지 않은 인스턴스는 밝은 무채색으로 시각화에 표시된다. 예를 들어, 예측과 실제 값이 크게 다른 케이스, 예측 값이 극단적으로 높거나 낮은 케이스 등을 드래그로 필터링할 수 있다.

생성 단계는 시스템에 의해 자동으로 수행된다. 시스템은 SMOTE 알고리즘을 사용하여 인스턴스의 입력 변수 값을 변형하여 새로운 인스턴스를 생성한다. 인스턴스의 입력 변수 값을 변형하여 추가한 방법은 다음과

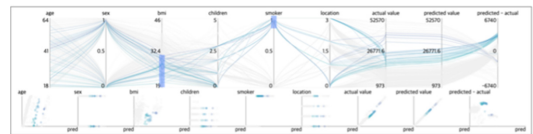


그림 7 다중 필터가 적용된 평행 좌표 그래프와 산점도. 이 그림에서는 BMI 지수가 19 이상 32 이하이면서 smoker(흡연 여부)가 참인 그룹이 선택되었다.

Fig. 7 Parallel coordinate graphs and scatterplots with multiple filters. In this figure, a group with a BMI index of 19 or more and 32 or less and a smoker was selected

같다. 첫째, 하나의 인스턴스(원본 인스턴스)를 골라 해당 인스턴스와 입력 변수 값의 분포가 유사한 인스턴스 K 개(변형 조건 인스턴스)를 선택한다. 이때, K는 유저 설정을 통해 변경할 수 있으며, 유사성 판단에는 맨해튼 거리 알고리즘을 사용한다. 둘째, 원본 인스턴스에서 변형할 변수와 참조할 변형 조건 인스턴스 하나를 선택하여 변수 값 변형을 진행한다. 만약 원본 인스턴스가 29 살 남성이고 변형 조건 인스턴스가 30 살 여성인데 변형할 변수가 성별이라면, 변형된 인스턴스는 29 살(원본 인스턴스 값 참조) 여성(변형 조건 인스턴스 값 참조)이 된다. 이 과정을 모든 가능한 두 개 이하의 변수 선택 조합에 대하여 반복한다. 이 과정을 변형 조건 인스턴스 K 개에서 반복한다. 마지막으로 원본 인스턴스로 모든 인스턴스가 활용될 때까지 위 작업을 반복한다. 이 과정을 통해 모든 가능한 두 개 이하의 변수 조합에 대하여 합리적인 값 범위 내에서의 변수 값 변형이 이뤄지게 된다. 인스턴스의 수가 N 개 변수의 수가 M 개라고 할 때, 총  $N * K * ((MC2) + (MC1))$  개의 변형 인스턴스가 생성된다.

이어서, 생성된 인스턴스를 변경된 입력 변수의 종류 및 변화의 방향을 바탕으로 인스턴스를 그룹 짓는다. 변화의 방향은 세 가지로 나뉘는데, 1) 수치형 변수의 값이 커짐, 2) 수치형 변수의 값이 작아짐, 3) 범주형 변수의 값이 변경됨이 있다. 이 그룹은 테이블 형태로 시각화된다. 테이블 시각화에서 하나의 열(column)은 개별 입력 변수, 타깃 변수의 실제 값, 기존 데이터의 예측 값, 입력 값이 변형된 후의 예측 값을 보여준다. 테이블에서 각 행(row)은 하나의 그룹을 의미한다. 각 그룹은 입력 값이 변형된 인스턴스들로 구성되어있다.

이 그룹은 그림 8 처럼 테이블 형태로 시각화된다. 테이블에서 각 열(column)은 개별 입력 변수, 타깃 변수의 실제 값, 기존 데이터의 예측 값, 입력 값이 변형된 후의 예측 값을 보여준다. 또한, 각 행(row)은 하나의 그룹을 의미한다. 각 그룹은 입력 값이 변형된 인스턴스들로 구성되어있다. 테이블에서 각 셀에는 해당 행의 그룹에 속하는 인스턴스가 해당 열에 해당하는 입력 변수에서 가지는 값의 분포를 표시하는 히스토그램이 시각

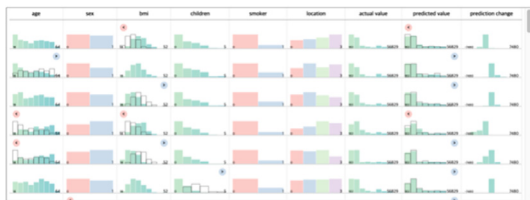


그림 8 제안된 a시스템의 테이블 시각화

Fig. 8 A table visualization in the proposed system

화된다. 변형이 이뤄지지 않은 입력 변수는 일반 색채 히스토그램만으로 표시된다. 수치형 변수는 값이 클수록 점점 진하게 표시되는 녹색 컬러 스케일을 활용했으며, 명목형 변수는 서로 쉽게 구분되는 색상으로 구성된 컬러 팔레트를 사용했다. 사용자는 색상을 이용하여 해당 그룹이 특정 변수에서 어떤 값을 많이 갖는지 빠르게 파악할 수 있다.

변형이 이뤄진 입력 변수는 색채가 있는 히스토그램과 테두리 선으로만 그려진 히스토그램 둘이 합쳐져 표시된다. 이때, 채색된 히스토그램은 원본 데이터의 값의 분포를 표시하며, 실선으로만 표시된 히스토그램은 값이 변형 생성된 데이터의 분포를 표시한다. 또한 변형이 생긴 입력 변수가 수치형 변수일 때는 셀 상단에는 동그란 칩 디자인 안에 화살표를 표시하여 입력 변수가 전반적으로 상승했는지, 또는 감소했는지 표시한다. 상승한 경우에는 푸른색 동그라미 안에 오른쪽 화살표가 표시되며, 감소한 경우 붉은색 동그라미 안에 왼쪽 화살표가 표시된다.

이 테이블 시각화는 입력 변수 값, 실제 값, 예측 값, 입력 변수의 변화 전후 예측 값의 변화량, 등을 바탕으로 정렬할 수 있다. 정렬된 테이블을 통해 사용자는 입력 변수 값의 변화가 모델 예측 값 변화에 어떤 영향을 주는지 파악할 수 있다(G3).

그림 9는 BMI가 상승하고 흡연 여부가 음에서 양으로 바뀐 그룹을 표시한다. 이 그룹은 입력 변수 변경 전보다 모델의 예측 값이 높아졌다. BMI와 흡연 여부 변화가 모델의 의료 비용 지출 예측 값을 높게 만드는 경향이 있음이 나타난다. 유저가 테이블의 행 위에 마우스 커서를 올리면 위와 같이 툴 팁 박스가 표시된다. 툴 팁 박스에는 변경된 입력 변수 및 예측 값 정보가 bmi+/smoker+/pred+ (BMI 상승, 흡연 여부 양으로 변경, 예측 값 상승)이 표시되고 있다.



그림 9 BMI가 높아지고, 흡연 여부가 거짓에서 참으로 바뀐 인스턴스 그룹을 표시한 테이블의 행

Fig. 9 A row in the table showing groups of instances with increased BMI and changed smoking from false to true

#### 4.3 개별 인스턴스 확인

사용자는 그림 10에 표시된 것과 같은 “자세히 보기 모달 윈도우”를 통해 개별 인스턴스 사례를 통해 사례를 탐색할 수 있다. 테이블 시각화는 개별 인스턴스의 값이 아니라 인스턴스 그룹의 통계치를 보여준다. 사용자는 개별 인스턴스 값을 보기 위해 더 자세히 보고 싶





그림 10 선택된 그룹에 속한 인스턴스 목록을 표시하는 모달 윈도우

Fig. 10 A modal window displays a list of instances belonging to the selected group

은 그룹을 선택할 수 있다.

마우스 클릭을 통해 그룹을 선택하면 모달 윈도우를 통해 개별 인스턴스 데이터를 보여준다. 이곳에서 사용자는 개별 인스턴스의 입력 변수 값과 예측 값의 변화를 확인할 수 있다. 사용자는 개별 인스턴스 목록을 활용하여 분석 결과를 타인에게 공유하는 데 사용할 수 있다(G4).

## 5. 유스 케이스 분석

### 5.1 유스 케이스 지침 설계

본 연구가 제안하는 시스템이 다양한 환경에 적용 가능한지 확인하기 위하여 다음 조건을 충족시키는 데이터 셋과 모델을 사용한 프로토타입을 제작했다. 첫째, 수치 예측 모델과 분류 예측 모델을 모두 실험한다. 둘째, 분석 도구의 도움 없이는 해석이 어려운 정도로 모델의 구조를 충분히 복잡하게 한다. 셋째, 수치형 데이터와 범주형 데이터가 복합적으로 구성된 데이터 셋을 사용한다. 이에 Insurance Premium Prediction 데이터 셋과 이를 학습한 인공 신경망 모델을 활용하고자 한다[33]. Insurance Premium Prediction 데이터 셋은 총 1,338명의 관찰 대상에 대한 나이, 성별, BMI, 흡연 여부, 거주지역, 의료 비용 지출 등 수치 및 범주로 표현 가능한 입력 변수를 가진다. 본 유스 케이스에서는 의료 비용 지출액을 나머지 입력 변수를 통해 예측하는 인공 신경망 모델을 개발하여 사용했다. 본 유스 케이스의 초기 화면은 그림 11과 같이 구성되어 있다.

이 사례 연구에서는 시스템의 사용 방법과 이점에 대한 전반적인 설명을 제공한다. 이 절에서는 설계 목표(G1~G4)에 대하여 질문(Q1~Q4)을 세우고, 그 질문에 답을 얻는 과정을 설명함으로써 본 연구가 제안하는 시스템의 사용성을 설명한다.



그림 11 유스 케이스에 해당하는 인터페이스는 Insurance Premium Prediction 데이터 셋과 이를 학습한 모델 인공 신경망 모델을 해석한다

Fig. 11 It interprets A model artificial neural network model learned by the Insurance Premium Prediction dataset

- Q1: 머신러닝 모델의 예측 값을 높게 만드는 입력 변수 목록은 무엇인가?
- Q2: 머신러닝 모델의 예측 값과 데이터의 실제 값이 크게 다른 인스턴스들은 공통적으로 어떤 입력 변수 값 분포를 가지고 있는가?
- Q3: Q1에서 선정된 변수 중 하나가 변경될 경우 머신러닝 모델이 예측한 의료 비용 지출 예측 값 변화는 어떠한가?
- Q4: Q1에서 선정된 변수 중 하나의 값이 큰 그룹에서, Q1에서 선정된 변수 중 다른 하나가 변경될 경우 흡연 여부 변경이 모델의 예측 값 미치는 영향을 보여주는 개별 인스턴스에는 무엇이 있는가?

### 5.2 유스 케이스 분석 수행

Q1: 머신러닝 모델의 예측 값을 높게 만드는 입력 변수 목록은 무엇인가?

이 질문은 그림 12의 사례처럼 평행 좌표 그래프의 predicted value 축의 값이 높은 인스턴스를 필터링하여 확인할 수 있다. Predicted value 축은 타깃 변수의 모델 예측 값을 표시한다. 따라서, 이 값이 높은 인스턴스를 추려내어, 이들의 변수 분포를 살펴보는 것이 질문에 대한 대답을 내는 데 도움이 된다. 사용자는 필터링된 인스턴스의 입력 변수 값 분포를 산점도에서 살펴볼 수

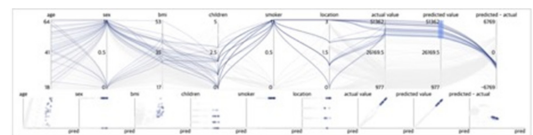


그림 12 타깃 변수의 모델 예측 값이 높은 인스턴스를 필터링한 모습

Fig. 12 Instances with high predicted values were filtered

있다. 다른 변수에서는 큰 경향성을 찾기 어렵지만, smoker(흡연 여부) 변수에서는 차이가 있다. 의료 비용 지출 예측 값이 높은 대부분의 인스턴스가 흡연자로 기록되어 있음이 산점도에 나타난다. 또한 흡연 여부보다는 미미한 경향성이지만 BMI 지수와 의료 비용 지출 예측 값에도 양의 상관관계가 있음이 보인다.

Q2: 머신러닝 모델의 예측 값과 데이터의 실제 값이 크게 다른 인스턴스들은 공통적으로 어떤 입력 변수 값 분포를 가지고 있는가?

이 질문도 평행 좌표 그래프 필터링과 산점도 시각화 분석을 통해 확인할 수 있다. 사용자는 이 질문에 대답하기 위해서, 그림 13처럼 실제 값과 모델 예측 값의 차이를 표시하는 축을 통해 인스턴스를 필터링할 수 있다. 예측 값과 실제 값이 다른 경우는 예측 값이 큰 경우와 작은 경우 둘로 나눌 수 있다. 이번 사례에서는 예측 값이 실제 값보다 큰 사례를 살펴본다. 평행 좌표 그래프와 산점도의 분포를 통해 흡연 여부가 참인 인스턴스에서 실제 값 보다 모델 예측 값이 크게 예측됨을 확인할 수 있다. 여기서 독특한 점은 대부분의 인스턴스가 실제 의료 비용 지출 값이 중간 값 정도에 분포하고 있다는 점이다. 이를 통해 사용자는 '모델이 중간 값 정도의 의료 비용을 지출하는 사람이 흡연할 경우 실제보다 크게 의료 비용을 예측한다.'고 추론할 수 있다.

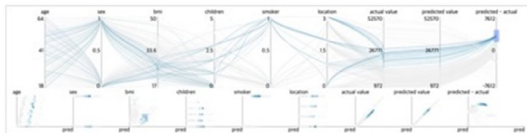


그림 13 타깃 변수의 모델 예측 값이 실제 값보다 큰 인스턴스를 필터링한 모습

Fig. 13 Instances with predicted value greater than actual value are filtered

Q3: Q1에서 선정된 변수 중 하나가 변경될 경우 머신러닝 모델이 예측한 의료 비용 지출 예측 값 변화는 어떠한가?

이 질문에 답하기 위해 하단 테이블 영역을 활용할 수 있다. 이 설명에서는 Q1에서 선정된 변수 중 하나로 smoker를 선택했다. 테이블 영역에서 smoker 열의 이름을 클릭해, 흡연 여부가 참에서 거짓 또는 거짓에서 참으로 변경된 그룹을 찾을 수 있다. 이들의 predicted value(모델 예측 값)에 흡연 여부의 변경이 모델 예측에 미치는 영향이 표시된다. 그림 14와 그림 15는 흡연 여부를 거짓에서 참으로 바꿀 경우 예측된 비용이 증가하고, 반대의 경우에는 감소하는 것을 보여준다. 사용자는 이를 통해 '흡연 여부가 모델의 의료 비용 예측에 영향

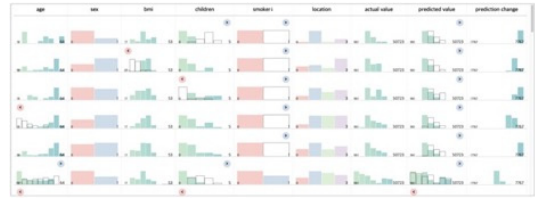


그림 14 흡연 여부가 거짓에서 참으로 변경된 그룹  
Fig. 14 Groups with smoking status changed from false to true

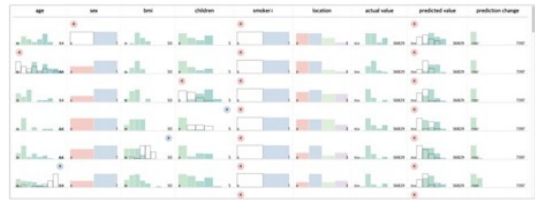


그림 15 흡연 여부가 참에서 거짓으로 변경된 그룹  
Fig. 15 Groups with smoking status changed from true to false

을 주며, 그 방향성은 양적(positive)이다.'라고 결론지을 수 있다. 이들은 다른 변수가 고정된 상황에서의 흡연 여부의 한계 효과(marginal effect)를 살펴본 것이기 때문에 평행 좌표 그래프와 산점도에서 확인한 것보다 더 확실할 수 있는 정보이다.

Q4: Q1에서 선정된 변수 중 하나의 값이 큰 그룹에서, Q1에서 선정된 변수 중 다른 하나가 변경될 경우 흡연 여부 변경이 모델의 예측 값 미치는 영향을 보여주는 개별 인스턴스에는 무엇이 있는가?

이 질문의 목적은 한 변수(A 변수)의 값을 특정 범위에 고정된 상태에서, 다른 변수(B 변수)의 값이 변화할 때 모델의 예측이 어떻게 바뀌는지를 살펴보는 작업을 진행하도록 하는 것이다. 이 사례에서는 A 변수로 BMI, B 변수로 smoker를 선정하였다.

이 질문에 답하기 위해서는 다음의 과정이 필요하다. 첫째, BMI가 높은 그룹을 평행 좌표 그래프에서 필터링한다. 이는 사용자의 마우스 드래그 인터랙션으로 가능하다. 둘째, 표에서 smoker 열을 클릭해 흡연 여부가 변경된 인스턴스를 맨 위로 정렬한다. 셋째, 정렬된 그룹 중 흡연 여부만 변경된 행(그림 16 속 테이블 시각화의 세 번째 행)을 마우스로 클릭하면 시스템은 이 조건을 갖춘 인스턴스 목록을 상세 보기로 띄워준다. 이곳에서 테이블 정렬 기능을 통해 변화 값이 크거나 작은 인스턴스를 대표 인스턴스로 선정할 수 있다. 입력 변수 변화 값이 큰 인스턴스를 선정할 경우, 예측 값 변화의 폭을 보여주는 데 유리하다. 반면, 입력 변수 변화 값이

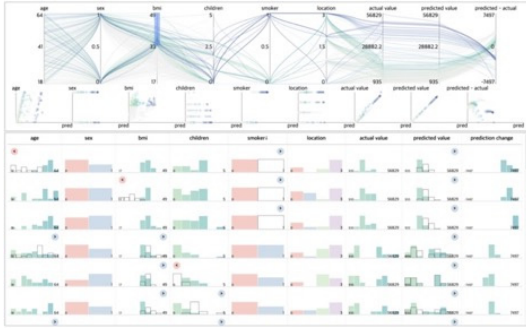


그림 16 BMI가 높은 그룹을 평행 좌표 그래프에서 필터링하고, 이들의 흡연 여부가 음에서 양으로 바꾼 그룹을 테이블 맨 위로 정렬시킨 모습

Fig. 16 Groups with high BMI are filtered from a parallel coordinate graph and groups whose smoking status changes from false to true are aligned at the top of the table

작은 인스턴스를 선정할 경우에는, 입력 변수의 작은 변화가 큰 예측 변화로 이어지지 않는지에 대한 안정성 검증을 수행할 수 있다. 대표 인스턴스는 사용자의 상황에 따라 다른 속성을 가지므로 본 시스템의 정렬 인터랙션이 유용하다.

## 6. 사용자 심층 인터뷰

### 6.1 사용자 심층 인터뷰 설계

본 연구에서는 입력 변수가 변형된 인스턴스 생성과 이를 탐색하는 필터링, 그룹, 정렬 등의 다양한 인터랙션이 사용자의 모델 해석에 도움이 된다는 연구결과를 바탕으로 시각적 분석 시스템을 설계 및 개발하였다. 제안된 시스템의 사용성 검증을 위하여, 3 명의 대상자를 바탕으로 심층 인터뷰를 진행하였다.

안재욱의 연구는 특정한 데이터를 분석하는 도구와 기법을 평가하기 위한 지표를 제안했는데 이는 포괄적 이해, 탐색 가능성, 유용성, 용이한 정도 등을 포함한다 [35]. 본 연구는 이 지표들을 머신러닝 모델의 해석을 위한 시각화 도구 평가에 적용했다.

- 1) 포괄적 이해: 해결해야 하는 문제 및 머신러닝 데이터 셋의 개요와 모델의 전반적인 결정 패턴을 파악하는 데 도움이 되는지 평가한다.
- 2) 탐색 가능성: 사용자가 모델 해석을 위해 어떤 입력 변수, 인스턴스, 인스턴스 그룹 등을 살펴봐야 하는지 아는 데 제안된 도구가 도움이 되는지 평가한다.
- 3) 유용성: 문제 해결을 위한 질문에 명확한 답을 찾아내는데 제안된 도구가 도움이 되는지 평가한다.

- 4) 용이한 정도: 제안된 도구를 타깃 사용자가 이해하고 활용하는 데 어려움이 있는지 조사한다.

인터뷰 대상자는 전문가 1명(이하 P1으로 칭함)과 비전문가 2명으로, 비전문가는 각각 실무 경험이 없는 전공자(이하 P2로 칭함)와 간단한 실무 경험이 있는 비전공자(이하 P3로 칭함)로 구성하였다. 전문가 인터뷰는 시스템의 기여점과 한계를 파악하는 데 주요 초점을 맞췄고, 비전문가 인터뷰는 타깃 사용자가 시스템을 이해하고 활용하는 데에 어려움과 혼동이 있는지 파악하는 데 초점을 맞췄다.

인터뷰에는 썹크 얼라우드 (think aloud) 기법이 활용되었다[36]. 이 기법은 사용자가 일련의 작업을 수행하는 동안 머릿속으로 생각하는 것을 말로 표현하도록 요청한다. 사용자의 발화를 통해 어떤 부분에서 편의와 혼란을 겪는지 이해하는 언어 프로토콜 분석 방법이다. 이 기법은 작업 수행 시 사용자의 기대, 경험, 문제점 등을 효과적으로 밝혀낼 수 있다는 장점이 있다.

비전문가 사용자에게 있어서 머신러닝 모델 해석은 난이도가 높은 작업이다. 이런 상황에서는, 사용자가 작업을 완료한 후 인터뷰 진행자의 질문을 통해 사용자의 경험을 적극적으로 조사하는 회고적 썹크 얼라우드 프로토콜(Retrospective Think Aloud Protocol, RTA) 방식이 권장된다[37]. 본 연구에서는 RTA 방식을 사용하여 유저의 사용 경험을 분석했다. 또한, 사용자가 시스템의 목적과 수행할 테스트를 충분히 이해할 수 있도록 준비를 할 수 있는 설명 세션을 충분히 제공했다.

심층 인터뷰 진행은 각 대상자 별로 약 1시간 30분이 소요되었다. 인터뷰는 설명 세션 45분과 사용 및 평가 세션 45분으로 나뉜다. 설명 세션에서는 연구 개요와 목적, 같은 목적을 공유하는 기존 연구 사례 그리고 설계된 시스템을 소개했다. 이때, 유스 케이스 분석에 활용된 의료 비용 예측 모델 분석 프로토타입을 바탕으로 시스템을 사용했다. 사용 및 평가 세션에서는 대상자 스스로 설계 목표(G1~G4)와 관련된 질문을 세우고 이에 대해 스스로 탐색하도록 하였다. 이때, 당뇨 여부 분류 모델 분석 프로토타입을 통해 사용자가 모델을 분석해 보게 하면서 시스템 전체와 개별 기능 모두에서 활용 난이도와 분석 유용성을 평가하도록 했다. 특히, 이 과정에서 본 연구의 주요 제안 사항인 실제 데이터와 변형된 데이터 모두를 활용하여 모델의 예측 패턴을 해석하는 일을 수행하는 데 본 연구가 제안하는 시스템의 효과를 살펴보았다.

사용자 인터뷰의 사용 및 평가 세션에서는 Pima Indian Diabetes 데이터 셋과 이를 학습한 로지스틱 회귀 모델을 분석하는 사례를 사용하였다[38]. Pima Indian Diabetes 데이터 셋은 임신 횟수, 포도당 부하 검사 수치, 혈압, 혈

청 인슐린 등의 의료 정보와 함께 768명의 관찰 대상자의 당뇨병 유무를 기록한 데이터이다. 본 인터뷰 세션에서는 이 데이터를 활용하여 의료 정보를 토대로 당뇨병 여부를 예측하는 로지스틱 회귀 모델을 개발하여 인터뷰에 활용하였다. 사용자 인터뷰에 사용된 인터페이스의 초기 화면은 그림 17과 같이 구성되었다.

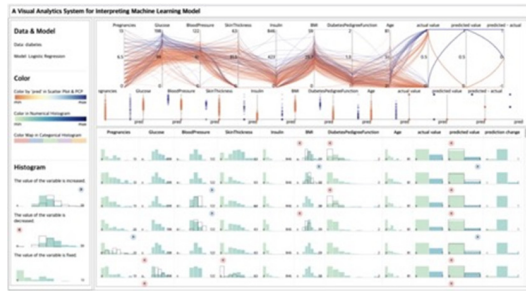


그림 17 제안된 시스템에 Pima Indian Diabetes 데이터 셋과 이를 학습한 로지스틱 회귀 모델이 적용된 모습

Fig. 17 Pima Indian Diabetes dataset and logistic regression model from which they were learned are applied to the proposed system

## 6.2 사용자 심층 인터뷰 수행 결과

사용자 인터뷰에서 모든 사용자는 본 연구가 제안하는 시스템을 이해하고 활용하여 머신러닝 모델을 분석하는 데 큰 어려움을 겪지 않았다. 인터뷰로부터 다음 세 가지 발견점을 얻을 수 있었다. 첫째, 본 연구가 제안하는 시스템은 사용자가 실제 현상이 아니라 모델의 작동 패턴을 해석하는 것에 초점을 맞춰 분석 작업을 하도록 유도하는 데 도움을 주었다. 사용자는 제안된 시스템을 통해 데이터 및 모델을 포괄적으로 이해하고 분석에 필요한 업무 목록을 수월하게 파악했다. 둘째, 본 연구가 제안하는 시스템은 머신러닝 비 전문가 사용자 층으로 하여금, 실제 데이터와 변형된 데이터를 각각 분석 목표에 맞춰 적절하게 사용하도록 도왔다. 사용자는 제안된 시스템을 활용해 원하는 분석 작업에 대한 명확한 답을 구할 수 있었다. 셋째, 본 연구가 제안한 시스템의 시각화 요소 사용 및 레이아웃 구성은 머신러닝 및 데이터 시각적 분석에 익숙하지 않은 사용자에게도 이해와 활용이 어렵지 않았다.

머신러닝 및 데이터 분석 분야의 전문 지식이 부족한 P2는 설명 세션에서 해석가능한 머신러닝 분야가 ‘실제 현상을 해석하는 것이 아니라 머신러닝 모델의 예측 결과를 해석’하는 목적을 가졌다는 점을 이해했다. 하지만, P2는 기존 연구 사례인 PD 플롯의 결과를 해석하면서 실제 현상과 모델 예측이 차이를 구분 짓는 데 혼란을

겪었다. 예를 들어, 온도가 높을수록 자전거 대여의 예측 값이 올라가는 시각적 표현을 자전거 대여의 실제 값이 올라가는 것으로 오인했다. 반면, 해당 대상자는 본 연구의 결과물을 통해 모델을 해석할 때에는 실제 값과 모델 예측 값을 쉽게 구분했다. 즉, 제안된 시스템은 사용자의 포괄적 이해와 용이한 정도에 도움이 되었다.

P2는 평행 좌표 그래프에 실제 값을 표시하는 축과 예측 값을 표시하는 축이 각각 존재하는 것이 이 두 값을 구분하는 데 도움이 되었다고 평가했다. 또한 P3는 평행 좌표 그래프에 예측 값과 실제 값의 차이 값을 표시하는 축이 있어서, 모델의 정확도를 높이는 입력 변수와 오류를 크게 만드는 입력 변수를 파악하는 데 도움이 되었다고 말했다. P3는 예측 값과 실제 값의 차이를 표시하는 축에서 예측 오류가 큰 인스턴스를 마우스 드래그 인터랙션으로 필터링하고, 산점도를 확인하여 필터링된 인스턴스가 다른 입력 변수에서 어떤 분포를 보이는지 파악하였다. 이 점을 미루어보아, 머신러닝 실무 경험이 부족한 사용자에게 평행 좌표 그래프의 축 활용이 모델 해석에 도움을 주는 점을 알 수 있었다. 이들 축의 활용을 통해 인터뷰 대상자들은 모델 해석을 위해 먼저 살펴봐야 할 인스턴스 목록을 추천했다. 뿐만 아니라, P3는 데이터 셋과 모델에 대한 설명이 없는 상황에서 필터링 및 정렬을 통해 모델 특성 파악 업무를 수월하게 진행하고, 평행 좌표 그래프와 산점도를 활용하여 극한값을 가지거나 특정 분포를 띄는 인스턴스를 필터링하는 등의 분석 작업을 수행했다. P3는 인터뷰에서 ‘시각적으로 튀는 요소가 궁금증을 유발했고, 이를 자연스럽게 먼저 살펴보게 되었다’고 설명했다. 이러한 인터뷰 결과를 통해 평행 좌표 그래프 및 산점도의 활용이 사용자에게 필요한 업무가 무엇인지 탐색하는 데 긍정적인 영향을 주었다고 평가할 수 있다. 제안된 시스템의 인터랙션 제공 방식이 탐색 가능성 및 용이한 정도가 긍정적인 영향을 주었다고 해석할 수 있다.

모든 인터뷰 대상자는 제안된 시스템이 실제 데이터와 변형된 데이터 모두를 활용하여 모델을 분석한다는 점을 이해하고, 이를 분석 업무에 활용했다. P2는 평행 좌표 그래프 및 산점도가 표시된 시스템 상단 영역에 실제 데이터가 표시되고, 테이블 형태로 히스토그램이 배치된 시스템 하단 영역에 변형된 데이터가 표시되는 점이 두 가지 데이터를 구분할 수 있는 요인이라고 언급했다. 실제 데이터와 변형된 데이터가 각각 물리적으로 구분된 공간에 표시되기 때문에 이를 구분 짓기 편하다는 의미이다. 이는 제안된 시스템의 시각적 표현 기법이 포괄적 이해 확보에 도움이 되었음을 시사한다.

P3는 레이아웃에 더해서 시각적 표현의 장점이 이 두



가지 데이터를 구분하게 만드는 데 도움이 된다고 평가했다. 히스토그램이 실제 값과 변형 생성된 값을 동시에 시각화하고, 이를 변형된 값의 방향성에 따라 파란색(값 커짐) 또는 빨간색(값 작아짐) 바탕을 가진 화살표와 함께 표현한 점을 장점으로 언급했다. 기존 연구에서는 변형된 값과 기존 값의 색상, 위치, 모양 등이 다르지 않아서 분석에 어려움이 있었던 한계점을 극복한 것이라고 볼 수 있다. 본 연구에서 화살표 및 색칠 여부가 다른 히스토그램 기법을 도입할 수 있었던 이유는 변형된 인스턴스를 변형된 변수의 종류 및 변화 방향(상승 및 하락 등)에 따라 그룹 지었기 때문이다. 그룹 기능과 레이아웃 배치와 시각적 요소 사용이 시스템의 유용성 및 용이한 정도 높이는 데 긍정적 효과를 줬다고 해석할 수 있다.

또한, 본 인터뷰를 통해 시스템 레이아웃 및 구성이 적절하게 되었음을 확인했다. P2와 P3는 별도의 지시 없이도 평행 좌표 그래프 확인, 산점도 확인, 필터링 인터랙션, 표, 정렬 인터랙션, 그룹 상세 보기 팝업 창 확인의 순서로 모델 해석 작업을 진행했다. 이는 4 장에서 소개한 분석 시스템이 작동하는 구조와 일치한다. 사용자가 기능을 순차적으로 사용하도록 유도할 수 있었던 이유는 각 기능이 페이지 상단에서 하단으로 순차적으로 배치되어 있을 뿐만 아니라, 각각의 기능이 오버 뷰(평행 좌표 그래프 및 산점도 확인), 관심 좁히기(평행 좌표 그래프 필터링 및 테이블 정렬), 자세히 보기(그룹 상세 정보 모달 윈도우)로 적절히 구분되어있기 때문이라고 볼 수 있다. 이는 '오버 뷰, 필터, 디테일'의 분리가 데이터 분석 시스템의 사용성을 높인다는 기존 시각화 레이아웃 연구의 제안과도 방향성을 같이한다[30]. 본 연구가 제안한 방식의 분석 시스템 구축이 사용자가 데이터와 머신러닝 모델을 포괄적으로 이해하고 필요한 업무가 무엇인지 탐색하는 데 도움을 주었다고 평가할 수 있다.

인터뷰를 통해 발견된 본 연구의 한계점 및 개선 요구 사항은 다음과 같다. 우선, 시각적 요소 해석의 어려움이 있었다. 평행 좌표 그래프에서 인스턴스를 표시하는 선이 꼬여 있거나 겹쳐 있는 경우가 다수 있어서 어떤 선이 어떤 선을 의미하는지 파악하기 어렵다는 피드백이 P1과 P3로부터 나왔다. 특히 인터뷰 P1은 다른 시각적 분석 시스템의 사례를 들어 시스템 사용성을 높이기 위해 축 순서의 변경, 축에 인스턴스의 빈도수 표시 등의 기능을 필요하다고 평가했다[39]. 다음으로 구체적인 값 파악이 어렵다는 점이 아쉬운 점으로 꼽혔다. P1은 데이터의 개요를 보다 정확하게 파악하기 위해 입력 변수간 상관계수를 보여주거나 인스턴스 그룹의 입력 변수 별 평균 및 분산 등 통계 정보가 필요하다고 지적

했다. 본 연구자는 이와 같은 피드백이 합당할 뿐만 아니라, 제안된 시스템에 쉽게 통합될 수 있는 기능으로 파악하였다. 이에 따라, 평행 좌표 그래프의 가시성 향상 및 통계치 정보 표시를 향후 개선 요구 사항으로 지정하였다.

## 7. 결론

본 연구는 데이터의 입력 변수가 모델의 예측 결과에 어떤 영향을 주는지를 바탕으로 머신러닝 모델을 해석하는 시각적 분석 시스템을 제안했다. 이를 위하여 SMOTE를 활용해 데이터 입력 변수 값을 변형하는 기술을 제안하고 변형된 데이터에 대한 모델 예측 값 변화 변화 추이를 시스템이 추적하도록 했다. 이 방식은 입력 데이터 값을 합리적인 범위 내에서 변화시킨다는 점에서 기존 연구의 한계를 극복했다.

또한, 제안된 시스템은 시각화 인터페이스를 통해 사용자의 모델 해석을 용이하게 했다. 머신러닝 예측 결과에 대해 필터, 그룹, 정렬 인터랙션을 적용했으며, 평행 좌표 그래프와 산점도, 테이블 시각화 등 다양한 시각화 양식을 조합했다. 이러한 기법이 모델 해석에 긍정적인 영향을 미침을 유스 케이스와 사용자 인터뷰를 통해 확인하였다. 본 연구의 기여는 머신러닝 비 전문가가 머신러닝 모델 해석 기법을 이해하고, 머신러닝 모델을 해석하는 데 도움이 되는 시각적 분석 시스템을 설계하고 평가했음에 있다.

이 연구의 한계점과 향후 개선 사항은 다음과 같다. 우선, 인터뷰에서 밝혀진 바와 같이, 평행 좌표 그래프의 가시성 향상 및 통계치 정보를 표시하여 시스템의 유용성을 높일 필요가 있다. 향후 연구에서는 이러한 한계점을 극복하기 위하여 정확한 수치 파악을 위한 통계치 표시 기능과 평행 좌표 그래프 및 산점도의 가시성을 높이는 방안을 도입할 필요성이 있다. 또한, 대규모 데이터에서의 시스템 확장성을 검증하지 못했다는 한계를 갖는다.

본 연구가 제안하는 시스템의 주요 목적은 입력 변수가 모델 출력에 미치는 영향을 판단하는 것이다. 이 주제를 연구 분야에서 개별 입력 변수가 모델 예측에 미치는 영향을 수치화하는 변수 중요도 연구가 주목받고 있다[40]. 향후 연구에서 이러한 변수 중요도 수치를 함께 시각화하면 사용자는 보다 정확하게 주요 변수를 파악할 수 있을 것이다.

## References

- [1] Molnar, Christoph, "Interpretable machine learning" Lulu.com, 2020.
- [2] Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A.,



- Specter, M., and Kagal, L. "Explaining explanations: An overview of interpretability of machine learning," *2018 IEEE 5th International Conference on data Science and Advanced Analytics (DSAA)*. IEEE, 2018.
- [3] Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning. arXiv preprint, arXiv:1702.08608, 2017.
- [4] Vilakone, P., Park, D. S., Xinchang, K., and Hao, F. "An efficient movie recommendation algorithm based on improved k-clique," *Human-centric Computing and Information Sciences*, Vol. 8, No. 1, pp. 1-15, 2018.
- [5] Lee, J. and Han, J. "Layer-wise Relevance Propagation (LRP) Based Technical and Macroeconomic Indicator Impact Analysis for an Explainable Deep Learning Model to Predict an Increase and Decrease in KOSPI," *정보과학회논문지*, Vol. 48, No. 12, pp. 1289-1297, 2021.
- [6] Jia, L., Wei, Z., Zhang, H., Wang, J., Jia, R., Zhou, M., ... and He, K. "An interpretable machine learning model based on a quick pre-screening system enables accurate deterioration risk prediction for COVID-19," *Scientific reports*, Vol. 11, No. 1, pp. 1-16, 2021.
- [7] Lim, Soohyun, et al. "A Study on Interface Design to Provide Effective Meeting Summary Information: Focused on Narrative Chart and Topic Map Visualization," *Archives of Design Research*, Vol. 34, No. 1, 2021.
- [8] Van Den Elzen, S., and Van Wijk, J. J., Van Den Elzen, Stef, and Jarke J. Van Wijk, "Baobabview: Interactive construction and analysis of decision trees," *2011 IEEE conference on visual analytics science and technology (VAST)*, IEEE, 2011.
- [9] Mitchell, Michael N., "Interpreting and visualizing regression models using Stata," Vol. 558. College Station, TX: Stata Press, 2012.
- [10] Kim, B., Wattenberg, M., Gilmer, J., Cai, C., Wexler, J., and Viegas, F., "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (tcav)" *International conference on machine learning*, PMLR, 2018.
- [11] Wang, Z. J., Turko, R., Shaikh, O., Park, H., Das, N., Hohman, F., ... and Chau, D. H. P., "CNN explainer: Learning convolutional neural networks with interactive visualization," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 2, 2020.
- [12] Ren, D., Amershi, S., Lee, B., Suh, J., and Williams, J. D., "Squares: Supporting interactive performance analysis for multiclass classifiers," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, No. 1, 2016.
- [13] Molnar, Christoph, Giuseppe Casalicchio, and Bernd Bischl, "Interpretable machine learning - a brief history," state-of-the-art and challenges. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Cham, 2020.
- [14] Park, H., Nam, Y., Kim, J. H., and Choo, J., "HyperTendril: Visual Analytics for User-Driven Hyperparameter Optimization of Deep Neural Networks," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 2, 2020.
- [15] Chanhee Park, Hyojin Kim, and Kyungwon Lee, "A Visualization System for Performance Analysis of Image Classification Models," *Electronic Imaging*, 2020.1, 2020.
- [16] Zhang, J., Wang, Y., Molino, P., Li, L., and Ebert, D. S., "Manifold: A model-agnostic framework for interpretation and diagnosis of machine learning models," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, No. 1, 2018.
- [17] Park, Y. and J. Y. Yun, "A Design Case Study of Artificial Intelligence Pipeline Visualization," *Archives of Design*, 2021.
- [18] Ming, Yao, et al. "Understanding Hidden Memories of Recurrent Neural Networks" *IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2017.
- [19] Kahng, M., Andrews, P. Y., Kalro, A., and Chau, D. H., "ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, No. 1, 2017.
- [20] Park, Chanhee, et al. "ComDia+: An interactive visual analytics system for comparing, diagnosing, and improving multiclass classifiers," *2019 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 2019.
- [21] Ming, Yao, Huamin Qu, and Enrico Bertini, "Rulematrix: Visualizing and understanding classifiers with rules" *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, No. 1, 2018.
- [22] Smilkov, D., Carter, S., Sculley, D., Viégas, F. B., and Wattenberg, M., "Direct- Manipulation Visualization of Deep Networks. Direct-manipulation visualization of deep networks," arXiv preprint arXiv:1708.03788, 2017.
- [23] Minsuk Kahng, Nikhil Thorat, Polo Chau, Fernanda Viégas, and Martin Wattenberg, "GAN Lab: Understanding Complex Deep Generative Models using Interactive Visual Experimentation," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, No. 1 (VAST 2018), Jan. 2019.
- [24] Ko, G., G. Lim, and H. Cho, "Survey on feature attribution methods in explainable AI," *Journal of KIISE*, Vol. 47, No. 12, 2020.
- [25] Friedman, Jerome H., "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, 2001.

- [26] Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E., "Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation," *Journal of Computational and Graphical Statistics*, Vol. 24, No. 1, 2015.
- [27] Cheng, Furui, Yao Ming, and Huamin Qu, "DECE: Decision Explorer with Counterfactual Explanations for Machine Learning Models," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 2, 2020.
- [28] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P., "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research* 16, 2002.
- [29] Kliegr, Tomáš, Štěpán Bahník, and Johannes Fürnkranz, "A review of possible effects of cognitive biases on interpretation of rule-based machine learning models," *Artificial Intelligence*, 2021.
- [30] Shneiderman, Ben, "The eyes have it: A task by data type taxonomy for information visualizations" *The craft of information visualization*. Morgan Kaufmann, 2003.
- [31] Amar, Robert, James Eagan, and John Stasko, "Low-level components of analytic activity in information visualization" *IEEE Symposium on Information Visualization*, 2005.
- [32] Aamodt, Agnar, and Enric Plaza, "Case-based reasoning: Foundational issues, methodological variations, and system approaches," *AI communications*, Vol. 7, No. 1, 1994.
- [33] Kaggle, "Insurance Premium Prediction", Kaggle [Online]. Available: <https://www.kaggle.com/noordeen/insurance-premium-prediction>, 2019.
- [34] Shi, Sheng, Xinfeng Zhang, and Wei Fan, "A modified perturbed sampling method for local interpretable model-agnostic explanation," *arXiv: 2002.07434*, 2020.
- [35] Ahn, Jae-wook, Catherine Plaisant, and Ben Shneiderman, A task taxonomy for network evolution analysis, *IEEE transactions on visualization and computer graphics*, Vol. 20, No. 3, 2013.
- [36] Van Someren, Maarten, Yvonne F. Barnard, and J. Sandberg, "The think aloud method: a practical approach to modelling cognitive," London: AcademicPress 11, 1994.
- [37] Eling, Sanne, Leo Lentz, and Menno de Jong, "Retrospective think-aloud method: Using eye movements as an extra cue for participants' verbalizations," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.
- [38] Kaggle, "Pima Indian Diabetes" Kaggle (<https://www.kaggle.com/uciml/pima-indians-diabetes-database/metadata>), 2016.
- [39] Steed, C. A., Shipman, G., Thornton, P., Ricciuto, D., Erickson, D., and Branstetter, M., "Practical application of parallel coordinates for climate model analysis," *Procedia Computer Science* 9, 2012.
- [40] Lundberg, Scott M., and Su-In Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems 2017*.



박 찬 희

2019 아주대학교 미디어학과 (학사). 2022 아주대학교 라이프미디어협동과정 (석사). 2021~현재 우아한형제들 사원. 관심분야: 데이터시각화, 데이터애널리틱스



이 경 원

1996 국민대학교 시각디자인학과 (학사) 2002 Computer Graphics & Interactive Media, Pratt Institute (석사). 2003~현재 아주대학교 소프트웨어융합대학 미디어학과 교수. 2009~2010 Visualization and Interface Design Innovation Lab, University of California at Davis 방문교수. 관심분야: 정보시각화, 인간-컴퓨터 상호 작용, 미디어아트