

DMQA open seminar

Applications of eXplainable Artificial Intelligence(XAI)

2022. 4. 8

Data Mining & Quality Analytics Lab.

발 표 자 : 황 석 철

hsc841123@korea.ac.kr

1. XAI 개요
2. SHAP(Shapley Additive exPlanations)
3. SHAP을 이용한 제품 설계
4. 정리



❖ 황 석 철

- 기계공학 전공 (구조해석, 최적설계)
- 고려대학교 산업경영공학과 Data Mining & Quality Analytics Lab.
- 지도 교수 : 김성범 교수님
- 석사 과정 (2021. 3 ~)

❖ 관심연구분야

- Machine Learning & Deep Learning, XAI

❖ 연락처

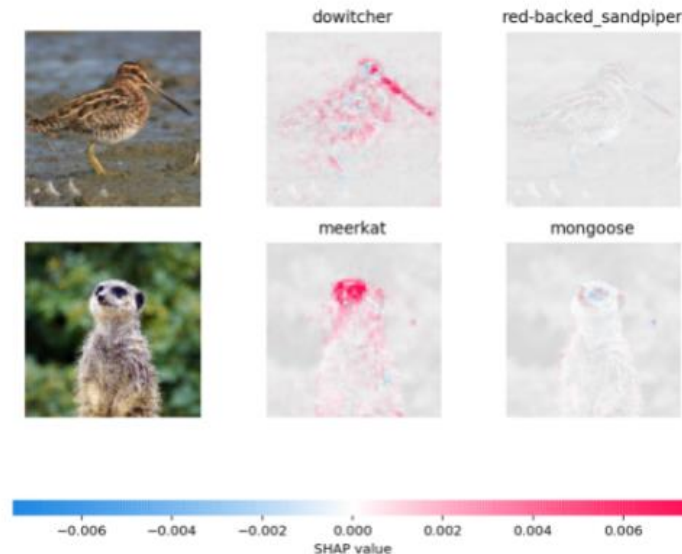
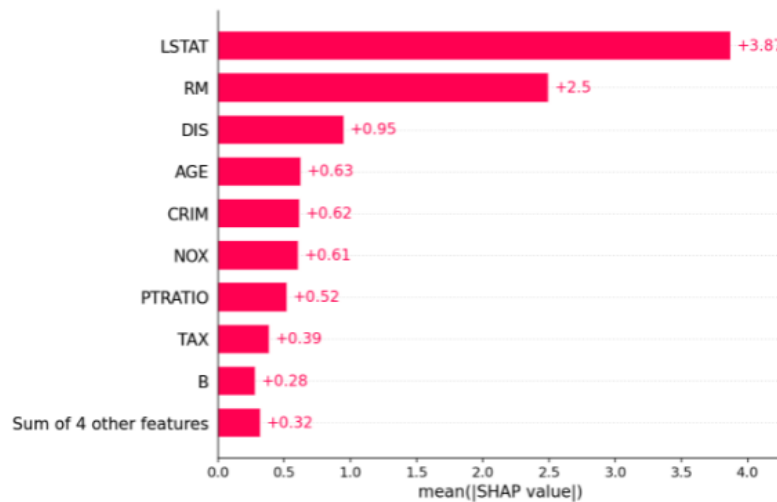
- hsc841123@korea.ac.kr

XAI 개요

1. XAI 개요

❖ XAI(eXplainable Artificial Intelligence, 설명 가능한 인공지능) 란?

- XAI는 인공지능 모델이 도출한 결과에 대해 어떤 근거로 의사 결정을 내렸는지 설명
- 인공지능 모델의 개선 혹은 도메인 전문가가 아닌 경우 신뢰성의 근거 확보에 효과적



이미지 출처 : <https://github.com/slundberg/shap>

1. XAI 개요

❖ XAI 역사

- XAI란 용어는 2004년 Michel van Lent, William Fisher, Michael Mancuso 가 제안
 - ✓ 모의 전투 프로그램에서 NPC와 플레이어의 전투 시 NPC의 행동 원인을 분석하고 설명함

Van Lent, M., Fisher, W., & Mancuso, M. (2004, July). An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of the national conference on artificial intelligence* (pp. 900-907). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

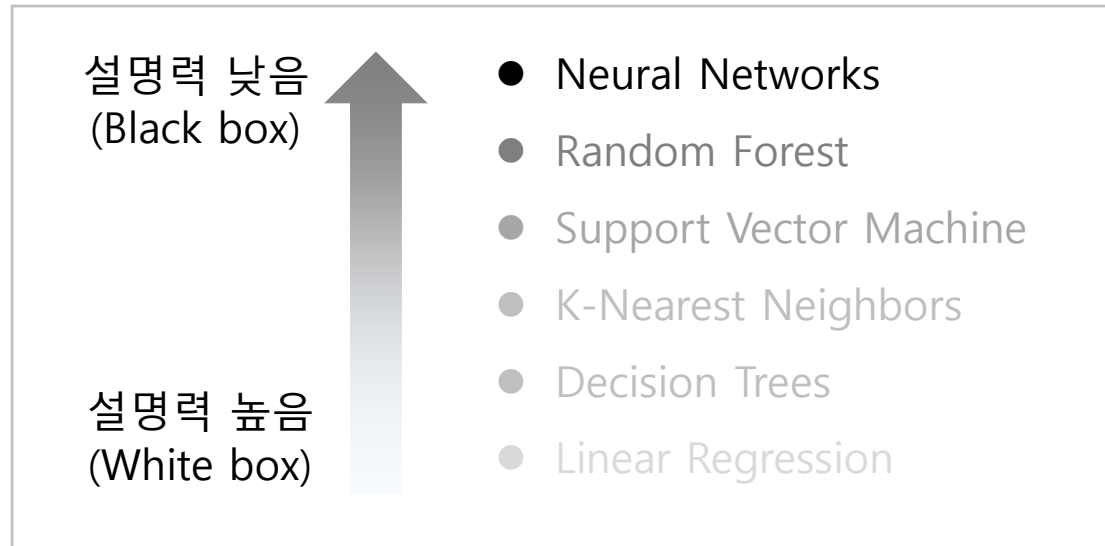
- XAI 개념은 1975년 처음 등장 Buchanan, Shortliffe
 - ✓ 의사들의 추론 과정을 확률적으로 모델링 하는 방법을 제안함
 - ✓ Rule based conditional probability approximation 을 통해 의사 결정 과정을 분석하는 방법을 제안

Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical biosciences*, 23(3-4), 351-379.

1. XAI 개요

❖ XAI는 설명력이 낮은 인공지능 모델의 설명을 제공

- 설명력 : 인공지능 모델이 내린 결정 근거를 사람이 이해할 수 있는 정도를 의미
 - ✓ 설명력이 높은 모델 (White box), 설명력이 낮은 모델 (Black box)



설명력 높음 (White box)

- Linear Regression

$$y = \beta_0 + 21.3x_1 + 10.1x_2$$

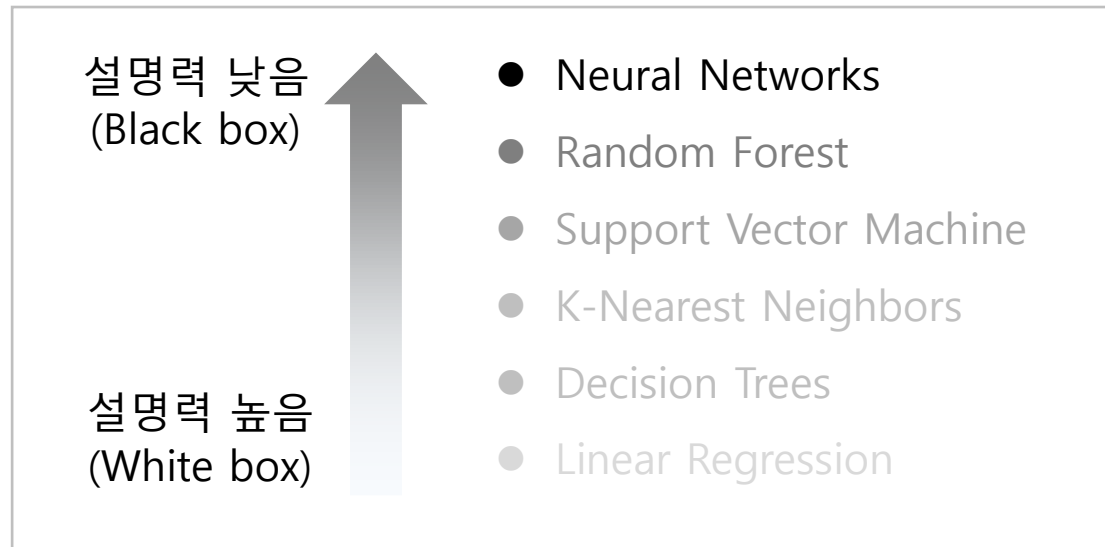
기울기 값을 통해 기여도 높은 변수 및 x 값에 따른 y 값 변경 방향 설명 가능

Morocho-Cayamcela, M. E., Lee, H., & Lim, W. (2019). Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access*, 7, 137184-137206.

1. XAI 개요

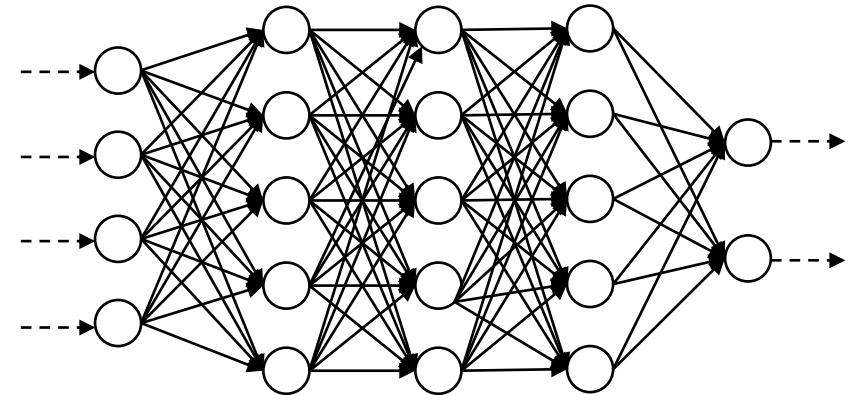
❖ XAI는 설명력이 낮은 인공지능 모델의 설명을 제공

- 설명력 : 인공지능 모델이 내린 결정 근거를 사람이 이해할 수 있는 정도를 의미
 - ✓ 설명력이 높은 모델 (White box), 설명력이 낮은 모델 (Black box)



설명력 낮음 (Black box)

- Neural Networks



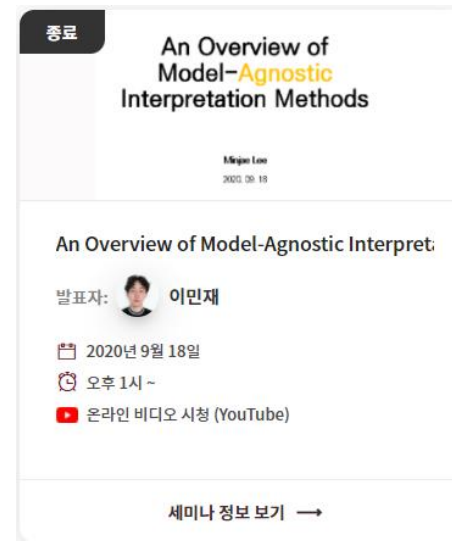
설명 어려움

Morocho-Cayamcela, M. E., Lee, H., & Lim, W. (2019). Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access*, 7, 137184-137206.

1. XAI 개요

❖ 대표적인 XAI 방법론

- LIME : Local Interpretable Model-agnostic Explanations
- **SHAP : Shapley Additive exPlanations**



<http://dmqa.korea.ac.kr/activity/seminar>

SHAP(Shapley Additive exPlanations)

2. SHAP(Shapley Additive exPlanations)

❖ SHAP(Shapley Additive exPlanations)

- Shapley Value과 피쳐(Feature)간 독립성을 기초로 제안된 방법론
(A unified approach to interpreting model predictions, 2017)
- Shapley Value를 활용하여 각 피쳐의 기여도를 계산하여 예측결과에 대한 설명을 제공

A Unified Approach to Interpreting Model Predictions

Scott M. Lundberg
Paul G. Allen School of Computer Science
University of Washington
Seattle, WA 98105
slund1@cs.washington.edu

Su-In Lee
Paul G. Allen School of Computer Science
Department of Genome Sciences
University of Washington
Seattle, WA 98105
suinlee@cs.washington.edu

Abstract

Understanding why a model makes a certain prediction can be as crucial as the prediction's accuracy in many applications. However, the highest accuracy for large

<https://youtu.be/f2XqxOny3NA>

SHAP
❖ Shapley additive explanation (SHAP) - Example

	X1 사용	X2 사용	X3 사용	예측 값
Case ①	X	X	X	28
Case ②	O	X	X	32
Case ③	X	O	X	30
Case ④	X	X	O	32
Case ⑤	O	O	X	33
Case ⑥	O	X	O	33
Case ⑦	X	O	O	32
Case ⑧	O	O	O	35

가중치: $\frac{1}{3}$
②-① = 4 $\frac{1}{3}$ 변수 3개 사용하는 경우 (1가지)
X1, X2, X3
가중치: $\frac{1}{6}$
⑤-③ = 1 $\frac{1}{6}$ 세 변수 중 한 변수 제외하는 경우 (3가지)
X1, X2, X3
⑧-⑦ = 3 $\frac{1}{3}$ 가중치: $\frac{1}{3}$

[핵심 머신러닝] Shapley Value (Explainable AI)
조회수 3,105회 · 2021. 7. 10. 77 좋아요 공유 저장 설정 ...

김성범(소장 / 인공지능융합연구실)
구독자 1,32만명

다보기

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

2. SHAP(Shapley Additive exPlanations)

❖ Shapley Value

- Shapley Value는 게임 이론(Game Theory)에 처음 등장, 각 플레이어의 기여도에 따라 보상을 할당하는 기준



Player 수 : 4명
총 보상 : 200만원

Player	A	B	C	D
Shapley Value	40.0	30.0	10.0	20.0
Reward	80만원	60만원	20만원	40만원

2. SHAP(Shapley Additive exPlanations)

❖ SHAP 활용 예제 <https://github.com/slundberg/shap>

- 보스턴 주택 가격 데이터
- 총 데이터 수 : 506개 (설명변수 : 13개 주택 영향 요인, 반응변수 : 1개 주택 가격)

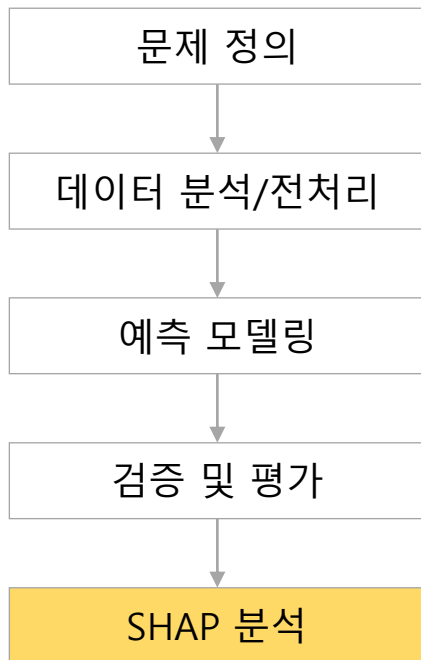
	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10

[01]	CRIM	자치시(town) 별 1인당 범죄율
[02]	ZN	25,000 평방피트를 초과하는 거주지역의 비율
[03]	INDUS	비소매상업지역이 점유하고 있는 토지의 비율
[04]	CHAS	찰스강에 대한 더미변수(강의 경계에 위치한 경우는 1, 아니면 0)
[05]	NOX	10ppm 당 농축 일산화질소
[06]	RM	주택 1가구당 평균 방의 개수
[07]	AGE	1940년 이전에 건축된 소유주택의 비율
[08]	DIS	5개의 보스턴 직업센터까지의 접근성 지수
[09]	RAD	방사형 도로까지의 접근성 지수
[10]	TAX	10,000 달러 당 재산세율
[11]	PTRATIO	자치시(town)별 학생/교사 비율
[12]	B	$1000(Bk-0.63)^2$, 여기서 Bk는 자치시별 흑인의 비율을 말함.
[13]	LSTAT	모집단의 하위계층의 비율(%)
[14]	MEDV	본인 소유의 주택가격(중앙값) (단위: \$1,000)

2. SHAP(Shapley Additive exPlanations)

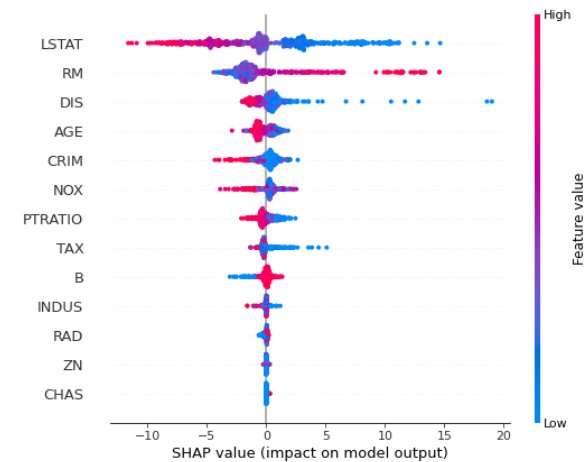
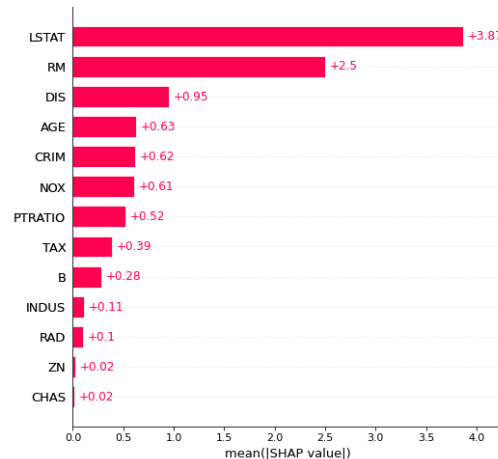
❖ SHAP 활용 예제 <https://github.com/slundberg/shap>

- SHAP 분석 절차 및 결과



중요도 높음

중요도 낮음



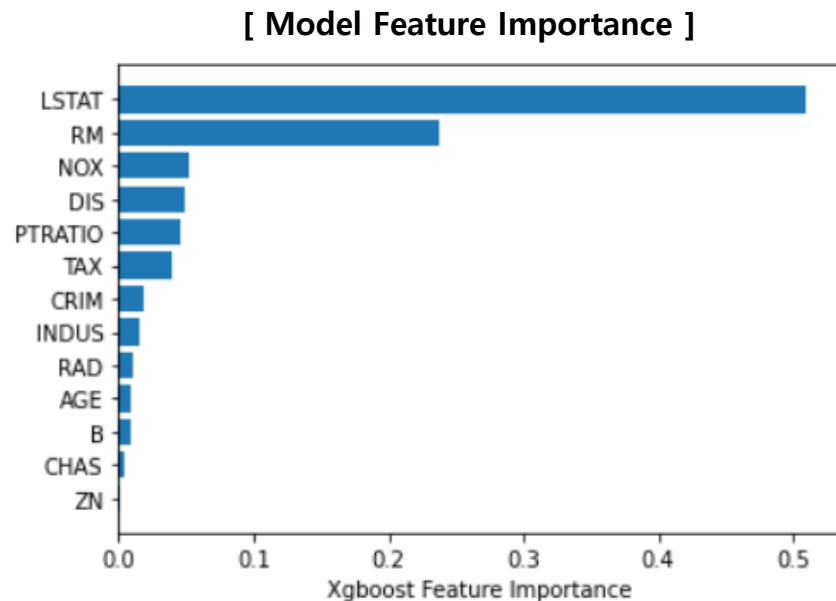
[SHAP 결과 설명]

- 집 값에 영향도가 높은 요인 : 하위계층 비율 > 방 수 > 업무지구 접근성 지수 > ...
- 하위계층 비율이 높을 수록 집값에 부정적인 영향(가격 하락)을 미침
방의 수가 많을 수록 집값에 긍정적인 영향(가격 상승)을 미침
10ppm 당 농축 일산화질소의 경우 중요도는 높으나 농도에 따른 영향은 중립

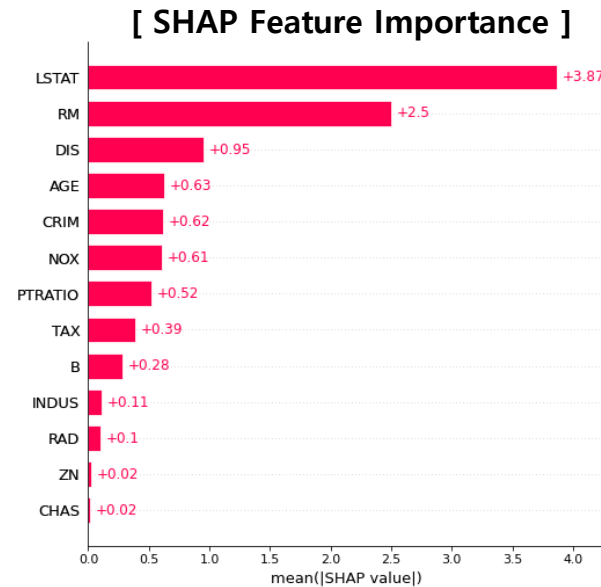
2. SHAP(Shapley Additive exPlanations)

❖ 참고 (Model Feature Importance 단점)

- 요인이 상호 의존적인 관계일 경우 결과 오차 발생
- 음의 영향도(부정적인 영향도)를 고려하지 않음 (ex 집 값 하락 요인)
- Feature Importance 계산 시 오류가 커질 경우 해당 값은 미포함



VS



Rank	Model	SHAP
1	LSTAT	LSTAT
2	RM	RM
3	NOX	DIS
4	DIS	AGE
5	PTRATIO	CRIM
6	TAX	NOX
7	CRIM	PTRATIO
8	INDUS	TAX
9	RAD	B
10	AGE	INDUS
11	B	RAD
12	CHAS	ZN
13	ZN	CHAS

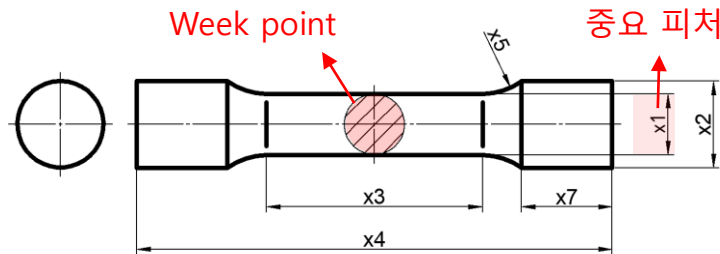
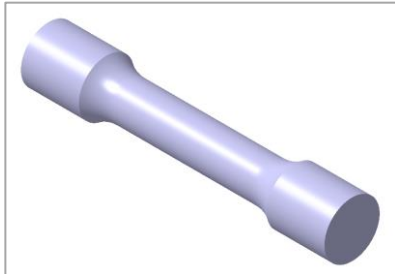
SHAP을 이용한 제품 설계

3. SHAP을 이용한 제품 설계 적용

❖ 배경 및 필요성

- 간단한 형상의 제품도 많은 설명변수(x , 설계인자)가 존재하며, 형상이 복잡해 질 수록 크게 증가함
- 제품의 설계를 위해서 반응변수 기여도가 높은 중요 설명변수를 찾고,

설명변수 값의 변화에 따른 반응변수(y) 경향을 파악하는 작업은 매우 중요함



- ✓ 설명변수 : 8개 (치수 7개, 재료 1개)
- ✓ 반응변수 : 1개 (Stress, 응력)



- ✓ 설명변수 : 140~145개 (치수, 재료, 시험조건)
- ✓ 반응변수 : 1~ 개 (타이어 성능 지표)

3. SHAP을 이용한 제품 설계 적용

❖ 제품 설계 및 변경 절차

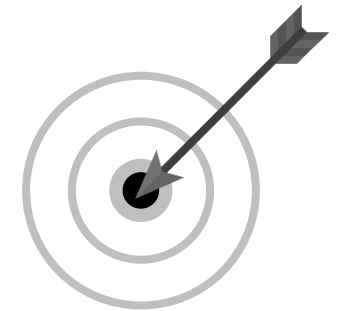
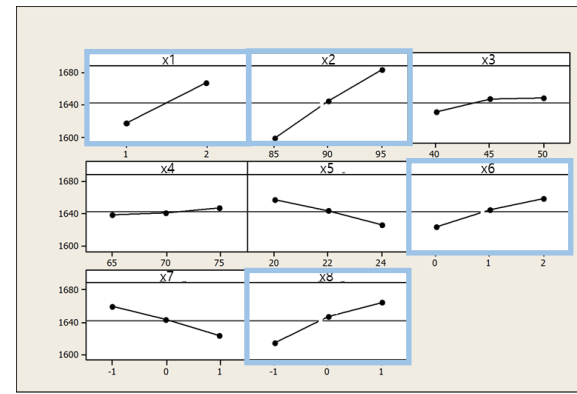
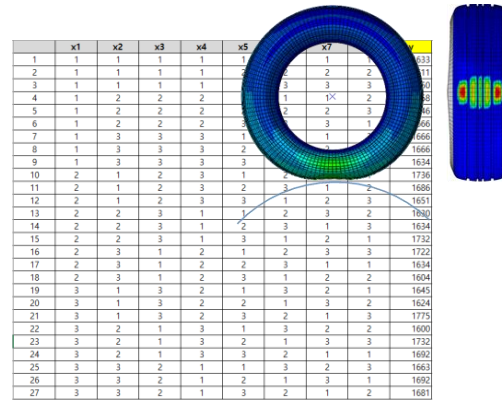
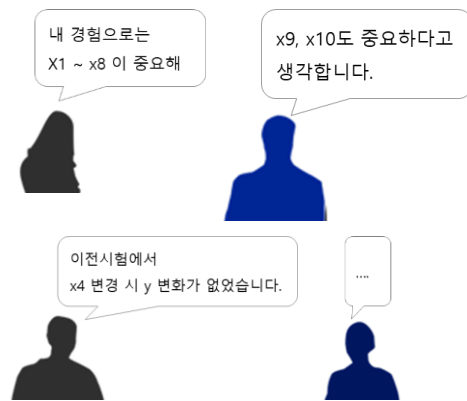
- 경험에 의해 중요 설명변수 선정, 실험수 감소를 위해 설명변수 최소화 필요
- 경험의 차이에 따른 불확실성 존재, 계획된 실험 수행을 위한 손실(비용, 시간)이 큼

주요 인자 선정

실험계획수립 및 실험

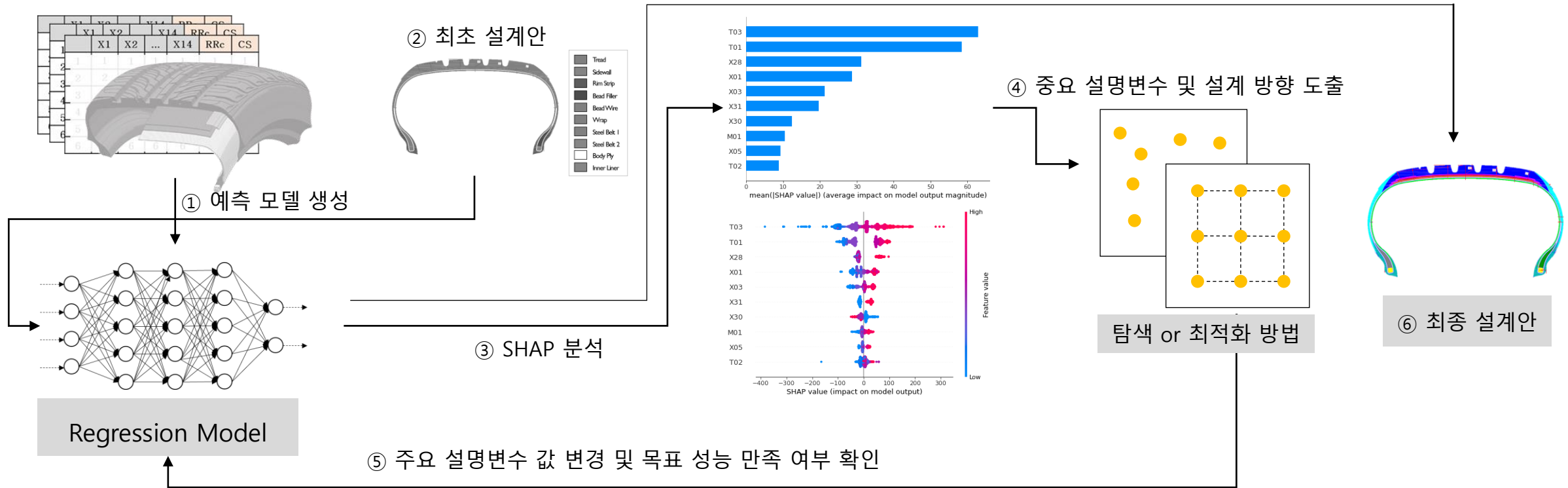
ANOVA 분석 및 설계 변경 인자 선정

최적화 / Try Error



3. SHAP을 이용한 제품 설계 적용

❖ SHAP 적용 제품 설계 및 변경



3. SHAP을 이용한 제품 설계 적용

❖ 타이어 설계 적용 사례

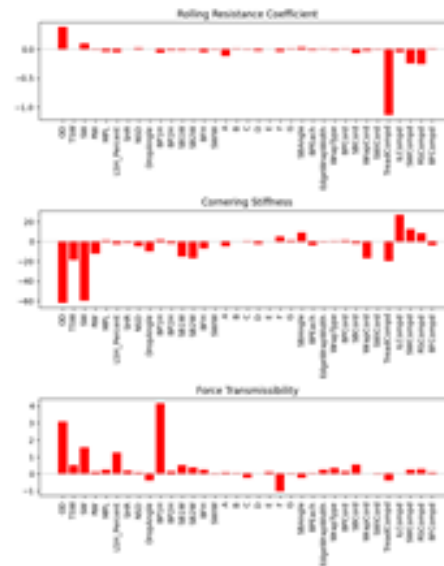
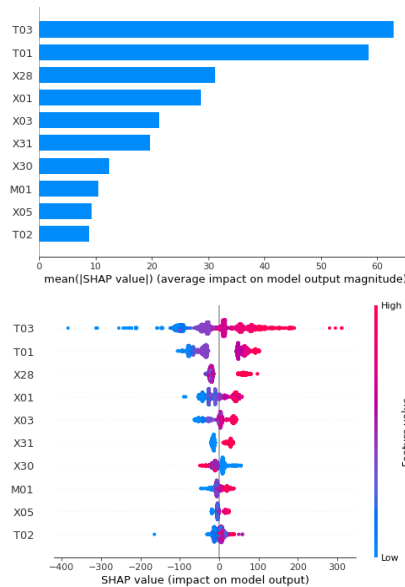
- 데이터 : 타이어 설계데이터 (타이어 형상 치수, 재료 물성값, 시험 조건, 시험 결과(연비, 핸들링, 소음/진동))
- 설명변수 : 타이어 형상 치수 , 재료 물성값, 시험 조건 , 56개
- 반응변수 : 연비(Rolling Resistance), 핸들링(Cornering Stiffness), 소음/진동(Force Transmissibility) 시험 결과
- 데이터 수 : 총 18,536개

항 목		특 징 명 (임의 명칭)
설명변수	타이어 형상 치수	X01, ..., X32
	재료 물성값	M01, ..., M20
	타이어 실내 시험 조건	T1, ..., T4
반응변수	실내 시험 결과	CS, RRc, FT

3. SHAP을 이용한 제품 설계 적용

❖ 타이어 설계 적용 사례

- 대상 타이어 : 165/70R14
- 목표 성능 : Cornering Stiffness 900 이상, RRc = 10 이하, Force Transmissibility = 85
- SHAP 분석 결과



[주요 설명 변수]

M01, X01, X03, X06, X22, X14

[설명변수 값에 따른 성능 변화]

	M01 (↑)	X01 (↑)	X03 (↑)	X06 (↑)	X22 (↑)	X14 (↑)
CS	↑	↓	↓	↑	↓	↑
RRc	↓	↑	↑	↑	↑	↓
FT	↓	↑	↑	↑	↑	↓

3. SHAP을 이용한 제품 설계 적용

❖ 타이어 설계 적용 사례

- 주요 설명 변수의 값을 최적화 방법을 적용하여 목표성능에 근접한 값을 도출함
- 핸들링 성능은 목표 성능 대비 2.7% 향상됨
- 연비 성능은 10.0 을 만족함
- 소음/진동 성능은 85를 만족함

	M01	X01	X03	X06	X22	X14		CS	RRc	FT
초기값	13.01	583	172	48.27	25	25	목표 성능	900이상	10이하	85이하
설계값	16.2	583	168	50.288	23	30	실제 성능	925.508	9.8	84.125

4. 정리

❖ XAI 개요 설명

- XAI는 인공지능 모델이 도출한 결과에 대한 근거를 설명 (설명력이 낮은 모델에 효과적)
(설명력 : 인공지능 모델이 내린 결정 근거를 사람이 이해할 수 있는 정도를 의미)

❖ SHAP 설명

- Shapley Value를 활용하여 각 피처의 기여도를 계산함으로써 예측결과에 대한 설명을 제공
- Model Feature Importance 와 SHAP 분석을 통해 도출된 Feature Importance의 차이점 설명
- SHAP 분석 예제

❖ SHAP을 활용한 제품 설계 적용 사례 소개

- [1] 안재현. (2020). XAI 설명 가능한 인공지능, 인공지능을 해부하다, 위키북스.
- [2] Van Lent, M., Fisher, W., & Mancuso, M. (2004, July). An explainable artificial intelligence system for small-unit tactical behavior. In *Proceedings of the national conference on artificial intelligence* (pp. 900-907). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
- [3] Shortliffe, E. H., & Buchanan, B. G. (1975). A model of inexact reasoning in medicine. *Mathematical biosciences*, 23(3-4), 351-379.
- [4] Morocho-Cayamcela, M. E., Lee, H., & Lim, W. (2019). Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access*, 7, 137184-137206.
- [5] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
- [6] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- [7] <https://github.com/slundberg/shap>

EOD