

논문 2026-1-10 <http://dx.doi.org/10.29056/jsf.2026.03.10>

머신러닝 시스템에서 데이터 품질 평가 기반의 데이터 중심 드리프트 관리 방법론

최옥주*†, 신원선**

A Data-centric Drift Management Methodology based on Data Quality Assessment in Machine Learning System

Okjoo Choi*†, Wonsun Shin**

요 약

머신러닝(ML) 시스템은 시간이 지남에 따라 발생하는 드리프트 현상으로 인해 예측 정확도와 신뢰성이 저하된다. 기존에는 주로 모델 재학습을 통해 성능 복구에 집중하였으나 드리프트의 근본 원인이 데이터에 있다는 점에 주목할 필요가 있다. 본 연구에서는 데이터 요구사항 기반의 품질 평가와 드리프트 분석을 통합하여 모델학습 및 예측 전 단계에서 선제적으로 드리프트를 관리하는 Q-DcDM 프레임워크를 제안한다. 본 프레임워크는 ML 생명주기 초기에 잠재적인 드리프트를 탐지하여 성능 저하를 예방하는 방법이다. 본 연구는 UCI Wine Quality와 CIFAR-10 데이터 셋을 통해 검증한 결과 모델 평균 성능이 각각 14.8%, 12.3%가 향상되었으며 모델 재학습 빈도 또한 유의미하게 감소된 것을 확인하였다.

Abstract

Machine learning (ML) systems often suffer from degradation in prediction accuracy and reliability due to data drift that occurs over time. Conventional approaches primarily focus on restoring performance through model retraining; however, it is essential to recognize that the root cause of drift lies in the data itself. In this study, we propose the Q-DcDM framework, which integrates data requirement-driven quality assessment with drift analysis to proactively manage drift before the model training and inference stages. The proposed framework enables early detection of potential drift in the initial phase of the ML lifecycle, thereby preventing performance degradation. Experimental results on the UCI Wine Quality and CIFAR-10 datasets demonstrate that the proposed approach improves average model performance by 14.8% and 12.3%, respectively, while also significantly reducing the frequency of model retraining.

한글키워드 : 데이터 품질, 품질 평가, 드리프트, 드리프트 탐지, 머신러닝 시스템

keywords : data quality, quality assessment, drift, drift detection, machine learning system

* 배재대학교 AI.소프트웨어공학부

** 배재대학교 스마트ICT융합학과

† 교신저자: 최옥주(email: okjoo.choi@pcu.ac.kr)

접수일자: 2026.03.09. 심사완료: 2026.03.17.

게재확정: 2026.03.20.

1. 서론

다양한 산업 분야에서 머신러닝(ML, Machine Learning) 시스템에 대한 의존도가 증가함에 따

라 ML 모델 배포 이후에도 모델의 정확성과 지속적인 성능을 유지하는 것이 중요해졌다[1]. ML 시스템 운영 환경에서 직면하는 성능 유지를 위한 주요 과제 중 하나가 드리프트(Drift) 현상이다. 드리프트는 입력 데이터의 통계적 특성이 시간이 지남에 따라 변하여 모델 성능을 저하시키는 현상이다[2]. 예를 들어, 과거에 안정적인 시장 데이터를 사용하여 회사의 주가를 예측하도록 학습된 모델은 처음에는 좋은 성능으로 운영되었으나 시간이 지나 시장 변동성이 증가함에 따라 데이터 분포의 변화로 인해 모델의 예측이 부정확해질 수 있다[3].

드리프트에 대한 기존의 대응 방식은 드리프트가 발생한 후 시스템 성능이 저하되면 ML 모델을 재학습하거나 미세 조정하는 등 모델 중심적인 전략을 채택하고 있다[4]. 그러나 이러한 방법은 일시적인 복구를 제공할 수 있으나, 운영 데이터가 빠르게 변화하는 동적 환경에서는 많은 시간과 비용이 요구된다. 이는 선제적이기 보다는 사후 대응적인 경향이 있어 수정 조치가 지연되고 ML 모델 유지 관리 주기가 비효율적이 될 수 있다. 이러한 ML 생명 주기 초기 단계의 데이터 검증 부족은 품질이 낮거나 유효하지 않은 데이터가 학습 파이프라인으로 유입되어 모델의 견고성과 일반화 능력을 저하시킨다.

본 연구의 목적은 품질 평가와 드리프트 사전 탐지 분석을 통해 ML 시스템이 동적이고 실제적인 운영 환경에 더욱 탄력적으로 대응할 수 있도록 하는 것이다. 이를 위해 ML 생명 주기 전반에 걸쳐 데이터를 모니터링하고 평가하는 품질 기반 데이터 중심 드리프트 관리(Q-DcDM) 프레임워크를 제안한다. Q-DcDM의 목표는 (1) 도메인 목적에 따라 데이터 품질 척도와 임계값을 설정하고, (2) 모델 학습 전에 품질 척도 및 통계 지표를 평가하여 드리프트 위험을 정량화하고, (3) 드리프트로 인한 성능 저하가 발생하기 전에

데이터 정제 또는 폐기와 같은 조기 개입을 가능하게 하는 것이다.

2. 관련 연구

2.1 드리프트 유형

2.1.1 데이터 드리프트

데이터 드리프트(Data Drift)는 ML 모델을 학습할 때와는 달리 입력 데이터(x)의 특징 분포가 변화하여 발생한다[4].

$$P_t(x) \neq P_{t+\Delta t}(x) \quad (\text{식-1})$$

- $P_t(x)$: 시간 t에서의 입력 데이터 분포
- $P_{t+\Delta t}(x)$: 시간 t+Δt에서의 입력 데이터 분포

예를 들어, 연령과 소득데이터를 기반으로 구매 예측을 학습한 ML 모델이 시간이 지남에 따라 고객의 연령과 소득 분포가 크게 변하는 경우 해당 데이터를 학습한 ML 모델의 예측 성능은 저하된다.

2.1.2 개념 드리프트

개념 드리프트(Concept Drift)는 ML 모델이 학습한 입력 데이터(x)와 목표 결과(y) 사이의 관계 변화로 시간 경과에 따른 조건부 분포($P(y|x)$)의 변화가 발생한다[5].

$$P_t(y|x) \neq P_{t+\Delta t}(y|x) \quad (\text{식-2})$$

- $P_t(y|x)$: 시간 t에서의 입출력 관계
- $P_{t+\Delta t}(y|x)$: 시간 t+Δt에서의 입출력 관계

예를 들어, 스팸메일을 탐지하도록 학습된 ML 모델이 시간이 지나 스팸의 유형이 변경되면서 이 모델은 더 이상 스팸을 정확하게 탐지할 수 없게 된다. 개념 드리프트는 발생 형태에 따라 급격한 드리프트, 점진적 드리프트, 누적 드리프트, 반복적 드리프트로 상세 분류한다.

2.1.3 사전확률 드리프트

사전확률 드리프트(Prior Probability Drift)는

출력된 클래스 사전확률이 시간에 따라 변하는 현상이다[6]. 입력과 출력간의 조건부 확률 $P(y | x)$ 는 변하지 않고 클래스 분포 자체가 달라지는 현상이다. 예를 들어, 스팸이 전체 메일의 5%였던 것이 시간이 지나면서 30%까지 증가하는 경우이다.

$$P_t(y) \neq P_{t+\Delta t}(y) \quad (\text{식-3})$$

- $P_t(y)$: 시간 t 에서의 클래스 분포
- $P_{t+\Delta t}(y)$: 시간 $t+\Delta t$ 에서의 클래스 분포

2.2 드리프트 탐지 기법

2.2.1 통계 기반 탐지 기법

1) PSI(Population Stability Index)

PSI는 학습 데이터 셋과 운영 데이터 셋이 동일한 분포인지 평가하는 척도이다[7]. 새롭게 입력되는 운영 데이터 셋에서 변수의 분포를 학습 데이터 셋과 비교하여 변수 분포의 변화를 측정한다. PSI 값이 낮을수록 두 데이터 셋의 분포가 더 유사하고 안정적이다.

2) KL(Kullback-Leibler) Divergence

KL Divergence는 예측값의 분포가 참(true value)인 기준 분포와 얼마나 다른지를 측정하는 값이다[8]. KL Divergence 결과는 항상 0 이상인 값을 반환한다. KL Divergence 결과값이 운영 데이터의 분포와 학습 데이터가 특정 임계값을 초과하면 드리프트가 발생했다고 판단한다.

3) KS(Kolmogorov-Smirnov) Test

KS Test는 두 데이터 셋이 동일한 분포를 따르는지 검정하기 위해 사용되는 비모수적 통계 검정기법이다[9]. KS-Test 결과값인 p-value가 0.05 미만이면 두 데이터 셋의 분포가 다르다는 것을 의미하고 드리프트가 탐지될 가능성이 있다고 판단한다.

4) MMD(Maximum Mean Discrepancy)

MMD는 두 확률분포 간의 차이를 정량화하는 커널 기반의 통계 거리 측정 기법이다[10]. 학습 데이터와 운영 데이터의 평균 임베딩 벡터 차이

를 계산하여 분포 차이를 탐지한다.

5) FID(Fréchet Inception Distance)

FID는 이미지 분포의 유사성을 측정하기 위한 통계 척도이다[11]. 두 이미지 데이터 셋(실제 이미지와 생성 이미지)에서 추출한 피처의 분포 평균과 공분산을 구한 후 이들 간의 Fréchet 거리를 계산한다. Fréchet 값이 0에 가까울수록 생성 이미지 분포가 실제 이미지와 유사하다고 판단한다.

2.2.2 스트리밍 기반 탐지 기법

1) DDM(Drift Detection Method)

DDM은 오류율과 표준 편차를 분석하여 스트리밍 데이터 셋의 드리프트 유무를 탐지한다[12]. 운영 데이터 셋 i 번째 피처에 대해 잘못된 예측을 할 확률인 오류율 p_i 와 표준 편차 $s_i = \sqrt{p_i \times (1 - p_i) / i}$ 를 정의한다. 운영 데이터 셋의 오류율이 증가하면 데이터 분포의 변경으로 인해 드리프트가 발생했다고 판단한다.

2) EDDM(Early DDM)

EDDM은 DDM과 유사하지만 오류율의 변화 감시가 아닌 두 개의 연속된 오류 사이의 거리를 평가한다[13]. 연속된 오류 사이의 평균 거리와 표준편차를 추적하여 오류 사이의 거리가 감소하면 드리프트가 발생했다고 판단한다.

3) HDDM(Hopping Drift Detection Method)

HDDM은 온라인 데이터 스트림에서 드리프트를 탐지하기 위한 기법이다[14]. 주된 특징은 “hopping” 윈도우 전략으로 다양한 길이의 윈도우에서의 패턴 변화를 기반으로 드리프트를 탐지한다. 분류기의 에러율 변화를 기반으로 하며 실시간 반응성이 높은 것이 장점이다.

4) PHT(Page-Hinkley Test)

PHT는 평균의 누적 변화량을 추적하여 갑작스러운 평균의 변화를 탐지하는 기법이다[15]. 특히 데이터의 평균 변화에 민감하게 반응하며 실시간 스트리밍 데이터에 적용하기 용이하다. 이

방법은 누적된 차이값이 특정 임계값을 넘는 경우 드리프트가 발생한 것으로 판단한다.

2.2.3 AutoML 기반 탐지 기법

Celik은 AutoML 환경에서 베이지안 최적화, 유전자 프로그래밍 등을 사용하여 개념 드리프트에 대응하기 위한 전략들을 제안하였다[16]. AMLAS는 ML 모델 선택, 하이퍼파라미터 튜닝을 자동화하여 일정 수준(예: 약 7%)의 성능이 떨어지면 모델을 실시간 적용시키는 프레임워크이다[17]. 센서데이터의 점진적 개념 드리프트를 보정하기 위해 메타러닝과 자동 앙상블 학습을 활용한 AutoML-DC모델이 제안되었다[18]. 이러한 연구들은 여전히 모델 재학습에 집중되어 있고 IoT환경이나 센서 데이터 등 한정된 범위에서의 대응방법이다.

3. 품질 기반 데이터 중심 드리프트 관리

본 연구에서는 ML 생명 주기 초기 단계에서 드리프트를 선제적으로 관리하기 위한 품질 기반 데이터 중심 드리프트 관리(Q-DcDM: Quality-based Data-centric Drift Management) 프레임워크를 제안한다. Q-DcDM 프레임워크는 그림 1과 같이 5개의 요소로 구성된다. (1) 품질 요구사항 정의, (2) 데이터 품질 평가, (3) 임계값 기반 데이터 셋 정제, (4) 드리프트 분석, (5) 드리프트 탐지 기법 추천

3.1 품질 요구 사항 정의

데이터 품질 요구사항 정의(Data Quality Requirement Definition)에서는 도메인 특성에 특화된 품질 평가를 목표 기반 데이터 품질 척도(DQM, Data Quality Metric)을 정의한다. 데이터 품질은 특정 비즈니스 목적이나 도메인 요구사항을 충족하는지를 평가하는 중요한 요소이다. 본

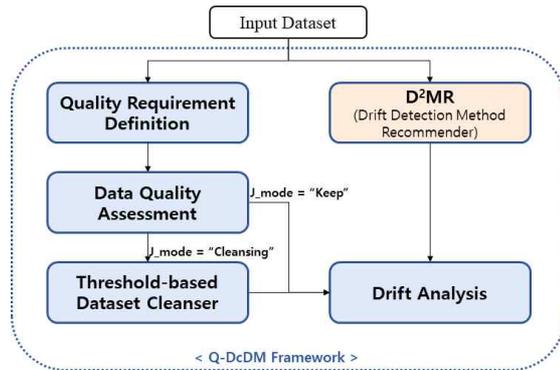


그림 1. Q-DcDM 프레임워크
Fig. 1. Q-DcDM Framework

연구에서는 목적 기반 데이터 품질 평가를 위해 데이터 품질 요구사항을 정의한다.

- (1) 도메인 특성에 맞는 품질 척도 정의
- (2) 각 품질 척도에 대한 임계값 설정
- (3) 임계값 위반 비율과 데이터 분류 설정

데이터 품질 척도는 사용되는 데이터 유형에 따라 다르게 설정한다. 정형 데이터의 경우는 정확성, 완전성, 일관성 및 중복성 네 가지 품질 척도를 선정한다. 이러한 척도는 ML 모델 학습에서 데이터 셋의 품질을 평가하기 위해 가장 널리 사용되고 있다. 정형 데이터의 품질 척도는 표준에서 정의되어 있으나 임계값은 도메인마다 다르게 설정할 수 있다. 표 1에서 임계값은 많은 문헌과 표준에서 제시해주는 값이다.

표 1. 품질 척도와 임계값(정형 데이터)
Table 1. Quality Metrics and Thresholds (Structured Data)

Metrics	Formula	Threshold
Accuracy (acc)	$acc = 1 - N_{be} / Total$ (N_{be} : No. of error values)	≥ 0.98
Completeness (com)	$com = 1 - N_{bl} / Total$ (N_{bl} : No. of missing values)	≥ 0.99
Consistency (cons)	$cons = N_{con} / Total$ (N_{con} : No. of consistent vals)	≥ 0.97
Redundancy (red)	$red = 1 - N_{bd} / Total$ (N_{bd} : No. of duplicates)	≤ 0.02

텍스트, 이미지, 오디오와 같은 비정형 데이터는 미리 정의된 스키마가 없기 때문에 기존의 품질 측정 기준을 적용하기 어렵다. 예를 들어, 비정형 데이터는 표 2와 같이 데이터 유형별 품질 측정 기준을 참조할 수 있다. 그러나 비정형 데이터의 품질은 데이터의 무결성, 콘텐츠 유용성, 통계적 일관성 등 도메인별 맞춤형 기준을 설정한다.

표 2. 품질 척도와 임계값(비정형 데이터)
Table 2. Quality Metrics and Thresholds (Unstructured Data)

Data Type	Metrics	Description	Threshold
Text	Readability Score	Assesses sentence clarity and grammar	≥ 60 readability score
	Language Consistency	Consistent language use across documents	$\geq 95\%$ consistent tokens
Image	Resolution Quality	Pixel density meets application requirement	$\geq 224 \times 224$ pixels
	Clarity/Blur Score	Measured using SSIM or Laplacian variance	SSIM ≥ 0.90 or LapVar ≥ 100
	Redundancy Detection	Duplicate image detection	$\leq 5\%$ duplicates
Audio	Signal-to-Noise Ratio	Quantifies audio clarity vs noise	SNR ≥ 20 dB
	Duration Validity	Consistent duration across samples	[1s - 10s]
	Transcription Accuracy	ASR transcription confidence	$\geq 95\%$ WER-based confidence

3.2 데이터 품질 평가

데이터 품질 평가(Data Quality Assessment)는 품질 척도 기반으로 품질 요구 사항을 준수하는지 평가한다.

- (1) 각 품질 척도 값 계산
- (2) 임계값 위반 비율 계산
- (3) 임계값 위반 비율에 따른 데이터 관리전략

품질 요구사항 정의 단계에서 설정한 척도와

임계값을 기반으로 척도 값을 계산하고 임계값 위반 비율을 계산한다. 임계값의 위반 비율에 따라 ML 시스템에서 데이터의 사용 기준을 설정한다. 이 기준은 도메인에 따라 데이터 특성에 따라 달라지기 때문에 일반화하기 어렵다. 본 연구에서 정형 데이터의 경우는 표 3과 같이 임계값 위반 비율과 이에 따른 데이터 분류 기준을 설정한다.

표 3. 임계값 위반 비율과 판단(정형 데이터)
Table 3. Threshold Violation Rate and Judgement

Violation Rate	Judgement
$\leq 10\%$	Keep
$> 30\%$	Discard
10% ~ 30%	Cleansing

임계값 위반 비율에 따라 데이터는 다음과 같이 세 가지로 분류한다.

(1) 보존(Keep): 품질 임계값 위반 비율이 낮아 모델 성능에 미치는 영향이 미미한 데이터는 ML 시스템의 예측 서비스에 사용될 수 있다.

(2) 정제(Cleansing): 위반 비율이 중간 수준인 데이터는 품질 저하 요인을 제거하는 정제 과정을 거친 후에서 재평가한다.

(3) 폐기(Discard): 위반 비율이 매우 높아 구조적 정보가 파괴된 데이터는 드리프트 발생 확률과 모델 오차를 유발할 가능성이 크므로 학습 및 예측 파이프라인에서 제외한다.

예를 들어, 정형 데이터의 경우 표 1에서 정의한 품질 척도인 정확도(accuracy)의 임계값 0.98을 위반한 비율이 10%이내이면 품질 평가 다음 단계인 드리프트 분석을 위해 데이터 셋을 보존한다. 그러나 임계값을 위반한 데이터가 30% 이상이면 그 데이터는 폐기하고 새로운 데이터 셋을 입력받는다. 임계값을 위반한 데이터가 10%

이상이고 30%이하이면 품질 기준을 위반한 데이터를 삭제 한 이후에 해당 데이터 셋은 다음 단계인 드리프트 분석을 위해 사용한다.

3.3 임계값 기반 데이터 셋 정제

임계값 기반 데이터 셋 정제(Threshold-based Dataset Cleansing)는 품질 평가 단계에서 “Cleansing” 판단을 받은 데이터 셋에서 임계값을 위반한 데이터를 제거하고 품질이 확보된 새로운 데이터 셋을 생성한다.

3.4 Drift Analysis

드리프트 분석(Drift Analysis)의 대상은 품질 평가 단계에서 “Keep” 판단을 받은 데이터 셋과 “Cleansing” 판단을 받은 데이터 셋을 정제한 새로운 데이터 셋이다. 이들 데이터 셋에 대해서 D²MR에서 추천한 탐지기법을 사용하여 드리프트 탐지여부를 평가한다. 평가 결과에 따라 ML 생명주기에서 다음 단계가 모델 재학습 또는 예측 단계를 결정한다.

3.5 드리프트 탐지 기법 추천

드리프트 탐지 기법 추천(D²MR, Drift Detection Method Recommender)은 저자의 기존 연구로서 데이터 특성을 분석하여 가장 적합한 드리프트 탐지 기법을 추천하는 프레임워크이다 [20]. 본 연구에서는 D²MR 프레임워크를 통합하여 Q-DcDM 프레임워크의 유효성을 높이고자 한다.

4. Case Study

본 연구에서는 정형 데이터 셋(UCI Wine Quality)과 비정형 데이터 셋(CIFAR-10)을 활용하여 Q-DcDM 프레임워크의 유효성을 검증한다.

표 4. Q-DcDM 수도코드
Table 4. Q-DcDM Pseudocode

Algorithm: Q-DcDM Framework
Input: D_base: original training dataset (baseline) D_ops: operational datasets {op1, op2, ..., opN} DQM: set of data quality metrics {DQM1, DQM2, ..., DQMn} T_quality: quality thresholds T_violation: violation-rate thresholds (Keep, Cleansing, Discard) ML_Model: deployed ML model Output: FinalDecision, ML_Model Define quality metrics according to DomainInfo Initialize FinalDecision $\leftarrow \emptyset$ For each dataset D_op in D_ops do For each metric m in DQM do q[m] \leftarrow ComputeQualityMetric(D_op, m) v[m] \leftarrow CheckThresholdViolation(q[m], T_quality[m]) End For violation_rate \leftarrow AggregateViolationRate(v) if violation_rate \leq T_violation.keep then decision \leftarrow "Keep" D_candidate \leftarrow D_op else if T_violation.keep < violation_rate \leq T_violation.discard then decision \leftarrow "Cleansing" D_candidate \leftarrow CleanseDataset(D_op) RecomputeQualityMetrics() else decision \leftarrow "Discard" FinalDecision[D_op] \leftarrow decision RequestNewDataset() end if detector \leftarrow D2MR_Recommend(D_candidate, DomainInfo) drift_score \leftarrow ComputeDrift(D_base, D_candidate, detector) if DriftDetected(drift_score) then ML_Model \leftarrow Retrain(ML_Model, D_candidate) action \leftarrow "Retrain" else action \leftarrow "Prediction" end if FinalDecision[D_op] \leftarrow action End For return FinalDecision, ML_Model

4.1 실험 설정

4.1.1 데이터 셋

1) UCI Wine Quality Dataset

UCI Wine Quality(이후, Wine) 데이터 셋은 약 4,898개의 샘플과 11개의 화학적 특성과 품질 등급이 포함되어 있는 수치형 변수로 구성된 정형 데이터 셋이다[20]. Wine 데이터 셋은 저차원의 정규분포가 강하게 나타나는 특성이 있어서 모든 유형의 드리프트의 발생이 가능하다.

2) CIFAR-10 Dataset

CIFAR-10 데이터 셋은 10개 클래스에 동일한 개수의 이미지로 구성된 비정형 데이터 셋이다[21]. CIFAR-10 데이터 셋은 저해상도(32x32) 컬러 이미지가 약 60,000개의 RGB 이미지로 구성되어 있으며 클래스 간 분포가 균일하고 시각적 스타일 변화가 주요한 드리프트 요인이 된다.

4.1.2 실험 절차

1) 운영 데이터 셋 생성

ML 시스템 운영환경에서 시간의 흐름에 따라 입력되는 운영 데이터를 생성한다. Wine과 CIFAR-10 데이터 셋에서 일부 변형된 각각의 10개의 운영 데이터 셋(예: op1, op2,...)을 생성한다. 이 데이터 셋은 실제 운영 환경에서 흔히 발생하는 ML 시스템의 교란 요인을 반영하여 특징 분포 변화, 노이즈 주입, 흐름 효과, 밝기 변화 및 기타 데이터의 교란 등을 포함하였다.

2) 품질 요구사항 설정과 평가

데이터 품질 척도 및 임계값을 설정하고 이를 기반으로 운영 데이터 셋의 품질 평가를 한다. 각 운영 데이터 셋의 임계값 위반 비율을 기준으로 해당 데이터 셋을 ML 시스템에서 활용할지 정제하여 사용할지 아니면 폐기할지 판단한다.

3) 드리프트 분석

품질 평가 단계에서 데이터 보존 또는 정제 판단을 받은 데이터는 드리프트 탐지 여부를 평가한다. 분석 결과가 드리프트 탐지 가능성이 나오면 ML 모델을 재학습하고 탐지가 되지 않으면 ML 생명주기의 예측으로 진행한다.

또한, Q-DcDM 프레임워크이 없는 경우의 초기 운영 데이터 셋을 대상으로 드리프트를 분석 결과가 Q-DcDM 기반의 드리프트 분석에 비해 모델의 재학습 횟수가 높다는 것을 확인한다. 운영 시스템에서 모델 재학습을 자주 수행하면 성능 저하가 초래되기 때문에 비교 분석을 통해 본 프레임워크의 유효성을 검증한다.

4) 모델 성능 비교

품질 평가에서 정제 판단을 받았으나 정제하지 않은 데이터 셋과 정제한 데이터 셋의 모델 성능을 비교하여 제안한 프레임워크의 유효성을 확인한다.

4.2 실험 결과

4.2.1 UCI Wine Quality Dataset

1) Wine 운영 데이터 셋 생성

Wine데이터 셋에 노이즈, 이상치 및 결측치 주입 등의 변환을 통해 10개의 운영 데이터 셋(op1_Wine~op10_Wine)을 임의적으로 생성한다.

2) 품질 요구사항 설정과 평가

① 품질 요구사항: Wine 데이터 셋은 품질 척도와 임계값을 표 1과 같이 선정한다. 임계값 위반 비율에 따른 ML 시스템에서의 데이터 사용 기준은 표 3과 같이 설정한다.

본 실험에서는 위반 비율과 판단을 다양한 문헌과 업무에서 사용하는 정형 데이터의 기준으로 하였으나 산업 도메인이나 조직에 따라 다르게 설정해야 한다. 예를 들어, 일반적인 산업 도메인

에 비해 국방, 항공, 의료 도메인의 경우 데이터를 폐기하는 임계값 위반 비율을 5%나 10%이내로 조정해야 한다.

② 품질 평가: 10개의 Wine 운영 데이터 셋에 4개의 품질 척도(accuracy, completeness, consistency, redundancy)를 평가하고 임계값 위반 비율을 계산한다. 품질 평가 결과는 표 5와 같다. op6_Wine 데이터 셋은 ML 시스템에서 활용가능하고 5개의 데이터 셋(op1_Wine, op5_Wine, op7_wine, op8_Wine, op9_Wine)은 정제로 판단되었다. 정제해야 하는 데이터 셋은 품질 임계값을 위반한 데이터를 제거하여 품질을 개선한 이후 사용한다. 품질 개선한 데이터 셋은 새로운 이름(op1-1_Wine, op5-1_Wine, op7-1_Wine, op8-1_Wine, op9-1_Wine)의 데이터 셋으로 생성한다. 위반 비율이 높은 4개의 데이터 셋(op2_Wine, op3_Wine, op4_Wine, op10_Wine)은 모델 학습에 부적합하므로 해당 데이터 셋은 폐기한다. 폐기한 데이터 셋을 사용할 경우 드리프트 발생 확률이 높고 모델 성능을 저하 시키는 요인이 된다.

표 5. Wine 데이터: 품질 평가
Table 5. Wine Data: Quality Assessment

Dataset	Violation Rate	Judgement	Dataset after Cleansing
op1_Wine	12%	Cleansing	op1-1_Wine
op2_Wine	20%	Discard	
op3_Wine	22%	Discard	
op4_Wine	24%	Discard	
op5_Wine	7%	Cleansing	op5-1_Wine
op6_Wine	4%	Keep	
op7_Wine	10%	Cleansing	op7-1_Wine
op8_Wine	13%	Cleansing	op8-1_Wine
op9_Wine	6%	Cleansing	op9-1_Wine
op10_Wine	25%	Discard	

3) 드리프트 분석

① Q-DcDM 기반 드리프트 분석

품질 평가에서 폐기 판단을 받은 운영 데이터 셋을 제외하고 6개의 데이터 셋에 대해 드리프트 가능성을 탐지한다. 이 때 사용된 KS-Test 기법은 기존 연구에서 실험한 D²MR에서 Wine 데이터 셋에 적합한 기법을 추천받아 사용한다. 본 실험에서는 최초의 Wine 데이터 셋과 6개의 운영 데이터 셋을 대상으로 분석하였다. KS-Test 기법은 p-value가 0.05 미만이면 드리프트 발생 가능성이 있는 것으로 간주한다. 그림 2에서 보는 바와 같이, 품질 평가에서 보존 판단을 받은 op6_Wine 데이터 셋은 드리프트 발생 가능성이 없는 것을 알 수 있다.

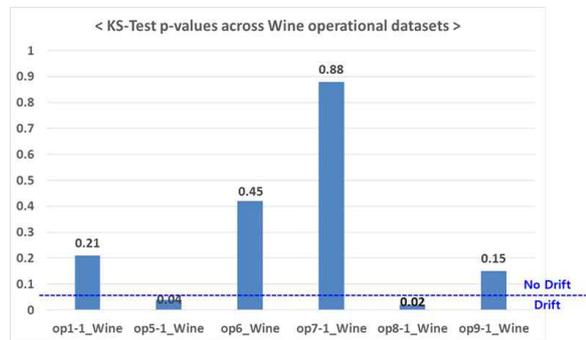


그림 2. Wine: Q-DcDM 기반 드리프트 분석
Fig. 2. Wine: Q-DcDM-based Drift Analysis

품질 평가에서 정제 판단을 받고 정제된 데이터 셋들은 일부는 드리프트 발생 가능성이 없는 것으로 확인된다. 그러나 정제된 데이터 셋 중에서 op5-1_Wine과 op8-1_Wine 데이터 셋은 드리프트 발생 가능성을 포함하고 있다. 이런 데이터의 경우는 드리프트 발생 요인이 잠재되어 있기 때문에 ML 모델을 재학습한다.

② Q-DcDM없이 수행한 드리프트 분석

그림 3은 품질 평가와 이에 따른 조치가 없었던 초기 운영 데이터 셋에 대한 드리프트 분석 결과이다. op6_Wine데이터 셋을 제외한 9개의 데이터 셋이 드리프트 발생 가능성이 있는 것으로 나타났다. 만약 이 데이터 셋들을 ML 시스템에서 사용하였다면 모델 성능이 저하되고 모델을 재학습하는 횟수가 증가하게 된다.

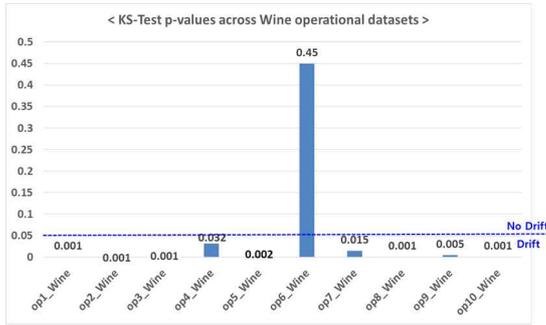


그림 3. Wine: Q-DcDM없는 드리프트 분석
Fig. 3. Wine: Drift Analysis without Q-DcDM

이와 같이 Q-DcDM 프레임워크에서는 데이터 셋이 두 개의 데이터 셋을 ML 모델을 재학습하는 반면, 기존 모델 중심 드리프트 관리에서는 9개의 데이터 셋을 재학습하게 된다.

4) 모델 성능 비교

그림 4는 품질 평가에서 정제 판단을 받았으나 정제하지 않은 데이터 셋과 이를 정제한 데이터 셋들의 성능 비교이다. 파란 막대는 정제되지 않은 초기 운영 데이터 셋이고 붉은 막대는 정제된 데이터 셋의 정확도이다. 모델의 평균 정확도는 데이터 정제 전 0.75에서 정제 후 0.898로 향상되었다. 따라서 품질 평가와 조기의 드리프트 분석에 따른 드리프트 관리가 드리프트 발생률을 감소하고 모델 성능과 밀접하다는 것을 알 수 있다.

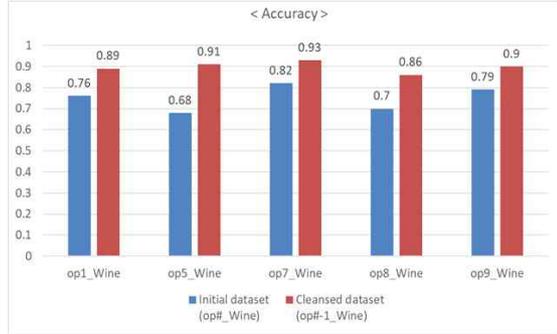


그림 4. Wine datasets의 모델 성능
Fig. 4. Model performance on Wine datasets

4.2.2 CIFAR-10 Dataset

1) 운영 데이터 생성

CIFAR-10 데이터 셋에 밝기 왜곡, 노이즈 주입, 흐림 현상, 레이블 불일치 등 다양한 이상 현상을 시뮬레이션하기 위해 10개의 운영 데이터 셋(op1_CIFAR ~ op10_CIFAR)을 생성한다.

2) 품질 요구사항 설정과 평가

① 품질 요구사항: CIFAR-10 데이터 셋의 품질 척도와 임계값, 임계값 위반 비율에 따른 사용 기준은 표 6과 같이 설정한다.

표 6. CIFAR-10 데이터: 임계값 위반 비율과 판단
Table 6. CIFAR-10 Data: Threshold Violation Rate and Judgement

Quality Metrics	Threshold	Violation Rate & Judgement
PSNR	≥ 30 dB	$\leq 15\%$: Keep $> 30\%$: Discard (10%, 30%]: Cleanse
SSIM	≥ 0.90	
NIQE	$\leq \text{median}(\text{baseline}) + 1.5 \times \text{IQR}(\text{baseline})$	
BRISQUE	$\leq \text{median}(\text{baseline}) + 1.5 \times \text{IQR}(\text{baseline})$	

② 품질 평가: 10개의 CIFAR-10 운영 데이터 셋에 4개의 품질 척도(PSNR, SSIM, NIQE,

BRISQUE)를 평가하고 임계값 위반 비율을 계산한다. 품질 평가는 표 7과 같은 결과를 얻었다.

보존 판단인 4개의 데이터 셋(op3_CIFAR, op4_CIFAR, op8_CIFAR, op10_CIFAR)은 ML 시스템에서 활용 가능하다. 3개의 데이터 셋(op1_CIFAR, op7_CIFAR, op9_CIFAR)은 품질 기준에 맞지 않은 데이터를 삭제한 후에 op1-1_CIFAR, op7-1_CIFAR, op9-1_CIFAR 데이터 셋을 생성한다. 위반 비율이 높은 3개의 데이터 셋(op2_CIFAR, op5_CIFAR, op6_CIFAR)은 폐기한다.

표 7. CIFAR-10 데이터 셋: 품질 평가
Table 7. CIFAR-10 Dataset: Quality Assessment

Dataset	Violation Rate	Decision	Dataset after Cleansing
op1_CIFAR	28%	Cleanse	op1-1_CIFAR
op2_CIFAR	35%	Discard	
op3_CIFAR	8%	Keep	
op4_CIFAR	12%	Keep	
op5_CIFAR	45%	Discard	
op6_CIFAR	52%	Discard	
op7_CIFAR	18%	Cleanse	op7-1_CIFAR
op8_CIFAR	5%	Keep	
op9_CIFAR	25%	Cleanse	op9-1_CIFAR
op10_CIFAR	2%	Keep	

3) 드리프트 분석

① 드리프트 탐지 기법: FID

본 단계에서는 D²MR에서 CIFAR-10 데이터 셋 특성을 분석하여 추천받은 FID 기법을 사용한다. 본 실험에서 원본은 최초의 CIFAR-10 데이터 셋이고 변형 데이터 셋은 표 7에서 굵은 글씨체의 7개의 데이터 셋이다. 즉, 품질 평가 단계에서 보존 판단 받은 데이터 셋과 정제한 데이터 셋이 드리프트 분석 대상이 된다.

② CIFAR-10에서의 FID 기준: 15

FID 결과 값이 0에 가까울수록 두 데이터 셋 간의 유사함을 의미한다. FID의 드리프트 탐지 기준은 절대적인 단일 기준값이 존재하지 않는다. 일반적으로 학계와 실무에서 통용되는 상대적 가이드라인이나 통계 기준은 표 8과 같다. 따라서 일반적으로 가이드라인에 따라 FID 값이 30 또는 40으로 기준을 잡고 그 이상이면 드리프트가 발생할 수 있다고 판단할 수 있다.

표 8. FID 수치 범위와 드리프트 상태
Table 8. FID numerical range and Drift status

FID range	Drift status	Description
0 ~ 10	Negligible	Nearly identical quality to the original.
10 ~ 30	Moderate	Model accuracy may decrease by 1-3%.
30 ~ 70	High	Significant degradation in model performance.
≥ 70	Critical	Dataset reliability is lost (Discard recommended.)

본 실험에서는 FID 기준값을 얻기 위해 CIFAR-10 데이터 셋을 특성을 확인하고자 FID 값과 모델 정확도의 관계를 분석하였다. 그림 5와 같이 모델 정확도가 급격하게 떨어지는 곡선(elbow point)의 시작점을 기준(FID=15)으로 설정하였다. 따라서 FID 값이 15이상이면 드리프트 발생 가능성이 있다고 판단하였다.

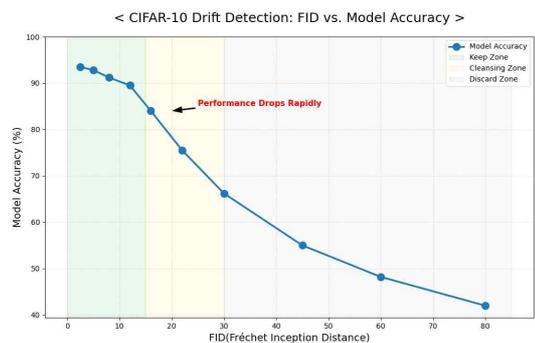


그림 5. CIFAR-10 dataset: FID vs. Accuracy
Fig. 5 CIFAR-10 dataset: FID vs. Accuracy

③ Q-DcDM 기반 드리프트 분석

품질 평가를 통해 7개의 데이터 셋을 대상으로 FID기법을 적용하여 분석하였다. 그림 6에서 보는 바와 같이 FID 값 15를 기준으로 대부분의 데이터 셋은 드리프트 발생 가능성이 없는 것으로 분석된다. op9-1_CIFAR 데이터 셋은 정제 후에도 FID 값이 18.75로 잠재적인 드리프트가 있을 수 있으므로 모델을 재학습한다.

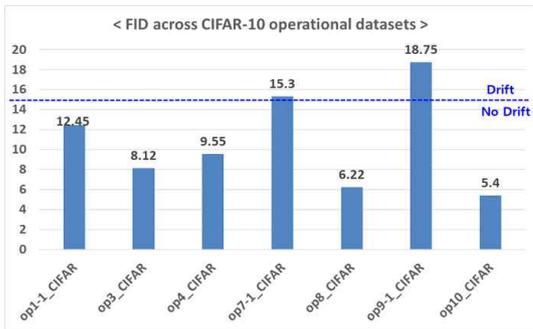


그림 6. CIFAR-10: Q-DcDM 기반 드리프트 분석
Fig. 6. CIFAR-10: Q-DcDM-based Drift Analysis

④ Q-DcDM없이 수행한 드리프트 분석

그림 7은 초기 운영 데이터 셋에 대한 드리프트 분석 결과이다. Q-DcDM 기반 드리프트 분석에서 모델 재학습 데이터 셋이 한 개인 반면 Q-DcDM 없이 수행한 드리프트 분석에서는 6개

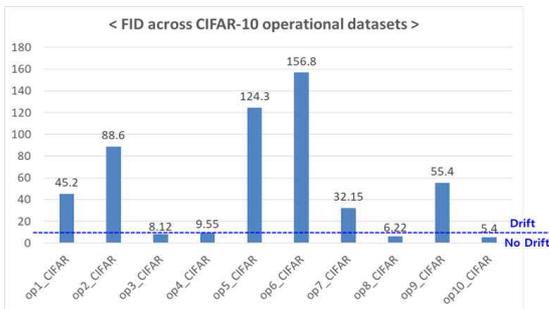


그림 7. CIFAR-10: Q-DcDM없는 드리프트 분석
Fig. 7. CIFAR-10: Drift Analysis without Q-DcDM

의 데이터 셋이 드리프트 발생 가능성이 있는 것으로 나타났다. 이는 모델 성능 저하와 모델을 재학습하는 횟수가 증가하게 되는 요인이 된다.

4) 모델 성능 비교

데이터 정제 전후의 데이터 셋에 대한 모델 성능은 그림 8과 같다.

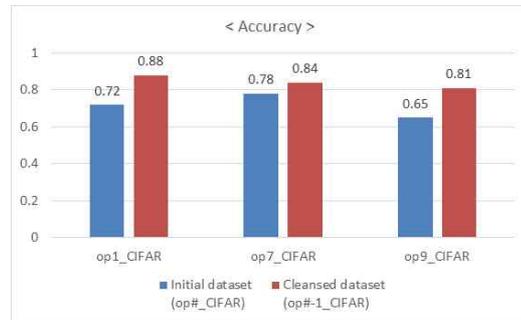


그림 8. CIFAR-10 데이터 셋의 모델 성능
Fig. 8. Model performance on CIFAR-10 datasets

파란 막대는 초기 운영 데이터 셋이고 붉은 막대는 정제 후 데이터 셋의 정확도이다. 모델의 평균 정확도는 데이터 정제 전 0.72에서 정제 후 0.843로 향상되었다. 이를 통해 ML 시스템에서 지속적인 신뢰성과 성능 유지를 위해 체계적인 품질 평가 및 초기 드리프트 관리는 필수적임을 알 수 있다.

4.3 실험 결과 요약

4.3.1 모델 성능 향상

실험 결과 상이한 유형의 두 데이터 셋에서 그림 9와 같이 각각 14.8%, 12.3%의 유의미한 성능 향상이 관찰되었다.

4.3.2 모델 재학습 횟수 감소

드리프트 분석 대상 데이터 셋을 줄이고 모델 재학습 횟수를 감소시키는 효과를 확인하였다.

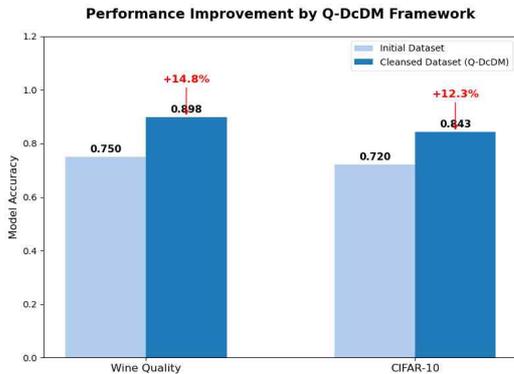


그림 9. Q-DcDM 프레임워크 기반 성능 향상
Fig. 9. Q-DcDM-based Performance Improvement

드리프트 분석을 통해 Wine 운영 데이터 셋은 9 개에서 한 개로, CIFAR-10 운영 데이터 셋은 6 개에서 한 개로 드리프트 발생 가능성이 있는 데이터 셋이 감소되었다.

4.4 토의

4.4.1 Q-DcDM과 ML 생명주기 통합

Q-DcDM 프레임워크는 그림 10과 같이 ML 파이프라인과 밀접하게 연계되어 있어서 통합 확장 가능하다. 품질 평가 단계에서 정제 판단을 받은 데이터 셋은 폐기하고 새로운 데이터를 수집한다. 또한 드리프트 분석에서 잠재적인 드리프트를 포함하고 있는 데이터 셋은 모델을 재학습하고 Q-DcDM 프레임워크에서 받는 데이터는 ML의 전처리 프로세스로부터 받는다. 이와 같이 Q-DcDM 프레임워크의 각 프로세스는 ML 생명주기 각 단계별로 밀접하게 연계되어 있다.

4.4.2 AutoML 기반 드리프트 탐지 기법

최근에 연구된 AutoML 기반 드리프트 탐지 기법과 Q-DcDM 프레임워크와 비교하였다. 기존 기법은 ML 모델의 하이퍼파라미터를 자동화하거나 메타러닝을 하는 방법 차이는 있지만 모델 재학습에 초점을 두고 있다.

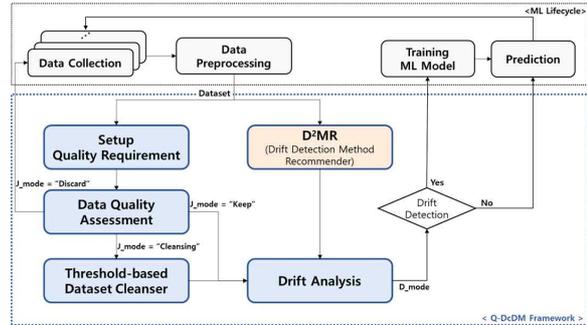


그림 10. Q-DcDM 프레임워크와 ML 생명주기
Fig. 10. Q-DcDM Framework and ML Lifecycle

표 9. AutoML기반 드리프트 탐지 기법 비교
Table 9. Comparison of AutoML-based Drift Detection

	AML4S	AutoML-DC	Q-DcDM
main method	automation of hyperparameter tuning	meta-learning automated ensemble learning	data quality evaluation
main focus	model retrain	model retrain	data
detection	data drift	concept drift	data/concept drift

5. 결론

본 연구에서는 ML 시스템의 운영환경에서 시간의 흐름에 따라 데이터 품질 저하와 데이터 분포 변화로 인해 발생하는 드리프트 문제를 해결하기 위하여 품질 기반 데이터 중심 드리프트 관리(Q-DcDM) 프레임워크를 제안하였다.

본 연구에서는 데이터 품질 평가와 데이터 분포 분석을 통합한 프레임워크를 설계하고 이를 통해 ML 생명주기 초기 단계에서 드리프트 발생 가능성을 사전에 탐지하고 대응할 수 있도록 하였다. 정형 데이터인 UCI Wine Quality와 비정형 데이터인 CIFAR-10 데이터 셋을 이용한 실험에서 제안한 프레임워크의 유효성을 검증하였다. 상이한 유형의 두 데이터 셋에서 각각 14.8%,

12.3%의 성능 향상이 되었다. 또한, 품질 평가를 선행함으로써 드리프트 분석 대상 데이터 셋의 양을 줄이고 모델 재학습 횟수를 감소시키는 효과도 확인하였다. 이러한 결과는 ML 시스템에서 지속적인 모델 성능과 신뢰성을 유지하기 위해 체계적인 데이터 품질 관리와 조기 드리프트 탐지 전략이 필수적임을 보여준다. 본 연구는 ML 시스템을 운영하는 모든 현장에서 활용 가능하나 특히 safety-critical SW를 운영해야 하는 국방이나 의료, 철도 분야에 더욱 유용할 것으로 예상된다.

본 연구에서 품질 척도나 임계값을 문헌에서 제시하는 값으로 설정하였으므로 산업 현장에 적용하기에 한계가 있다. 따라서 향후에는 산업 현장의 데이터를 확보하여 검증할 예정이다. 또한 도메인 제약 사항을 반영하거나 시계열, 영상 등 다양한 유형의 멀티모달 데이터 확장과 AutoML 기반 드리프트 대응 전략을 결합한 지능형 데이터 중심 ML 운영 관리 프레임워크에 대한 연구가 요구된다.

이 논문은 정부(과학기술정보통신부)의
재원으로
정보통신기획평가원-지역지능화혁신인재양
성사업의 지원을 받아 수행된 연구임
(IITP-2026-RS-2022-00156334)

참 고 문 헌

- [1] C. Zhang et al., “Understanding deep learning requires rethinking generalization”, ICLR 2017, DOI: 10.48550/arXiv.1611.03530
- [2] Firas Bayram et al., “From Concept Drift to Model Degradation: An Overview on Performance-Aware Drift Detectors”, Knowledge-Based Systems, vol. 245, Jun. 2022, DOI: 10.1016/j.knosys.2022.108632
- [3] A. Tsymbal, “The Problem of Concept Drift: Definitions and Related Work”, Computer Science Department, Trinity College, Dublin, 106(58), 2004, https://www.researchgate.net/publication/228723141_The_Problem_of_Concept_Drift_Definitions_and_Related_Work
- [4] Tri Minh Triet Pham et al., “Time to Retrain? Detecting Concept Drifts in Machine Learning Systems”, arXiv, 2024, DOI: 10.48550/arXiv.2410.09190
- [5] Jie Lu et al., “Learning under Concept Drift: A Review”, IEEE Transactions on Knowledge and Data Engineering, 31(12), pp.2346-2363, Dec. 2019, DOI: 10.1109/TKDE.2018.2876857
- [6] Moreno-Torres et al., “A unifying view on dataset shift in classification”, Pattern Recognition, 45(1), pp.521-530, Jan 2012, DOI: 10.1016/j.patcog.2011.06.019
- [7] B. Yurdakul and J. Naranjo, “Statistical Properties of Population Stability Index”, Journal of Risk Model Validation, 14(4), pp.89-100, 2020, DOI: 10.21314/JRMV.2020.227
- [8] S. Kullback and R. A. Leibler, “On information and sufficiency”, Institute of Mathematical Statistics, 22(1), pp.79-86, 1951. <https://www.jstor.org/stable/2236703>
- [9] J. Gama et al., “A Survey on Concept Drift Adaptation”, ACM Computing Surveys, 46(4), Article:44, pp.1-37, Mar. 2014. DOI: 10.1145/2523813
- [10] A. Gretton, et al., “A Kernel Two-Sample Test”. Journal of Machine Learning Research, vol. 13, pp.723-773, 2012, DOI: 10.5555/2188385.2188410
- [11] Min Jin Chong and David Forsyth, “Effectively Unbiased FID and Inception Score and where to find them”, CVPR 2020, DOI: 10.48550/arXiv.1911.07023
- [12] J. Gama et al., “Learning with Drift

Detection”, Lecture Notes in Computer Science (LNAI 3171), pp.286-295, 2004, DOI: 10.1007/978-3-540-28645-5_29

[13] M. Baena-García et al., “Early Drift Detection Method”, In Proceedings of the International Workshop on Knowledge Discovery from Data Streams, vol. 6, pp. 77-86, 2006

[14] A. Pesaranghader et al. “Reservoir of Diverse Adaptive Learners and Stacking Fast Hoeffding Drift Detection Methods for Evolving Data Streams”. Machine Learning, vol. 107, pp.1711-1743, 2018, DOI: 10.1007/s10994-018-5719-z

[15] E.S. Page, “Continuous Inspection Schemes”. Biometrika, 41(1/2), pp.100-115, 1954, DOI: 10.2307/2333009

[16] B. Celik and J. Vanschoren, “Adaptation Strategies for Automated Machine Learning on Evolving Data”, IEEE Transactions on Pattern Analysis and Machine Intelligence”, 43(9), pp. 3067-3078, 2021, DOI: 10.1109/TPAMI.2021.3062900

[17] E. Kalaitzidis et al., “AMLAS: An AutoML Pipeline for Data Streams”, MDPI Machine learning & Knowledge extraction, 7(3), pp1-33, July. 2025, DOI: <https://doi.org/10.3390/make7030087>

[18] M. Schaller et al., “AutoML for Multi-Class Anomaly Compensation of Sensor Drift”, arXiv, Feb. 2025, DOI: <https://doi.org/10.48550/arXiv.2502.19180>

[19] O. Choi, “D2MR: A Framework for Recommending Drift Detection Method via Data Characteristic-based Rule Mapping in Machine Learning System”, Journal of Software Assessment and Valuation, 21(3), pp.47-60, 2025 , DOI: <http://dx.doi.org/10.29056/jsav.2025.09.05>

[20] UCI Wine Quality Dataset, [Internet], <https://archive.ics.uci.edu/dataset/186/wine+quality>

[21] CIFAR-10 Dataset, [Internet], <https://www.cs.toronto.edu/~kriz/cifar.html>

저 자 소 개



최옥주(Okjoo Choi)

2008.02 숙명여자대학교 컴퓨터과학과 박사
 1990.8-1996.3 LG생산기술원 주임원연구원
 1996.7-2009.8 한국오라클 수석건설턴트
 2009.9-2022.12 카이스트 전산학부 연구교수
 2023.10-2024.2 강원대학교 산학협력중점교수
 2024.3-현재 배재대학교 AI·SW공학부 조교수
 <주관심분야> 빅데이터 분석, 데이터품질, 소프트웨어품질, 프로젝트 관리, AI안정성



신원선(Wonsun Shin)

1997년 숙명여자대학교 전산학과 석사
 2022년 충북대학교 전파통신공학과 박사수료
 2022년~현재 배재대학교 스마트ICT융합학과 박사과정
 1996년 숙명여자대학교 아태여성정보통신원 연구원
 2015년 공릉컴 평생교육원 이사
 2020년~현재 주식회사 비전21테크 대표
 관심분야: 자동음성인식, 감정인식, 데이터 품질