

예상 질문

1. 프로젝트 목적/기획 관련 질문

- **Q1. 왜 '서울시'만 대상으로 삼았나요?**
 - A: 데이터 수집의 현실적 제약 때문이었고, 모델 구축의 기초 단계로 서울을 선택했습니다. 향후 전국 단위로 확장할 수 있는 기반을 마련하는 데 목적이 있습니다.
 - **Q2. 알뜰주유소 가격이 일반 주유소보다 높은 이유는 뭔가요?**
 - A: 운영주체마다 가격 책정 기준이 다르고, 일부 지역에서는 경쟁 구도가 약해 알뜰주유소라도 가격 인하 노력이 약한 경우가 있기 때문입니다.
 - **Q3. 이 모델이 실제 정책에 어떤 식으로 활용될 수 있나요?**
 - A: 예측된 적정 가격을 기준으로 '가격 상한선 가이드라인' 제시, 또는 지역별 인센티브 정책의 기반으로 활용할 수 있습니다.
-

2. 데이터 관련 질문

- **Q4. 휘발유 가격 0원인 데이터는 왜 발생하나요?**
 - A: 해당 주유소에서 휘발유를 판매하지 않거나, 오류 또는 결측 데이터일 가능성이 높습니다.
 - **Q5. 왜 공시지가를 선택했나요? 휘발유 가격과 어떤 관련이 있나요?**
 - A: 공시지가는 주유소가 위치한 지역의 경제 수준을 간접적으로 보여주는 지표이며, 부동산 가치와 상업지역일수록 유가가 높은 경향이 있어 반영했습니다.
 - **Q6. 환율과 국제유가는 어떤 기준으로 시차를 두고 반영했나요?**
 - A: 국제유가나 환율 변화가 주유소 가격에 바로 반영되지는 않기 때문에, 실험적으로 14일 전 시점의 데이터를 사용해 가장 높은 상관성을 확인했습니다.
-

3. 모델링 관련 질문

- **Q7. 왜 딥러닝 모델은 사용하지 않았나요?**
 - A: 현재 데이터 규모가 딥러닝의 장점을 충분히 살릴 만큼 크지 않고, 해석 가능한 모델이 중요하다고 판단해 전통적인 ML 모델들을 먼저 적용했습니다.
- **Q8. 왜 Random Forest가 가장 성능이 좋았다고 판단했나요?**
 - A: MAE, MSE, R^2 등의 평가 지표를 기준으로 가장 안정적인 예측 성능을 보였고, 과적합도 적었습니다. 또한 변수 중요도 해석이 용이하다는 점도 장점이었습니다.
- **Q9. KNN은 왜 성능이 낮았나요?**

- A: 데이터 차원이 높아질수록 거리 기반 계산이 불안정해지고, K 설정에 따라 민감하게 반응하기 때문에 예측 정확도가 낮게 나왔습니다.

4. 결과 및 해석 관련 질문

- **Q10.** 변수 중요도 분석 결과, 가장 큰 영향을 미친 변수는 무엇이었나요?
 - A: 모델에 따라 다르지만, 대체로 공시지가, 환율, 국제유가, 법정동명이 높은 영향도를 보였습니다.
- **Q11.** Heatmap 상에서 상관계수가 낮은 변수는 제거하지 않았나요?
 - A: 네, 상관계수만으로는 변수 제거를 판단하지 않았고, 실제 모델 성능에 기여하는지를 기준으로 판단했습니다.


5. 한계와 개선 방향 관련 질문

- **Q12.** 이 모델을 실시간 시스템으로 구현하려면 무엇이 필요할까요?
 - A: 실시간 데이터 수집 API 구축, 모델 자동 재학습 시스템, 시계열 모델 통합이 필요합니다.
- **Q13.** 모델이 다년간 데이터를 사용했다면 더 나은 결과가 나왔을까요?
 - A: 네, 특히 유가는 계절성과 트렌드가 존재하기 때문에, 다년간 데이터를 사용한 시계열 기반 모델이 더 정밀할 수 있습니다.

고난도·비판적 예상 질문

[정책/경제 관점 질문]

Q1. “적정 가격”이라는 개념은 주관적이지 않나요? 머신러닝이 정말 그걸 정량화할 수 있을까요?

-  이 질문은 모델이 “정책 수단”으로 쓰일 수 있는지에 대한 의심입니다.

답변 팁:

저희가 말하는 '적정 가격'은 절대적인 의미가 아니라, 경제적·시장 데이터 기반의 예측값을 말합니다.

이는 정책 수립 시 참고할 수 있는 '객관화된 기준값'으로서 기능하며, 단일 결정이 아닌 의사결정의 한 요소로 활용될 수 있습니다.

Q2. 공시지가와 휘발유 가격 사이에 인과관계가 있다고 보시나요? 단순 상관 아닌가요?

- 🔍 상관 ≠ 인과라는 머신러닝 모델의 근본적 한계를 찌르는 질문입니다.

답변 팁:

맞습니다. 저희 모델은 인과관계를 직접 추론하지 않습니다.

하지만 공시지가는 지역의 소비 수준과 상업적 밀집도를 반영하기 때문에, 가격 형성에 일정 영향이 있을 것으로 가정했습니다.

궁극적으로는 인과추론 기반의 모델로 확장할 여지가 있습니다.

🧠 [모델링/AI 윤리 질문]

Q3. 동일한 조건에서 특정 지역이 비싸게 예측된다면, 그 예측은 지역 차별적일 수도 있지 않나요?

- 🔍 AI 모델의 윤리성과 공정성에 대한 질문입니다.

답변 팁:

중요한 포인트입니다. 저희는 단순히 과거 데이터를 기반으로 학습했기 때문에, 이미 존재하는 시장 불균형이 예측값에 반영될 수 있습니다.

향후에는 공정성 제약(**fairness constraint**) 또는 정책 기준 상한선을 설정해 윤리적 제약도 고려할 계획입니다.

Q4. Random Forest 모델의 해석력은 높다고 했지만, 실제 변수 중요도는 어떻게 검증했나요?


- 🔍 단순히 모델이 중요하다고 말한 변수를 얼마나 신뢰할 수 있는가에 대한 질문입니다.

답변 팁:

내부적으로는 `feature_importances_` 속성과 `permutation importance`를 병행해서 검토했습니다. 단일 모델의 편향을 줄이기 위해, 여러 모델에서 반복적으로 중요도가 높게 나타난 변수들만 종합적으로 분석했습니다.

[데이터 및 설계 비판 질문]

Q5. 셀프 여부와 상표를 단순 범주형으로 처리했는데, 그 안의 복잡한 맥락은 반영되지 않지 않나요?

-  셀프=싸다, 상표=정책영향 등 이면적 요소들을 날카롭게 파고드는 질문입니다.

답변 팁:

네, 그 점은 인지하고 있습니다. 현재는 단순 인코딩 기반이지만, 향후에는 주유소 위치(**GIS**), 경쟁 강도, 브랜드 충성도 등 맥락 데이터를 추가해 더 정교한 피처로 확장할 수 있다고 봅니다.

Q6. 14일 전 환율을 쓰셨다고 했는데, 그 시차는 어떤 근거로 정하신 건가요? 임의적인 선택 아닌가요?

-  모델 입력 설계의 정당성에 대해 묻는 질문입니다.


답변 팁:

다양한 시차를 실험했고, 7일, 14일, 30일 중에서 **MAE** 기준 가장 성능이 안정적인 시차가 14일이었습니다.

앞으로는 시계열 모델 기반의 자동 시차 최적화도 고려하고 있습니다.

[현실 적용 관련 질문]

Q7. 모델이 도출한 예측값이 실제 주유소 운영자에게 어떤 식으로 영향을 미칠 수 있을까요?

-  정책과 실무 사이의 괴리를 짚는 질문입니다.

답변 팁:

이 모델은 '주유소별 가격 제안'이 아닌, 지역 기반의 시장 가격 가이드라인 제공에 중점을 둡니다.

즉, 운영자가 참고할 수 있는 공공 기반의 정보제공 시스템에 활용될 수 있습니다.

Q8. 지금 만든 모델을 실제 서비스로 구현하려면, 가장 큰 현실적 허들은 뭐라고 보세요?

- 🔍 실무화 가능성에 대한 핵심 질문입니다.

답변 팁:

가장 큰 허들은 실시간 데이터 수집과 자동 재학습 시스템 구축입니다.

특히, 주유소별 데이터는 민감한 정보일 수 있어, 데이터 공개 범위와 **API** 연동 가능성이 현실적인 제약입니다.

7조 프로젝트 Sheet

1. 이상치 클러스터링

- a. 데이터 분석 대상
 - i. 이상치(정규분포 86% 넘는)
- b. 물리적범위
 - i. 서울시
- c. 시간적범위
 - i. 2024년 10월 1일 ~ 2024년 12월 31일

2. 휘발유 가격으로 경유가격 예측 모델

- a. 독립변수(X)
 - i. 주유소 일일 휘발유 가격 평균(단위: 원/L)
 - ii. 환율(원/USD)
 - iii. 원유가격(Dubai, USD/배럴)
 - iv. 계절
 - v. 행정동별 공시지가
- b. 종속변수(y)
 - i. 주유소 경유가격(단위: 원/L)
- c. 물리적범위
 - i. 서울
- d. 시간적범위
 - i. 봄데이터: 춘분(2024년 3월 20일)
 - ii. 여름데이터: 하지(2024년 6월 21일)
 - iii. 가을데이터: 추분(2024년 9월 23일)
 - iv. 겨울데이터: 동지(2024년 12월 21일)
- e. 예상: 경유가격은 환율의 영향과 밀접한 관련이 있으나 휘발유 가격과는 유의한 상관관계가 없을 듯함. 그러나 휘발유 또한 환율의 영향을 받으므로 휘발유 가격으로 경유가격 예측이 가능한 것처럼 보일 수는 있을 듯

f. 참고자료

- i. <국내 휘발유, 경유 가격변동성 분석: 설명요인과 지렛대효과(2010)>, 김형건 외, 한국산업경제학회

3. 정부 정책사업 자영 알뜰 주유소 적정가(회귀) - **Final** 주제

a. 독립변수(X)

- i. 법정동명(원핫인코딩)
- ii. 법정 동단위 공시지가
- iii. 환율(원/USD) - 14일전 기준
- iv. 원유가격(Dubai, USD/배럴) - 14일전 기준
- v. 월(1~12월, 원핫인코딩)
- vi. 셀프 주유소 여부(원핫인코딩)
- vii. 주유소 상표(알뜰 포함)

b. 종속변수(y)

- i. 휘발유 가격

c. 물리적 범위

- i. 서울시

d. 시간적범위

- i. 2024년 1월 1일 ~ 2024년 12월 31일

e. 참고자료: https://www.knoc.co.kr/sub04/sub04_8_1.jsp

(추가연구) 시간 데이터 확장

- 기존 데이터는 계절의 하루뿐이므로 데이터가 적음
- 봄: 2024년 4월 19일 곡우 봄비가 내리는 날
- 여름: 2024년 7월 22일 대서 1년 중 가장 더운 날
- 가을: 2024년 10월 23일 상강 서리가 내리기 시작하는 날
- 겨울: 2024년 12월 7일 대설 1년 중 눈이 가장 많이 내리는 날

(추가연구) 근접한 타 주유소의 데이터를 기반으로 알뜰주유소 요금 산정 모델

- 기존의 '동' 기준의 요금 설정에서 벗어나, 알뜰주유소 위치 정보를 활용하여 인접 주유소 가격과의 상관관계로 적정 가격 산출
- 방식(예) 인접 주유소 10개의 주유소 유가를 기준으로 알뜰주유소 요금 산정

(폐기)

4. 이상치 탐지 모델(회귀)

지역별로 기름값이 차이가 나는 이유

(순위별 - 전국으로 봤을 때! 서울경기권으로 한정되었을 때는 다를 수 있음)

1. 물류 및 유통 비용

- 정유소와의 거리: 정유소가 있는 울산, 여수, 서산(대산) 근처는 물류비가 적어 가격이 낮음
- 운송비 증가 지역: 강원 산간 지역, 제주도, 울릉도 등은 기름을 운송하는 데 추가 비용이 발생해 가격이 높음
- 도심 내 물류비용: 서울 시내에는 교통 체증이 심해서 기름 수송비가 더 들어감. 화물차가 주유소로 기름을 공급하는 데 시간이 오래 걸리면 물류비가 증가하고, 이 비용이 기름값에 반영

2. 주유소 간 경쟁 정도

- 도시 vs. 시골: 수도권, 경기 외곽처럼 주유소 경쟁이 심한 곳은 가격이 낮아지고, 군 단위·시골 지역은 주유소 수가 적어 가격이 높게 유지
- 셀프 주유소 영향: 서울·수도권은 셀프 주유소가 많아 경쟁이 심한 반면, 제주도 등 일부 지역은 셀프 주유소가 적어 가격이 높을 수 있음

3. 지역 특성 및 수요 차이

- 관광지(제주·강릉·부산): 관광객들은 가격에 둔감해 주유소에서 가격을 높게 책정하는 경우가 많음
- 산업단지(울산·창원·포항): 물류·운송업이 많아 대량 공급이 이뤄지며 가격이 낮은 편
- 고속도로 휴게소: 리터당 150~200원 비싼 경향이 있음
- 강남, 용산 같은 부촌 지역은 고급 외제차 비율이 높고, 정유사 브랜드 주유소(직영점) 선호도가 높음. 소비자들이 가격보다도 품질을 더 중요시
- 땅값과 운영비: 서울은 임대료, 인건비 등 운영비가 전국에서 가장 높아지고, 이는 기름값에 반영. 강남, 종로 같은 중심지는 땅값이 비싸서 기름값도 높게 책정되는 경향

4. 정책 및 세금 차이

- 유류세는 전국 동일하지만, 일부 지자체(제주도)는 유류비 지원 정책(농·어업용 면세유, 대중교통 유가 보조금 등)이 있음
- 수도권 환경 규제: 전기차·CNG 보급 증가로 휘발유 수요가 줄어들면서 가격에 영향을 줌

5. 브랜드·정유사 차이

- 정유사 직영 vs. 알뜰주유소: 정유사 직영 주유소는 브랜드 가치 때문에 가격이 높고, 정부 지원 알뜰주유소는 50~100원 저렴한 경향

=====

주유소의 이상치 가격을 가지고 -> 이상치가 나타나는 주유소의 특징이 그 주유소 위치의 공시지가와 관련이 있다고 판단. -> 우선 이상한 주유소와 공시지가 관계도를 학습. -> 다른 지역의 가격이 유독 비싼 주유소들을 샘플링 -> 이 주유소의 공시지가 예측모델로 한번 예측 -> 실제 이 주유소지역의 공시지가가 비싼지 정답과 비교.

왜 이상치만 뽑아냈냐..? 공시지가가 비싼 것들은 가격이 비싼 주유소와 평균적으로 관련이 더 있었기 때문.

가격이 비싸지 않은 주유소는 공시지가와 크게 관련이 없고 그냥 전국평균을 따라간거 같음.

진행 프로세스 (시행착오)

- 데이터 수집 과정 (데이터 선택)
 - 지도 api 불러오기
 - kakao api 관련 코드가 예전 코드라 실행이 안됨
 - kakao map api 사이트에서 활성화 요청 버튼 on
 - 앱키가 4가지가 있는데 확인해보니 JavaScript 키가 잘 작동한 것 확인
 - 처음에 나왔을 때는 글자가 깨져서 나옴
 - UTF-8로 하면 한글은 나오는데 안나오는 부분이 있어서 확인해 보니, (법정동) 표시 있는 주소에 해당하는 좌표만 안나옴
 - 그래서 (법정동)부분 제거하고 좌표 가져오도록 함.
 - 좌표가 안나오는 부분은 지도에 아예 없는 것으로 확인
-

PPT 구성 초안 (우선 CRISP-DM 기준으로 정리) (파란색 = 수정 필요)

표지, 목차, 비즈니스 이해, 데이터 이해, 데이터 준비, 모델링, 평가, 배포, 참고 문헌

표지 (1장)

- 프로젝트 제목: 서울시 행정구역별 평균 유가의 이상치 탐지 및 상관관계 분석
 - 부제: 공공데이터(정형)를 활용한 AI 모델링 프로젝트
 - 팀원: 7조 - 김은서, 이동인, 이종화, 최예셀 (가나다 순)
 - 발표일: 2025.04.10
-

목차 (1장)

1. 비즈니스 이해 (Business Understanding)
 2. 데이터 이해 (Data Understanding)
 3. 데이터 준비 (Data Preparation)
 4. 모델링 (Modeling)
 5. 평가 (Evaluation)
 6. 배포 (Deployment)
 7. 참고 문헌 (References)
-

Business Understanding 비즈니스 이해 - 03.31 ~ 04.01 (4~5장)

- 1) 프로젝트 배경 (1장)
 - a) 유가는 지역 경제와 소비자 물가에 직접적인 영향을 미치는 주요 요소
 - b) 일부 지역에서는 유가가 평균보다 비정상적으로 형성되는 경우 발생
 - c) 특정 경제적·사회적 요인과 관련 가능성이 있음
 - d) 공공데이터를 활용하여 이상치 탐지 및 공시지가와의 관계 분석
- 2) 프로젝트 목표 및 성공 기준 (1장)
 - a) 서울시 행정구역별 유가 이상치 탐지 및 패턴 분석
 - b) 공시지가와 유가 간의 상관관계 분석
 - c) 데이터 기반 정책적 시사점 도출 및 시각화 제공
- 3) 상황 평가
: 리소스, 요구사항, 가정, 제약 조건, 위험, 비용 및 이점 등 평가
- 4) 데이터 마이닝 목표 설정 (데이터 마이닝을 통해 달성하고자 하는 목표 명확히)
 - a) 유가 데이터에서 이상치를 탐지하고 원인 분석

b) 공시지가와의 관계를 회귀 분석 및 상관관계 분석을 통해 도출

5) 프로젝트 계획 수립 (2~3장) **WBS**

- a) 데이터 수집 및 전처리 (04.01 ~ 04.02)
- b) 데이터 분석 및 이상치 탐지 모델링 (04.03 ~ 04.06)
- c) 상관관계 분석 및 시각화 (04.07 ~ 04.09)
- d) 최종 발표 (04.10)

Data Understanding 데이터 이해 - 03.31 ~ 04.02 ()

1) 초기 데이터 수집 (1장)

: 분석을 위해 필요한 데이터를 수집하고, 이 과정에서 어떤 데이터를 어떻게 수집했는지 기록

- a) 유가 데이터
 - i) 출처: 오피넷 (한국석유공사) ([싼 주유소 찾기 오피넷](#))
 - ii) 경로: 국내유가통계 → 유가내려받기
 - iii) 주요 내용: 주유소별 유가, 지역별 평균 유가, 날짜별 가격 변동 정보 등
- b) 공시지가 데이터
 - i) 출처: 서울열린데이터광장
([서울시 개별공시지가 정보> 데이터셋> 공공데이터 | 서울열린데이터광장](#))
 - ii) 경로: 서울시 개별공시지가 정보 → 데이터셋
 - iii) 주요 내용: 지역별 개별공시지가, 변동률, 지가 상승 요인 등

2) 데이터 설명 및 탐색 (1장)

: 수집된 데이터의 기본적인 특성(데이터의 구조, 변수 종류, 데이터 타입, 크기 등) 기술 (데이터를 시각화하고 통계적 방법으로 탐색. 데이터의 패턴, 추세, 이상치, 상관관계 등을 발견)

- 데이터 구조 분석: 변수 종류, 데이터 타입, 크기 등 기술

- 기초 통계 분석: 평균, 표준편차, 최솟값/최댓값 확인

- 시각화를 통한 데이터 패턴 확인 (히스토그램, 상자그림 등)

a) 유가 데이터 (예시)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 13150 entries, 0 to 13149
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  -
0   번호        13150 non-null  object
1   지역        13149 non-null  object
2   상호       13149 non-null  object
3   주소        13149 non-null  object
4   기간        13149 non-null  float64
5   상표        13149 non-null  object
6   셀프여부    13149 non-null  object
7   고급회발유  13149 non-null  float64
8   회발유      13149 non-null  float64
9   경유        13149 non-null  float64
10  실내등유    13149 non-null  float64
dtypes: float64(5), object(6)
memory usage: 1.1+ MB
```

b) 공시지가 데이터 (예시)

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 143975 entries, 0 to 143974
Data columns (total 13 columns):
#   Column      Non-Null Count  Dtype
---  -
0   시도명      143975 non-null  object
1   시군구명    143975 non-null  object
2   법정동명    143975 non-null  object
3   토지코드    143975 non-null  int64
4   공시지가(원/㎡) 143975 non-null  int64
5   시군구코드  143975 non-null  int64
6   법정동코드  143975 non-null  int64
7   필지구분코드 143975 non-null  int64
8   필지구분명  143975 non-null  object
9   분번       143974 non-null  float64
10  부번       143974 non-null  float64
11  기준년도    143974 non-null  float64
12  기준년월    143974 non-null  object
dtypes: float64(3), int64(5), object(5)
memory usage: 14.3+ MB
```

3) 데이터 품질 검증 ()

: 데이터의 품질을 검증하고 문제점을 기록.

결측값, 이상치, 중복값, 오류, 일관성 문제 등 검토 및 처리 계획 수립

Data Preparation 데이터 준비 - 04.03 ~ 04.06 ()

- 1) 데이터 선택
: 분석에 사용할 데이터를 선택하고, 포함/제외할 데이터에 대한 근거 마련
 - 2) 데이터 정제 (Data Cleaning)
: 데이터의 결측값, 오류, 중복 등을 처리하고 정제
 - 3) 데이터 구성 (Construct Data)
: 파생 속성(derived attributes)을 생성하거나 새로운 레코드를 만들
 - 4) 데이터 통합 (Integrate Data)
: 여러 데이터 소스를 병합
 - 5) 데이터 형식화 (Format Data) = 스케일링 (정규화 및 표준화 적용)
: 데이터를 분석에 적합한 형식으로 재구성
 - 6) 데이터셋 설명: 최종 데이터셋에 대한 설명 작성
-

Modeling 모델링 - 04.07 ~ 04.09 ()

- 1) 모델링 기법 선택
: 문제에 적합한 모델링 기법 선택 및 이유
 - a) 이상치 탐지 모델
 - i) KNN (K-Nearest Neighbors)
: 특정 지역의 유가가 평균보다 비정상적으로 높은지 탐지
 - ii) DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
: 밀도 기반 클러스터링 기법 활용, 지역별 이상값 탐지
 - b) 공시지가 - 유가 관계 분석 모델
 - i) 회귀 분석 (Linear Regression)
: 공시지가를 독립변수(X), 유가를 종속변수(y)로 설정하여 관계 분석
 - ii) 피어슨 상관계수 (Pearson Correlation Coefficient)
: 공시지가와 유가 간의 관계 강도 및 방향 측정
- 2) 테스트 설계 생성
: 모델을 테스트하기 위한 설계를 만들
- 3) 모델 구축
: 모델을 구축하고, 매개변수를 설정
- 4) 모델 평가: 모델의 성능을 평가하고, 필요한 경우 매개변수를 수정

Evaluation 평가 ()

- 1) 모델 시연
 - a) 이상치 탐지 모델 평가 지표: **Precision, Recall, F1-score** 활용
 - b) 상관관계 분석 평가: 피어슨 상관계수 및 **RMSE (Root Mean Square Error)** 분석
- 2) 결과 해석 및 평가: 데이터 마이닝 결과를 비즈니스 성공 기준과 비교하여 평가
 - a) 특정 지역에서 유가 이상치가 발생하는 패턴 도출
 - b) 공시지가와 유가 간의 관계 시각화 제공
- 3) 프로세스 검토: 전체 프로세스를 검토하고 개선할 점을 찾음
- 4) 다음 단계 결정: 추가 작업 또는 배포를 결정

Deployment 배포 ()

- 1) 배포 계획 수립
: 모델을 실제 환경에 배포하기 위한 계획을 세움
- 2) 모니터링 및 유지보수 계획
: 배포 후 모델의 성능을 모니터링하고 유지보수하는 계획을 수립
- 3) 최종 보고서 작성
: 프로젝트의 최종 보고서와 발표 자료를 작성
- 4) 프로젝트 검토
: 프로젝트 경험을 문서화하고, 향후 프로젝트를 위한 교훈을 도출
- 5) 결론
: 핵심 요약 및 결과 정리

references 참고 문헌 (1장)

1. 한국석유공사 오피넷 (<https://www.opinet.co.kr/>)
2. 서울열린데이터광장 (<https://data.seoul.go.kr/>)
3. 데이터 마이닝 기법 및 머신러닝 모델 관련 논문 및 문서

QnA (1장)